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

Dimos Kambouroudis* David McMillan† Katerina Tsakou‡

November 13, 2019

Abstract

We examine the role of implied volatility, leverage effect, overnight returns and volatility of realized volatility in forecasting realized volatility by extending the heterogeneous autoregressive (HAR) model to include these additional variables. We find that implied volatility is important in forecasting future realized volatility. In most cases a model that accounts for implied volatility provides a significantly better forecast than more sophisticated models that account for other features of volatility, but exclude the information backed out from option prices. This result is consistent over time. We also assess whether leverage effect, overnight returns and volatility of realized volatility carry any incremental information beyond that captured by implied volatility and past realized volatility. We find that while overnight returns and leverage effect are important for some markets, the volatility of realized volatility is of limited value for most stock markets.

Keywords: volatility forecasting; HAR model; realized volatility; implied volatility indices; leverage effect; overnight returns; GARCH.

*d.s.kambouroudis@stir.ac.uk - Department of Accounting and Finance, University of Stirling, Stirling, FK9 4LA, UK

†david.mcmillan@stir.ac.uk - Department of Accounting and Finance, University of Stirling, Stirling, FK9 4LA, UK

‡katerina.tsakou@swansea.ac.uk - Department of Accounting and Finance, School of Management, Swansea University, Bay Campus, Fabian Way, Swansea, SA1 8EN, UK (Corresponding author)

1 Introduction

Volatility plays an important role in both theoretical and empirical finance. Since volatility is a key input in asset pricing, derivative pricing and risk management, forecasting volatility is a critical task. The availability of high-frequency data has enabled the development of non-parametric daily estimators of volatility known as realized volatility (RV) based on the sum of squared intraday returns (Andersen & Bollerslev, 1998). Due to its non-latent character, RV is not only used as a measure of daily volatility but also to explore the predictability of RV. Indeed, empirical results in Andersen et al. (2003) clearly show that reduced form time series models for RV outperform the popular GARCH and related stochastic volatility models in forecasting future volatility.[1]

At the same time, several studies have suggested the importance of explicitly incorporating implied volatility (IV) in the traditional GARCH and stochastic volatility models for the purpose of forecasting. In particular, it appears that IV contains additional information about future volatility beyond that captured in model based volatility forecasts.[2] There are also a few studies that suggest that IV forecasts outperform GARCH based forecasts, suggesting that IV forecasts are strong at least when the alternative is historical models based only on lower frequency returns.[3] However, little attention has been paid on whether extracting information from option prices is useful in forecasting realized volatility.

In this paper we examine the importance of IV as an additional source of volatility information. Implied volatility measures the volatility implied by option prices on the underlying asset and are considered to be the market's forecasts of the volatility of the stock market. Thus, IV is a forward looking measure of the expected stock market volatility, and a general indication of the risk aversion of the market. The construction of Chicago Board Options Exachange's (CBOE) VIX volatility index computed using the S&P500 option prices, the so-called "investor fear gauge", has become an important market indicator (Whaley, 2000). It is probably the most widely used example of option-implied information. Its popularity as a hedging instrument for investors has spawn the introduction of several other volatility indices around the world. The issue is whether IV retains additional information about future RV. To address this issue, we apply the heterogeneous autoregressive (HAR) model proposed by Corsi (2009). We augment the HAR model to include an IV index as an additional regressor.

[1] The superiority of high-frequency based models is empirically demonstrated, among others, in Martens & Zein (2004), Pong et al. (2004) and Koopman et al. (2005).

[2] See, among others, Blair et al. (2001), Frijns et al. (2010), Yu et al. (2010), Kambouroudis et al. (2016).

[3] See, Poon & Granger (2003) for an extensive review on the topic.

The straightforward estimation and superior forecast performance of the HAR model make it widely used for forecasting RV. Since its introduction, several extensions of the HAR model have been proposed in order to capture stylized facts of volatility. In this paper, we also examine the importance of explicitly incorporating three features of volatility. First, we allow for the leverage effect, i.e. the tendency of volatility to increase more after a negative shock than a positive shock of the same magnitude. Martens et al. (2009) provide evidence that accounting for the leverage effect leads to improvements in forecast accuracy. Corsi & Renò (2012) extend the HAR model to allow for the leverage effect. They show that the impact of leverage effect on RV is significant for the S&P500 index. Similar results are provided in Wang et al. (2015) for the Chinese stock market.

Second, we allow for the effect of overnight returns. Thus far, the evidence on the importance of overnight returns for volatility forecasting is less conclusive. Gallo (2001) extends the GARCH model with overnight returns and finds mixed results on their predictive power. On the other hand, Tsiakas (2008) and Tseng et al. (2012) show that news released at nighttime has predictive content for the subsequent daytime volatility.

Third, we consider that the volatility of RV is time-varying. Following the empirical evidence of Corsi et al. (2008) RV exhibits volatility clustering. Using data on S&P500 index, they show that allowing for time-varying volatility of RV leads to a substantial improvement of the accuracy of the volatility forecasts.

The paper makes at least three contributions. First, we provide a comprehensive analysis of the predictive content of IV indices for forecasting RV in a cross-section of 10 international stock indices. To the best of our knowledge, the two previous studies that have examined the role of IV have exclusively focused on the S&P500 index. Specifically, Busch et al. (2011) and Oikonomou et al. (2019) find implied volatility extracted using different methods contain incremental information beyond that in high-frequency RV for the S&P500 index.[4] Second, we examine whether allowing for the leverage effect, overnight returns and volatility of RV improves the accuracy of volatility forecasts. In particular, we are interested in examining whether these features retain additional information apart from that contained in IV and RV. Third, we evaluate the relative forecast performance of our models not only on average over the forecast period but also locally. We assess whether the forecast performance of the models changes over time.

Our results highlight the importance of IV information for forecasting realized volatility. The inclusion of IV in the HAR model substantially improves forecasts for all 10 international stock

[4]Note that Busch et al. (2011) evaluates the role of IV also on the $/DM exchange rate and 30 year T-bond futures.

markets we study. Taking into account separately the leverage effect and overnight returns improves the forecast accuracy. However, we find that explicitly accounting for the volatility of realized volatility does not significantly improve forecasts. A model that includes the information of both RV and IV outperforms the models that account for one or more of the other features. This result is consistent over time. Finally, the empirical results show that augmenting the inclusion of the leverage effect, overnight returns and volatility of realized volatility in the HAR-IV model marginally improve its performance. Overall, our results show that IV is informative for predicting RV for all stock indices we include in our study, while the other features of volatility we account for marginally improve the forecasts for some equity markets.

The remainder of the paper is organized as follows. In Section 2, we outline realized volatility and describe the volatility model that we use to evaluate the role of implied volatility, leverage effect, overnight returns and volatility of realized volatility on realized forecasts. The forecast evaluation techniques we use to assess the relative forecast performances are presented in the same section. Section 3 describes the data. In Section 4, we present the empirical results, and Section 5 concludes.

2 Empirical framework

This section starts with a brief introduction to realized volatility. We then describe the HAR models of Corsi (2009) and its various extensions that will be used to assess the role of implied volatility, leverage effect, overnight returns and volatility of realized volatility on realized forecasts. We conclude this section with a description of the techniques we rely on 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 $m \rightarrow \infty$, that is in case of continuous sampling, RV_t converges to the true integrated variance. Since then, several alternatives to the standard RV have been proposed (see, Barndorff-Nielsen & Shephard (2004), Hansen & Lunde (2006) among others). However, the standard RV calculated based in the 5-minute intraday returns remains the most popular in practical applications (Bollerslev et al., 2009).

2.2 Modeling realized volatility

There are two main classes of models for estimating and forecasting RV. They both aim to capture the long-memory characteristic of the volatility series. The first considers the ARMA and fractionally integrated ARMA (ARFIMA) model for RV (see, Andersen et al., 2003; Koopman et al., 2005; Martens et al., 2009, among others). The second considers the heterogeneous autoregressive (HAR) model for RV developed by Corsi (2009). The HAR model has been widely used to model and forecast RV, because of its simplicity and ability to reproduce volatility persistence, while it does not formally belong to the long memory models.

