# Forecasting Realized Volatility Measures with Multivariate and Univariate Models: The Case of The US Banking Sector*

Gianluca Cubadda†
Università di Roma "Tor Vergata"

Alain Hecq‡
Maastricht University

Antonio Riccardo§
Interactive Data Kler'S

February 25, 2018

**Abstract**

This paper aims at evaluating the forecasting performances of a set of univariate fractional white noise processes versus multivariate factor models for realized volatility measures. We do not provide a new horse race comparaison of different approaches. Instead, through forecast accuracy analyses, we wish to evaluate the underlying mechanisms that has generated realized volatilities of 13 banking sector asset returns.

JEL: C32

Keywords: Common volatility, long-memory processes, HAR models, index models, forecasting.

*The authors thank blabla. The usual disclaimers apply.

†Universita' di Roma "Tor Vergata", Dipartimento di Economia e Finanza, Via Columbia 2, 00133 Roma, Italy. Email: gianluca.cubadda@uniroma2.it.

‡Maastricht University, Department of Quantitative Economics, P.O.Box 616, 6200 MD Maastricht, The Netherlands. Email: a.hecq@maastrichtuniversity.nl.

§Interactive Data Kler'S, Via Cristoforo Colombo, 149, 00147 Rome, Italy . Email: a.riccardo90@gmail.com

# 1 Introduction

Several realized volatility measures, such as the realized variance or the bipower variation, display long-memory features with the typical slow decay pattern of the empirical autocorrelation function.[1] Although long-memory processes are commonly observed in other fields than in financial econometrics (see Baillie, 1996 and references therein) and that several models of long range dependence have been proposed in the literature (Haldrup and Vera-Valdés, 2017), the fractional integration process of order $d$, denoted $\mathsf{I}(d)$, has been extensively studied in econometrics and statistics since at least Granger (1980) and Granger and Joyeux (1980). An example of an $\mathsf{I}(d)$ process is the fractional white noise $y_t = (1-L)^{-d}\varepsilon_t$, where $L$ denotes the lag operator, $-0.5 < d < 0.5$ and $\varepsilon_t$ is a white noise sequence. For $0 < d < 0.5$, the process is long-memory with positive autocorrelation decaying at a hyperbolic rate. For $-0.5 < d < 0$ the sum of absolute values of the autocorrelations tends to a constant and the process is said to be antipersistent. The class of fractionally integrated processes extends to ARFIMA$(p, d, q)$ cases where $\varepsilon_t$ admits a covariance stationary ARMA representation.

Several estimators of the long-memory parameters of series have been proposed in the literature among which the log periodogram regression of Geweke and Porter-Hudak (1983), the Local Whittle Likelihood Estimator of Robinson (1995) as well as the usual maximum likelihood estimator of an ARFIMA$(p, d, q)$ process that we for instance use in Section 2 of our paper. At the estimation level, Corsi (2009) has proposed a univariate Heterogeneous Autoregressive model (hereafter HAR) as an alternative way to approximate the long range dependence observed in volatility series. For daily series, the HAR is a parsimonious restricted autoregressive model of lag order 22 with daily, weekly and monthly effects. The HAR model can easily be estimated by OLS and it performs well in forecasting exercises.

The literature on the sources of long-memory is quite large, from the aggregation across heterogeneous series argument raised by Granger (1980) to the impact of structural changes that spuriously lead to the detection as a fractional integrated process. Chevillon, Hecq and Laurent (2018, CHL18 hereafter) provide an alternative explanation that we consider in this paper. Indeed CHL18 investigate the mechanisms underlying the long-memory feature generated from a vector autoregressive model (VAR hereafter). They start by assuming that the dynamic interactions between $n$ daily realized volatility measures $Y_t^{(day)} \equiv \left(Y_{1,t}^{(day)}, \ldots, Y_{n,t}^{(day)}\right)'$ is generating by a VAR(1) such that

$$Y_t^{(day)} = \Phi_1 Y_{t-1}^{(day)} + u_t,\ t = 1...T \tag{1}$$

with $u_t$ a $n-$dimensional martingale difference sequence and $\Phi_1$ a $n \times n$ matrix of coefficients. Then for $n \to \infty$, namely when the number of series increases, and under some regularity conditions on values of $\Phi_1$ that are commonly established on real data, they prove that each marginalized individual series (i.e. the final equation representation) is a fractional white noise $(1-L)^d Y_{it}^{(day)} = u_{it}$ processes with the same $d$ parameter. This last observation about the similarities of the values $d$ for different assets is in accordance with Andersen, Bollerslev, Diebold and Labys (2001) who found $d = 0.4$ for

[1] Among many others, see e.g. Andersen, Bollerslev, Diebold and Labys, 2001 for an application to exchange rates, Corsi , Mittnik , Pigorsch and Pigorsch (2008) as well as Hillebrand and Meideiros (2016) for stock price applications.

most exchange rate realized volatilities. CHL2018 provide two specific examples of those conditions on $\Phi_1$. In the first parameterization, the $\Phi_{1,n}$ matrix[2] has diagonal elements converging to 1/2 as $n \to \infty$, and with vanishing off-diagonal elements. Importantly, the off-diagonal elements decrease at an $O\left(n^{-1}\right)$ rate and the sum of each row equals 1 at all $n$. This means in practice that there exist contagion effects, that individually are tiny but jointly potentially important.Note that sparse type regressions (e.g. Lasso) would probably erroneously put a zero at many of those small off-diagonal elements. In the second example, one innovation dominates the others in terms of magnitude.

