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
1226-3192/

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

Table 1
Acceptance rate of ADF test for model (1) when DGP is $FI(d)$.

n	d					
	0.5	0.6	0.7	0.8	0.9	1
1250	0.06	0.27	0.56	0.78	0.90	0.95
2500	0.00	0.07	0.33	0.67	0.88	0.95
5000	0.00	0.01	0.16	0.55	0.84	0.95

integration $FI(d), d = 0.5, 0.6, \ldots, 1$, for sample size $n = 1250, 2500, 5000$, (5 years, 10 years, 20 years). The $FI(d)$ process is simulated using the procedure described in Section 3. The ADF test is constructed from HAR model (1) and is $ADF = (\hat{\phi}_d + \hat{\phi}_w + \hat{\phi}_m - 1)/se(\hat{\phi}_d + \hat{\phi}_w + \hat{\phi}_m)$. The simulation results are almost the same as those of the ADF test constructed from the AR(22) model. Table 1 shows acceptance rate of the ADF test based on 10,000 replications. It reveals that the unit root hypothesis is accepted at high frequency for d close to 1. Diebold and Kilian (2000) demonstrated that pretesting unit root generally improves forecast accuracy. From Table 1, we know that, for d close to 1, pretesting results would be almost the same as imposing unit root. This indicates that, for data sets from nonstationary fractional integration with d close to 1, a unit root HAR model (2) is more suitable than a stationary HAR model (1) for forecasting long-memory data such as realized volatility.

It is useful to reparameterize (2) into an augmented Dickey–Fuller-type regression

$$y_{t+1} - y_t = \phi_0 + \phi_w(y_t^w - y_t) + \phi_m(y_t^m - y_t) + \epsilon_{t+1}, \tag{3}$$

from which we estimate ϕ_w, ϕ_m and their standard errors by OLS regression. For an estimate of ϕ_d and its standard error, we can apply the delta method to $\phi_d = 1 - \phi_w - \phi_m$ using the estimation results for $(\phi_w, \phi_m)'$ obtained from fitting (3). Instead, one may use another reparameterization $y_{t+1} - y_t^m = \phi_0 + \phi_d(y_t - y_t^m) + \phi_w(y_t^w - y_t^m) + \epsilon_{t+1}$ for $\hat{\phi}_d$ and its standard error. The reparameterization method may be simpler for practitioners than the delta-method if they use statistical softwares.

There is a large literature acknowledging long range dependence and proposing methods which accommodate the long range dependence, for example, Barigozzi, Brownlees, Giampiero, Gallo, and Veredas (2014) and Brownlees and Gallo (2010). The proposed IHAR model fits in this more general strand of the literature.

3. A Monte Carlo forecast comparison

This section compares forecast performance of the proposed IHAR model with that of the HAR model for a class of long-memory processes: fractional integration $FI(d)$ given by $(1 - B)^d y_t = a_t$ with fractional integration parameter $d = 0.6, 0.7, 0.8, 0.9$, where a_t is a white noise process. We choose the fractional integration because it is frequently used for modeling financial volatility, see Kellard, Dunis, and Sarantis (2010) for example. In addition to the long memory structure, we consider another structure of GARCH(1,1) for the error process in order to address the conditional heteroscedasticity of realized volatility pointed out by Corsi, Mittnik, Pigorsch, and Pigorsch (2008) and others. We note the observations of Corsi (2009, Figure 5) and Hwang and Shin (2013, Figure 1) that residuals from HAR fittings reveal strong volatility clustering, see also the lower block of Table 5.

The artificial data y_t are recursively computed using a large lag AR approximation

$$y_t = -\sum_{j=1}^{L} \pi_j y_{t-j} + a_t, \quad a_t = \sigma_t \epsilon_t, \qquad \sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2, \qquad \pi_j = \Gamma(j-d)/\{\Gamma(-d)\Gamma(j+1)\}$$

with $L = 1000$ to the $FI(d)$ model. The π coefficients are recursively computed by $\pi_j = \pi_{j-1}(j-1-d)/j, j = 1, 2, 3, \ldots, \pi_0 = 1$. In order to remove the start-up effect, $\{y_t, t = -m, -m+1, \ldots, 0, 1, \ldots, n\}$ are generated with $y_{-m} = y_{-m-1} = \cdots = y_{-m-L} = 0$ and $m = 50{,}000$ and $\{y_t, t = 1, \ldots, n\}$ is used for analysis. For the GARCH parameters, we consider two cases $(\alpha_0 = 1, \alpha_1 = 0, \beta_1 = 0)$ and $(\alpha_0 = 0.1, \alpha_1 = 0.5, \beta_1 = 0.4)$ which will be denoted by IID and GARCH, respectively. Note that IID and GARCH errors are homoscedastic and conditionally heteroscedastic, respectively.

