STOCK INDEX REALIZED VOLATILITY FORECASTING IN THE PRESENCE OF HETEROGENEOUS LEVERAGE EFFECTS AND LONG RANGE DEPENDENCE IN THE VOLATILITY OF REALIZED VOLATILITY

DIMITRIOS P. LOUZIS1, SPYROS XANTHOPOULOS-SISINIS2 and APOSTOLOS P. REFENES3

Abstract

In this article, we account for the presence of heterogeneous leverage effects and the persistence in the volatility of stock index realized volatility. The Heterogeneous Autoregressive (HAR) realized volatility model is extended in order to account for asymmetric responses to negative and positive shocks occurring at distinct frequencies, as well as, for the long range dependence in the heteroscedastic variance of the residuals. Compared with established HAR and Autoregressive Fractionally Integrated Moving Average (ARFIMA) realized volatility models, the proposed model exhibits superior in-sample fitting, as well as, out-of-sample volatility forecasting performance. The latter is further improved when the realized power variation is used as a regressor, while we show that our analysis is also robust against microstructure noise.

JEL codes: C13; C22; C51; C53

Keywords: Volatility Forecasting; High Frequency Data; Leverage Effects.

1 Introduction

“A volatility model must be able to forecast volatility; this is the central requirement in almost all financial applications.”, Engle and Patton (2001). Indeed, everyday core business functions such as Basel II capital adequacy calculations, risk management, capital allocation, derivatives pricing and hedging, rely on accurate volatility estimation and forecasting. A plethora of volatility implementations have been proposed in the open literature, e.g. see Poon and Granger (2003) for a good review. In Andersen and Bollerslev (1998), the authors showed that the daily unobserved volatility could be adequately approximated by the sum of squared intraday returns, the so-called, realized volatility6. As evidence appeared that realized volatility possessed long memory, a number of researchers employed the Autoregressive Fractionally Integrated Moving Average (ARFIMA) specification for its modelling (e.g. see Andersen et al., 2003; Giot and Laurent, 2004; Koopman et al., 2005; Degiannakis, 2008; Angelidis and Degiannakis, 2008; Martens et al., 2009, Degiannakis and Floros, 2010).

An alternative implementation, based on the Heterogeneous Market Hypothesis and Muller’s et al. (1997) HARCH model, the Heterogeneous Autoregressive Realized Volatility model (HAR-RV, referred to as HAR henceforth) was also proposed by Corsi (2009). The HAR model utilized volatility components of different time resolutions in order to capture the long memory property of realized volatility in a more straightforward manner. Its tractable estimation and good volatility forecasting performance encouraged its use in several econometric studies e.g., see Andersen et al. (2007) on stock, exchange rate and bond price volatility forecasting, Forsberg and Ghysels (2007) and Martens et al. (2009) on volatility forecasting and Clements et al. (2008) on Value at Risk applications.

We contribute to this growing literature by introducing a logarithmic HAR model with asymmetries, or leverage effects7, modelled here as lagged standardized returns and absolute standardized returns (analogous to an EGARCH-type structure), occurring at distinct time horizons: daily, weekly and monthly. Moreover, in order to capture any remaining long range dependence in the volatility of realized volatility, we propose a Fractionally Integrated GARCH (FIGARCH) implementation for the conditional heteroscedasticity of the residuals,. We also apply the Realized Power Variation (RPV), proposed by Barndorff-Nielsen and Shephard (2004) as a regressor, which has been shown to be a robust to jumps, more persistent and accurate predictor of future volatility than realized volatility. As far as we are aware, this is the first time a HAR model with RPV regressors is combined with heterogeneous asymmetric effects and a FIGARCH specification for the residuals. Finally, the robustness of our findings to microstructure noise is assessed using the Two-Times Scale (TTS) volatility estimator of Zhang et al. (2005) which consistently estimates the integrated variance in the presence of microstructure noise.

The proposed model is initially estimated using two ten year data sets from the S&P 500 and DJIA stock indices. We find that against eight alternative HAR and ARFIMA models, the proposed model produces superior in-sample fitting revealing that not only

1 Financial Engineering Research Unit, Department of Management Science and Technology, Athens University of Economics and Business, 47A Evelpidon Str., 11362 Athens, Greece
2 Bank of Greece, Financial Stability Department, Market Risk Section, 3 Amerikis Str. GR 105 64, Athens, Greece. Note: the views expressed in this article, do not necessarily represent the views of the Bank of Greece
3 Financial Engineering Research Unit, Department of Management Science and Technology, Athens University of Economics and Business, 47A Evelpidon Str., 11362 Athens, Greece
4 E-mail: xanthopo@aeub.gr
5 Financial Engineering Research Unit, Department of Management Science and Technology, Athens University of Economics and Business, 47A Evelpidon Str., 11362 Athens, Greece
6 The authors would like to thank Prof. R. Tsay for his helpful comments.
7 “Bad news” in a stock market (i.e. negative returns) tend to increase future volatility more than “good news” (i.e. positive returns). This asymmetry between negative and positive returns is referred to as asymmetric or leverage effect. In theory, the leverage of the company increases as its stock price goes down, i.e. the company uses more debt than owned capital to finance its business activities. This increases the risk of investing in this stock which in turn increases its volatility.
past negative daily, but also weekly and monthly negative shocks yield a greater impact on current volatility than positive ones, suggesting a heterogeneous component structure in asymmetric effects. Moreover, an interesting contribution of past monthly positive shocks is also identified. Although the inclusion of leverage effects in the HAR regression reduces both the skewness and the heteroskedasticity of the error term, it does not eliminate the ARCH effects. Through Exact Local Whittle (ELW) and Maximum Likelihood Estimation (MLE) integration order estimations, the suspected long range dependence in the volatility of realized volatility is also verified.

The out-of-sample one day ahead, five and twenty-two days ahead forecasting performance is then evaluated for seven established loss functions, as well as with Hansen’s (2005) Superior Predictive Ability (SPA) test. The proposed specification minimizes the majority of the loss functions, for both indices and for all the forecasting horizons. Its volatility forecasting performance is further improved when the RPV is included as a regressor, while its superiority is also confirmed by the SPA test p-values. Finally, the TTS estimated realized volatility forecasting results underline its robustness against the microstructure noise in the returns process.

The remaining of this article is organized as follows: In Section 2 we introduce the realized volatility measures and the mathematical notations and definitions used throughout this article. In Section 3 we present the HAR based models. The data set, descriptive statistics and the in-sample maximum likelihood models estimation are shown in Section 4. In Section 5, we present the realized volatility forecasting evaluation methodology and results. Section 6 summarizes and concludes this article.

