

Neural network heterogeneous autoregressive models for realized volatility

Jaiyool Kim^a, Changryong Baek^{1,a}

^aDepartment of Statistics, Sungkyunkwan University, Korea

Abstract

In this study, we consider the extension of the heterogeneous autoregressive (HAR) model for realized volatility by incorporating a neural network (NN) structure. Since HAR is a linear model, we expect that adding a neural network term would explain the delicate nonlinearity of the realized volatility. Three neural network-based HAR models, namely HAR-NN, HAR(∞)-NN, and HAR-AR(22)-NN are considered with performance measured by evaluating out-of-sample forecasting errors. The results of the study show that HAR-NN provides a slightly wider interval than traditional HAR as well as shows more peaks and valleys on the turning points. It implies that the HAR-NN model can capture sharper changes due to higher volatility than the traditional HAR model. The HAR-NN model for prediction interval is therefore recommended to account for higher volatility in the stock market. An empirical analysis on the multinational realized volatility of stock indexes shows that the HAR-NN that adds daily, weekly, and monthly volatility averages to the neural network model exhibits the best performance.

Keywords: realized volatility (RV), heterogeneous autoregressive (HAR) model, neural network (NN), long memory

1. Introduction

The rapid development of technology for handling high-frequency transaction data in finance has opened a new era in the volatility modeling domain. Rather than considering the closing price to obtain traditional returns, the sum of intraday square returns called realized volatility (RV) is considered a more precise approximation of volatility. In this regard, this study considers the statistical volatility modeling of financial markets based on RV.

The RV also shows stylized facts that are similar to a traditional daily log-return as a proxy for volatility. It also includes high persistency, time-varying conditional variance, and non-Gaussianity; see for example, Andersen *et al.* (2001a, 2001b), Barndorff-Nielsen and Shephard (2002a, 2002b), Corsi (2009) and the references therein. Therefore, long memory models such as the autoregressive fractionally integrated moving average (ARFIMA) model have been applied to RV data. On the other hand, Corsi (2009) proposed a simple linear model called the heterogeneous autoregressive (HAR) model composed of three heterogeneous components of RV obtained at a different time interval. They suggested daily, weekly, and monthly averages in HAR, and it is widely reported as exhibiting a superior forecasting performance with its simple estimation procedure. However, the recent increase in computing capacity shows the need to improve HAR models. One apparent method of improvising is to generalize HAR to non-linear models using a neural network (NN).

¹ Corresponding author: Department of Statistics, Sungkyunkwan University, 25-2 Sungkyunkwan-ro, Jongno-Gu, Seoul 03063, Korea. E-mail: crbaek@skku.edu

This is not the first study to combine a neural network with the HAR model. It has already been considered in Hillebrand and Medeiros (2010), McAleer and Medeiros (2011). However, these studies focused on ensembling estimates using bagging, and hence it has a slightly different perspective. In this study, we delve deeper into the neural network modeling of HAR for RV. This is because the application of a neural network entails tuning parameters, such as the number of layers and hidden units, to ensure that the bias-variance trade-off between the number of parameters determines forecasting performance.

The HAR model is a constrained autoregressive (AR) model; therefore, we work with the feed-forward AR neural network. We consider a three-neural-network based HAR models, depending on which variables enter the model. We considered the original HAR model with daily, weekly, and monthly averages, and the infinite HAR model suggested by Hwang and Shin (2014) for a better approximation of high persistence. Finally, the hybrid of the first and second models are considered to estimate a moderate number of parameters. The proposed models are fully tested in terms of out-of-sample forecasting over 11 major stock indices globally.

The paper is organized as follows. In Section 2, the RV and HAR models are briefly recalled; the neural-network based HAR models are proposed in Section 3. We evaluate the out-of-sample forecasting errors of the proposed models over 11 multinational RVs in Section 4. Some properties of residuals obtained from fitting NN-based HAR models to real data set are investigated in Section 5. Empirical findings on the neural network HAR models are discussed and concluded in Section 6.

2. Realized volatility and HAR model

Let the price of the financial index, such as stock, and exchange rate of interest at t^{th} day be P_t . Subsequently, the (daily) log-return is defined as

$$r_t = \log \frac{P_t}{P_{t-1}} = \log P_t - \log P_{t-1}.$$

The volatility model can be represented as

$$r_t = \sigma_t \cdot \epsilon_t, \quad \epsilon_t \sim \mathcal{WN}(0, 1),$$

to ensure that the log-return changes over time and is indexed by σ_t . However, the rapid development of data observation and storage technologies has facilitated the acquisition of log-return in almost real-time. Therefore, we can consider the continuous analogue of the volatility model in high frequency. Subsequently, the log-return for Δ frequencies, which are cuts of M equidistant intervals in one day, is given as:

$$r_{(t-j\Delta)} = \log P_{(t-j\Delta)} - \log P_{(t-(j+1)\Delta)}. \quad (2.1)$$