Artificial long-memory data sets $y_t, t = 1, \ldots, n$ are simulated with independent standard normal errors ϵ_t generated by RNNOA, a FORTRAN subroutine in IMSL library for $n = 1250, 2500, 5000$. For each sample $\{y_t, t = 1, \ldots, n\}$, h-step ahead forecasts $\hat{y}_{n+h|n}$ are computed from estimated HAR model and estimated IHAR model for $h = 1, 5, 22, 66, 250$ which correspond to a day, a week, a month, a quarter, and an year respectively. Note that one year forecast is not rare in financial risk analysis. For example, one may be interested in one-year VaR (value at risk) for a financial position for which one need volatility forecasts of the log-returns up to one-year ahead. This procedure is repeated $K = 10{,}000$ times independently to produce a set of K independent h-step ahead forecasts $\hat{y}_{n+h|n}^{(k)}, k = 1, \ldots, K$.

For each model HAR and IHAR, we first compute the forecast MAE(h) (mean absolute error) of the h-step ahead forecasts $\hat{y}_{n+h|n}$ computed from the model as given by $MAE(h) = \sum_{k=1}^{K} |\hat{y}_{n+h|n}^{(k)} - y_{n+h}|/K$. Table 2 reports MAE efficiency of the IHAR forecast relative to the HAR forecast given by

$$\text{Relative MAE efficiency} = \text{MAE of the HAR forecasts/MAE of the IHAR forecasts}. \tag{4}$$

Table 2
MAE efficiency of the IHAR forecasts relative to the HAR forecasts.

h	DGP: FI(d)–IID				DGP: FI(d)–GARCH			
	d							
	0.6	0.7	0.8	0.9	0.6	0.7	0.8	0.9
$n = 1250$ (5 years)								
1	0.995	0.997	0.999	1.000	0.998	1.000	1.001	1.002
5	0.994	0.999	1.003	1.005	0.995	1.003	1.010	1.014
22	0.974	0.992	1.009	1.020	0.975	0.997	1.016	1.031
66	0.964	1.000	1.029	1.050	0.975	1.016	1.046	1.068
250	0.900	0.959	1.018	1.074	0.905	0.972	1.035	1.096
$n = 2500$ (10 years)								
1	0.999	1.001	1.001	1.001	1.000	1.001	1.001	1.001
5	0.995	1.000	1.003	1.003	0.994	1.000	1.004	1.007
22	0.977	0.992	1.003	1.010	0.976	0.994	1.007	1.016
66	0.967	0.995	1.012	1.023	0.976	1.005	1.022	1.035
250	0.961	1.023	1.060	1.080	0.973	1.034	1.065	1.084
$n = 5000$ (20 years)								
1	0.998	0.999	1.000	1.000	0.997	0.999	1.000	1.000
5	0.995	0.999	1.000	1.001	0.993	0.998	1.002	1.002
22	0.971	0.987	0.996	1.002	0.975	0.993	1.003	1.009
66	0.982	1.001	1.010	1.014	1.005	1.023	1.027	1.028
250	1.018	1.072	1.073	1.061	1.032	1.089	1.087	1.073

Note: Efficiency value larger than 1 implies better IHAR forecast than HAR forecast.

Table 3
Basic statistics of the RVs.

	Size(n)	Min.	Median	Max.	Mean	St.dev	Skewness	Kurtosis
S&P500	4627	11.17	69.18	760.48	81.20	51.69	3.33	20.59
NASDAQ	4629	20.60	99.21	847.85	123.95	80.44	2.32	8.77
JPY/USD	4765	0.77	36.98	536.91	40.48	19.49	6.00	109.65
EUR/USD	4299	1.26	37.40	286.51	39.95	16.00	2.25	17.97

Note: Unit = bp, 1bp = 0.01%. The scale BP is applied only for Min, Median, Max, Mean, and St.dev.

From Table 2, we see efficiency gain of the IHAR model over the HAR model for strong long memory cases. Consider first the strong long memory cases of $d = 0.8, 0.9$. We note that relative efficiency is generally greater than 1. The efficiency is larger for larger h or for GARCH error. For $d = 0.7$, efficiency of the IHAR model is almost the same as that of the HAR model. On the other hand, for $d = 0.6$, we see some efficiency loss for the IHAR model relative to the HAR model.

Interestingly, even for the 1-step forecast of $h = 1$, we observe better IHAR forecasts than HAR forecasts for d close to 1, which is more significant for the conditionally heteroscedastic GARCH error than for the homoscedastic IID error. The extra improvement for the GARCH error case is more significant for d close to 1, for larger h, or for smaller T.

From this Monte Carlo efficiency study, we can say that the IHAR model produces better long-term forecasts than the HAR model for a strong long memory processes while it looses forecast efficiency for weak long memory processes. Conditional heteroscedasticity of error, i.e. conditionally heteroscedastic volatility of realized volatility, gives extra improvements for the IHAR forecasts.

