Journal of the Korean Statistical Society 45 (2016) 371–380

Contents lists available at ScienceDirect

Journal of the Korean Statistical Society

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

An integrated heteroscedastic autoregressive model for forecasting realized volatilities

Soojin Cho, Dong Wan Shin *

Department of Statistics, Ewha University, Seoul, Republic of Korea

ARTICLE INFO

Article history:
Received 29 September 2015
Accepted 29 December 2015
Available online 1 February 2016

AMS 2000 subject classifications:
primary 62M10
secondary 62M20

Keywords:
Conditional heteroscedasticity
Fractional integration
HAR model
High frequency data
Long-memory
Volatility forecasting

ABSTRACT

A new strategy for forecasting realized volatility (RV) is proposed for the heteroscedastic autoregressive (HAR) model of Corsi (2009). The strategy is constraining the sum of the HAR coefficients to one, resulting in an integrated model, called IHAR model. The IHAR model is motivated by stationarity of estimated HAR model, downward biases of estimated HAR coefficients, and over-rejection of ADF test for long-memory processes. Considerable out-of-sample forecast improvements of the IHAR model over the HAR model are demonstrated for RVs of 4 financial assets: the US S&P 500 index, the US NASDAQ index, the Japan yen/US dollar exchange rate, and the EU euro/US dollar exchange rate. Forecast improvement is also verified in a Monte Carlo experiment and in an empirical comparison for an extended data set. The forecast improvement is shown to be a consequence of the fact that the IHAR model takes better advantage of the long memory of RV and the conditional heteroscedasticity of RV than the HAR model.

1. Introduction

Forecasting volatility is essential for financial pricing, asset allocation, and risk management. Among many volatility measures, realized volatility (RV) based on intra-day high frequency asset observations is one of the major interests building a large amount of results in the recent literature. A good review on RV is provided by McAleer and Medeiros (2008) in which we find a review on forecasting RV in Section 5.

The HAR (Heteroscedastic AutoRegressive) model proposed by Corsi (2004, 2009) is very useful in forecasting financial realized volatility. The HAR model is conceptually appealing because it represents volatilities of different short-term and long-term market participants via daily, weekly, and monthly volatility components. We find many successful applications and extensions of the model. Among many others, we refer Andersen, Bollerslev, and Diebold (2007) and Corsi, Pirino, and Reno (2010) for models with jump; McAleer and Medeiros (2008) for models having leverage effect; Busch, Christensen, and Nielsen (2011) for models with implied volatility and jumps; Hwang and Shin (2014) for an infinite order model; Hwang and Shin (2013, 2015) and Song and Shin (2015) for structural breaks; and Yun and Shin (2015) for the issue of overnight in RV forecasting.

The HAR model represents efficiently long-memories of financial volatilities by employing the efficient regressors of the one-day lag, one-day lagged weekly moving average, and one-day lagged monthly moving average of realized volatility. However, we note that the HAR model is an AR(22) model and estimated HAR models are usually stationary. For example,

* Corresponding author. Tel.: +82 2 3277 2614; fax: +82 2 3277 3606.
E-mail address: shindw@ewha.ac.kr (D.W. Shin).

http://dx.doi.org/10.1016/j.jkss.2015.12.004

the estimated models for the USD/CHF exchange rate and the US T-bond realized volatilities in Corsi (2009) are all stationary because the sums 0.91, 0.81 of the three HAR coefficients are all smaller than 1.

We point out that the estimated-stationarity matters in forecasting long-term realized volatilities. It is well known in the literature that financial volatilities are long memory, see for example, the seminal papers by Andersen, Bollerslev, Diebold, and Ebens (2001) and Andersen, Bollerslev, Diebold, and Labys (2001a,b). Especially the latter two papers discussed long-memory properties of realized exchange rate volatilities and realized stock price volatilities. In the estimated stationary HAR model for financial volatilities, autocorrelation functions (ACFs) for large lags decay faster at exponential rates than algebraic rates of the usual ACFs of financial volatilities. Therefore, some of the long memories in the financial volatilities remain unexplained by the estimated stationary HAR models. Owning to the unexplained long-memories, long-term volatility forecasts regress to the global mean more rapidly than they should be, resulting in efficiency losses in long-term forecasts. More discussions on the unexplained memory problems will be made in Section 2.