2 Realized volatility measures

In Andersen and Bollerslev (1998), the authors defined realized variance as the sum of squared intraday returns and proved that an unbiased and less noisy estimator for the daily unobserved volatility, than the squared daily returns proxy (for a good review on realized volatility see McAller and Medeiros, 2008). Let us define the mth intraday return for day t as

\[ r_{m,t} = 100 \times \left( \log(P_{m,t}) - \log(P_{m,t-1}) \right) \]

with \( m = 1, ..., M \), where M is the total number of intraday returns. The first and last price levels observed at day t are denoted as \( P_{0,t} \) and \( P_{M,t} \) respectively, while the overnight return, or “sleepage”, is defined as

\[ r_{0,j} = 100 \times \left( \log(P_{0,j}) - \log(P_{M,j}) \right) \]

Since these close-to-open price levels are often in practice quite different, the overnight returns could bias the realized variance estimation and hence the following scaling is applied:

\[ \sigma_{RV,j}^2 = \left( \sigma_{in}^2 + \sigma_{ov}^2 \right) / \sigma_{in}^2 \sum_{n=1}^{M} r_{m,n}^2 \]

and \( \sigma_{in}^2 \) and \( \sigma_{ov}^2 \) are the “open-to-close” and “close-to-open” sample variances respectively, which in turn are computed from

\[ \sigma_{in}^2 = \frac{1}{T} \sum_{t=1}^{T} \left( \log(P_{M,t}) - \log(P_{0,t}) \right)^2 \]

and

\[ \sigma_{ov}^2 = \frac{1}{T} \sum_{t=1}^{T} \left( \log(P_{0,t}) - \log(P_{M,t}) \right)^2 \]

(see Martens, 2002; Koopman et al., 2005 and Degiannakis, 2008). The realized volatility is simply the square root of the realized variance, i.e. \( \sigma_{RV} = \sqrt{\sigma_{RV,j}^2} \). Since realized volatility is an observable variable, standard time series techniques can be used for forecasting purposes.

The intraday sampling frequency used in this article is five minutes, which for liquid assets like the S&P 500 and the DJIA stock indices it has been found to be the highest sampling frequency with acceptable market microstructure bias (see Andersen et al., 2001a; Koopman et al., 2005; Corsi et al., 2008 and Degiannakis, 2008). Moreover, in order to verify the robustness of our findings in the presence of microstructure noise, we also calculate the realized variance with the Two-Times Scale volatility estimator presented below.

2.1 The Two-Times Scale (TTS) estimator

The TTS estimator (Zhang et al., 2005) utilizes two realized variance estimates, one calculated from low frequency sampled returns (e.g. fifteen minutes) and one calculated from higher frequency sampled returns (e.g. one minute). Then, by subsampling the return series, it reduces the variance of the low frequency realized variance.

Consider for example the fifteen minutes intervals 8:30–8:45, 8:45–9:00, ..., used for producing the fifteen minutes intraday returns. Similarly, the intervals 8:31–8:46, 8:46–9:01, ... or 8:32–8:47, 8:47–9:02, ... could also be used and so on. Hence, the full one minute returns grid can be partitioned into \( \lambda = 1, ..., \Lambda \) (here \( \Lambda = 15 \)) non-overlapping subgrids. If \( M_{(\lambda)} \) is the total number of intraday observations in each subsample set \( \lambda \), then the low frequency realized variance estimator for the subsample \( \lambda \) is given by:

\[ \sigma_{RV\{\lambda\}}^2 = \sum_{m=1}^{M_{(\lambda)}} r_{m,n}^2 \]

Likewise, the high frequency realized variance estimator is given by:

\[ \sigma_{RV\{-f\}j}^2 = \sum_{n=1}^{M_{(j)}} r_{m,n}^2 \]

where \( M_{(j)} \) is the number of observations when the higher sampling frequency is used, i.e. the number of observations in the full grid. Since the microstructure noise induced bias of the low frequency estimators is a function of the noise variance in the return processes, which is in turn consistently estimated by the high frequency realized variance estimator, we can use the latter in order to eliminate the low frequency estimator bias. The TTS estimator is then given by:

\[ \sigma_{RV-TTS,j}^2 = \frac{1}{\Lambda} \sum_{\lambda=1}^{\Lambda} \sigma_{RV\{\lambda\}}^2 - \frac{\overline{M}}{M_{(j)}} \sigma_{RV\{-f\}j}^2 \]

where the first summation is the average of the realized variance computed over the \( \lambda \) subsamples and \( \overline{M} = \frac{1}{\Lambda} \sum_{\lambda=1}^{\Lambda} M_{(\lambda)} = \frac{M_{(j)} - \Lambda + 1}{\Lambda} \) is the average number of observations in the subsamples. Here, the TTS volatility estimates are computed using a fifteen minute sampling interval as in Martens et al. (2009).
3 The Heterogeneous Autoregressive (HAR) models for realized volatility

Based on the Heterogeneous Market Hypothesis (Muller et al., 1993) and the HARЧ model (Muller et al., 1997), Corsi (2009) proposed an approximate long memory model for realized volatility, the Heterogeneous Autoregressive Realized Volatility model, denoted as HAR henceforth. The author suggested that a significant contributor to the market’s heterogeneity was the presence of three types of market agents with different time investment horizons: short (daily), medium (weekly) and long term (monthly) investment horizons. Short-term traders (such as hedge funds, FX and statistical arbitrage traders) typically adjust their market positions intraday, swiftly reacting on any relevant new information. Medium and long-term investors (such as commercial banks and pension funds) have longer holding periods and restructure their trading portfolios according to lower frequency information flow. Hence, the same informational content is distinctly assimilated across market participants, inducing different reaction times to the same market events. This asymmetry leads to a hierarchical structure of volatility components with distinguishable frequencies, where low (e.g. monthly) frequency volatility components should yield a greater impact on the overall volatility than high (e.g. daily) frequency volatility components. The economic rationale is that short-term investors interpret the level of long-term volatilities as predictions of future volatility and adjust their trading strategies accordingly, while short-term volatility is irrelevant to investors with longer holding periods. Corsi showed that by aggregating daily, weekly and monthly volatility components in an autoregressive structure, one could capture the heterogeneity of realized volatility, whilst approximating its long range dependence properties.

Here, in order to mitigate any positivity restrictions on the model’s parameters and error term (e.g. see Andersen et al., 2003), we will use the logarithm of the realized volatility in the HAR implementation:

\[ \ln r_{v, t}^{(d)} = a_0 + a_1 \ln r_{v, t-1}^{(d)} + a_2 \ln r_{v, t-1}^{(w)} + a_3 \ln r_{v, t-1}^{(m)} + u_t \] (1)

where \( \ln r_{v, t}^{(d)} \) is the daily log-realized volatility, \( \ln r_{v, t}^{(w)} \) is the weekly log-realized volatility and \( \ln r_{v, t}^{(m)} \) is the monthly log-realized volatility. The error term is denoted as \( u_t \).