On the other hand though, many papers have also documented the existence of co-movements in the volatility of asset returns. In integrated markets, common factors in volatility (see e.g. Engle, Ng and Rothshild, 1990, Diebold and Nerlove, 1989) are the result of a common reaction of investors, policy makers or central banks to news/shocks in some macroeconomic and financial variables. However, one important implication underlying CHL2018, beyond the values of the coefficient matrix $\Phi_1$, is that $rank(\Phi_1) = n$ because there are non-zero diagonal elements in $\Phi_1$ and off-diagonal ones are small but different from zero. This would obviously contradicts the presence of a particular form of commonalities, named common features in volatility[3], observed inter alia. by Engle and Marcucci (2006), Engle and Susmel (1993), Hecq, Laurent and Palm (2016) and Anderson and Vahid (2007) to quote a few.

Discriminating between those two approaches in a potentially high dimensional setting might be unfeasible using conventional likelihood ratios or Wald tests. Consequently, this paper compares the forecasting performances of two different modeling strategies: $(i)$ on the one hand we consider a set of univariate models potentially derived from a system with hidden correlations $(ii)$ on the other hand we take several multivariate models, possibly with common factors. For the former framework we model the long-memory feature on individual series using both the maximum likelihood estimation of autoregressive fractionally integrated moving average processes (ARFIMA$(p, d, q)$ and the fractional white noise ARFIMA$(0, d, 0)$ as a special case) as well as the HAR models basically estimated by ordinary least squares. For the second multivariate strategy, it should be noted that we must be able to capture the long-memory features observed in the series. Consequently, we first look at a multivariate version of the HAR model called the Vector HAR (VHAR henceforth, see Bubák, Kočenda and Žikeš, 2011). Then we study the performance of the VHAR Index model (VHARI henceforth; Cubadda, Guardabascio, Hecq 2016) in which we restrict the VHAR using a common index structure. We will come back in Section 3 on the precise meaning of those specifications. Note that we use these multivariate models and not, for instance, generic factor models based on principal component analysis. There are two reasons for that. First the VHARI is nested within the unrestricted VHAR, which is in turn restricted versions of a VAR with 22 daily lags. Hence, the restrictions underlying the VHAR and the VHARI could in principle be tested for, whereas the factor structure is typically postulated in dynamic factor models. Second, at the representation theory level, the common factors obtained from the VHARI preserve the same temporal cascade structure as in the univariate HAR with the weekly (monthly) index being equal to the weekly

[2] The subscript $n$ denotes that the dimension of $\Phi_1$ increases with the system.

[3] Defined intuitively to simplify matters as $rank(\Phi_1) = k << n$.

(monthly) moving average of the daily index. This is an important property that is not shared by most of the alternative factor methods (e.g., principal components, canonical correlations, etc).

The rest of the paper is as follows. Section 2 motivates our study by first looking at the presence of long-memory features in the volatility in price return series of thirteen major US banks. We estimate individual HAR and ARFIMA models on the whole sample. We also look at a simple VAR(1) of series $Y_t^{(day)}$ to have a first clue about the values of $\Phi_1$. Section 3 presents and briefly review the VHAR and the VHARI models. Section 4 compares the forecasting performance of the different approaches. Section 5 concludes.

# 2 Data description

We have considered intraday data extracted from the NYSE "trade and quote" (TAQ) dataset downloaded from Thomson Reuters. It contains the 250 most liquid assets quoted on New York Stocks Exchange covering the period from 03/01/2006 to 31/12/2014 not including weekends and holidays, for a sample of 2265 trading days. From the dataset, we focus in this study on the prices of thirteen major banks. These are, in alphabetical order of the acronyms, (1) BAC: Bank of America Corporation, (2) BBT: BB&T Corporation, (3) BK: Bank Of New York Mellon Corporation (The), (4) C: Citigroup Inc., (5) COF: Capital One Financial Corporation, (6) JPM: J P Morgan Chase & Co, (7) KEY: KeyCorp, (8) PNC: PNC Financial Services Group, Inc. (The), (9) RF: Regions Financial Corporation, (10) STI: SunTrust Banks, Inc., (11) STT: State Street Corporation, (12): USB U.S. Bancorp and (13) WFC: Wells Fargo & Company.

Data have been cleaned following the procedure proposed by Barndorff-Nielsen et al (2009). It consists of the following different steps:

*Steps applied to all data*

P1. Delete entries with a time stamp outside the 9:30 am to 4 pm window when the exchange is open.

P2. Delete entries with a bid, ask or transaction price equal to zero.

*Steps applied only to quote data*

Q1. When multiple quotes have the same timestamp, replace all these with a single entry with the median bid and median ask price.

Q2. Delete entries for which the spread is negative.

Q3. Delete entries for which the spread is more than 50 times the median spread on that day.

Q4. Delete entries for which the mid-quote deviated by more than 10 mean absolute deviations from a rolling centered median (excluding the observation under consideration) of 50 observations (25 observations before and 25 after).

*Steps applied only to trade data*

T1. If multiple transactions have the same time stamp: use the median price.

T2. Delete entries with prices that are above the ask plus the bid-ask spread. Similar for entries with prices below the bid minus the bid-ask spread.

T3. Delete entries for which the price deviated by more than 10 mean absolute deviations from a rolling centered median (excluding the observation under consideration) of 50 observations (25 observations before and 25 after). [4]

At the end of the procedure, the number of trades has been reduced from around 225 millions to slightly more than 105 millions. Notice that, although only trade prices are considered in the analysis, the above cleaning procedure also involves quote data. The reason of this choice is to obtain the more coherent trade prices as possible.