4. A real data set out-of-sample forecast comparison

We compare out-of-sample forecast performance of the IHAR model with that of the HAR model as well as those of random walk model and fractional integration model. We choose daily realized volatilities (RVs) of 2 stock price indices and 2 foreign exchange rates: the US S&P 500 index (S&P500), the US NASDAQ index, the Japan yen/US dollar exchange rate (JPY/USD), and the EU euro/US dollar exchange rate (EUR/USD). The data period is Jan. 1 1999–May 26 2015 for the EUR/USD and Jan. 1 1997–May 26 2015 for the other assets. The data sets are purchased from tickdatamarket (www.tickdatamarket.com). The RV of a given asset for a given day is the square root of the sum of squares of 5 min intra-day log-returns. Plots of the RV series are displayed in Fig. 1. We will apply models (1) and (2) for forecasting $y_t = RV_t$.

4.1. Preliminary analysis

Basic statistics of the RVs are reported in Table 3. We see some asymmetries for all the 4 RVs as measured by (min, median, max) = (11.17, 69.18, 760.48) and skewness = 3.33 for S&P500 for example. We also observe non-normality of sharp central tendency as measured by kurtosis = 20.59 for S&P500 for example.

Table 4
Fractional integration parameter ($\hat{d}$) and tests for nonstationarity for the RVs.

	$\hat{d}$	$se(\hat{d})$	$\hat{d}/se(\hat{d})$	$(\hat{d}-0.5)/se(\hat{d})$
S&P500	0.588	0.029	20.485	3.063
NASDAQ	0.538	0.028	18.903	1.350
JPY/USD	0.402	0.029	14.107	−3.433
EUR/USD	0.398	0.031	12.742	−3.284

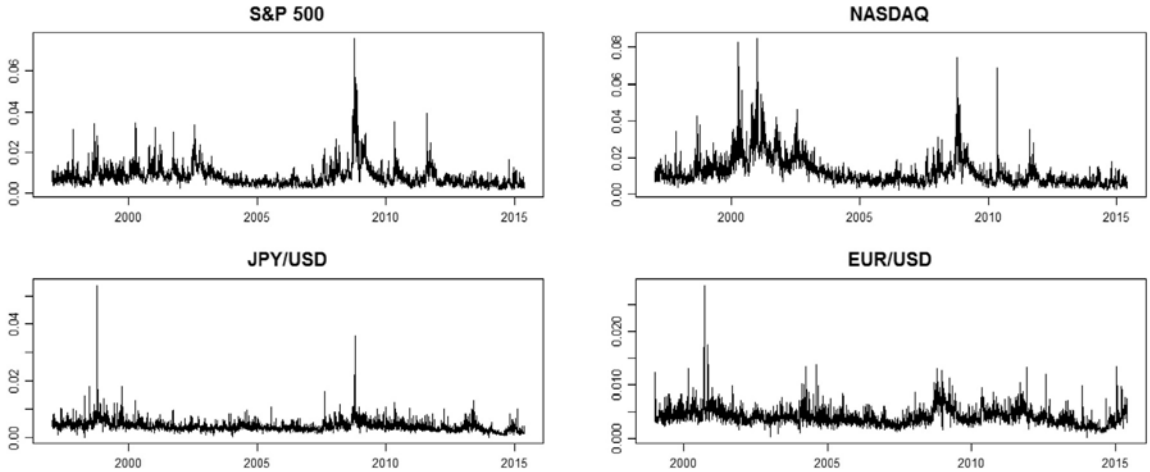


Fig. 1. Time series plots of the realized volatility series.

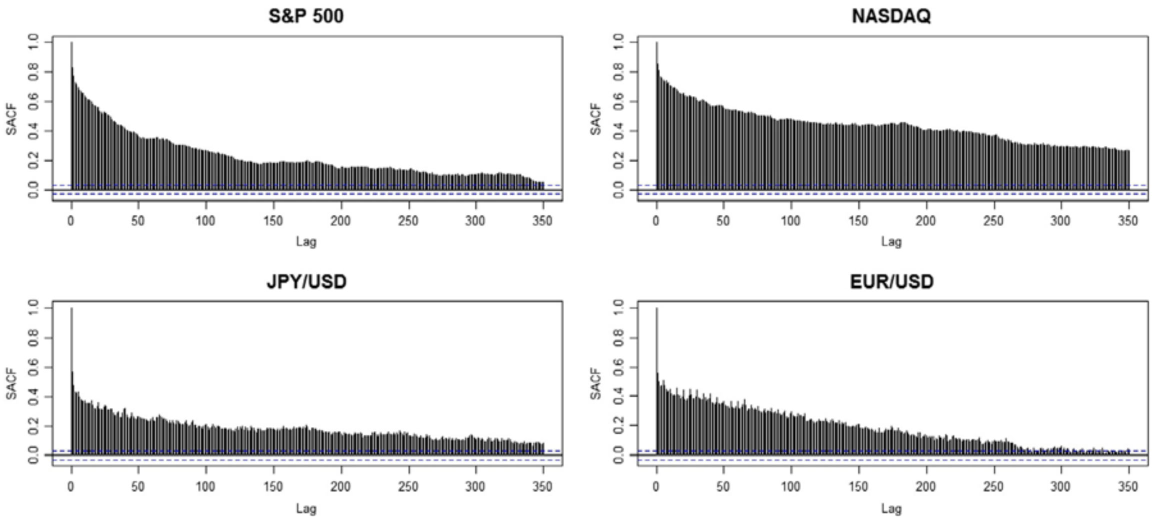


Fig. 2. Sample ACF plots of the realized volatility series.