To illustrate the model, Corsi and Reno (2009) also included past negative daily, weekly and monthly returns as regressors in the HAR model, aiming to capture the leverage effects in the volatility structure, one could capture the heterogeneity of realized volatility, whilst approximating its long range dependence properties. Hence, the same informational content is distinctly assimilated across market participants, inducing different reaction times to the same market events. This asymmetry leads to a hierarchical structure of volatility components with distinguishable frequencies, where low (e.g. monthly) frequency volatility components should yield a greater impact on the overall volatility than high (e.g. daily) frequency volatility components. The economic rationale is that short-term investors interpret the level of long-term volatilities as predictions of future volatility and adjust their trading strategies accordingly, while short-term volatility is irrelevant to investors with longer holding periods. Corsi showed that by aggregating daily, weekly and monthly volatility components in an autoregressive structure, one could capture the heterogeneity of realized volatility, whilst approximating its long range dependence properties.

In Corsi et al. (2008), the authors proposed a GARCH(\(p,q\)) error process in order to account for the time varying conditional heteroscedasticity of the normally distributed HAR errors, i.e. the so called “volatility of realized volatility”:

\[ u_t = \sigma_{v,t} \varepsilon_t \] (2)

\[ \varepsilon_t | I_{t-1} \sim N(0,1) \quad I_{t-1} \text{ is the information available until } t-1 \]

\[ \sigma_{v,t}^2 = \omega + \alpha(L)u_t^2 + \beta(L)\sigma_{v,t-1}^2 \] (3)

where \( \varepsilon_t | I_{t-1} \sim N(0,1) \quad I_{t-1} \text{ is the information available until } t-1 \quad L \text{ is the lag operator } \quad \alpha(L) = \alpha + \alpha_2 L^2 + \ldots + \alpha_\ell L^{\ell} \quad \beta(L) = \beta_1 L + \beta_2 L^2 + \ldots + \beta_q L^q \]

Corsi and Reno (2009) also included past negative daily, weekly and monthly returns as regressors in the HAR model, aiming to capture the leverage effects in the volatility process, plus a jump component. Finally, the authors in Andersen et al. (2007) proposed a HAR model with a jump component and found that the latter had restricted persistence compared with the continuous part of the quadratic variation, i.e. its contribution to forecasting volatility was limited.

3.1 The Asymmetric HAR-(Fi)GARCH models

In this article, we propose extending the HAR specification towards three directions. Firstly, we adopt a more flexible EGARCH-type structure for implementing the asymmetries in the volatility process. We expand the HAR model of Equation (1) in order to include standardized and absolute standardized returns aggregated over different time resolutions. Here, we include the complete returns dataset in the analysis, thus allowing for asymmetric responses to both negative and positive shocks. Secondly, through an FIGARCH specification, we account for the long memory of the residual’s variance in Equation (4). Finally, we use the Realized Power Variation (RPV) as a regressor, which has been shown to be robust to jumps and a more persistent and accurate predict

Initially, the asymmetric dynamics of past daily positive and negative returns are introduced. The Asymmetric (daily) HAR (hereafter ADHAR) model with daily leverage effects is defined as follows:

\[ \ln r_{v, t}^{(d)} = a_0 + a_1 \ln r_{v, t-1}^{(d)} + a_2 \ln r_{v, t-1}^{(w)} + a_3 \ln r_{v, t-1}^{(m)} + \beta_d(\varepsilon_{v,t}^{(d)2} + \varepsilon_{v,t-1}^{(d)2} + \varepsilon_{v,t-2}^{(d)2} + \ldots + \varepsilon_{v,t-\ell}^{(d)2}) + u_t \] (5)

where \( \varepsilon_{v,t}^{(d)} \) are the daily standardized returns. Equation (5) can be extended in order to account for the heterogeneity in asymmetric effects, i.e. asymmetric volatility reactions not only to past daily but also to weekly and monthly standardized returns. The Asymmetric HAR (AHAR) is given by:

\[ \ln r_{v, t}^{(d)} = a_0 + a_1 \ln r_{v, t-1}^{(d)} + a_2 \ln r_{v, t-1}^{(w)} + a_3 \ln r_{v, t-1}^{(m)} + \beta_d(\varepsilon_{v,t}^{(d)2} + \varepsilon_{v,t-1}^{(d)2} + \varepsilon_{v,t-2}^{(d)2} + \ldots + \varepsilon_{v,t-\ell}^{(d)2}) + u_t \] (6)

where \( \varepsilon_{v,t}^{(d)} = \sum_{i=1}^{h} r_{v,t-i} / \sqrt{\sum_{i=1}^{h} \sigma_{v,t-i}^{2}} \) are the daily (\( h = d = 1 \)), weekly (\( h = w = 5 \)) and monthly (\( h = m = 22 \)) standardized returns. The response of the logarithmic realized variance to past positive and negative standardized returns is given by:
The leverage effects are captured by the coefficient \( \theta_j \), which is expected to be negative and statistically different from zero, should past negative shocks yield a greater impact on future volatility.

Although accounting for leverage effects in Equation (6), may lead to some reduction in the skewness of the errors, the heteroscedasticity in the residuals is expected to remain due to the variance of the realized volatility estimator (Corsi et al., 2008). A straightforward approach, is to implement a GARCH\((p,q)\) error process to account for the conditional heteroscedasticity of the HAR residuals, in an AHAR-GARCH model.

We suspect however that the residuals could still retain the long memory property of realized volatility. Motivated by the findings of Beltratti and Morana (2005) (see Section IV), we propose to model the residuals with a Fractionally Integrated GARCH specification (FIGARCH\((m,d,q)\), see Baillie et al., 1996), implemented as:

\[
\sigma_{t,j}^2 = \omega + \beta(L)\sigma_{t,j}^2 + \left(1 - \beta(L) - \varphi(L)(1-L)^d\right)\varepsilon_t^2
\]

The FIGARCH model captures the long memory behavior of the variance process through the long memory, or fractional differencing parameter, \( d \), and is essentially an ARFIMA implementation of the squared residuals in Equation (6). For values of the differencing parameter \( d \) between 0 and 1, the autocorrelation of the volatility process exhibits a slow hyperbolic rate of decay and as the term \((1-L)^d\) in Equation (7) is an infinite summation, the FIGARCH obtains an infinite order specification:

\[
(1-L)^d = \sum_{k=0}^{\infty} \Gamma(d+k)\Gamma(d-k+1) L^k
\]

\[
= 1-d_{1}L-\frac{1}{2}d_{1}(1-d_{1})L^{2}-\frac{1}{2}d_{1}(1-d_{1})(2-d_{1})L^{3}-... \tag{8}
\]

where \( \Gamma(\cdot) \) denotes the gamma function. In practice, the above summation is truncated at 1000 lags, as suggested in Baillie et al. (1996).