Then, prices have been sampled at 5-minute frequency using the previous point interpolation method and from the correspondent returns, two different realized volatility measures have been computed: the 5-minute Realized Variance (RV) and the 5-minute Median Truncated Realized Variance (MedRV) such as

$$RV_t \equiv \sum_{j=1}^{M} r_{t,j}^2,$$

$$MedRV_t \equiv \frac{\pi}{6 - 4\sqrt{3} + \pi} \left( \frac{M}{M-2} \right) \sum_{j=2}^{M-1} \mathrm{med} \left( |r_{t,j}||r_{t,j-1}||r_{t,j+1}| \right)^2 .$$

where $r_{t,j}$ are the high frequency intra-day returns, observed for $M$ intra-day 5-min periods we have considered each day. Figures 1 and 2 display the levels as well as the log levels of the 13 realized variance series, similar patterns emerge for the MedRV and are not reported to save space. We plot both levels and log-levels of RV series. We can see later that this distinction has an impact on the interpretation of the factors that we extract from those variables. Indeed although taking the logs seems natural to get variables with properties closer to the Gaussian distribution, the aggregation of the levels is easier when the goal is to obtain an index from the (weighted) sum of individual volatilities. In Figures 1 and 2, "whole sample" denotes the period 03/01/2006 to 31/12/2014. In the forecasting exercise of Section 4, we will only report forecasting performances for the post-crisis period 03/01/2008 to 31/12/2014. While our study stays valid for highly volatile periods, a model confidence set approach was not able to statistically distinguish the different clusters when such a huge crisis period is included in the sample. This issue has also been noticed by Hecq, Laurent and Palm (2012) for instance.

The slow decay of the ACF is obvious on every series and is not reported here. Table 1 illustrates that for the log of the 13 realized volatility series[5], both the HAR and the fractional white noise

[4] Both Q4 and T3 are very closely related to the procedure by Brownlees and Gallo (2006). Indeed, the median is used in place of the trimmed sample mean, $\bar{p}_i(k)$, and the mean absolute deviation from the median in place of $s_i(k)$.

[5] Similar results are obtained on MedRV. Also we only report the results for the logs of the series while we compare forecasting performances of both levels and log levels in the forecasting section.

Figure 1: Levels of realized variance series - whole sample

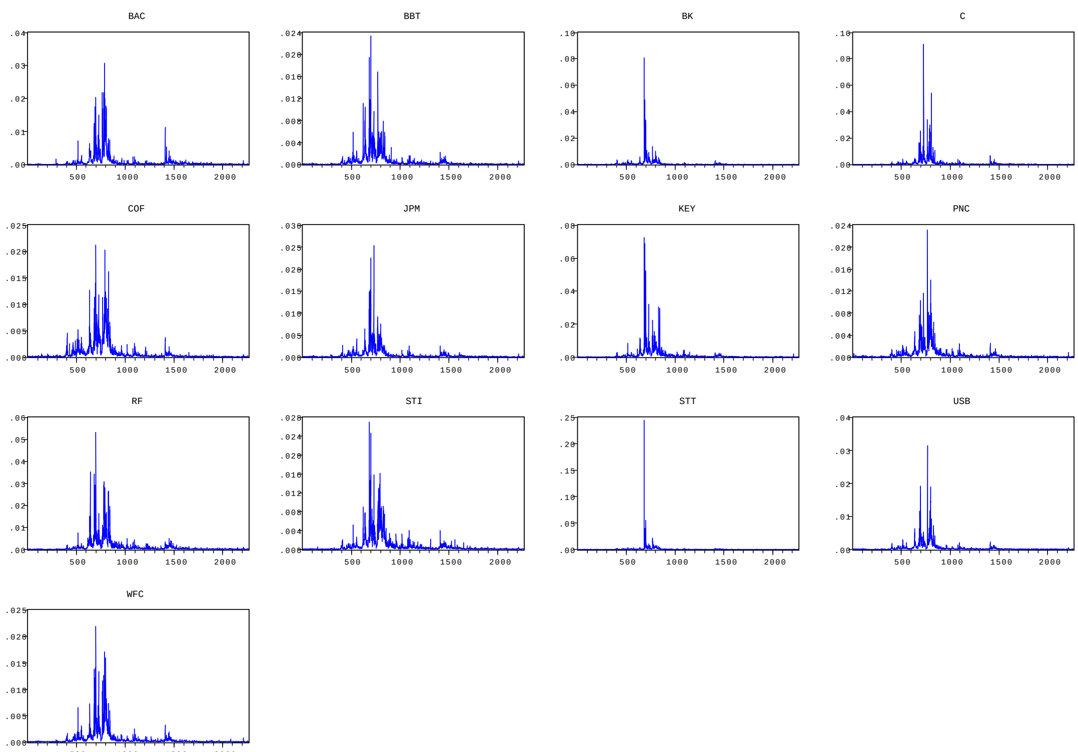
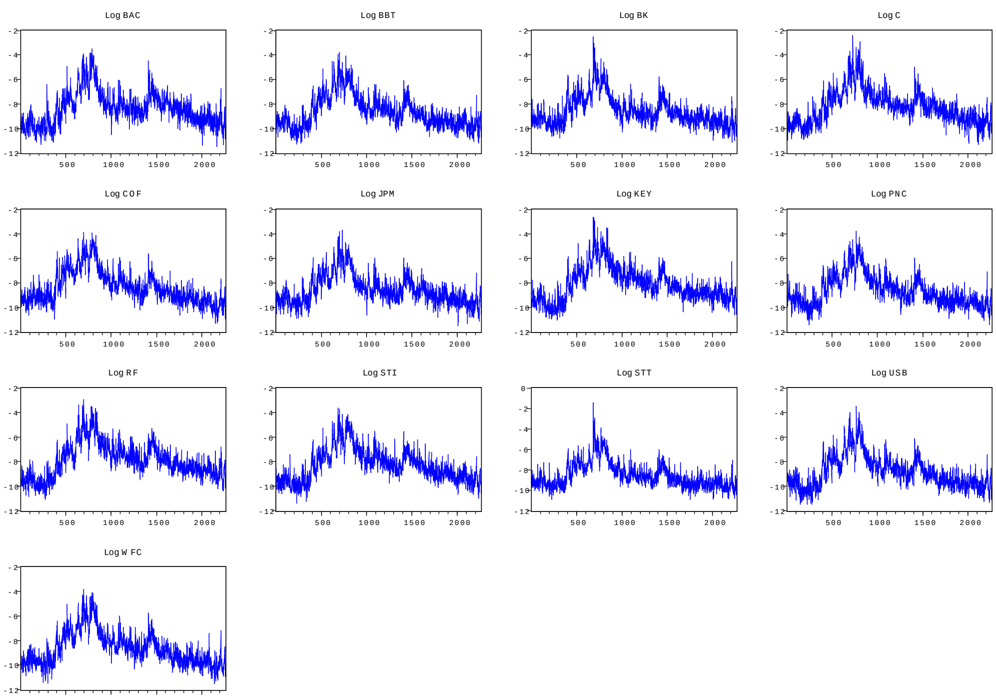