Long-memory features of the RVs are analyzed by the sample autocorrelation function (SACF) in Fig. 2 and estimated fractional integration parameter $\hat{d}$ in Table 4. The SACFs decline to zero very slowly implying persistent long memories. The fractional integration parameter d is estimated by the method of Geweke and Porter-Hudak (1983), called GPH, for which we use the tuning parameter $n^{0.75}$ as Kellard et al. (2010) did. The RVs of the stock prices have estimated d values somewhat greater than 0.5, indicating nonstationarity while the RVs of the exchange rates have $\hat{d}$ somewhat smaller than 0.5 indicating stationarity.

Tests for nonstationarity are conducted using the GPH tests as reported in Table 4. For the S&P500, the GPH test $(\hat{d}-0.5)/se(\hat{d})$ rejects the null hypothesis $d \leq 0.5$ against nonstationarity $d > 0.5$ at 1% level. For the other assets, the GPH test $(\hat{d}-0.5)/se(\hat{d})$ does not reject stationarity.

Sample estimation results for RVs are reported in Table 5. The IHAR model reveals much smaller constant estimates with smaller standard errors than those of the HAR model. The daily coefficients of the IHAR models are similar to those

Table 5
In-sample estimation results of the HAR and the IHAR models.

	S&P500		NASDAQ		JPY/USD		EUR/USD	
	HAR	IHAR	HAR	IHAR	HAR	IHAR	HAR	IHAR
Estimation results								
Const. $\times 10^4$	4.166***	−0.030	4.507***	−0.036	4.316***	−0.037	3.383***	0.012
	(0.852)	(0.4)	(1.162)	(0.571)	(0.723)	(0.219)	(0.680)	(0.183)
Day	0.447***	0.452***	0.443***	0.446***	0.265***	0.268***	0.170***	0.172***
	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)	(0.018)	(0.018)
Week	0.329***	0.329***	0.287***	0.288***	0.263***	0.265***	0.332***	0.335***
	(0.026)	(0.026)	(0.027)	(0.027)	(0.031)	(0.031)	(0.035)	(0.035)
Month	0.173***	0.220***	0.233***	0.266***	0.364***	0.466***	0.414***	0.493***
	(0.021)	(0.019)	(0.022)	(0.02)	(0.031)	(0.026)	(0.034)	(0.030)
Residual analysis by GARCH(1,1) fitting								
$\hat{\alpha}_0 \times 10^7$	1.338***	1.420***	2.550***	2.603***	0.499***	0.398***	0.348***	0.347***
	(0.199)	(0.204)	(0.366)	(0.360)	(0.094)	(0.076)	(0.052)	(0.051)
$\hat{\alpha}_1$	0.145***	0.144***	0.198***	0.192***	0.092***	0.087***	0.061***	0.064***
	(0.013)	(0.013)	(0.014)	(0.014)	(0.009)	(0.008)	(0.006)	(0.007)
$\hat{\beta}_1$	0.843***	0.842***	0.810***	0.812***	0.893***	0.903***	0.917***	0.915***
	(0.012)	(0.013)	(0.012)	(0.012)	(0.011)	(0.010)	(0.008)	(0.008)

Note: $(\hat{\alpha}_0, \hat{\alpha}_1, \hat{\beta}_1)$ are the GARCH(1,1) parameters of the residual. Numbers in parentheses are standard errors.
*** Significance at the 0.1% level.

of the HAR models and so are the weekly coefficients. On the other hand, the monthly coefficients of the IHAR models are non-negligibly greater than those of the HAR models. This indicates that the monthly term plays a more important role in the IHAR model than in the HAR model for forecast, which produces a much better long-term IHAR forecasts than the HAR forecasts as will be investigated in the following subsection.

The residuals have strong conditional heteroscedasticity as in Corsi et al. (2008). The GARCH parameters $(\hat{\alpha}_1, \hat{\beta}_1)$ are highly significant and indicate strong persistency in volatility movements as indicated by the near unity sums $\hat{\alpha}_1 + \hat{\beta}_1$. As discussed in the Monte Carlo experiment in Section 3, this point is related with improved forecast performance of the IHAR model relative to the HAR model. This issue will be more discussed in the following subsection.