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 differences in time horizon due to the fact that market participants have a large spectrum of trading frequency. The main idea is that participants 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:

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

where ε_{t+h} is an innovation term and $y_t^{(d)}$, $y_t^{(w)}$, $y_t^{(m)}$ are the three 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 realized variance, i.e. $y_{t+h} = log(RV^2_{t+h})$ instead of the realized volatility in the original formulation of the HAR model in Corsi (2009). This choice is motivated by the fact that Andersen et al. (2003) note that the distribution of the logarithmic realized variance is much closer to normal, and Andersen et al. (2007) find similar results when realized variance, 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 than the one in Corsi (2009). The original HAR model regresses RV on the previous 1-day, 5-day and 22-day average realized volatility. To avoid overlapping terms and, thus, ease interpretation, we follow Patton & Sheppard (2015) so that the weekly component is constructed based on log RV between lag 2 and 5 and the monthly component consists of log RV between lag 6 and 22. Thus, to formalize the structure of our HAR model for RV, let $y^{(d)}_t = logRV^2_t$, $y^{(w)}_t = \frac{1}{4}\sum_{i=1}^{4} logRV^2_{t-i}$ and $y^{(w)}_t = \frac{1}{17}\sum_{i=5}^{21} logRV^2_{t-i}$ be the daily, weekly and monthly components. One of the reasons for the popularity of the HAR model in Eq. (4) is its simplicity. Once the three volatility components have been constructed, the HAR model is estimated by simple OLS.

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

$$y_{t+h} = \alpha_0 + \alpha_d y^{(d)}_t + \alpha_w y^{(w)}_t + \alpha_m y^{(m)}_t + \beta' \boldsymbol{X} + \varepsilon_{t+h} \tag{5}$$

so that the model allows for $k \times 1$ vector $\boldsymbol{X}$ of exogenous variables and includes implied volatility, leverage effect and overnight returns. [5]

To examine the information content of implied volatility on RV, we add three components of implied volatility in the HAR model (hereafter HAR-IV) by setting $\beta' \boldsymbol{X} = \beta_d logIV^{(d)}_t + \beta_w logIV^{(w)}_t + \beta_{mlog} IV^{(m)}_t$ in Eq. (5). The log IV components are defined analogously to the RV components.[6]

[5]We do not explicitly consider the role and results for jumps components in this paper. Important contributions in this area include Barndorff-Nielsen & Shephard (2005), Andersen et al. (2007) and Corsi & Renò (2012). Of note, Busch et al. (2011) finds 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 & 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. Using the bi-power variation, we do consider jumps and find that while accounting for jumps improves performance over the basic HAR model for some indices, it does not improve performance compared to 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.

[6]To the best for our knowledge, only Buncic & Gisler (2016) augment the HAR model with the three IV components of the VIX index for forecasting RV. However, the objective of this study is to examine whether US-based

Corsi & Renò (2012) extend the HAR model by adding positive and negative daily, weekly and monthly returns to capture the leverage effect. The HAR-L specification includes positive and negative daily returns by setting $\beta' \boldsymbol{X} = \beta_1 r_t^+ + \beta_2 r_t^-$, that is RV reacts asymmetrically to previous daily positive and negative returns.[7]

We capture the arrival of news at nighttime by adding the overnight returns as explanatory variable in the HAR model. Similar to Wang et al. (2015) we set $\beta' \boldsymbol{X} = \beta_1 r_{over,t}$. The overnight returns, $r_{over,t}$, is defined as the differences of the logs of the open price on day $t+1$ and the close price on the previous day t.

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

Finally, we also extend the models to capture two or more features of RV simultaneously. Thus, we have sixteen HAR specifications to model and forecast RV.

2.3 Forecast evaluation

The ability of the models described in Subsection 2.2 to accurately forecast RV is assessed using the QLIKE loss function. 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. For this reason, Patton (2011) identifies a family of loss functions that are 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 Eq. (3). Patton (2011) demonstrates the robustness of the QLIKE loss function that is computed as:

$$QLIKE: \; L(\hat{\sigma}_t^2, h_{t|t-k}) = \frac{\hat{\sigma}_t^2}{h_{t|t-k}} - log \frac{\hat{\sigma}_t^2}{h_{t|t-k}} - 1 \tag{6}$$

where $h_{t|t-k}$ is the volatility point forecast based on $t-k$ information ($k>0$) and $\hat{\sigma}_t^2$ is the proxy for the conditional volatility (here, RV). Brownlees et al. (2011), Sevi (2014) and Tian et al. (2017),

volatility information can be used to improve RV forecast in a large cross-section of international equity markets.

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

among others, use the QLIKE loss function 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 Giacomini-White (GW) pairwise test proposed by Giacomini & White (2006). The MCS procedure determines the set, $\widehat{M}^*_{1-\alpha}$, that contains the best models from a full set of models, M, at a given confidence level, $(1-\alpha)$. Starting with the full set of models, 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, $\widehat{M}^*_{1-\alpha}$, containing the best models from M at the $(1-\alpha)$ confidence level.

Following Hansen et al. (2011), we employ two test statistics for testing the null hypothesis of equal predictive ability; the range statistics $T_R = \underset{i,j \in M}{max} \frac{|\bar{d}_{ij}|}{\sqrt{\widehat{var}(\bar{d}_{ij})}}$ and the semi-quadratic statistics $T_{SQ} = \sum\limits_{i,j \in M} \frac{(\bar{d}_{ij})^2}{\widehat{var}(\bar{d}_{ij})}$, where $d_{ij,t}$ is the loss differential between models i and j in the set, $\bar{d}_{ij} = \frac{1}{T_h}\sum_{t=N}^{T-h} d_{ij,t}$. The MCS procedure yields p-values for each of the models 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 $\hat{p}_i \geq \alpha$.

Giacomini-White (GW) test is a test of conditional predictive ability proposed by Giacomini & White (2006). The test evaluates the forecasting performance of two competing models, accounting for parameter uncertainty. In short, let $L(y_t; \hat{y_t})$ denote the forecast loss where y_t is the 'true' value and $\hat{y_t}$ is the predicted value. The difference in loss of model i relative to a benchmark model o is defined as

$$d_{i,t} = L(y_t; \hat{y_{o,t}}) - L(y_t; \hat{y_{i,t}}) \tag{7}$$

The issue is whether the two models have equal predictive ability. That is, the null hypothesis tested is $H_0 : E(d_{i,t+\tau} \mid h_t) = 0$, where h_t is some information set. The CPA test statistic is then computed as a Wald statistic

$$CPA_t = T(T^{-1}\sum_{t=1}^{T-\tau} h_t d_{i,t+\tau})' \hat{\Omega}_T^{-1} (T^{-1}\sum_{t=1}^{T-\tau} h_t d_{i,t+\tau}) \sim \chi^2_1 \tag{8}$$

where $\hat{\Omega_T}$ is the Newey and West (1987) HAC estimator of the asymptotic variance of the $h_t d_{i,t+\tau}$.

The MCS procedure and the GW test help to 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 itself change over time (Giacomini & Rossi, 2010). To evaluate whether the relative predictive performance of the models is time-varying we compute the cumulated sum of the QLIKE errors and test the local performance of the models using the

Giacomini & Rossi (2010) fluctuation test.

The cumulative forecast error is

$$cumFE_t = \sum_{\tau=T_0}^{t} FE_\tau \tag{9}$$

as a function of $t = T_0, ..., T$, where T is the number of out-of-sample observations and FE_τ is the forecast error based on the QLIKE loss function. The $cumFE$ sequence allow us to assess the changes in the forecast performance of the models over time. The model that yields the lowest cumulative QLIKE forecast errors produces the most accurate forecast over time.