In order to achieve better long-term forecasts, we propose a new strategy of constraining the sum of HAR coefficients to 1. The constrained model is a unit-root model and is nonstationary. We call the model an "integrated HAR", IHAR, model. The IHAR model will be more motivated in Section 2 by over-rejection of the ADF test against fractional integration, by level shift in realized volatility, and by downward biases of the estimated HAR coefficients. The forecast advantage of the IHAR model over the HAR model will be investigated in a Monte-Carlo study of Section 3.

In Section 4, out-of-sample forecast performances of the proposed IHAR model will be compared with those of the HAR model and other two models of random walk and fractional integration for 4 RVs based on high frequency data sets: 2 US stock price indices and 2 foreign exchange rates relative to US dollar for, roughly, the last two decades. The comparison reveals considerable forecast improvement of the IHAR model over the HAR model and the other two models. Similar favorable results are observed for the IHAR forecasts in an extended comparison in Section 5 for an expanded data set consisting of all the 20 index series in the realized library of Oxford-Man Institute.

2. An Integrated HAR model

We first discuss some problems in terms of memory properties of the HAR model and next propose the IHAR model to overcome the problems. The HAR model of Corsi (2009) is

$$y_{t+1} = \phi_0 + \phi_d y_t + \phi_w y_t^w + \phi_m y_t^m + \epsilon_{t+1}, \tag{1}$$

where ϵ_t is a sequence of regression error and $y_t^w = (y_t + \cdots + y_{t-4})/5$, $y_t^m = (y_t + \cdots + y_{t-21})/22$ are the weekly and monthly moving averages of y_t.

We claim that, even though the HAR model explains successfully a large part of long-memory in financial volatilities, some non-negligible part of long-memory remains unexplained. The unexplained long-memory problems of HAR models are discussed in terms of stationarity of the estimated HAR model and biases of estimated HAR coefficients. Firstly, since model (1) is a special case of AR(22) model, the autocorrelation function (ACF) decays to zero exponentially for large lag if $\phi_d + \phi_w + \phi_m < 1$ and other technical conditions for stationarity hold. Obviously, the exponential decay of the ACF is faster than the algebraic decay of the ACFs of long-memory processes for financial volatilities. Secondly, the estimated HAR coefficients are biased downward. We note in Shaman and Stine (1988) and Tanaka (1984) that the ordinary least squares (OLS) estimate $\hat{\phi}_1$ based on a sample of size n for an AR(1) model $y_{t+1} = \phi_0 + \phi_1 y_t + \epsilon_{t+1}$ has the downward bias $E[\hat{\phi}_1 - \phi_1] = -(1 + 3\phi_1)/n + O(n^{-2})$. Similarly, the sum of estimated HAR coefficients is also downwardly biased. Both the exponential decay of the ACF of the HAR model and the downward bias of the estimated HAR coefficients imply that the estimated HAR models are less long memory than they should be for forecasting long-memory volatilities. Therefore, long-term forecasts from the HAR models remain to be improved.

Usually, since financial volatilities are long memory, the sum of the estimated HAR coefficients is close to 1. For examples, we have 0.93, 0.96, 0.95 for the KOSPI (Korean stock price index), the Korea won—US dollar exchange rate, and the US S&P500 realized volatilities, respectively, in Park and Shin (2014); 0.97 for the US S&P500 realized volatility in Busch et al. (2011). Noting that these estimates are underestimated, together with "less-long-memory" property of the estimated stationary HAR models, we consider an alternative model having 1 for the sum of the HAR coefficients for forecasting long-memory volatilities. The model is an integrated model as given by

$$y_{t+1} = \phi_0 + \phi_d y_t + \phi_w y_t^w + \phi_m y_t^m + \epsilon_{t+1}, \qquad \phi_d + \phi_w + \phi_m = 1, \tag{2}$$

which will be called an IHAR model.

Some recent papers such as Hwang and Shin (2013), Song and Shin (2015), and Varneskov and Perron (2015) report presence of level shifts in volatility which makes the volatility process nonstationary. Such nonstationarity would be more well-captured by the IHAR model than the HAR model.

The IHAR forecast model may be more motivated by the high acceptance rate of the ADF (Augmented Dickey–Fuller) test against long-memory nonstationary fractional integrations. Bisaglia and Procidano (2002) and many others reported that the ADF tests fail to detect nonstationarity of fractional integration $FI(d)$

$$(1 - B)^d y_t = a_t$$

with $0.5 < d < 1$, where B is the back-shift operator such that $By_t = y_{t-1}$ and a_t is a white noise. We report a Monte-Carlo acceptance rate of the level 5% ADF test for model (1) when data are generated from nonstationary fractional