3.2 The Realized Power Variation

Recently, the Realized Power Variation (RPV) proposed by Barndoff-Nielsen and Shephard (2004), has been found to produce superior realized volatility forecasts when implemented as a regressor in a HAR model. The RPV of order \( p \), is defined as:

\[
RPV(p) = \mu^p M \sum_{n=1}^{M} |\varepsilon_n|^p \tag{9}
\]

where \( 0 < p < 2 \), \( \mu = E[\varepsilon]^p = 2^{p/2} \Gamma(1/2)(p+1)/\Gamma(1/2) \) with \( \varepsilon \overset{iid}{\sim} N(0,1) \). For values of \( p \) between 0 and 2 and as \( M \to \infty \) it holds that: \( RPV(p) \to IPV = \int_0^{\infty} \sigma^p(s)ds \), where IPV denotes the integrated power variation.

Note that when \( p = 2 \), the RPV is by definition equal to the realized volatility defined earlier (i.e. \( RPV(2) = \sigma_{\text{RV},t}^2 \)).

Forsberg and Ghysels (2007), Ghysels et al. (2006) and Ghysels and Sinko (2006) demonstrated the ability of realized absolute variation, i.e. RPV(1), to produce superior volatility forecasts compared to the squared return volatility measures. They argued that the RPV is a better predictor of realized volatility because of its robustness to jumps, its smaller sampling error and its improved predictability. In Liu and Maheu (2009) and Fuentes et al. (2009), the authors showed that an RPV of order other than one can significantly improve the accuracy of volatility forecasts. Here, following Liu and Maheu (2009), we use an RPV of order 1.5 as a regressor in the HAR models presented above. Hence, the simple HAR-RPV model is defined as:

\[
\text{IPV}_t = \alpha_0 + \alpha_1 \text{IPV}_{t-1} + \alpha_2 \text{IPV}_{t-2} + \alpha_3 \text{IPV}_{t-3} + \alpha_4 \text{IPV}_{t-4} + \varepsilon_t
\]

where \( \text{IPV}_t = \log(\sigma_{\text{RV},t}^2) \) is the daily logarithm and \( \text{IPV}_{t,J} = (1/h)[\text{IPV}_{t} + \text{IPV}_{t-1} + \text{IPV}_{t-2} + ... + \text{IPV}_{t-h+1}] \) with \( h = w = 5 \) and \( h = m = 22 \) being the weekly and monthly RPV components respectively. The other HAR models are analogously defined.

4 The data set, descriptive statistics and model estimation

The data set was obtained from Tick Data and consists of five minutes previous tick interpolated prices, \( \{P_{m,t}\} \), for the S&P 500 and the DJIA cash indices over a ten year period, from 1.1.1997 to 12.31.2006. After adjustments for holidays and half-holidays, there were \( T = 2,508 \) trading days per index, with six and a half trading hours per day, interpreted as \( M = 78 \) intraday returns. Each full data set, was divided into \( T' = 1,508 \) in-sample observations, from 31.01.1997 to 30.12.2002 and \( n = T - T' = 1,000 \) out-of-sample observations.

The descriptive statistics for the daily logarithmic returns, daily standardized returns, realized variance and logarithmic realized variance for the two full data sets are shown in Table 1. Both original return series have negative skewness and fat tails, a departure from normality which can be attributed to mainly negative price shocks near the end of 1997 and 1998, all through 2000 and towards the end of 2002. The skewness and kurtosis of the standardized returns and of the logarithmic variance series suggest that the

---

\(^a\) In this case, the RPV is not robust to jumps and converges to the integrated volatility plus the jump component.

\(^b\) The daily logarithmic returns are calculated as \( r_t = 100 \times \left( \log(P_{m,t}) - \log(P_{m,t-1}) \right) \) where \( P_{m,t}, \ P_{m,t-1} \) is the closing price of day \( t, (t-1) \).
respective distributions are approximately normal. The Lilliefors test also yields evidence in favor of the null for the standardized returns and the logarithmic realized variance of the S&P 500 index and the standardized returns of the DJIA index. However, the Jarque Bera (JB) and Anderson Darling tests, reject the gaussianity assumption for all series. These conflicting results suggest that the standardized returns and the logarithmic realized variance distributions are not perfectly Gaussian, a conclusion similar as in Andersen et al. (2001a) and Andersen et al. (2001b).

Table 1: Descriptive statistics and stylized facts for the S&P 500 and DJIA indices

<table>
<thead>
<tr>
<th></th>
<th>S&amp;P 500</th>
<th>DJIA</th>
</tr>
</thead>
<tbody>
<tr>
<td>( r_i )</td>
<td>0.026</td>
<td>0.049</td>
</tr>
<tr>
<td>( r_i / \sigma_{RV,i} )</td>
<td>0.088</td>
<td>0.083</td>
</tr>
<tr>
<td>( \sigma^2_{RV,i} )</td>
<td>0.971</td>
<td>0.622</td>
</tr>
<tr>
<td>( \log(\sigma^2_{RV,i}) )</td>
<td>-0.433</td>
<td>-0.474</td>
</tr>
<tr>
<td>( r_i )</td>
<td>0.026</td>
<td>0.033</td>
</tr>
<tr>
<td>( r_i / \sigma_{RV,i} )</td>
<td>0.074</td>
<td>0.045</td>
</tr>
<tr>
<td>( \sigma^2_{RV,i} )</td>
<td>1.053</td>
<td>0.688</td>
</tr>
<tr>
<td>( \log(\sigma^2_{RV,i}) )</td>
<td>-0.341</td>
<td>-0.372</td>
</tr>
</tbody>
</table>

Normality tests

<table>
<thead>
<tr>
<th>Test</th>
<th>S&amp;P 500</th>
<th>DJIA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jarque-Bera</td>
<td>952</td>
<td>1145</td>
</tr>
<tr>
<td>[p-value]</td>
<td>[0.000]</td>
<td>[0.000]</td>
</tr>
<tr>
<td>Lilliefors</td>
<td>0.0493</td>
<td>0.069</td>
</tr>
<tr>
<td>[p-value]</td>
<td>[0.000]</td>
<td>[0.000]</td>
</tr>
<tr>
<td>Anderson-Darling</td>
<td>14.529</td>
<td>1.754</td>
</tr>
<tr>
<td>[p-value]</td>
<td>[0.000]</td>
<td>[0.000]</td>
</tr>
</tbody>
</table>

The fractional integration order of the logarithmic realised variance series is calculated using the Exact Local Whittle (ELW) estimator proposed by Shimotsu and Phillips (2005). The estimator relies on the frequency domain representation of the observed series, as expressed by its Discrete Fourier Transform (DFT) and evaluated at \( m \) Fourier frequencies from the spectrum’s origin. A widely adopted functional form for \( m = g(T) = T^\mu \) with \( 0 < \mu < 1 \), while the empirical evidence suggests that values for \( \mu \) in the interval \( [0.5, 0.6] \) limit the bias and variance of the integration order estimate. The resulting estimator is consistent and asymptotically distributed as \( N(0, \frac{1}{4}) \) for \( d_{RV} \), as long as the difference between the upper and lower bound of the search interval for \( d_{RV} \) is smaller than 4.5.