In addition to computing the $cumFE$ to identify the forecast performance of the models over time, we also apply the Giacomini & Rossi (2010) fluctuation test to formally test whether the models' relative performance is time-varying. Giacomini & Rossi (2010) test the null hypothesis that two models, k and l, have equal predictive ability at each point in time estimated over a rolling window of data by proposing the fluctuation test statistic:

$$F_{t,h,m} = \widehat{\sigma}^{-1} m^{-1/2} \sum_{j=t-m/2}^{t+m/2-1} \Delta L_{h,j} \tag{10}$$

for $t = R + 1 + m/2, ..., T - m/2 + 1$, where $\Delta L_{h,j} = (e_{t+h}^{(k)} - e_{t+h}^{(l)})$ is the loss differential between models k and l, where $\widehat{\sigma}^2$ is the HAC standard error estimator of σ^2. The fluctuation test is similar to the GW test described before computed over a rolling out-of-sample window. Specifically, the unconditional test is computed for each rolling window. If the $max_t F_{t,h,m}$ is higher than the critical value, k_a, the null hypothesis of local equal predictive ability between the two models is rejected.

3 Data

Our dataset consists of daily realized variance data, open and close prices and implied volatility indices for 10 international equity markets. The daily realized variance measures that we employ are based on five-minute intraday returns and are obtained from the Oxford-Man Institute's Quantitative Finance Realized Library (Heber et al., 2009).[8] The daily open and close prices, and the implied volatility indices are collected from Datastream. Since the various implied volatility indices

[8] http://realized.oxford-man.ox.ac.uk/data/download. Our realized variances are from Libray Version 0.2. The Oxford-Man Institute's Quantitative Finance Realized Library contains a selection of daily non-parametric estimated of volatility. The choice of the five minute RV is mostly due to its simplicity and robustness. Recently, Liu et al. (2015) considered over 400 different volatility estimators and concluded that it is difficult to significantly beat 5-minure RV.

have been listed on different dates, we consider the period from February 2, 2001 to February 28, 2017, in order to study the indices over the same time period.

In particular, 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 implied volatility 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.[9]

Table 1 provides the descriptive statistics for the (log) RV and IV as well as for daily returns of all data we use in our study. 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 of STOXX50 with kurtosis 4.83 exhibiting 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. They are fairly close to being normally distributed with the exception of the log IV series of CAC40 and Nikkei225 that have 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.

4 Results

Table 2 presents the out-of-sample forecast evaluation results for all ten indices that we consider using a rolling estimation window with the initial in-sample fitting period from February 2, 2001 to December 31, 2005. In order to assess the one-step-ahead forecasts we report the forecast errors of the benchmark HAR model in Eq. (4) and the fifteen HAR specifications augmented with one or more explanatory variables (Eq. 5) based on the QLIKE loss function. The relative error obtained by the ratio of the QLIKE error for each model to the QLIKE error of the benchmark HAR is

[9] 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 & Tian, 2005). As the volatility indices are quoted in annualized percentages and in order 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.

provided in brackets. The model that yields the lowest loss, and thus 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. The relative errors in brackets indicate the poor performance of the benchmark HAR model, as in most cases the relative forecast error is below one. Thus, accounting for one or more of the features of volatility appears to improve forecast accuracy. First, augmenting the benchmark HAR model with each of the various stylized facts individually (first part of Table 2) enhances the forecast accuracy across all indices. Specifically, explicitly incorporating the implied volatility index in the HAR model improves strikingly 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 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.

Second, in general, of all features, accounting for implied volatility 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, i.e. HAR-IV, HAR-L, HAR-O and HAR-G we find that the biggest predictive gains are driven by implied volatility. In addition, the parsimonious HAR-IV specification yields remarkably lower loss than specifications that exclude implied volatility, but include simultaneously two or more of the other stylized facts, such as HAR-LO, HAR-LG, HAR-OG, HAR-LOG.

Third, while the substantial forecast improvement of the HAR model is driven by the inclusion of implied volatility, the other features of realized volatility play some role as well. Comparing the forecast error of the HAR-IV with more sophisticated models that include not only implied volatility, but also leverage effects, overnight returns and/or the volatility of realized volatility we find that the latter are marginally favored across all indices. Overall, the HAR-IVOG performs best for the S&P500, DJIA and Nikkei225, while the HAR-IVLO is the best model for the Nasdaq100 and most of the European indices with the exception of the DAX and FTSE100 index that it is the second best model with the first being the HAR-IVLOG and HAR-IVL respectively.

To compare the relative forecast performance of the models we apply the MCS test. Based on the QLIKE loss function Table 3 presents the p-values from the MCS test at the 10% and 25% significance level. The MCS findings reveal the same picture as Table 2. Across all indices only HAR specifications that account for implied volatility belong to the MCS as the p-values are larger than 10% indicating that models that take into account the information of implied volatility are superior to the other HAR specifications. This indicates that the implied volatility is important in forecasting future realized volatility. The results are consistent using both the range statistics and the semi-quadratic statistics. As for the other stylized facts of realized volatility, they appear to contain incremental information about future realized volatility beyond that captured in HAR-IV.

In sum, the main result of Tables 2 and 3 is that a substantial part of the predictive gains is driven by implied volatility backed out from option prices. In most cases, the parsimonious HAR-IV preforms better than models that do not account for IV even if they are more sophisticated, i.e. contain more than one of the other stylized features of RV simultaneously.

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

To further evaluate the strength of this result we examine whether the forecast performance of the HAR-IV relative to the models that exclude implied volatility information, i.e. 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 the Giacomini and Rossi (2010) test of local forecast performance of the models. The cumulative forecast error, as defined in Eq. (9), are plotted in Figure 1.[10] First, we observe that the HAR-IV specification yields the lowest cumulated sum of the QLIKE errors. This indicates that the HAR-IV produces the most accurate forecast consistently over time and, thus, including implied volatility substantially improves the forecast performance of the HAR model. The only exceptions are for the AEX, SMI and Nikkei225 indices where the HAR-IV, HAR-L and HAR-O models yield the lowest QLIKE errors interchangeably over time. Second, we find that the basic HAR and HAR-G specifications provide the worst forecasts over time.

Further graphic evidence of the superiority of HAR -IV is provided in Figure 2 based on the

[10]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.

results of Giacomini and Rossi (2010) test. The test compares the relative performance of two competing models over time. We evaluate the performance of HAR-IV against the basic 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 it 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 one of the competing models 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, specifically, below the lower dashed line. This means that the HAR-IV model consistently outperforms the basic HAR model and 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 is 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. Overall, the superiority of the HAR-IV model is consistent across time regardless how turbulent or tranquil the period is.

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

It is evident from the results reported in Tables 2, 3 and Figures 1 and 2 that the HAR model that accounts for the implied volatility index outperforms not only the HAR model, but also of the models that capture well-known features of realized volatility, such as leverage effect, overnight returns and the volatility of realized volatility. In this subsection, we evaluate whether these features of realized volatility carry any additional information to future realized volatility beyond that contained in implied volatility and realized volatility. For this reason we augment the HAR-IV model with one or more of these features simultaneously.

Table 4 reports the Giacomini and White pairwise test results for all indices. The p-values reported in the tables are based on the mean differences between the row model and the column model. The null hypothesis of equal forecasting performance between the row and column models in terms of QLIKE forecast error. The signs in brackets indicate which model performs best. A positive sign shows that the row model forecast yields larger loss than the column model forecast, which implies that the column model is significantly superior. Similarly, a negative sign denotes that the row model forecast performs significantly better than the column model forecast, since the latter produces larger loss.

Several interesting conclusions emerge from Table 4. First, augmenting the HAR-IV with overnight returns significantly enhances forecast accuracy for all indices with the exception of the DJIA and SMI index. Second, incorporating the volatility of the realized volatility in the HAR-IV significantly improves the forecast performance of the model only for the S&P500 and DJIA index. For all other indices, the HAR-IV outperforms the HAR-IVG. In some cases, the null hypothesis of equal predictive ability between the HAR-IV and HAR-IVG is rejected at the 5% level. In other cases, the p-value of the test is high, which indicates that the volatility of realized volatility does not contain any significant information.

Third, adding leverage effect, one of the most salient features of realized volatility, adds surprisingly little in terms of forecast accuracy. Specifically, the HAR-IVL model only significantly outperforms the HAR-IV for the SMI index. 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 of the HAR-IV and, thus, leverage effect appears not to play a significant role in forecasting realized volatility.