4.2. Forecast comparison

Let an asset be given which is one of the S&P500, the NASDAQ, the JPY/USD, or the EUR/USD. Forecast performances of the HAR model (1), the IHAR model (2), random walk (RW) model $y_{t+1} = y_t + \epsilon_{t+1}$, and fractional integration (FI) model $(1-B)^d y_{t+1} = \phi_0 + \epsilon_{t+1}$ are compared for predicting the last m observations where m is 15% of the whole observation, $m = 0.15n$, and n is the number of observations of the asset given in Table 3. Starting from the day corresponding to $t = 0.85n$, for each $T = 0.85n, 0.85n+1, \ldots, n-h$, out-of-sample h-step ahead forecasts $\hat{y}_{T+h|T}$ are constructed from each of the 4 models estimated using data set $\{y_1, \ldots, y_T\}$, $h = 1, 5, 22, 66, 250$. The performance criteria MAE(h), RMSE(h), and MAPE(h) are computed for each model and h, which are

$$MAE(h) = M^{-1} \sum_{T=0.85n}^{n-h} |\hat{y}_{T+h|T} - y_{T+h}|, \qquad RMSE(h) = \sqrt{M^{-1} \sum_{T=0.85n}^{n-h} (\hat{y}_{T+h|T} - y_{T+h})^2},$$

$$MAPE(h) = 100 \times M^{-1} \sum_{T=0.85n}^{n-h} |\hat{y}_{T+h|T} - y_{T+h}|/y_{T+h},$$

where $M = n - h + 1 - 0.85n = 0.15n - h + 1$. The MAE efficiency of the IHAR forecast relative to the HAR forecast is given by (4) and other efficiencies are defined similarly.

Table 6 reports efficiencies of the IHAR forecasts relative to the forecasts based on the other models HAR, RW, and FI. We observe that, for all the 4 RVs, the IHAR model tends to produce uniformly best forecasts for all h considered here in all the three performance criteria MAE, RMSE, MAPE.

The left block of Table 6 shows efficiencies of the IHAR forecasts relative to the HAR forecasts. In all the 3 performance measures, we see that all the efficiency values are greater than 1, indicating better forecast performance of the IHAR model over the HAR model, and that the efficiency values increase as h increases.

In MAE performance, the efficiency values for (S&P500, NASDAQ, JPY/USD, EUR/USD) are: (1.514, 1.692, 1.231, 1.354), substantially greater than 1 for the three month forecasts of $h = 66$; (1.136, 1.152, 1.233, 1.181), meaningfully greater than 1 for the monthly forecast of $h = 22$; (1.032, 1.042, 1.111, 1.078), somewhat greater than 1 for the weekly forecast of $h = 5$;

Table 6
MAE, RMSE, and MAPE efficiencies of the IHAR forecasts relative to benchmark models HAR, random walk (RW), and fractional integration (FI).

h	HAR/IHAR			RW/IHAR			FI/IHAR		
	MAE	RMSE	MAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE
S&P 500									
1	1.021***	1.002	1.063	1.084***	1.118***	1.057	0.996	0.998	1.001
5	1.032^{a}	1.005	1.112	1.138***	1.162***	1.126	1.069**	1.084**	1.067
22	1.136**	1.028	1.314	1.167***	1.173***	1.152	1.129***	1.134***	1.124
66	1.514***	1.309**	1.814	1.172***	1.169***	1.152	1.152***	1.148***	1.139
250	2.152***	1.748***	2.511	1.163***	1.138**	1.171	1.155	1.132**	1.169
NASDAQ									
1	1.024***	1.005	1.063	1.069**	1.111***	1.047	0.997	0.993	1.012
5	1.042*	1.015	1.116	1.151***	1.179***	1.137	1.059*	1.082**	1.061
22	1.152**	1.062	1.310	1.163***	1.179***	1.157	1.118***	1.128***	1.124
66	1.692***	1.418***	1.984	1.195***	1.169***	1.180	1.166***	1.140***	1.163
250	2.670***	2.172***	3.238	1.108***	1.104**	1.123	1.101	1.093*	1.128
JPY/USD									
1	1.061***	1.007	1.154	1.233***	1.261***	1.163	1.040***	1.018**	1.091
5	1.111***	1.024^{a}	1.263	1.269***	1.295***	1.211	1.115***	1.118***	1.135
22	1.233*	1.079	1.502	1.265***	1.297***	1.210	1.188***	1.193***	1.188
66	1.231	1.083	1.543	1.133**	1.186*	1.148	1.092*	1.125^{a}	1.133
250	1.322	1.108	1.354	1.077	1.146**	1.096	1.056	1.113*	1.084
EUR/USD									
1	1.044***	1.002	1.108	1.271***	1.253***	1.228	1.071***	1.019	1.126
5	1.078***	1.008	1.188	1.271***	1.254***	1.230	1.127***	1.086*	1.158
22	1.181*	1.042	1.379	1.348***	1.336***	1.311	1.251***	1.229***	1.259
66	1.354	1.138	1.728	1.170***	1.178**	1.166	1.142**	1.132*	1.177
250	1.125	1.004	1.437	1.015	1.048	1.025	1.000	1.024	1.032

Note: Efficiency value larger than 1 indicates better forecast accuracy of the IHAR forecast than the benchmark model forecast.
[a] Significance at the 10% level by the Diebold–Mariano test.
[*] Significance at the 5% level by the Diebold–Mariano test.
[**] Significance at the 1% level by the Diebold–Mariano test.
[***] Significance at the 0.1% level by the Diebold–Mariano test.