The log realized variance ELW integration order estimates for both indices are shown in Fig. 1 for \( \mu \in [0.4, 0.7] \). The \( d_{RV} \) estimates vary between 0.55 and 0.6, suggesting that the realized variance follows a covariance non-stationary fractionally integrated process.

Finally, the descriptive statistics and stylized facts for the two indices for the TTS estimated realized variance indicate no significant departure from the aforementioned squared intraday returns observations and are available upon request.

Fig. 1. S&P 500 and DJIA indices log realized variance ELW fractional differencing parameter estimates

4.1 Estimation of the HAR based models

Before proceeding with the parameters estimation\(^{10} \), the optimum lag order for the AHAR-FIGARCH model was first determined. The lag structure combinations which minimized the AIC and SIC criteria were an AHAR-FIGARCH\((1,d_{RV},1)\) and an AHAR-FIGARCH\((0,d_{RV},0)\) for the S&P 500 and DJIA indices respectively. The coefficient estimates for all the HAR based models, as well as the respective in-sample diagnostics are summarized in Table 2 for the S&P 500 index (estimations results for the DJIA are available upon request). The discussion of the estimation results in terms of the volatility components, the leverage effects and the presence of long memory in the residuals is next presented.

\(^{10}\text{All estimates were deduced by numerical optimization of the log likelihood function (Maximum Likelihood Estimation, MLE) and they were conducted with the Ox Metrics G@RCH 4.2 package developed by Laurent and Peters (2002).} \)
Table 2: The HAR based realized volatility models estimation results for the S&P 500 index

<table>
<thead>
<tr>
<th></th>
<th>HAR</th>
<th>HAR-GARCH</th>
<th>ADHAR</th>
<th>AHAR</th>
<th>AHAR-GARCH</th>
<th>FIGARCH</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a_0)</td>
<td>-0.006</td>
<td>-0.008</td>
<td>-0.074*</td>
<td>-0.174*</td>
<td>-0.177*</td>
<td>-0.178*</td>
</tr>
<tr>
<td></td>
<td>(0.013)</td>
<td>(0.013)</td>
<td>(0.022)</td>
<td>(0.032)</td>
<td>(0.031)</td>
<td>(0.031)</td>
</tr>
<tr>
<td>(a_{(d)})</td>
<td>0.301*</td>
<td>0.293*</td>
<td>0.213*</td>
<td>0.140*</td>
<td>0.139*</td>
<td>0.144*</td>
</tr>
<tr>
<td></td>
<td>(0.030)</td>
<td>(0.035)</td>
<td>(0.032)</td>
<td>(0.032)</td>
<td>(0.031)</td>
<td>(0.032)</td>
</tr>
<tr>
<td>(a_{(w)})</td>
<td>0.422*</td>
<td>0.454*</td>
<td>0.507*</td>
<td>0.363*</td>
<td>0.381*</td>
<td>0.375*</td>
</tr>
<tr>
<td></td>
<td>(0.050)</td>
<td>(0.052)</td>
<td>(0.050)</td>
<td>(0.053)</td>
<td>(0.053)</td>
<td>(0.052)</td>
</tr>
<tr>
<td>(a_{(m)})</td>
<td>0.170*</td>
<td>0.146*</td>
<td>0.171*</td>
<td>0.352*</td>
<td>0.344*</td>
<td>0.345*</td>
</tr>
<tr>
<td></td>
<td>(0.044)</td>
<td>(0.044)</td>
<td>(0.042)</td>
<td>(0.046)</td>
<td>(0.045)</td>
<td>(0.045)</td>
</tr>
</tbody>
</table>

\(\gamma_{(d)}\) --- --- -0.137* -0.089* -0.083* -0.083*
\(\gamma_{(w)}\) --- --- --- -0.106* -0.104* -0.103*
\(\gamma_{(m)}\) --- --- --- -0.056* -0.057* -0.058*

\(\omega\) 0.280* 0.030 0.255* 0.240* 0.005 0.017
\(d_u\) --- --- --- --- --- 0.222*

\(\alpha_i\) --- 0.057** --- --- 0.021** ---
\(\beta_i\) --- 0.832* --- --- 0.955* 0.736*
\(\varphi_i\) --- --- --- --- --- 0.570* (0.085)

<table>
<thead>
<tr>
<th></th>
<th>LogL</th>
<th>AIC</th>
<th>SIC</th>
<th>Skewness</th>
<th>Excess Kurt.</th>
<th>JB</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-1.163</td>
<td>1.573</td>
<td>1.591</td>
<td>0.333</td>
<td>0.529</td>
<td>45</td>
</tr>
<tr>
<td></td>
<td>-1.153</td>
<td>1.563</td>
<td>1.588</td>
<td>0.309</td>
<td>0.559</td>
<td>43</td>
</tr>
<tr>
<td></td>
<td>-1.095</td>
<td>1.483</td>
<td>1.508</td>
<td>0.307</td>
<td>0.610</td>
<td>51</td>
</tr>
<tr>
<td></td>
<td>-1.049</td>
<td>1.427</td>
<td>1.466</td>
<td>0.308</td>
<td>0.766</td>
<td>59</td>
</tr>
<tr>
<td></td>
<td>-1.042</td>
<td>1.420</td>
<td>1.467</td>
<td>0.306</td>
<td>0.828</td>
<td>65</td>
</tr>
<tr>
<td></td>
<td>-1.040</td>
<td>1.419</td>
<td>1.469</td>
<td>0.294</td>
<td>0.784</td>
<td>59</td>
</tr>
</tbody>
</table>

\(Q(5)\) [0.560] [0.657] [0.095] [0.748] [0.829] [0.837]
\(Q(50)\) [0.244] [0.277] [0.182] [0.424] [0.229] [0.507]
\(Q^2(5)\) [0.001] [0.259] [0.031] [0.319] [0.392] [0.794]
\(Q^2(50)\) [0.212] [0.800] [0.657] [0.860] [0.976] [0.978]
\(ARCH-LM(2)\) [0.0007] [0.4438] [0.005] [0.098] [0.257] [0.800]
\(ARCH-LM(5)\) [0.0055] [0.6079] [0.057] [0.416] [0.752] [0.961]

Notes: Standard errors are presented in parentheses. * and ** indicate statistical significance at 1%, and 5% significance levels respectively. LogL is the optimized value of the log likelihood. A(SIC) is the Akaike (Schwartz) Information Criterion. \(Q(h)\) and \(Q^2(h)\) are the Ljung-Box statistics for \(h\)th order serial correlation for the standardized and squared standardized residuals respectively. The p-values for the Jarque Bera (JB), \(Q(\cdot)\), \(Q^2(\cdot)\) and the ALCH-LM(\(\cdot\)) test statistics are depicted in brackets.