Comparing more sophisticated HAR-IV specifications we find that the most sophisticated model does not significantly outperform the more restricted specifications, which means that on average adding some features of volatility do not significantly improve the forecast accuracy, and a more parsimonious model is superior. Specifically, augmenting HAR-IVO with leverage effect (HAR-IVLO) gives the best forecast model for the Nasdaq, CAC40, DAX and SMI index. For these indices, the HAR-IVLO model performs significantly better than all the other competing models. The HAR-IVOG produces the best forecast for the S&P500 and DJIA, whereas adding leverage effect does not significantly improve the forecast. For the STOXX50, AEX, Nikkei225 and FTSE100 indices an even more parsimonious model performs best. Specifically, for the STOXX50, AEX, Nikkei225 indices, the HAR-IVO provide the best forecast, while for the FTSE100 the HAR-IV is preferred.

5 Conclusion

This paper examines the role of implied volatility, leverage effect, overnight returns and volatility of realized volatility in forecasting future realized volatility in 10 international stock markets.

We assess the role of these variables by augmenting the HAR model of Corsi (2009). The empirical results show that adding each of these variables on the HAR model improves its forecast performance with the exception of the volatility of realized volatility that does not lead to a significant improvement of the forecast performance of the HAR model.

The most important finding is the strong role of implied volatility as an additional source of volatility information. Specifically, the HAR-IV model in most cases provides significantly better forecasts than the more sophisticated models that exclude the information backed out from option prices. This result is consistent over time regardless how turbulent or tranquil the period is. We also assess whether leverage effect, overnight returns and volatility of realized volatility carry any incremental information beyond that captured by implied volatility and past realized volatility by augmenting the HAR-IV model with these features of volatility. Perhaps surprisingly, while overnight returns appear to be beneficial for almost all stock markets and leverage effect adds some incremental information in four markets, the volatility of realized volatility has some value for only two US indices, but much less for the remaining eight stock markets. This confirms our result of the strong role of implied volatility and the fact that adding some features of volatility do not always improve the forecast accuracy, and a more parsimonious model is superior.

Overall, our results show that implied volatility is very informative for predicting realized volatility for all stock indices we include in our study, while the other features of volatility we account for marginally improve the forecasts for some equity markets.

Acknowledgements

We are grateful to participants at the International Conference on Computational and Financial Econometrics (2018), the Forecasting Financial Markets Conference (2018), the International Association for Applied Econometrics Conference (2018), the University of Stirling and Swansea University for helpful comments and discussions.

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.

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. Forecasting and empirical methods in finance and macroeconomics.

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

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. Ã. (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.

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.

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.

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 0.2.

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

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.

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. Forecasting Returns and Risk in Financial Markets using Linear and Nonlinear Models.

Martens, M. & Zein, J. (2004). Predicting financial volatility: High-frequency time-series forecasts vis-Ã -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.

Oikonomou, I., Stancu, A., Symeonidis, L., & Simen, C. W. (2019). The information content of short-term options. *Journal of Financial Markets*.

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

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.

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.

Poon, S.-H. & Granger, C. W. J. (2003). Forecasting volatility in financial markets: A review. *Journal of Economic Literature*, 41(2), pp. 478–539.

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

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.

Tseng, T.-C., Lai, H.-C., & Lin, C.-F. (2012). The impact of overnight returns on realized volatility. *Applied Financial Economics*, 22(5), 357–364.

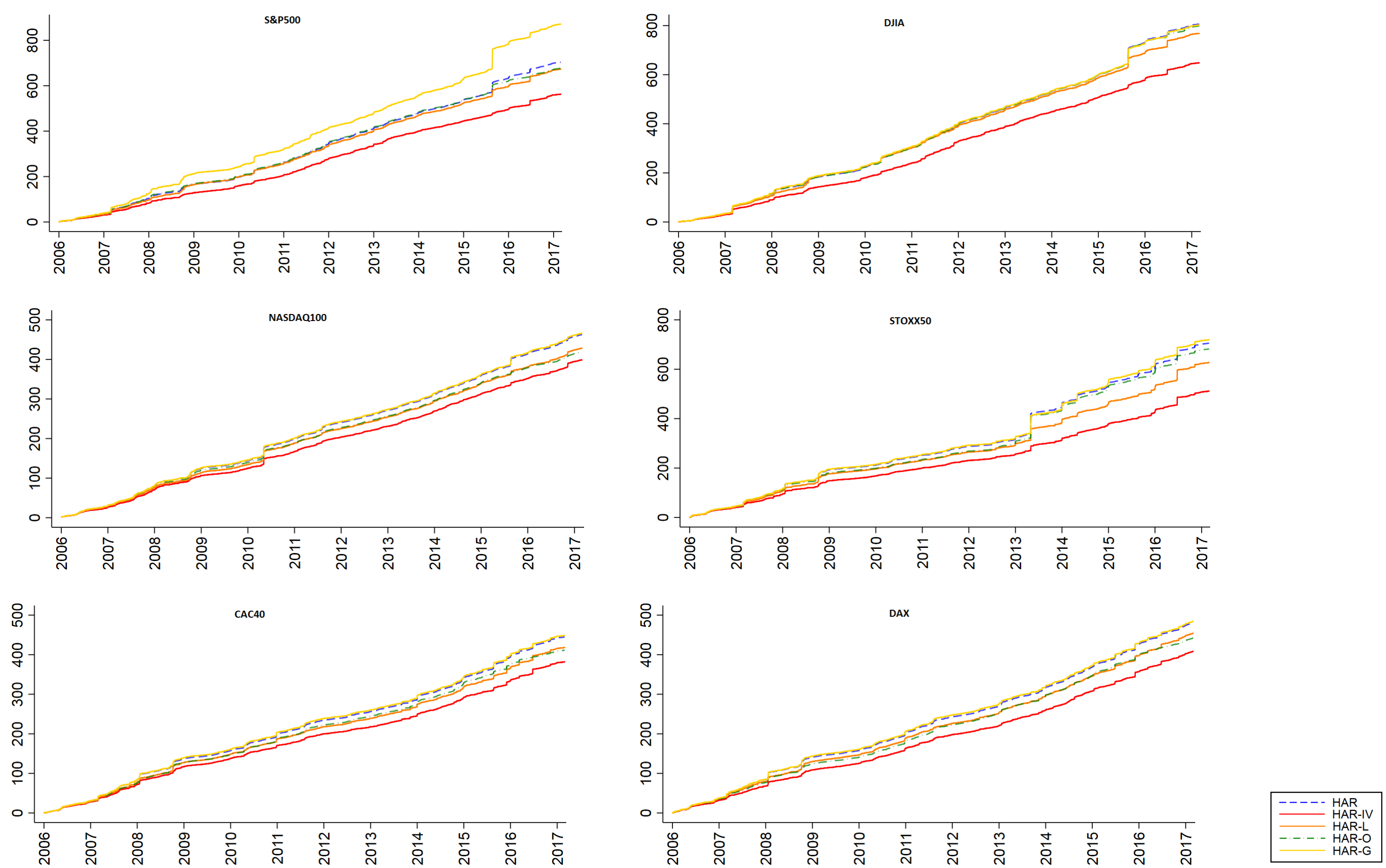
Tsiakas, I. (2008). Overnight information and stochastic volatility: A study of european and us stock exchanges. *Journal of Banking & Finance*, 32(2), 251 – 268.

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.

Whaley, R. E. (2000). The investor fear gauge. *The Journal of Portfolio Management*, 26(3), 12–17.