(1.021, 1.024, 1.061, 1.044), still greater than 1 in the 1-step forecast. We see similar relative performances in the other RMSE and MAPE performances.

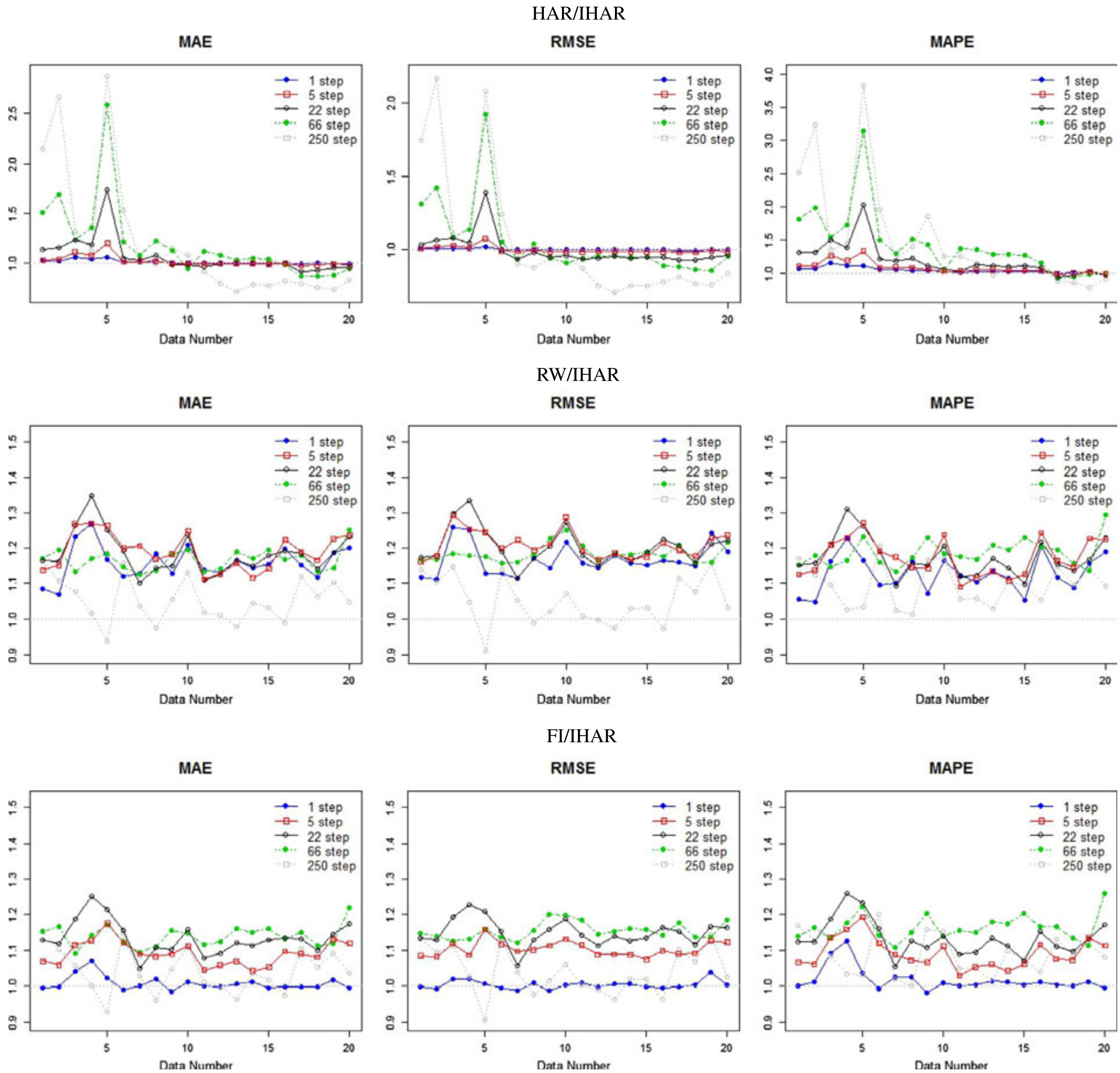
Even for the 1-step forecast, the IHAR model is better than the HAR model. A reason for this improvement is explained by observing the highly significant estimated GARCH(1,1) parameters of the residuals in the bottom block of Table 5, whose sums are all close to 1. Therefore, as observed in the Monte Carlo comparison in Section 3, conditional heteroscedasticity in the residual would have produced better 1-step forecast for the IHAR model than for the HAR model.

Table 6 also reports statistical significances of the MAE and RMSE efficiencies of the IHAR forecasts over the HAR forecasts computed from the Diebold–Mariano test of Diebold and Mariano (1995). For the long-run variance estimates of the mean loss-differences in the Diebold–Mariano tests for h-step ahead forecasts, we use sample autocovariances of the loss-differences for lags up to $(h-1)$ with unit kernel. Consider first the RVs of the stock prices S&P and NASDAQ: in the MAE performance, we see statistical significances for all h; in the RMSE performance, we see high significance for $h = 66, 250$ but no significance for $h = 1, 5, 22$. Consider next the RVs of the JPY/USD and the EUR/USD: in the MAE performance, we note statistical significances for $h = 1, 5, 22$; in the RMSE performance, we see some significance for $h = 5$.