4.1.1 The volatility components

For the HAR, HAR-GARCH and ADHAR models, the coefficient of the lagged weekly volatility component, \(a_{(w)}\), bears the greatest impact on current volatility, followed by the daily volatility component, while the monthly one influences the total volatility significantly less. Hence, the day-ahead volatility appears to be the aggregate effect of short and mostly medium term volatility components and much less of longer term volatility factors. However, when the heterogeneity of the leverage effects is taken into consideration in the AHAR, AHAR-GARCH and AHAR-FIGARCH models, the balance shifts drastically. The contribution of the lagged daily volatility component is approximately halved, that of the weekly one is slightly moderated, while the contribution of the monthly volatility component more than doubles. Now, past longer horizon volatility events appear to primarily shape the indices future volatility, a conclusion which is in agreement with the economic rationale laid out in Section 3.
4.1.2 The leverage effects

For both indices, accounting for daily leverage effects in the AdHAR model leads to a reduction in the residuals’ skewness and a significant improvement in the goodness of fit indicators. However, there is still evidence of ARCH effects in the residuals, as the ARCH-LM tests suggest. For the DJIA index, the inclusion of weekly and monthly standardized returns in the AHAR model reduces the skewness of the errors, while there is also evidence in favor of the rejection of the ARCH effects hypothesis in the DJIA residuals. This is also reflected in the GARCH coefficients estimates of the AHAR-GARCH model for the DJIA index, where none of them are statistical significant at a 5% significance level.

The coefficients of the lagged daily, weekly and monthly standardized returns, \( \delta_{t} \), in the AHAR, AHAR-GARCH and AHAR-FIGARCH models are all statistically significant at a 1% significance level, confirming that future market volatility will react asymmetrically not only to yesterday’s negative returns, but also to past weekly and monthly returns. Their negative weighting also suggests that past negative shocks induce more market volatility than past positive ones.

The latter is clearly depicted in Fig. 2, where the impact of past daily, weekly and monthly shocks on future realized volatility is shown. The slope of the impact on logarithmic volatility is equal to \( \tilde{\delta}_{t} - \tilde{\gamma}_{t} \) when standardized returns are negative, i.e. \( z_{t}^{(i)} < 0 \), and \( \tilde{\delta}_{t} + \tilde{\gamma}_{t} \) when standardized returns are positive, i.e. \( z_{t}^{(i)} > 0 \), where \( \tilde{\delta}_{t} \) and \( \tilde{\gamma}_{t} \) are the estimates of \( \delta_{t} \) and \( \gamma_{t} \), respectively. It is clear that past negative return events, irrespective of the time horizon, subscribe to future volatility variations more than positive ones. The volatility contribution hierarchy is analogous to that of the volatility components, with the weekly standardized return being the prevailing contributor to the overall volatility, followed by the monthly one. However, it is interesting to note that past positive monthly shocks will also tend to increase volatility, while only past positive daily and weekly shocks will have a negative impact on volatility. As far as we are aware, this is a novel finding, underlining the importance of including in the analysis the complete returns dataset and not just past negative returns.

Fig. 2. The impact on volatility from lagged negative and positive daily, weekly and monthly standardized returns for the S&P 500 and DJIA indices

4.1.3 Long range dependence in the residuals

The presence of long memory in the residuals’ variance is depicted in Fig. 3, where the ELW fractional differencing parameter estimates, \( d_{\mu} \), for the HAR model squared residuals are shown for \( \mu \in [0.4, 0.7] \). For both indices, the ELW \( d_{\mu} \) estimates are always less than 0.5 and statistically significant, thus confirming that not only realized volatility, but also the “volatility of realized volatility” is autocorrelated for longer time periods. The presence of a covariance stationary long memory process in the HAR residuals confirms the suitability of the proposed FIGARCH implementation for their modelling. The AHAR-FIGARCH residuals \( d_{\mu} \) MLE estimates shown in Table 2 are very close to the respective HAR squared residuals \( d_{\mu} \) ELW estimates, as expected.

Fig. 3. S&P 500 and DJIA indices HAR squared residuals ELW fractional differencing parameter estimates
Overall, the proposed AHAR-FIGARCH model yields the best data fit, as measured by the AIC and SIC criteria, reducing also the skewness of the residuals. Nonetheless, the excess kurtosis values suggest that the use of a more fat tailed distribution than the normal, might have been more appropriate.

Finally, for both indices, the TTS realized volatility estimation results indicate no qualitative difference confirming the robustness of our findings in the presence of microstructure noise (estimation results are available upon request).

5 Realized volatility forecasting and evaluation

In order to evaluate the realized volatility forecasting performance, a rolling window of $T^*$ observations, $\left[1+i:(T^* + i)\right]_{i=0}^{n-1}$ was used to re-estimate the models and produce $n$ out-of-sample day-ahead realized volatility forecasts calculated as:

$$E\left(\sigma^j_{RV,t/l_i}^2 / l_i \right) = \hat{\sigma}^j_{RV,t/l_i} = \left(\exp\left(l\hat{\nu}^j_{t/l_i} + 0.5\sigma^j_{RV,t/l_i}\right)\right)_{1/2}$$

where $l\hat{\nu}^j_{t/l_i}$ is the day-ahead logarithmic realized variance forecast and $\sigma^j_{RV,t/l_i}$ is the model $j$ residuals volatility$^{11}$, while $n=1,000, 996$ and $979$ observations for the day ahead, five days and twenty-two days ahead realized volatility forecasts respectively, spanning from the 31$^{st}$ December 2002 to the 29$^{th}$ December 2006. For the five and twenty two days ahead realized volatility forecasts, the corresponding realized volatilities were computed as the square root of the sum of daily realized variances over each forecasting period.

5.1 Realized volatility forecasting evaluation

In order to evaluate the model’s out-of-sample realized volatility forecasting performance over the three forecasting horizons, we used the seven loss functions shown in Table 3 below.

<table>
<thead>
<tr>
<th>Table 3: The loss functions</th>
</tr>
</thead>
<tbody>
<tr>
<td>$MSE = (1/n) \sum_{i=1}^{n} \left( \sigma_{RV,j} - \hat{\sigma}^j_{RV,t/l_i} \right)^2$</td>
</tr>
<tr>
<td>$MAE = (1/n) \sum_{i=1}^{n} \left</td>
</tr>
<tr>
<td>$MAPE = (1/n) \sum_{i=1}^{n} \left( \sigma_{RV,j} - \hat{\sigma}^j_{RV,t/l_i} \right)$</td>
</tr>
<tr>
<td>$MLAE = (1/n) \sum_{i=1}^{n} \left( \log \left( \sigma_{RV,j} - \hat{\sigma}^j_{RV,t/l_i} \right) \right)$</td>
</tr>
<tr>
<td>$QLIKE = (1/n) \sum_{i=1}^{n} \left( \log \left( \hat{\sigma}^j_{RV,t/l_i} / \hat{\sigma}^j_{RV,t/l_i} \right) + \sigma_{RV,j} \right)$</td>
</tr>
<tr>
<td>$R2LOG = (1/n) \sum_{i=1}^{n} \left( \log \left( \sigma_{RV,j} / \hat{\sigma}^j_{RV,t/l_i} \right) \right)^2$</td>
</tr>
<tr>
<td>$HMSE = (1/n) \sum_{i=1}^{n} \left( \left( \sigma_{RV,j} / \hat{\sigma}^j_{RV,t/l_i} \right) - 1 \right)^2$</td>
</tr>
</tbody>
</table>