Figure 2: Logs of realized variance series - whole sample



models fit pretty well that feature. We provide OLS estimates (but the intercept to save space) of the following HAR equation:

$$\begin{aligned}
y_{i,t}^{(day)} &= \alpha_{i0} + \alpha_{i1}^{(day)} y_{i,t-1}^{(day)} + \alpha_{i2}^{(w)} y_{i,t-1}^{(w)} + \alpha_{i3}^{(m)} y_{i,t-1}^{(m)} + \epsilon_{i,t}, \\
y_{i,t}^{(day)} &= \ln Y_{i,t}^{(day)} = \ln RV_{i,t}, \\
y_{i,t}^{(w)} &= \frac{1}{5} \sum_{k=0}^{4} y_{i,t-kday}^{(day)}, \quad y_{i,t}^{(m)} = \frac{1}{22} \sum_{j=0}^{21} y_{i,t-kday}^{(day)}, \\
&\text{for } i = 1...13 \text{ returns}
\end{aligned}$$

where * denotes a rejection of the null hypothesis at a 1% significance level.

We also report the $\bar{R}^2$ as well as the $p-value$ of the Ljung-Box $\chi^2$ test for the null hypothesis that the first 10 lags are zero. Note that the maximum likelihood estimates of $\hat{d}$ are rather close to 0.5. It should be noted that ML bounds the parameter estimate at that value. We have also considered the estimation for the series in first differences before we recover "unbounded" values $\hat{d}+1$. In every case we obtain estimated values for $d+1$ between 0.5 and 0.54 with a significant difference to 0.5 in only one case at 1%.

Finally, let us have a look at VAR coefficient matrices in order to figure out how close we are from CHL18 observations. Estimating VARs for the log of the 13 realized variances as well as for the log of MedRV we obtain VAR(5), VAR(1) and VAR(2) for respectively AIC, BIC and HQIC using $p\max = 22$ days. Table 2 provides results for the VAR(1) coefficient matrix chosen by SBC for the logs of the realized variances. We indeed observe as in the model proposed by CHL18 a large value for the diagonal elements and relatively small and often non significant off-diagonal elements. Moreover we notice that we strongly reject the overall Granger non-causality hypothesis in each equations. P-values in each equation $< 0.0001$ for the null hypothesis that all the variables but the lags of the dependent variable are jointly equal zero.

Next section introduces different forms of multivariate VHAR models.

# 3 The Vector Heterogeneous Autoregressive model and its factor extensions

The Vector Heterogeneous Autoregressive model (VHAR, see inter alia, in Bubák et al. (2011), and Souček and Todorova (2013)) reads as follows for the levels of the $n$ series:

$$Y_t^{(day)} = \beta_0 + \Phi^{(day)} Y_{t-1day}^{(day)} + \Phi^{(w)} Y_{t-1day}^{(w)} + \Phi^{(m)} Y_{t-1day}^{(m)} + \varepsilon_t, \quad t = 1, 2, ..., T, \tag{2}$$

where $(day)$, $(w)$, and $(m)$ denote, respectively, time horizons of one day, one week (5 days a week), and one month (assuming 22 days within a month) such that

$$Y_t^{(w)} = \frac{1}{5} \sum_{j=0}^{4} Y_{t-jday}^{(day)}, \quad Y_t^{(m)} = \frac{1}{22} \sum_{j=0}^{21} Y_{t-jday}^{(day)}.$$

Innovations $\varepsilon_t$ are i.i.d. with $\mathrm{E}(\varepsilon_t) = 0$, $\mathrm{E}(\varepsilon_t \varepsilon_t') = \Sigma$ (positive definite), finite fourth moments. Clearly the VHAR (2) involves $(n^2 \times 3)$ parameters and is therefore much more parsimonious than a VAR with 22 unrestricted daily lags. Still a VHAR is an interesting model to consider when no factors are present as it is able to generate long-memory like features and has contagion effects between volatilities. System (2) can easily be estimated my multivariate least square regressions, which means using OLS equation by equation if no cross equation restrictions are present.

Let us further assume that (2) can be rewritten as follows

$$Y_t^{(day)} = \beta_0 + \beta^{(day)} \omega' Y_{t-1day}^{(day)} + \beta^{(w)} \omega' Y_{t-1day}^{(w)} + \beta^{(m)} \omega' Y_{t-1day}^{(m)} + \varepsilon_t, \tag{3}$$

where $\omega$ is a $n \times q$ full-rank matrix. In terms of parameters, (3) needs $4(n \times q) - q^2$ instead of $n^2 \times 3$ in (2). Following Reinsel (1983), we label (3) as the VHAR-index (VHARI) model. To some extent, the VHARI model is related to the pure variance model of Engle and Marcucci (2006) in the sense that a reduced-rank restriction is imposed to the mean parameters of a multivariate volatility model. However, a fundamental difference between (3) and the common volatility model (see also Hecq, Laurent and Palm, 2016) stems from the fact that the former has in general a different left null space for the loading matrices of the factors $\beta = \left[\beta^{(day)} : \beta^{(w)} : \beta^{(m)}\right]$. Obviously, common volatility is allowed in the VHARI model if there exists a full-rank $n \times s$ (with $s < q$) matrix such that $\delta' \beta = 0$.