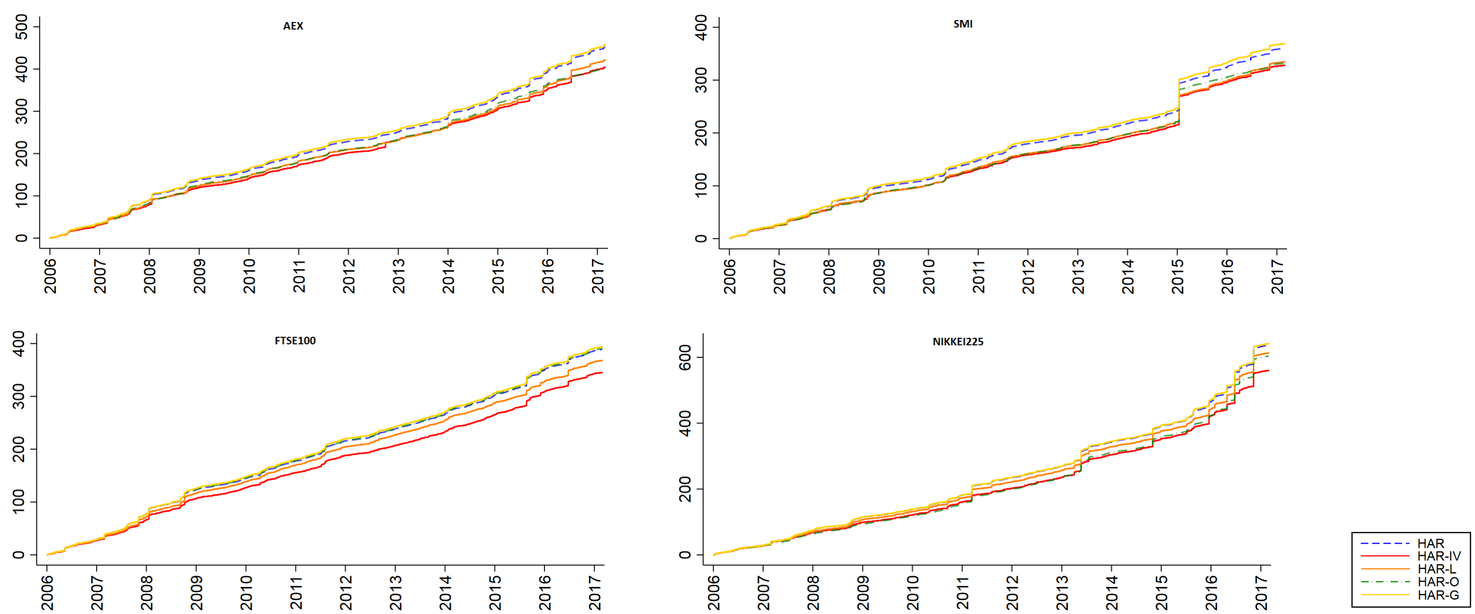
Yu, W. W., Lui, E. C., & Wang, J. W. (2010). The predictive power of the implied volatility of options traded OTC and on exchanges. *Journal of Banking & Finance*, 34(1), 1–11.