We may explain the improved long-term forecasts from the IHAR models over those from the HAR models by three reasons. One reason is that the long-memories of the RVs are so much more persistent than the memories represented by the estimated stationary HAR models that the IHAR models with unit root are more consistent with the long memory RVs than the estimated HAR models. On the other hand, in the estimated HAR models, the long-term forecasts regress to the global means at more rapid speeds than those implied by the long-memories of the RVs, resulting in the relative poor long-term forecasts. Another reason is that, as verified in Section 3, conditionally heteroscedastic volatility of RV shown in the lower block of Table 5 is a feature which the IHAR model takes better advantage of than the HAR model in forecasting. A third reason is that, as mentioned in the later paragraph of the preceding subsection, the improved monthly coefficient estimates in the IHAR model capture the strong long-memory dynamics of the RVs better than those of the HAR model.

The middle block of Table 6 shows efficiencies of the IHAR forecasts relative to RW forecasts, which are the last value of RVs for all h. Note that random walk is a simple AR(1) model with a unit root. For $h = 1, 5, 22, 66, 250$, for all 4 RVs, IHAR forecasts are better than RW forecasts in all measures MAE, RMSE, and MAPE.

The right block of Table 6 displays efficiencies of the IHAR forecasts relative to FI forecasts. We note general uniform dominance of IHAR forecasts over the FI forecasts.



Note: Efficiency value larger than 1 indicates better IHAR forecast than the corresponding benchmark model forecast.

Fig. 3. Relative efficiencies of h-step ahead out-of-sample IHAR forecasts for the 20 indexes in Oxford-Man library.

We can say that the IHAR model produces the best forecasts among the 4 models for all h considered here in all measures. For short-term forecast of $h = 1, 5$, we see that HAR is the next best one followed by FI and next by the worst RW. For long-term forecast of $h = 66, 250$, FI is the second best, RW is the third best and HAR seems to be the worst.

5. An extended comparison

This section makes a forecast comparison of the 4 models, HAR, IHAR, RW, and FI, for all the index related RV series contained in the realized library of Oxford-Man Institute which are freely available on the website (http://realized.oxford-man.ox.ac.uk/). The names of the index series are listed in Table 7. Data period is Jan. 3 2000–Dec. 15 2015. Sampling interval for RV is 5 min. Efficiencies of the IHAR forecasts relative to the HAR, RW, FI forecasts are computed in the same manner as those for Section 4.2 and are displayed in Fig. 3.

From Fig. 3, we see that the IHAR model has substantial forecast advantage over all the 3 other models HAR, RW, and FI for the extended RV data sets. More detailed investigations follow.

Table 7
Data number and description in Fig. 3.

Number	Asset	Number	Asset
1	S&P 500	11	Swiss Market Index
2	NASDAQ	12	AEX Index
3	JPY/USD	13	DAX
4	EUR/USD	14	CAC 40
5	KOSPI Composite Index	15	Euro STOXX 50
6	Russel 2000	16	Hang Seng
7	Nikkei 225	17	IBEX 35
8	FTSE 100	18	FTSE MIB
9	DJIA	19	Bovespa Index
10	IPC Mexico	20	All Ordinaries

The top part of Fig. 3 shows the efficiencies of IHAR forecasts relative to HAR forecasts. We observe

1. In MAPE performance, the IHAR model is substantially better than the HAR model for all the 20 assets.
2. In MAE performance, for the first 10 assets, the IHAR model is better than the HAR model; for the last 10 assets the IHAR model and the HAR model have similar 1, 5, 22 step forecast accuracy while the IHAR model has worse 66, 250 step forecasts than the HAR model.
3. In the RMSE performance, no model dominates the other model in forecast performance.

The middle and bottom parts of Fig. 3 show the efficiencies of the IHAR forecasts relative to the RW forecasts and relative to the FI forecasts, respectively. We observe the efficiency values are almost uniformly significantly larger than 1, implying uniformly significantly better the IHAR forecasts relative to the RW forecasts and relative to the FI forecasts for all h in all the 3 measures MAE, RMSE, and MAPE.

6. Conclusion

We have proposed an integrated HAR (IHAR) model whose HAR coefficient sum is one. Forecast advantages of the IHAR model over the existing HAR model and two benchmark models of random walk (RW) and fractional integration (FI) are demonstrated for 4 realized volatilities of US S&P500 index, NASDAQ index, the Japan yen/US dollar exchange rate, and the EU euro/US dollar exchange rate. The improvement is more conspicuous for long-term forecasts. The superior IHAR forecasts relative to the 3 other HAR, RW, FI forecasts are also justified by a Monte-Carlo simulation and by an empirical comparison for all 20 index volatilities in a realized volatility library of Oxford-Man Institute. The improved forecasts are due to the fact that, compared with the HAR model, the proposed IHAR model has better addressing of the long memory feature of the realized volatility and of the conditional heteroscedastic feature of the realized volatility. The spirit of imposing unit root may be applied to vector valued financial volatilities for which a vector error correction HAR model may have forecast advantage over a vector HAR model considered by Busch et al. (2011). Especially, it would be a good topic of future research to study vector error correction HAR model to improve the vector HAR forecasts of a vector of realized volatility and implied volatility by Busch et al. (2011).