MSE and MAE stand for the Mean Square Error and Mean Absolute Error standard loss functions respectively. MAPE is the Mean Absolute Percentage Error and MLAE is the Mean Logarithm of Absolute Errors. The QLIKE criterion proposed by Bollerslev, Engle and Nelson (1994) is the loss implied by a Gaussian likelihood. The R2LOG loss function of Pagan and Schwert (1990) is equivalent to the MSE, but for the logarithm of realized volatility. This loss function applies a greater penalty when forecasting errors occur in low volatility periods, than when they occur in high volatility periods. Using the MSE (R2LOG) criteria is equivalent to testing whether the R-square, $R^2$, of the Mincer-Zarnowitz regressions between the (logarithmic) realized volatility and the forecasted volatility is equivalent to testing whether the $R^2$ of the variance of the benchmark model, is always greater than that of its counterparts (Marcucci, 2005). Finally, the HMSE is the Heteroscedastic Mean Squared Error proposed by Bollerslev and Ghysels (1996).

The predictive ability of the realized volatility models was also assessed via Hansen’s (2005) Superior Predictive Ability (SPA) test. The SPA test relies on a predetermined loss function in order to test whether the null hypothesis that the benchmark model is not outperformed by any of its competitors, is rejected or not. The forecasting performance of the benchmark model, model 0, with respect to model $k$ is deduced from: $f_{i,k} = L_{0,j} - L_{k,j}, \ k = 1..J, t = 1..N$, where $L_{s,j} = L\left( \sigma_{RV,j}, \hat{\sigma}^j_{RV,t/l_i} \right), \ j = 0, k,$ is the predetermined forecast loss function of the benchmark model and of model $k$ respectively. Under the null hypothesis and assuming stationarity for $f_{i,k},$ we expect that on average the forecasting loss function of the benchmark model will be smaller, or at least equal to that of model $k$. Thus, the null hypothesis can be stated as: $H_0: \max_{k=1..J} \mu_k = E\left( f_{i,k} \right) \leq 0$ and can be tested through the following test statistic: $T^{SPA}_n = \max_{k=1..J} \left\{ \sqrt{n} f_{i,k} / \sqrt{\var\left( \sqrt{n} f_{i,k} \right)} \right\}$, where $f_{i,k} = (1/n) \sum_{t=i-1}^{n} f_{i,k}$ and $\var\left( \sqrt{n} f_{i,k} \right)$ is the variance of $\sqrt{n} f_{i,k}$. Both $\var\left( \sqrt{n} f_{i,k} \right)$ and the test statistic $p$-values are consistently estimated via stationary bootstrapping as in Politis and Romano (1994).

$^{11}$ The transformation in Equation (10) is derived from the realized variance lognormality assumption: A random variable $x_i$ is lognormally distributed if $y_i = \log x_i$ is normally distributed. Then, the expectation of $y_i$ is $E(y_i) = \exp\left( \mu + 0.5\sigma^2 \right)$, with $\mu$ and $\sigma^2$ denoting the mean and the variance of $y_i$, respectively, e.g. see Beltratti and Morana (2005) and Giot and Laurent (2004).
The SPA test analysis focused only on the MSE and QLIKE loss functions as these two measures have been shown to be robust against volatility proxy noise (see Patton, 2006). Since realized volatility is a proxy for the true unobservable volatility, the aforementioned two loss functions can yield consistent model rankings, without negating however the informative power of the other loss functions in Table 3, as they were put to use above. Finally and only for completeness purposes, we included in the analysis a GARCH model, the Fractionally Integrated Exponential GARCH (FIEGARCH) model (see Bollerslev and Mikkelsen, 1996), which accounts for both leverage effects and long memory.

5.2 The models’ loss function performance

In Table 4, the Table 3 loss functions results, as well as each model’s relative performance rankings (in parenthesis) are shown for the S&P 500 index (forecasting results for the DJIA are available upon request). In this study, we also include some relevant long memory ARFIMA models, in order to provide a straightforward comparison to the HAR based model proposed above (description and estimation results for the ARFIMA models are available upon request). Across all forecasting horizons, the proposed AHAR-RPV-FIGARCH model nearly always ranks first amongst the alternative models, minimizing the respective loss functions, with the exception of the DJIA index twenty-two days ahead forecast where the AHAR-RPV-GARCH model lists first. Note, that for the DJIA index where the persistence in the residuals’ variance is smaller than the one for the S&P 500 index (e.g. see the Tables 2 and Fig. 1), the benefit of implementing an explicit long memory volatility specification for the residuals is moderated, especially for longer term forecasts. The AHAR-RPV-GARCH model otherwise typically ranks second for the one and five days ahead forecasting horizons, followed by the AHAR-FIGARCH and the AHAR-GARCH models. For the S&P 500 index and for the twenty-two days ahead horizon, the AHAR-FIGARCH ranks second (first for the QLIKE and HMSE loss functions), followed by the AHAR-RPV-GARCH and AHAR-GARCH models.

As for the rest HAR and ARFIMA model variations, the HAR models with leverage effects (AdHAR, AHAR) typically outperform the more advanced ARFIMA models (ARFIMA-FIGARCH and ARFIMAX-FIGARCH). However, for the S&P500 index twenty-two days ahead forecasts, the aforementioned performance ranking is reversed. In turn, the AHAR model consistently outperforms the AdHAR one, underlining again the importance of considering the heterogeneity in the leverage effects.

It is clear nonetheless that even though the inclusion of daily, weekly and monthly (absolute) standardized returns in the HAR regression reduces the heteroscedasticity in the residuals (see the ARCH-LM tests in Table 2), a GARCH, or significantly more so, a FIGARCH implementation for the residuals invariably enhances the forecasting performance. Finally, the ARFIMA model performs, par from a few exceptions, better than the HAR-GARCH model. The basic HAR model is the worst performer shown here, but not the worst overall, as it is always better than the FIEGARCH model, which is not shown here due to space considerations (the results are available from the authors upon request).