Beyond the important aspect in terms of parsimony that is shared with many factor models, there are two further motivations for using (3). First, the indexes $f_t^{(day)} = \omega' Y_{t-1day}^{(day)}$ obtained from (3) satisfy the property

$$f_t^{(w)} = \frac{1}{5} \sum_{j=0}^{4} f_{t-jday}^{(day)}, \quad f_t^{(m)} = \frac{1}{22} \sum_{j=0}^{21} f_{t-jday}^{(day)}. \tag{4}$$

as for the observed univariate realized volatilities. Hence, the temporal cascade structure of the HAR model is preserved meaning that the weekly (monthly) index is equal to the weekly (monthly) moving average of the daily index. This would not be generally the case with either traditional reduced-rank regression models as in Engle and Marcucci (2006) or principal component methods.

Second, premultiplying both side (3) by $\omega'$ yields

$$f_t^{(day)} = \omega' \beta_0 + \omega' \beta^{(day)} f_{t-1day}^{(day)} + \omega' \beta^{(w)} f_{t-1day}^{(w)} + \omega' \beta^{(m)} f_{t-1day}^{(m)} + \omega' \varepsilon_t, \tag{5}$$

which shows that the indexes themselves follow a VHAR model. When $q = 1$ the unique index is generated by an univariate HAR model. This property is not shared by alternative methods to aggregate time series (e.g., averages, principal components, canonical correlations, etc.) since the resulting linear combination would generally follow a rather complicated ARMA structure; see Cubadda, Hecq and Palm (2009), Hecq, Laurent and Palm (2016) and the references related to the final equation representation of multivariate models therein.

In order to estimate the parameters of model (3), we resort to a switching algorithm (see details about the estimation technique and Monte Carlo evaluations in Cubadda, Guardabascio and Hecq, 2017) that is widely applied in cointegration analysis (see Boswijk and Doornik, 2004, and the references therein). The strategy consists in alternating between estimating $\omega$ for a given value of $\beta$ and $\Sigma$, and estimating $\beta$ and $\Sigma$ for a given value of $\omega$. In details, the procedure goes as follows:

1. Conditional to an (initial) estimate of the $\omega$, estimate $\beta$ and $\Sigma$ by OLS on (3).

2. Premultiplying both the sides of (3) by $\Sigma^{-1/2}$ one obtains

$$\Sigma^{-1/2}(Y_t^{(d)} - \beta_0) = \Sigma^{-1/2}\beta^{(d)}\omega' Y_{t-1d}^{(d)} + \Sigma^{-1/2}\beta^{(w)}\omega' Y_{t-1d}^{(w)} + \Sigma^{-1/2}\beta^{(m)}\omega' Y_{t-1d}^{(m)} + \Sigma^{-1/2}\varepsilon_t.$$

   Applying the Vec operator to both sides of the above equation and using the property $\text{Vec}(ABC) = (C' \otimes A)\text{Vec}(B)$ one gets

$$\begin{aligned}\text{Vec}\left[\Sigma^{-1/2}(Y_t^{(d)} - \beta_0)\right] &= \left(Y_{t-1d}^{(d)\prime} \otimes \Sigma^{-1/2}\beta^{(d)}\right)\text{Vec}(\omega') + \left(Y_{t-1d}^{(w)\prime} \otimes \Sigma^{-1/2}\beta^{(w)}\right)\text{Vec}(\omega') \\ &+ \left(Y_{t-1d}^{(m)\prime} \otimes \Sigma^{-1/2}\beta^{(m)}\right)\text{Vec}(\omega') + \text{Vec}\left(\Sigma^{-1/2}\varepsilon_t\right), \qquad (6)\end{aligned}$$

   from which we can finally estimate by OLS the $\omega$ coefficients conditional to the previously obtained estimates of the parameters $\beta$ and $\Sigma$.

3. Switch between steps 1 and 2 till numerical convergence occurs.

As shown by Boswijk (1995), the proposed switching algorithm has the property to increase the Gaussian likelihood in each step. Note that a numerical stability problem may arise when the number of series is very large. A possible solution is to resort to a properly "regularized" estimate of the autocorrelation matrix function of series $Y_t^{(d)}$ instead on the natural one that is implicitly used in our procedure (see Bernardini and Cubadda (2015) for details).

In order to identify the number of factors $q$, one can use the usual information criteria proposed by Schwarz (BIC), Hannan-Quinn (HQIC) and Akaike (AIC). We propose some variants of them that are based on the theoretical framework developed by Takeuchi (1976). In short, this author extends the AIC by relaxing the strong assumption that the set of the candidate models includes the true model. This extension is relevant in our case for at least two reasons. First, HAR processes are generally seen as an approximation to long-memory processes (Corsi, 2009). Second, the residuals of HAR models are typically non-Gaussian and heteroskedastic (e.g., Corsi, Audrino and Reno, 2012; Corsi, Mittnik, Pigorsch and Pigorsch, 2008), whereas our switching algorithm aims at maximizing the Gaussian likelihood. In the Appendix, we develop a Takeuchi-type modification to the traditional information criteria for our VHARI models. We denote the modified criteria as MAIC, MHQIC, and MBIC.