Acknowledgments

The authors are very grateful for two anonymous referees whose constructive comments improve the paper significantly. This research was supported by the National Research Foundation of Korea (NRF) (2015-1006133).

References

Andersen, T. G., Bollerslev, T., & Diebold, F. X. (2007). Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility. *The Review of Economics and Statistics*, *89*(4), 701–720.
Andersen, T. G., Bollerslev, T., Diebold, F. X., & Ebens, H. (2001). The distribution of realized stock return volatility. *Journal of Financial Economics*, *61*, 43–76.
Andersen, T. G., Bollerslev, T., Diebold, F. X., & Labys, P. (2001a). *Modeling and forecasting realized volatility (No. w8160)*. National Bureau of Economic Research.
Andersen, T. G., Bollerslev, T., Diebold, F. X., & Labys, P. (2001b). The distribution of realized exchange rate volatility. *Journal of the American Statistical Association*, *96*(453), 42–55.
Barigozzi, M., Brownlees, C., Giampiero, M., Gallo, G., & Veredas, D. (2014). Disentangling systematic and idiosyncratic dynamics in panels of volatility measures. *Journal of Econometrics*, *182*, 364–384.
Bisaglia, L., & Procidano, I. (2002). On the power of the augmented Dickey–Fuller test against fractional alternatives using bootstrap. *Economics Letters*, *77*(3), 343–347.
Brownlees, C., & Gallo, G. (2010). Comparison of volatility measures: a risk management perspective. *Journal of Financial Econometrics*, *8*, 29–56.
Busch, T., Christensen, B. J., & Nielsen, M. (2011). The role of implied volatility in forecasting future realized volatility and jumps in foreign exchange, stock, and bond markets. *Journal of Econometrics*, *160*(1), 48–57.
Corsi, F. (2004). A simple long memory model of realized volatility. Manuscript, University of Southern Switzerland.
Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, *7*(2), 174–196.
Corsi, F., Mittnik, S., Pigorsch, C., & Pigorsch, U. (2008). The volatility of realized volatility. *Econometric Reviews*, *27*(1-3), 46–78.
Corsi, F., Pirino, D., & Reno, R. (2010). Threshold bipower variation and the impact of jumps on volatility forecasting. *Journal of Econometrics*, *159*(2), 276–288.

Diebold, F. X., & Kilian, L. (2000). Unit-root tests are useful for selecting forecasting models. *Journal of Business & Economic Statistics*, *18*, 265–273.
Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, *13*(3), 134–144.
Geweke, J., & Porter-Hudak, S. (1983). The estimation and application of long-memory time series models. *Journal of Time Series Analysis*, *4*, 221–238.
Hwang, E., & Shin, D. W. (2013). A CUSUM test for a long memory heterogeneous autoregressive model. *Economics Letters*, *121*, 379–383.
Hwang, E., & Shin, D. W. (2014). Infinite-order, long-memory heterogeneous autoregressive models. *Computational Statistics & Data Analysis*, *76*, 339–358.
Hwang, E., & Shin, D. W. (2015). A CUSUMSQ test for structural breaks in error variance for a long memory heterogeneous autoregressive model. *Statistics & Probability Letters*, *99*, 167–176.
Kellard, N., Dunis, C., & Sarantis, N. (2010). Foreign exchange, fractional cointegration and the implied-realized volatility relation. *Journal of Banking & Finance*, *34*(4), 882–891.
McAleer, M., & Medeiros, M. C. (2008). Realized volatility: A review. *Econometric Reviews*, *27*(1-3), 10–45.
Park, S., & Shin, D. W. (2014). Modeling and forecasting realized volatilities of Korean financial assets featuring long memory and asymmetry. *Asia-Pacific Journal of Financial Studies*, *43*(1), 31–58.
Shaman, P., & Stine, R. A. (1988). The bias of autoregressive coefficient estimators. *Journal of the American Statistical Association*, *83*(403), 842–848.
Song, H., & Shin, D. W. (2015). Long-memories and mean breaks in realized volatilities. *Applied Economics Letters*, *22*, 1273–1280.
Tanaka, K. (1984). An asymptotic expansion associated with the maximum likelihood estimators in ARMA models. *Journal of the Royal Statistical Society*, *46*(B), 58–67.
Varneskov, R.T., & Perron, P. (2015). Combining long memory and level shifts in modeling and forecasting the volatility of asset returns, Working Paper.
Yun, S., & Shin, D. W. (2015). Forecasting the realized variance of the log-return of Korean won US dollar exchange rate addressing jumps both in stock-trading time and in overnight. *Journal of the Korean Statistical Society*, *44*(2), 390–402.