### Table 3: Forecast loss functions for the S&P 500 index

<table>
<thead>
<tr>
<th></th>
<th>1 day ahead</th>
<th>5 days ahead</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MSE</td>
<td>MAE</td>
</tr>
<tr>
<td>HAR</td>
<td>0.02754 (11)</td>
<td>0.13093 (11)</td>
</tr>
<tr>
<td>HAR-GARCH</td>
<td>0.02698 (8)</td>
<td>0.12915 (9)</td>
</tr>
<tr>
<td>AdHAR</td>
<td>0.02567 (6)</td>
<td>0.12470 (6)</td>
</tr>
<tr>
<td>AHAR</td>
<td>0.02471 (5)</td>
<td>0.12124 (5)</td>
</tr>
<tr>
<td>AHAR-GARCH</td>
<td>0.02436 (4)</td>
<td>0.12035 (4)</td>
</tr>
<tr>
<td>AHAR-FIGARCH</td>
<td>0.02429 (3)</td>
<td>0.12011 (3)</td>
</tr>
<tr>
<td>AHAR-RPV-GARCH</td>
<td>0.02418 (1)</td>
<td>0.11979 (1)</td>
</tr>
<tr>
<td>AHAR-RPV-FIGARCH</td>
<td>0.02741 (10)</td>
<td>0.13019 (10)</td>
</tr>
<tr>
<td>ARFIMA</td>
<td>0.02679 (7)</td>
<td>0.12813 (8)</td>
</tr>
<tr>
<td>ARFIMA-FIGARCH</td>
<td>0.02710 (9)</td>
<td>0.12676 (7)</td>
</tr>
<tr>
<td>ARFIMAX-FIGARCH</td>
<td>0.02710 (9)</td>
<td>0.12676 (7)</td>
</tr>
</tbody>
</table>

12 The application of ARFIMA models for realized volatility modelling and forecasting purposes was first proposed by Andersen et al. (2001a) and Andersen et al. (2003), based on the analysis of Granger (1980) and Granger and Joyeux (1980).
From the results synopsis presented above, the following conclusions can be drawn: Firstly, regardless if an RPV regressor is used or not, accounting for asymmetric effects is as important as accounting for their heterogeneity: the AHAR model always outperforms the AdHAR one and the latter consistently outperforms the HAR. Secondly, as expected, we confirm that considering the conditional heteroscedasticity of the realized volatility residuals is essential in volatility modelling and forecasting applications, even with the presence of asymmetric effects in the volatility equation. However, it is now evident that the heteroscedasticity is better accounted for with a FIGARCH implementation which captures the long memory of the variance residuals identified in Section V. All models with residuals FIGARCH specifications outperform the respective models with GARCH ones, according to almost all the loss functions, irrespective if it is a HAR or ARFIMA model, the index or time horizon (except from the twenty-two day ahead DJIA index forecast), with an RPV regressor implementation or not. Finally, the RPV is a better predictor of realized volatility than the squared returns measure, significantly improving the volatility forecasting performance when added as a regressor. Both RPV implementations almost invariably outperform their squared returns counterparts.

### 5.2.1 The TTS volatility forecast rankings

The out-of-sample volatility forecasting analysis for the HAR based models is once again evaluated using the TTS realized volatility estimates (forecasting results with the TTS estimator are available upon request). For both indices and across all forecasting horizons, the AHAR-FIGARCH model outperforms all the other models as it minimizes almost all the loss functions, with the exception of the MALE criterion for the S&P 500 index for the one and twenty-two days ahead forecasts. These results confirm that the aforementioned findings are robust to the realized volatility calculation bias which is induced by microstructure noise in the return process. However, in tests run by the authors but not shown here, we noted that the inclusion of the RPV regressor does not significantly improve the model’s forecasting ability, when the dependent variable is the TTS realized volatility.

### 5.3 The SPA test results

The SPA test p-values for the MSE and QLIKE loss functions are shown in Table 4 and align with the aforementioned findings. When the AHAR-RPV-FIGARCH or the AHAR-RPV-GARCH model is the benchmark model, the null hypothesis of superior performance is strongly accepted (at a 10% significance level), for both indices and both loss functions, for the short and mid term forecasting horizons (one and five days ahead). When an alternative realized volatility model is chosen as the benchmark model, then the null hypothesis is rejected, implying that another model, or models, produces statistically significant better forecasts.

For the twenty-two days ahead forecast horizon, the null hypothesis is also accepted for the ARFIMA-FIGARCH and ARFIMAX-FIGARCH models, indicating that for longer term forecasts the ARFIMA models are competitive to the HAR models, potentially benefitting by their genuine long memory structure. Overall, the asymmetric RPV models exhibit the best out-of-sample forecasting performance, yielding for both functions the highest p-values for almost all the forecast horizons.

<table>
<thead>
<tr>
<th></th>
<th>S&amp;P 500</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1 day ahead</td>
<td>5 days ahead</td>
<td>22 days ahead</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>MSE</td>
<td>QLIKE</td>
<td>MSE</td>
<td>QLIKE</td>
<td>MSE</td>
<td>QLIKE</td>
</tr>
<tr>
<td>AHAR-FIGARCH</td>
<td>0.33087</td>
<td>0.43052</td>
<td>0.13882</td>
<td>0.13186</td>
<td>0.02946</td>
<td>0.02866</td>
</tr>
<tr>
<td>AHAR-RPV-FIGARCH</td>
<td>0.33225</td>
<td>0.43173</td>
<td>0.13925</td>
<td>0.13209</td>
<td>0.02957</td>
<td>0.02872</td>
</tr>
</tbody>
</table>

Table 4: Superior Predictive Ability (SPA) test p-values
6 Conclusions

Building on the HAR realized volatility model proposed by Corsi (2009), we captured the leverage effects in the volatility process using (absolute) standardized returns of daily, weekly and monthly frequencies. The HAR error’s heteroscedasticity and long memory was also accounted for with a FIGARCH implementation. Moreover, we introduced the RPV as a regressor, which has been shown to be robust to jumps, has a smaller sampling error and is more predictable. In order to examine the robustness of our findings to microstructure noise, we also calculated the realized variance with a TTS estimator and then re-evaluated our models.

The proposed AHAR-FIGARCH model produced the best in-sample fitting against the alternative HAR and ARFIMA based realized volatility models. The estimation results confirmed the appropriateness of our modelling as heterogeneity in the asymmetric effects was established, along with a long range dependence in the volatility’s residuals. Rankings of each model’s forecasting performance for seven established loss functions were also produced. Overall, the proposed model with the RPV as a regressor (i.e. the AHAR-RPV-FIGARCH), minimized the majority of the forecast loss functions, across all forecasting horizons and indices. The SPA test p-values also confirmed that the AHAR-RPV-GARCH model was not, for the most part, outperformed by any other model. The TTS estimated realized volatility forecasting results demonstrated that the proposed model specification is also robust against the microstructure noise in the returns process.

The published evidence so far concurs that despite the predominant economic significance of producing accurate volatility forecasts, there is no single “ideal” volatility model for all markets and for all financial applications. Here, we showed that by accounting in a HAR specification for the heterogeneity of asymmetric effects, the long memory property of the volatility of realized volatility and by using the RPV as a regressor in the realized volatility process, the stock index volatility forecasting performance can be significantly improved. We have no reason to doubt that similar improvements can also be realizable for other liquid stock indices. However, further investigation is necessary into the performance of the proposed specification in key financial applications like risk management, but also for capital allocation, derivatives pricing and hedging and for other asset classes, such as bonds and currencies.

7 References