# 4 Forecasting

Table 3 and 4 compare the VHAR with different VHARI factor restrictions and individual fractional white noise processes estimated either by MLE or from the HAR approximation. The latter model being used as the benchmark. The number of indexes $q$ has been determined using information criteria. One important purpose of the empirical analysis is to test the adequacy of the VHARI restrictions compared to both univariate HAR model and the VHAR model. Hence, for both RV and MedRV, it has been performed a direct out-of-sample $h-$step ahead forecast for $h = 1, 5, 22$, using a rolling window of 500 observations. As in Cubadda, Hecq and Guardabascio (2017), the goodness of the forecasts has been evaluated trough the average relative mean squared forecast errors (ARMSFE). It is defined as

$$ARMSFE_m = \frac{1}{n}\sum_{i=1}^{n}(\frac{MSFE_{m,i}}{MSFE_{HAR,i}}) \times 100$$

where $m$ denotes the model (e.g. VHARI or VHAR) and $n$ represents the assets, 13 in our example.In Tables 3 and 4, we also report the quartiles of the number of factors distribution that are obtained by the various information criteria. In addition, it has been performed the model confidence set (MCS) suggested by Hansen et al. (2011) to find the set of multivariate models which forecast equally well. The analysis consists in testing for the null hypothesis of equal predictive ability (EPA) at the 20% level and it is implemented using a block bootstrap scheme with 5000 resamples. In tables, the models within the superior set are denoted with a star.

From the values of the ARMSFE in Table 3 it emerges that individual processes outperforms multivariate ones at any time horizons. Considering the MCS, results are less clear cut. For instance with $h = 1$, every model is in the same cluster although multivariate models have a ARMSFE at about 30% higher than the HAR model. In Table 4, taking the logs provides a clearer pattern. For $h = 1$ ane $h = 5$ univariate models, boh using HAR and FWN, form a first cluster whose acuracy is significantly different from multivariate versions. There are no significant differences for $h = 22$.

# 5 Conclusions

Hesitating between two forms of multivariate models generating the behavior of the volatility of 13 asset returns, we are tempted to conclude that we would face a high dimensional process with high diagonal elements and small off-diagonal ones. MCS does not find significant differences between levels of the series. However, univariate models outperforms multivariate ones when series are in logs. This would favor the large dimensional VAR with small contagion effects as the underlying generating mehanism of realized volatility measures.

# References


[1] Andersen , T., Bollerslev, T., Diebold, F.X. and P. Labys (2001), The distribution of realized exchange rate volatility . *Journal of the American Statistical Association* 96 ( 453 ): 42 – 55 .

[2] Anderson, H., and F. Vahid (2007), Forecasting the Volatility of Australian Stock Returns: Do Common Factors Help?, *Journal of Business and Economic Statistics*, 25, 76-90.

[3] Baillie, R.T., (1996), Long memory processes and fractional integration in econometrics, *Journal of Econometrics* 73, 5–59.

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

[5] Barndorff-Nielsen, O.E. and N. Shephard (2002), Econometric analysis of realized volatility and its use in estimating stochastic volatility models, *J. R. Statist. Soc. B*, 64, 253–280.

[6] Bernardini E. and G. Cubadda (2015), Macroeconomic Forecasting and Structural Analysis through Regularized Reduced-Rank Regression, *International Journal of Forecasting*, 31, 682-691.

[7] Bauwens, L., Hafner, Ch. and S. Laurent (2012), Volatility Models, *Handbook of Volatility Models and Their Applications* (eds. Bauwens, L., Hafner, Ch. and S. Laurent), Wiley, Ch. 1.

[8] Boswijk, H.P., (1995), Identifiability of cointegrated systems, *Tinbergen Institute Working Paper*, 95/78.

[9] Boswijk, H.P, and J.A. Doornik (2004), Identifying, estimating and testing restricted cointegrated systems: An overview, *Statistica Neerlandica*, 58, 440-465.

[10] Bubák, V., Kočenda, E. and F. Žikeš (2011), Volatility transmission in emerging European exchange markets, *Journal of Banking & Finance*, 35, 2829-2841.

[11] Chevillon, G., Hecq, A. and S. Laurent (2018), Generating univariate fractional integration within a large VAR(1), forthcoming *Journal of Econometrics*.

[12] Corsi, F. (2009), A Simple Approximate Long-Memory Model of Realized Volatility, *Journal of Financial Econometrics*, Vol. 7, No.2, 174-196.

[13] Corsi, F., Audrino, F. and R. Renò (2012), HAR Modeling for Realized Volatility Forecasting, *Handbook of Volatility Models and Their Applications* (eds. Bauwens, L., Hafner, Ch. and S. Laurent), Wiley, Ch. 15.

[14] Corsi, F., S. Mittnik, C. Pigorsch, and U. Pigorsch (2008), The volatility of realized volatility, *Econometric Reviews*, 27, 46-78.

[15] Cubadda, G., Hecq A. and F.C. Palm (2009), Studying Co-movements in Large Multivariate Models Prior to Modeling, *Journal of Econometrics,* 148, 25-35.

[16] Diebold, F.S. and M. Nerlove (1989), The dynamics of Exchenge rate volatility: a multivariate latent-factor ARCH model, *Journal of Applied Econometrics,* 4, 1-22.

[17] Engle, R.F, Ng, V.K. and M. and Rothshild (1990), Asset pricing using the FACTOR-ARCH model, *Journal of Econometrics*, 45, 213-37.

[18] Engle, R.F. and J. Marcucci (2006), A Long-run Pure Variance Common Features Model for the Common Volatilities of the Dow Jones, *Journal of Econometrics* 132, 7-42.

[19] Engle, R.F. and R. Susmel (1993), Common Volatility in International Equity Markets, *Journal of Business and Economic Statistics*, 11, 167–176.