Figure 1: Cumulated sum of the QLIKE forecast error



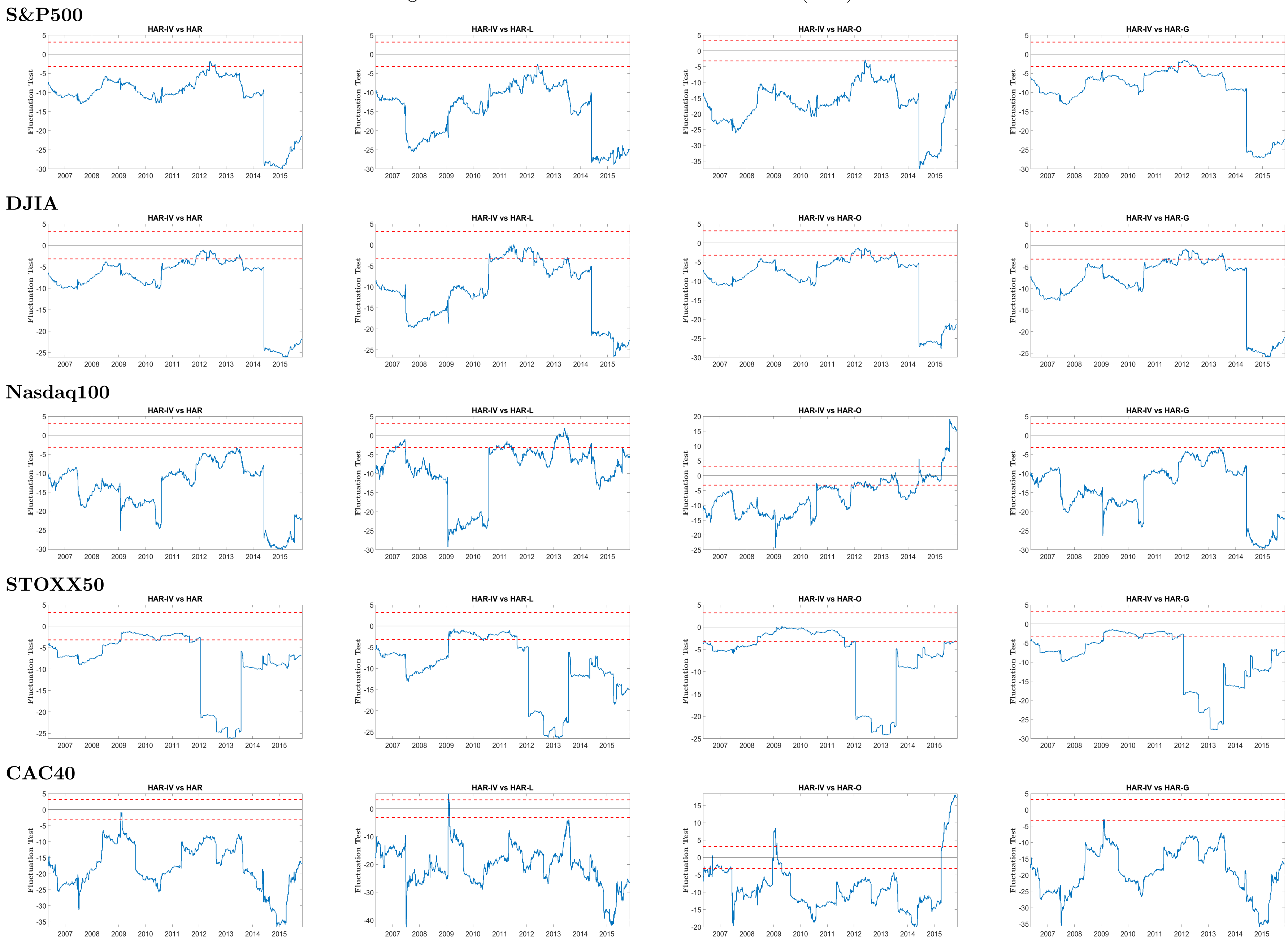
- Continued on next page -

Figure 1 – continued from previous page



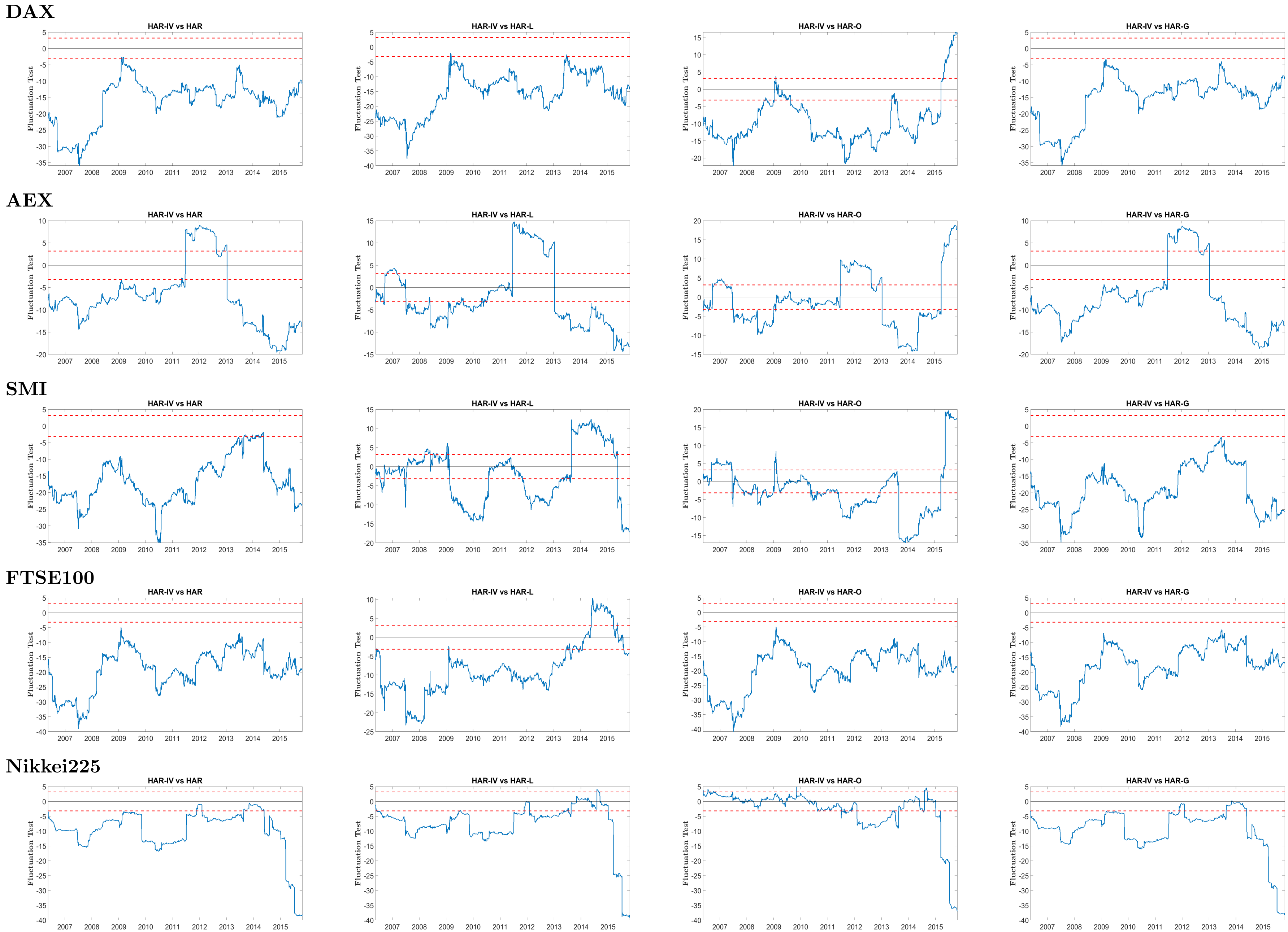
Note: The figures present the cumulative QLIKE forecast errors of the HAR model (blue dashed line), HAR-IV (red line), HAR-L (orange line), HAR-O (green dashed-dotted line) and HAR-G (yellow line) for all stock markets as defined in Eq. 9 over the forecast period (January 1, 2006 - February 28, 2017). 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.

Figure 2: Fluctuation test of Giacomini & Rossi (2010)



- Continued on next page -

Figure 2 – continued from previous page



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 indicates the fluctuation test statistic. The HAR-IV specification is significanly better (worse) than one of the competing models at the 5% level when the solid line is below (above) the lower (upper) dashed line.

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	-9.828	1.114	0.460	3.403	0.806	0.774	0.743	0.806	0.353	0.178
DJIA	-9.831	1.096	0.516	3.526	0.774	0.745	0.715	0.774	0.364	0.192
Nasdaq100	-9.735	1.022	0.424	3.087	0.843	0.795	0.764	0.843	0.289	0.164
STOXX50	-9.274	1.029	0.136	4.830	0.769	0.724	0.706	0.769	0.325	0.219
CAC40	-9.405	0.974	0.410	3.229	0.831	0.789	0.768	0.831	0.319	0.202
DAX	-9.282	1.045	0.421	3.211	0.835	0.796	0.778	0.835	0.328	0.212
AEX	-9.607	1.021	0.515	3.186	0.844	0.805	0.783	0.844	0.323	0.197
SMI	-9.864	0.923	0.863	3.757	0.859	0.824	0.806	0.859	0.328	0.207
FTSE100	-10.011	1.017	0.590	3.271	0.844	0.809	0.790	0.844	0.335	0.207
Nikkei225	-9.606	0.906	0.337	3.484	0.783	0.727	0.694	0.783	0.293	0.175
log(IV)										
S&P500	-8.915	0.749	0778	3.407	0.984	0.970	0.958	0.984	0.078	0.049
DJIA	-9.045	0.734	0.831	3.421	0.984	0.972	0.961	0.984	0.107	0.051
Nasdaq100	-8.484	0.840	0.798	2.764	0.990	0.981	0.973	0.990	0.029	0.050
STOXX50	-8.466	0.707	0.614	3.092	0.986	0.972	0.960	0.986	0.019	0.048
CAC40	-8.591	0.706	0.177	6.987	0.967	0.949	0.928	0.967	0.204	-0.002
DAX	-8.515	0.710	0.767	3.260	0.988	0.975	0.964	0.988	-0.016	0.045
AEX	-8.634	0.790	0.728	3.101	0.987	0.976	0.966	0.987	0.068	0.058
SMI	-8.975	0.715	0.968	3.743	0.989	0.977	0.966	0.989	-0.066	0.033
FTSE100	-8.937	0.764	0.688	3.166	0.984	0.971	0.958	0.984	0.050	0.040
Nikkei225	-8.380	0.618	0.634	4.312	0.981	0.964	0.949	0.981	0.044	0.039
Returns										
S&P500	$1.46 * 10^{-4}$	0.012	-0.226	12.188	-0.082	-0.045	0.026	-0.082	-0.052	0.018
DJIA	$1.69 * 10^{-4}$	0.011	-0.064	11.964	-0.080	-0.039	0.039	-0.080	-0.045	0.032
Nasdaq100	$2.09 * 10^{-4}$	0.016	-0.019	8.550	-0.049	-0.060	0.015	-0.049	-0.062	0.009
STOXX50	$-7.16 * 10^{-5}$	0.015	-0.046	7.673	-0.028	-0.043	-0.046	-0.028	-0.043	-0.049
CAC40	$-3.19 * 10^{-5}$	0.015	-0.019	8.143	-0.030	-0.038	-0.051	-0.030	-0.039	-0.053
DAX	$1.44 * 10^{-4}$	0.015	-0.052	7.590	-0.010	-0.024	-0.021	-0.010	-0.024	-0.021
AEX	$-4.34 * 10^{-5}$	0.015	-0.093	9.388	-0.004	-0.010	-0.057	-0.004	-0.010	-0.057
SMI	$2.96 * 10^{-5}$	0.012	-0.171	9.817	0.033	-0.051	-0.039	0.033	-0.052	-0.036
FTSE100	$3.94 * 10^{-5}$	0.012	-0.155	9.498	-0.043	-0.043	-0.055	-0.043	-0.045	-0.059
Nikkei225	$9.25 * 10^{-5}$	0.015	-0.387	9.349	-0.043	-0.005	-0.013	-0.043	-0.007	-0.013

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 February 2, 2001 to February 28, 2017. The last six columns of the table provide the first to third order of the autocorrelation function (ACF) and partial ACF (PACF).

