International Journal of Forecasting 33 (2017) 337–344

Contents lists available at ScienceDirect

International Journal of Forecasting

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

A vector heterogeneous autoregressive index model for realized volatility measures

Gianluca Cubadda [a,*], Barbara Guardabascio [b], Alain Hecq [c]

[a] Universita' di Roma "Tor Vergata", Dipartimento di Economia e Finanza, Via Columbia 2, 00133 Roma, Italy

[b] ISTAT, DCSC-SER/C, Viale Liegi 13, 00198 Roma, Italy

[c] Maastricht University, Department of Quantitative Economics, P.O.Box 616, 6200 MD Maastricht, The Netherlands

ARTICLE INFO

Keywords:
Common volatility
HAR models
Index models
Combinations of realized volatilities
Forecasting

ABSTRACT

This paper introduces a new model for detecting the presence of commonalities in a set of realized volatility measures. In particular, we propose a multivariate generalization of the heterogeneous autoregressive model (HAR) that is endowed with a common index structure. The vector heterogeneous autoregressive index model has the property of generating a common index that preserves the same temporal cascade structure as in the HAR model, a feature that is not shared by other aggregation methods (e.g., principal components). The parameters of this model can be estimated easily by a proper switching algorithm that increases the Gaussian likelihood at each step. We illustrate our approach using an empirical analysis that aims to combine several realized volatility measures of the same equity index for three different markets.

1. Introduction

The presence of co-movements in volatility measures is usually explained by common reactions of investors, policy makers or central banks to news relating to certain macroeconomic and financial variables. Engle and Marcucci (2006) find evidence indicating the presence of common ARCH factors (Engle & Susmel, 1993) between 435 pairs obtained from 30 stocks of the Dow Jones industrial index. However, their statistical approach might suffer from severe size distortions when applied in a multivariate setting (see Cubadda & Hecq, 2011; Hecq, Laurent, & Palm, 2016). Anderson and Vahid (2007) propose the examination of information criteria for determining the presence and number of principal component factors out of 21 Australian weekly stock return volatilities. It turns out that this latter approach is probably more robust to the presence of jumps and fat tails than the canonical correlation framework of Engle and Marcucci (2006). However, these contributions assume the dynamics of the system to be very parsimonious, contrary to the observed time series properties of daily volatility measures. For instance, the univariate heterogeneous autoregressive model (HAR; see Corsi, 2009) captures the long range dependence observed in daily time series using a restricted autoregressive model of order 22.

This paper proposes a new model for analyzing the joint behaviors of a set of daily volatility measures. We start out with a multivariate version of the HAR, namely the vector HAR (VHAR henceforth, see Bubák, Kočenda, & Žikeš, 2011). Next, we test, and consequently restrict, the VHAR by means of a multivariate autoregressive index model (Reinsel, 1983). In particular, we impose proper reduced rank restrictions on the coefficient matrices of the VHAR to obtain the vector heterogeneous autoregressive index model (VHARI henceforth).

* Corresponding author.
E-mail addresses: gianluca.cubadda@uniroma2.it (G. Cubadda), guardabascio@istat.it (B. Guardabascio), a.hecq@maastrichtuniversity.nl (A. Hecq).

http://dx.doi.org/10.1016/j.ijforecast.2016.09.002
0169-2070/

The VHARI is nested within the unrestricted VHAR, which in turn is a restricted version of a vector autoregressive model (VAR) of order 22. The VHARI provides a parsimonious model whose forecasting performance can be compared with those of either less restricted multivariate models (e.g., VHAR or VAR(22)) or univariate HAR equations. At the representation theory level, the common factors obtained from the VHARI, namely the indexes, preserve the same temporal cascade structure as in the HAR; i.e., the weekly (monthly) index is equal to the weekly (monthly) moving average of the daily index. This is an important property of the VHARI that is not shared by most of the alternative aggregation methods (e.g., principal components, canonical correlations, etc.). Moreover, in a VHARI with one common component, a specification that is not rejected by the data in the empirical section of this paper, the unique index is generated by an univariate HAR model. This is not generally the case for alternative aggregation strategies either.