[20] Fengler, M. and K. Gisler (2015), A variance spillover analysis without covariances: What do we miss?, *Journal of International Money and Finance*, 51, 174-195.

[21] Geweke, J., Porter-Hudak, S., (1983), The estimation and application of long memory time series models, Journal of Time Series Analysis, 4, 221–238.

[22] Granger, C.W.J. (1980), Long memory relationships and the aggregation of dynamic models, *Journal of Econometrics* 14 (2), 227–238.

[23] Granger, C.W.J. and R. Joyeux (1980), An introduction to long-memory time series models and fractional differencing. Journal of Time Series Analysis 1 (1), 15–29.

[24] Haldrup, N., Vera-Valdés, E. (2017), Long memory, fractional integration, and cross sectional aggregation, *Journal of Econometrics* , vol.199, 1-11.

[25] Hansen, P.R., Lunde, A. and J. Nason (2011), The model confidence set, *Econometrica*, 79, 453-497.

[26] Heber, G., Lunde, A., Shephard, N. and K. Sheppard (2009), Oxford-Mann institute's realized library version 0.1. *Oxford-Man Institute, University of Oxford*.

[27] Hecq, A., Laurent, S. and F. Palm (2012), Common Intraday Periodicity, *Journal of Financial Econometrics*, Vol.10, Issue 2, 325–35.

[28] Hecq, A., Laurent, S. and F. Palm (2016), On the Univariate Representation of BEKK Models with Common Factors, *Journal of Time Series Econometrics*, 8, 91-113.

[29] Liu, L., Patton, A. and K. Sheppard (2015), Does anything beat 5-minute RV? A comparison of realized measures across multiple asset classes, *Journal of Econometrics*, 187, 293-311.

[30] Patton, A. (2011), Data-based ranking of realised volatility estimators, *Journal of Econometrics*, 161, 284-303.

[31] Patton, A. and K. Sheppard (2009), Optimal combinations of realised volatility estimators, *International Journal of Forecasting*, 25, 218-238.

[32] Reinsel, G. (1983), Some results on multivariate autoregressive index models, *Biometrika*, 70, 145-156.

[33] Robinson, P.M., (1995), Gaussian semiparametric estimation of long range dependence, *Annals of Statistics*, 23, 1630–1661.

[34] Souček, M. and N. Todorova (2013), Realized volatilities transmission between crude oil and equity futures markets: A multivariate HAR approach, *Energy Economics*, 40, 586-597.

Table 1: Univariate models for log of realized volatility

| | | $d$ | $\alpha_1$ | $\alpha_2$ | $\alpha_3$ | $\bar{R}^2$ | $Q(10)$ |
|---|---|---|---|---|---|---|---|
| BAC | ARFIMA(0,$d$,0) | 0.497* | – | – | – | 0.83 | 0.306 |
| | HAR | – | 0.491* | 0.263* | 0.219* | 0.83 | 0.594 |
| BBT | ARFIMA(0,$d$,0) | 0.496* | – | – | – | 0.85 | 0.049 |
| | HAR | – | 0.416* | 0.347* | 0.215* | 0.85 | 0.108 |
| BK | ARFIMA(0,$d$,0) | 0.496* | – | – | – | 0.82 | 0.109 |
| | HAR | – | 0.449* | 0.307* | 0.214* | 0.82 | 0.103 |
| C | ARFIMA(0,$d$,0) | 0.497* | – | – | – | 0.85 | 0.054 |
| | HAR | – | 0.441* | 0.339* | 0.195* | 0.84 | 0.000 |
| COF | ARFIMA(0,$d$,0) | 0.497* | – | – | – | 0.86 | 0.003 |
| | HAR | – | 0.421* | 0.367* | 0.190* | 0.86 | 0.037 |
| JPM | ARFIMA(0,$d$,0) | 0.497* | – | – | – | 0.82 | 0.135 |
| | HAR | – | 0.498* | 0.266* | 0.205* | 0.82 | 0.782 |
| KEY | ARFIMA(0,$d$,0) | 0.497* | – | – | – | 0.85 | 0.107 |
| | HAR | – | 0.444* | 0.305* | 0.230* | 0.86 | 0.252 |
| PNC | ARFIMA(0,$d$,0) | 0.496* | – | – | – | 0.84 | 0.058 |
| | HAR | – | 0.432* | 0.342* | 0.201* | 0.84 | 0.028 |
| RF | ARFIMA(0,$d$,0) | 0.497* | – | – | – | 0.86 | 0.000 |
| | HAR | – | 0.418* | 0.356* | 0.203* | 0.86 | 0.000 |
| STI | ARFIMA(0,$d$,0) | 0.496* | – | – | – | 0.85 | 0.005 |
| | HAR | – | 0.415* | 0.353* | 0.209* | 0.85 | 0.011 |
| STT | ARFIMA(0,$d$,0) | 0.490* | – | – | – | 0.82 | 0.015 |
| | HAR | – | 0.371* | 0.384* | 0.216* | 0.82 | 0.049 |
| USB | ARFIMA(0,$d$,0) | 0.497* | – | – | – | 0.85 | 0.014 |
| | HAR | – | 0.452* | 0.343* | 0.179* | 0.85 | 0.347 |
| WFC | ARFIMA(0,$d$,0) | 0.498* | – | – | – | 0.85 | 0.006 |
| | HAR | – | 0.481* | 0.286* | 0.209* | 0.85 | 0.294 |