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	0.2505	0.2870	0.1647	0.2464	0.1557	0.1695	0.1581	0.1285	0.1379	0.2329
HAR-IV	0.2004 (0.7998)	0.2310 (0.8048)	0.1420 (0.8620)	0.1804 (0.7251)	0.1337 (0.8587)	0.1440 (0.8493)	0.1415 (0.8950)	0.1170 (0.9109)	0.1224 (0.8871)	0.2049 (0.8798)
HAR-L	0.2397 (0.9567)	0.2736 (0.9533)	0.1526 (0.9259)	0.2200 (0.8887)	0.1464 (0.9402)	0.1601 (0.9447)	0.1475 (0.9326)	0.1194 (0.9291)	0.1305 (0.9458)	0.2242 (0.9628)
HAR-O	0.2411 (0.9621)	0.2843 (0.9904)	0.1492 (0.9055)	0.2398 (0.9668)	0.1440 (0.9249)	0.1558 (0.9195)	0.1412 (0.8931)	0.1185 (0.9222)	0.1385 (1.0038)	0.2207 (0.9477)
HAR-G	0.3103 (1.2385)	0.2856 (0.9951)	0.1660 (1.0074)	0.2514 (1.0198)	0.1571 (1.0088)	0.1707 (1.0073)	0.1599 (1.0110)	0.1316 (1.0245)	0.1396 (1.0120)	0.2346 (1.0076)
HAR-IVL	0.2024 (0.8077)	0.2343 (0.8162)	0.1405 (0.8527)	0.1794 (0.7214)	0.1323 (0.8494)	0.1429 (0.8430)	0.1370 (0.8664)	0.1137 (0.8851)	**0.1208 (0.8759)**	0.2048 (0.8795)
HAR-IVO	0.1926 (0.7687)	0.2265 (0.7892)	0.1301 (0.7896)	0.1704 (0.6792)	0.1201 (0.7712)	0.1308 (0.7714)	0.1244 (0.7865)	0.1074 (0.8355)	0.1226 (0.8890)	0.1936 (0.8312)
HAR-IVG	0.1993 (0.7955)	0.2295 (0.7997)	0.1424 (0.8645)	0.1833 (0.7376)	0.1341 (0.8613)	0.1440 (0.8494)	0.1419 (0.8975)	0.1187 (0.9235)	0.1232 (0.8931)	0.2049 (0.8799)
HAR-LO	0.2304 (0.9198)	0.2691 (0.9377)	0.1391 (0.8445)	0.2124 (0.8506)	0.1339 (0.8597)	0.1472 (0.8683)	0.1309 (0.8277)	0.1107 (0.8612)	0.1310 (0.9495)	0.2131 (0.9150)
HAR-LG	0.2395 (0.9558)	0.2740 (0.9547)	0.1530 (0.9286)	0.2239 (0.9049)	0.1471 (0.9444)	0.1603 (0.9456)	0.1484 (0.9384)	0.1208 (0.9398)	0.1313 (0.9516)	0.2248 (0.9653)
HAR-OG	0.2410 (0.9620)	0.2834 (0.9873)	0.1503 (0.9122)	0.2513 (1.0126)	0.1445 (0.9283)	0.1561 (0.9212)	0.1420 (0.8978)	0.1202 (0.9354)	0.1402 (1.0165)	0.2216 (0.9517)
HAR-IVLO	0.1941 (0.7748)	0.2295 (0.7995)	**0.1284 (0.7795)**	**0.1696 (0.6761)**	**0.1187 (0.7622)**	0.1301 (0.7678)	**0.1204 (0.7616)**	**0.1049 (0.8164)**	0.1211 (0.8780)	0.1937 (0.8318)
HAR-IVLG	0.2014 (0.8035)	0.2332 (0.8123)	0.1410 (0.8557)	0.1822 (0.7333)	0.1325 (0.8507)	0.1428 (0.8423)	0.1377 (0.8706)	0.1149 (0.8946)	0.1215 (0.8811)	0.2046 (0.8788)
HAR-IVOG	**0.1917 (0.7649)**	**0.2251 (0.7844)**	0.1305 (0.7924)	0.1730 (0.6898)	0.1203 (0.7724)	0.1306 (0.7705)	0.1244 (0.7869)	0.1079 (0.8398)	0.1235 (0.8954)	**0.1935 (0.8309)**
HAR-LOG	0.2299 (0.9174)	0.2699 (0.9404)	0.1393 (0.8454)	0.2187 (0.8752)	0.1341 (0.8614)	0.1471 (0.8679)	0.1312 (0.8297)	0.1116 (0.8686)	0.1318 (0.9557)	0.2133 (0.9159)
HAR-IVLOG	0.1931 (0.7708)	0.2284 (0.7958)	0.1287 (0.7812)	0.1721 (0.6866)	0.1189 (0.7636)	**0.1297 (0.7654)**	0.1207 (0.7631)	0.1054 (0.8205)	0.1218 (0.8832)	0.1936 (0.8315)

Note: The table reports the quasi-likelihood (QLIKE) error defined in Eq. 6 for all ten 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 QLIKE error for each model to the QLIKE error of the benchmark HAR is provided in brackets. The model that yields the lowest loss, and thus the model with the best forecast performance is indicated in bold for every index.

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	0.002	0.001	0.006	0.005	0.000	0.000	0.005	0.002	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.001
HAR-IV	0.018	0.008	0.147*	0.012	0.000	0.000	0.012	0.002	0.007	0.004	0.000	0.000	0.001	0.003	0.010	0.007	0.193*	0.143*	0.080	0.021
HAR-L	0.002	0.001	0.006	0.003	0.000	0.000	0.005	0.002	0.000	0.000	0.000	0.000	0.002	0.004	0.018	0.008	0.002	0.003	0.000	0.000
HAR-O	0.002	0.001	0.009	0.005	0.000	0.000	0.012	0.002	0.000	0.000	0.000	0.000	0.001	0.001	0.010	0.001	0.000	0.000	0.000	0.009
HAR-G	0.002	0.002	0.006	0.005	0.000	0.001	0.005	0.002	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.001
HAR-IVL	0.021	0.008	0.056	0.012	0.000	0.001	0.012	0.003	0.007	0.006	0.001	0.001	0.021	0.010	0.038	0.008	1.000**	1.000**	0.032	0.019
HAR-IVO	0.038	0.017	0.288**	0.012	0.086	0.068	0.310**	0.310**	0.045	0.058	0.083	0.074	0.040	0.052	0.038	0.008	0.132*	0.123*	0.985**	0.991**
HAR-IVG	0.038	0.011	0.288**	0.012	0.000	0.000	0.012	0.002	0.007	0.001	0.000	0.000	0.001	0.001	0.010	0.002	0.027	0.062	0.080	0.082
HAR-LO	0.002	0.001	0.006	0.003	0.016	0.014	0.005	0.002	0.007	0.001	0.000	0.000	0.021	0.025	0.038	0.008	0.002	0.001	0.080	0.025
HAR-LG	0.002	0.001	0.006	0.004	0.000	0.000	0.005	0.002	0.000	0.000	0.000	0.000	0.001	0.001	0.018	0.007	0.000	0.000	0.000	0.000
HAR-OG	0.002	0.001	0.006	0.005	0.000	0.000	0.012	0.002	0.000	0.000	0.000	0.000	0.001	0.001	0.010	0.002	0.000	0.000	0.000	0.005
HAR-IVLO	0.038	0.011	0.288**	0.012	1.000**	1.000**	1.000**	1.000**	1.000**	1.000**	0.089	0.074	1.000**	1.000**	1.000**	1.000**	0.410**	0.410**	0.978**	0.991**
HAR-IVLG	0.038	0.011	0.147*	0.012	0.000	0.000	0.012	0.002	0.007	0.005	0.001	0.001	0.021	0.009	0.018	0.008	0.193*	0.143*	0.080	0.020
HAR-IVOG	1.000**	1.000**	1.000**	1.000**	0.086	0.041	0.090	0.131*	0.045	0.051	0.089	0.074	0.040	0.046	0.038	0.008	0.017	0.031	1.000**	1.000**
HAR-LOG	0.002	0.001	0.006	0.003	0.016	0.011	0.005	0.002	0.007	0.000	0.000	0.000	0.021	0.010	0.018	0.008	0.000	0.000	0.080	0.019
HAR-IVLOG	0.038	0.017	0.288**	0.012	0.086	0.081	0.204*	0.198*	0.296**	0.296**	1.000**	1.000**	0.321**	0.321**	0.038	0.008	0.193*	0.123*	0.985**	0.991**

Note: The table reports the p-values of the MCS (Hansen et al., 2011) test in terms of the QLIKE criterion for our models accross all indices. The p-values based on the range statistics (T_R) and the semi-quadratic 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 eveluation perod covers January 1, 2006 - February 28, 2017.