The rest of the paper proceeds as follows. Section 2 presents the VHAR and VHARI models, with their implications. Statistical inference is discussed in detail in Section 3. Note that we use a switching algorithm to maximize the Gaussian likelihood of a given VHARI specification. Hence, in principle, the adequacy of our set of restrictions can be checked using either information criteria or likelihood ratio tests. However, this strategy cannot be implemented for factors obtained through principal component analysis, for instance. Moreover, in the same vein as Takeuchi (1976), we propose some modified versions of the usual information criteria that are better suited for non-Gaussian series. Section 4 contains a Monte Carlo simulation exercise that documents the small-sample properties of our modelling strategy. Section 5 uses the suggested framework to combine ten realized volatility measures of the same equity index for three different markets using data from the Oxford-Man Institute of Quantitative Finance. Finally, Section 6 concludes.

2. Model representation

2.1. The vector heterogeneous autoregressive model

Our starting point for capturing the dynamic interactions within a set of n daily realized volatility measures $Y_t^{(d)} \equiv \left(Y_{1,t}^{(d)}, \ldots, Y_{n,t}^{(d)}\right)'$ is a multivariate version of the univariate HAR model (Corsi, 2009), as was used by Bubák et al. (2011) and Souček and Todorova (2013), *inter alia*.

The vector $Y_t^{(d)}$ can include either the same kind of volatility measure (e.g., the realized variance)[1] for different markets in a study of volatility co-movements or several volatility measures (realized variance, bipower variation, etc.)[2] for the same market in order to construct an optimal linear combination like that of Patton and Sheppard (2009). The latter analysis is pursued in Section 5 of this paper.

The vector heterogeneous autoregressive model (VHAR) can be written as follows:

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

where (d), (w), and (m) denote time horizons of one day, one week (five days in a week), and one month (assuming 22 days in a month) respectively, such that

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

Here, the innovations ε_t are i.i.d. with $E(\varepsilon_t) = 0$, $E(\varepsilon_t \varepsilon_t') = \Sigma$ (positive definite), and finite fourth moments.

Beyond the fact that the HAR is a popular forecasting tool, two considerations that arise from our empirical analysis have led us to refer to Eq. (1) as a starting point. First, having estimated unrestricted VAR(p) models on a set of different volatility measures for each of the markets at hand, it emerges that we reject the null of no error autocorrelation for lags p of five or higher (using heteroskedasticity-robust LR tests). This means that the data have a greater dependence on the past. In principle, one could increase the VAR order considerably, but the curse of dimensionality remains a problem even when the sample size is as large as is the case in typical financial applications. Hence, Eq. (1) is a good compromise in terms of parameter proliferation, since a VAR(22) has $N^2 \times 22$ mean parameters, whereas the model in Eq. (1) needs $N^2 \times 3$ of them. Second, for the set of realized volatilities considered, the coefficient matrices $\Phi^{(d)}$, $\Phi^{(w)}$ and $\Phi^{(m)}$ are far from being diagonals, and consequently a set of individual HAR models does not seem appropriate.

The next subsection introduces additional meaningful restrictions to Eq. (1), namely the steps required to go from the VHAR to the VHARI.

2.2. The VHAR-index model

Let us further assume that Eq. (1) can be rewritten as:

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

where ω is a $n \times q$ full-rank matrix. In terms of parsimony, Eq. (2) needs $4(n \times q) - q^2$ parameters instead of $n^2 \times 3$ in Eq. (1). Following Reinsel (1983), we label Eq. (2) the VHAR-index (VHARI) model. To some extent, the VHARI modeling is related to the pure variance model of Engle and Marcucci (2006), in the sense that a reduced-rank restriction is imposed on the mean parameters of a multivariate volatility model. However, one fundamental difference between Eq. (2) and the common volatility model (see also Hecq et al., 2016) stems from the fact that the former generally has a different left null space for the loading matrices of the indexes $\beta = \left[\beta^{(d)} : \beta^{(w)} : \beta^{(m)}\right]$. Obviously, common volatility is allowed in the VHARI model if there exists a full-rank $n \times s$ (with $s < q$) matrix such that $\delta'\beta = 0$.

[1] The realized covariances may also be included in $Y_t^{(d)}$, see Fengler and Gisler (2015).

[2] The realized variances are computed using $RV_t \equiv \sum_{i=1}^{M} r_{t,i}^2$, where $r_{t,i}$ are the high frequency intra-day returns, observed for M intra-day periods each day. For instance, when the market is open between 9 a.m. and 4 p.m., $M = 79$ for 5-min returns. The bipower variation $BV_t \equiv \frac{\pi}{2}\frac{M}{M-1}\sum_{i=2}^{M} |r_{t,i}||r_{t,i-1}|$ is one of the measures of the integrated volatility that is designed to be robust to jumps. See, *i.a.*, Barndorff-Nielsen and Shephard (2004) and Bauwens, Hafner, and Laurent (2012).