Table 2: VAR(1) estimated matrix for log of realized volatility

$$\hat{\Phi}_1 =$$

$$\left(\begin{array}{ccccccccccccc}
0.523 & -0.046 & -0.025 & 0.018 & -0.062 & 0.020 & -0.017 & -0.077 & 0.124 & 0.063 & -0.048 & -0.018 & -0.0 \\
-0.086 & 0.310 & 0.070 & -0.076 & 0.068 & 0.034 & 0.032 & -0.018 & 0.006 & 0.022 & 0.052 & -0.011 & -0.0 \\
0.036 & 0.072 & 0.393 & -0.035 & 0.035 & 0.115 & -0.013 & -0.016 & -0.043 & 0.000 & 0.122 & 0.004 & 0.0 \\
0.117 & 0.010 & 0.012 & 0.540 & 0.067 & 0.074 & 0.100 & 0.047 & 0.077 & 0.112 & 0.041 & 0.073 & 0.1 \\
0.045 & 0.183 & 0.098 & 0.131 & 0.497 & 0.140 & 0.049 & 0.164 & 0.074 & 0.153 & 0.130 & 0.109 & 0.2 \\
0.056 & 0.018 & 0.189 & -0.007 & 0.017 & 0.388 & -0.101 & 0.001 & -0.103 & -0.071 & 0.062 & 0.006 & 0.0 \\
0.030 & 0.107 & 0.032 & 0.095 & 0.030 & 0.021 & 0.450 & 0.124 & 0.177 & 0.128 & 0.070 & 0.093 & 0.0 \\
-0.055 & 0.009 & -0.045 & -0.043 & 0.089 & -0.062 & 0.021 & 0.303 & -0.051 & -0.022 & 0.056 & 0.070 & 0.0 \\
0.170 & 0.096 & 0.013 & 0.118 & -0.024 & 0.028 & 0.186 & 0.051 & 0.484 & 0.156 & 0.003 & 0.087 & 0.0 \\
0.064 & -0.039 & -0.027 & 0.078 & -0.024 & -0.042 & 0.053 & -0.013 & 0.159 & 0.300 & -0.067 & -0.025 & 0.0 \\
0.000 & 0.037 & 0.149 & 0.022 & 0.045 & 0.037 & 0.055 & 0.084 & -0.034 & -0.037 & 0.349 & 0.065 & 0.0 \\
0.031 & 0.096 & 0.013 & -0.034 & 0.033 & 0.035 & 0.086 & 0.125 & 0.049 & 0.079 & 0.050 & 0.357 & 0.0 \\
-0.003 & 0.083 & 0.023 & 0.114 & 0.189 & 0.094 & 0.044 & 0.129 & 0.009 & 0.046 & 0.082 & 0.125 & 0.4
\end{array}\right.$$

Table 3: Forecast comparison for levels of MedRV - post crisis period

| | ARMSFE | | | $\hat{q}$ | | | |
|---|---|---|---|---|---|---|---|
| Method/Criterion | $h = 1$ | $h = 5$ | $h = 22$ | [Q1 | Q2 | Q3] | Mode |
| VHARI/BIC | 137,9* | 108,3 | 110,5* | [ 1 | 2 | 3 ] | 3 |
| VHARI/HQIC | 136,3* | 112,0 | 106,6* | [ 3 | 4 | 6 ] | 3 |
| VHARI/AIC | 134,6* | 111,0 | 109,9* | [ 7 | 8 | 9 ] | 8 |
| VHAR | 135,5* | 118,0 | 111,9* | | | | |
| VHARI (R)/BIC | 135,0* | 102,4 | 107,7* | [ 1 | 2 | 2 ] | 2 |
| VHARI (R)/HQIC | 131,1* | 104,0 | 105,5* | [ 1 | 3 | 3 ] | 3 |
| VHARI (R)/AIC | 130,3* | 110,9 | 105,3* | [ 13 | 13 | 13 ] | 13 |
| VHAR (R) | 130,3 | 110,9 | 105,3 | | | | |
| FWN | 97,3* | 100,1* | 95,6* | | | | |
| HAR | 100* | 100* | 100* | | | | |

Note: $h$ is the forecasting horizon. ARMSFE is the average of the mean square forecast errors relative to the HAR univariate forecasts.
$Q_i$ indicates the $i$-th quartile of the number of factors distribution.
In bold, the method with the lowest ARMSFE for each $h$.
The methods within the superior set of models have been denoted with a star.

Table 4: Forecast comparison for logs of MedRV - post crisis period

| | ARMSFE | | | $\hat{q}$ | | | |
|---|---|---|---|---|---|---|---|
| Method/Criterion | $h = 1$ | $h = 5$ | $h = 22$ | [Q1 | Q2 | Q3] | Mode |
| VHARI/BIC | 121,6 | 108,4 | 111,5* | [ 1 | 2 | 2 ] | 2 |
| VHARI/HQIC | 117,9 | 110,1 | 120,6 | [ 3 | 3 | 4 ] | 3 |
| VHARI/AIC | 117,7 | 110,8 | 118,3* | [ 6 | 6 | 7 ] | 6 |
| VHAR | 118,8 | 117,2 | 114,2* | | | | |
| VHARI (R)/BIC | 137,6 | 115,1 | 106,9* | [ 1 | 1 | 2 ] | 1 |
| VHARI (R)/HQIC | 122,7 | 111,7 | 112,8* | [ 2 | 3 | 4 ] | 2 |
| VHARI (R)/AIC | 115,2 | 111,2 | 111,4* | [ 5 | 8 | 9 ] | 8 |
| VHAR (R) | 115,3 | 111,3 | 111,5* | | | | |
| FWN | 99,8* | 99,2* | 101,3* | | | | |
| HAR | 100* | 100* | 100* | | | | |

Note: $h$ is the forecasting horizon. ARMSFE is the average of the mean square forecast errors relative to the HAR univariate forecasts.
$Q_i$ indicates the $i$-th quartile of the number of factors distribution.
In bold, the method with the lowest ARMSFE for each $h$.
The methods within the superior set of models have been denoted with a star.