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

** Indicates that the model belongs to the MCS at the 95% confidence level.

Table 4: Giacomini-White test

	HAR-IVL	HAR-IVO	HAR-IVG	HAR-IVLO	HAR-IVLG	HAR-IVOG	HAR-IVLOG		HAR-IVL	HAR-IVO	HAR-IVG	HAR-IVLO	HAR-IVLG	HAR-IVOG	HAR-IVLOG
S&P500								*DJIA*							
HAR-IV	0.028$^{(-)}$	0.018$^{(+)}$	0.000$^{(+)}$	0.047$^{(+)}$	0.395	0.011$^{(+)}$	0.027$^{(+)}$	HAR-IV	0.365	0.188	0.000$^{(+)}$	0.681	0.567	0.098$^{(+)}$	0.551
HAR-IVL		0.012$^{(+)}$	0.001$^{(+)}$	0.028$^{(+)}$	0.001$^{(+)}$	0.007$^{(+)}$	0.016$^{(+)}$	HAR-IVL		0.122	0.131	0.208	0.000$^{(+)}$	0.069$^{(+)}$	0.137
HAR-IVO			0.036$^{(-)}$	0.032$^{(-)}$	0.020$^{(-)}$	0.002$^{(+)}$	0.261	HAR-IVO			0.358	0.415	0.202	0.000$^{(+)}$	0.563
HAR-IVG				0.086$^{(+)}$	0.030$^{(-)}$	0.020$^{(+)}$	0.049$^{(+)}$	HAR-IVG				0.800	0.319	0.197	0.698
HAR-IVLO					0.049$^{(-)}$	0.006$^{(+)}$	0.001$^{(+)}$	HAR-IVLO					0.322	0.183	0.001$^{(+)}$
HAR-IVLG						0.012$^{(+)}$	0.027$^{(+)}$	HAR-IVLG						0.116	0.213
HAR-IVOG							0.052$^{(-)}$	HAR-IVOG							0.358
Nasdaq100								*STOXX50*							
HAR-IV	0.104	0.000$^{(+)}$	0.006$^{(-)}$	0.000$^{(+)}$	0.480	0.000$^{(+)}$	0.000$^{(+)}$	HAR-IV	0.540	0.000$^{(+)}$	0.078$^{(-)}$	0.000$^{(+)}$	0.304	0.015$^{(+)}$	0.008$^{(+)}$
HAR-IVL		0.000$^{(+)}$	0.053$^{(-)}$	0.000$^{(+)}$	0.009$^{(-)}$	0.000$^{(+)}$	0.000$^{(+)}$	HAR-IVL		0.001$^{(+)}$	0.109	0.000$^{(+)}$	0.106	0.051$^{(+)}$	0.017$^{(+)}$
HAR-IVO			0.000$^{(-)}$	0.032$^{(+)}$	0.000$^{(-)}$	0.038$^{(-)}$	0.037$^{(+)}$	HAR-IVO			0.000$^{(-)}$	0.338	0.000$^{(-)}$	0.067$^{(-)}$	0.067$^{(-)}$
HAR-IVG				0.000$^{(+)}$	0.183	0.000$^{(+)}$	0.000$^{(+)}$	HAR-IVG				0.000$^{(+)}$	0.410	0.000$^{(+)}$	0.000$^{(+)}$
HAR-IVLO					0.000$^{(-)}$	0.052$^{(-)}$	0.000$^{(-)}$	HAR-IVLO					0.000$^{(-)}$	0.074$^{(-)}$	0.080$^{(-)}$
HAR-IVLG						0.000$^{(+)}$	0.000$^{(+)}$	HAR-IVLG						0.000$^{(+)}$	0.000$^{(+)}$
HAR-IVOG							0.032$^{(+)}$	HAR-IVOG							0.384
CAC40								*DAX*							
HAR-IV	0.239	0.013$^{(+)}$	0.130	0.006$^{(+)}$	0.162	0.013$^{(+)}$	0.007$^{(+)}$	HAR-IV	0.208	0.001$^{(+)}$	0.708	0.000$^{(+)}$	0.128	0.000$^{(+)}$	0.000$^{(+)}$
HAR-IVL		0.031$^{(+)}$	0.191	0.013$^{(+)}$	0.480	0.033$^{(+)}$	0.014$^{(+)}$	HAR-IVL		0.002$^{(+)}$	0.285	0.001$^{(+)}$	0.362	0.001$^{(+)}$	0.001$^{(+)}$
HAR-IVO			0.010$^{(-)}$	0.053$^{(+)}$	0.025$^{(-)}$	0.145	0.069$^{(+)}$	HAR-IVO			0.001$^{(-)}$	0.194	0.002$^{(-)}$	0.610	0.041$^{(+)}$
HAR-IVG				0.005$^{(+)}$	0.143	0.010$^{(+)}$	0.005$^{(+)}$	HAR-IVG				0.000$^{(+)}$	0.169	0.000$^{(+)}$	0.000$^{(+)}$
HAR-IVLO					0.010$^{(-)}$	0.077$^{(-)}$	0.269	HAR-IVLO					0.001$^{(-)}$	0.436	0.116
HAR-IVLG						0.026$^{(+)}$	0.011$^{(+)}$	HAR-IVLG						0.001$^{(+)}$	0.000$^{(+)}$
HAR-IVOG							0.068$^{(+)}$	HAR-IVOG							0.131
AEX								*SMI*							
HAR-IV	0.102	0.004$^{(+)}$	0.545	0.001$^{(+)}$	0.119	0.004$^{(+)}$	0.001$^{(+)}$	HAR-IV	0.002$^{(+)}$	0.005$^{(+)}$	0.000$^{(-)}$	0.011$^{(+)}$	0.031$^{(+)}$	0.006$^{(+)}$	0.011$^{(+)}$
HAR-IVL		0.083$^{(+)}$	0.062$^{(-)}$	0.006$^{(+)}$	0.290	0.075$^{(+)}$	0.006$^{(+)}$	HAR-IVL		0.002$^{(+)}$	0.001$^{(-)}$	0.061$^{(+)}$	0.000$^{(-)}$	0.002$^{(+)}$	0.021$^{(+)}$
HAR-IVO			0.003$^{(-)}$	0.123	0.053$^{(-)}$	0.472	0.146	HAR-IVO			0.003$^{(-)}$	0.009$^{(+)}$	0.002$^{(-)}$	0.010$^{(-)}$	0.034$^{(+)}$
HAR-IVG				0.000$^{(+)}$	0.054$^{(+)}$	0.002$^{(+)}$	0.000$^{(+)}$	HAR-IVG				0.003$^{(+)}$	0.002$^{(+)}$	0.003$^{(+)}$	0.003$^{(+)}$
HAR-IVLO					0.003$^{(-)}$	0.072$^{(-)}$	0.580	HAR-IVLO					0.032$^{(-)}$	0.003$^{(-)}$	0.004$^{(-)}$
HAR-IVLG						0.048$^{(+)}$	0.003$^{(+)}$	HAR-IVLG						0.002$^{(+)}$	0.014$^{(+)}$
HAR-IVOG							0.081$^{(+)}$	HAR-IVOG							0.010$^{(+)}$
FTSE100								Nikkei225							
HAR-IV	0.355	0.460	0.002$^{(-)}$	0.466	0.500	0.024$^{(-)}$	0.549	HAR-IV	0.806	0.005$^{(+)}$	0.892	0.027$^{(+)}$	0.766	0.004$^{(+)}$	0.023$^{(+)}$
HAR-IVL		0.279	0.128	0.385	0.079$^{(-)}$	0.087$^{(-)}$	0.102	HAR-IVL		0.003$^{(+)}$	0.868	0.008$^{(+)}$	0.457	0.002$^{(+)}$	0.006$^{(+)}$
HAR-IVO			0.235	0.361	0.429	0.002$^{(-)}$	0.503	HAR-IVO			0.007$^{(-)}$	0.604	0.004$^{(-)}$	0.564	0.533
HAR-IVG				0.228	0.207	0.464	0.369	HAR-IVG				0.033$^{(+)}$	0.807	0.004$^{(+)}$	0.028$^{(+)}$
HAR-IVLO					0.553	0.121	0.085$^{(-)}$	HAR-IVLO					0.012$^{(-)}$	0.641	0.647
HAR-IVLG						0.140	0.448	HAR-IVLG						0.002$^{(+)}$	0.009$^{(+)}$
HAR-IVOG							0.196	HAR-IVOG							0.618

Note: The tables reports the p-values of the conditional Giacomini-White (Giacomini & White, 2006) test. The null hypothesis is that the row model and column model perform equally well is tested in terms of squared 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 eveluation perod covers January 1, 2006 - February 28, 2017.