The integrated volatility for one day is given as the square root of the quadratic variation as:

$$\text{IV}_t^{(d)} = \sqrt{\int_t^{t+1d} \sigma^2(w)dw}, \quad (2.2)$$

and Andersen *et al.* (2003) show that it can be well approximated by the RV given by

$$\text{RV}_t^{(d)} = \sqrt{\sum_{j=0}^{M-1} r_{(t-j\Delta)}^2}. \quad (2.3)$$

Likewise, the daily return RV also shares similar stylized facts such as high persistence in auto-correlations also known as long memory, non-Gaussianity, and heterogeneous conditional variance. Among many models incorporating such features, high-persistence is the key feature in understanding and modeling the RV. For example, the long memory model, such as ARFIMA, is a popular model considered in the literature. Alternatively, Corsi (2009) suggested a very simple linear model approximating the long memory model. Define the weekly and monthly RV (denoted by $RV_t^{(w)}$ and $RV_t^{(m)}$, respectively) as

$$RV_t^{(w)} = \frac{1}{5} (RV_t^{(d)} + RV_{t-1d}^{(d)} + \dots + RV_{t-4d}^{(d)}), \tag{2.4}$$

$$RV_t^{(m)} = \frac{1}{22} (RV_t^{(d)} + RV_{t-1d}^{(d)} + \dots + RV_{t-21d}^{(d)}), \tag{2.5}$$

where $RV_t^{(d)}$ is considered as today's RV. Subsequently, the HAR model is given by

$$RV_{t+1d}^{(d)} = c + \beta^{(d)}RV_t^{(d)} + \beta^{(w)}RV_t^{(w)} + \beta^{(m)}RV_t^{(m)} + \omega_{t+1d}, \quad \omega_{t+1d} \sim \mathcal{WN}(0, \sigma^2). \tag{2.6}$$

Therefore, the HAR model captures the long memory feature by considering local averages of the immediate past, moderate (weekly), and long (monthly) history. It is also observed that the HAR is a constrained AR(22) model; therefore, the estimation and inference is straightforward when using autoregressive moving average (ARMA) modeling approaches.

Hwang and Shin (2014) proposed infinite order HAR, HAR(∞), by allowing the infinite order autoregressive terms, namely

$$RV_{t+1d}^{(d)} = c + \sum_{j=0}^{\infty} \beta_j RV_{t-jd}^{(d)} + \omega_{t+1d}, \quad \omega_{t+1d} \sim \mathcal{WN}(0, \sigma^2). \tag{2.7}$$

This model is a long memory model under suitable conditions on parameters $\{\beta_j, j \in \{N \cup \{0\}\}\}$.

3. Neural network-based HAR models

In this section, we propose the neural network-based HAR models. The basic framework is the feed-forward AR(p)-NN model with a single hidden layer and q hidden units. The AR(p)-NN model is represented as

$$y_t = \beta_0' \mathbf{z}_t + \sum_{j=1}^q \beta_j H(\mathbf{z}_t) + \epsilon_t, \quad \epsilon_t \stackrel{iid}{\sim} \mathcal{N}(0, \sigma^2), \tag{3.1}$$

where $\beta_0 = (\beta_{00}, \beta_{01}, \beta_{02}, \dots, \beta_{0p})'$, and $\mathbf{z}_t = (1, y_{t-1}, y_{t-2}, \dots, y_{t-p})'$. \mathbf{z}_t in (3.1) is an autoregressive vector and $H(\cdot)$ is a suitable activation function. For example, sigmoid active function gives

$$H(\mathbf{z}_t) = \left(1 + \exp(-\gamma_j' \mathbf{z}_t)\right)^{-1},$$

and hyperbolic tangent function gives

$$H(\mathbf{z}_t) = \tanh(-\gamma_j' \mathbf{z}_t),$$

where $\gamma_j = (\gamma_{j0}, \gamma_{j1}, \dots, \gamma_{jp})'$ is the weight vector of \mathbf{z}_t . It must be noted that the feed-forward AR(p)-NN model is a non-linear model with an activation function that plays the role of a basis function. For more details regarding estimation, see Kock and Teräsvirta (2011).

We consider the three-neural-network-based HAR models. The first model is a natural extension of (3.1) to HAR, with

$$\mathbf{z}_t = \left(1, \text{RV}_t^{(d)}, \text{RV}_t^{(w)}, \text{RV}_t^{(m)}\right)'.$$

For example, the sigmoid activation function gives the following

$$\begin{aligned} \text{RV}_{t+1d}^{(d)} &= \beta_{00} + \beta_{0d}\text{RV}_t^{(d)} + \beta_{0w}\text{RV}_t^{(w)} + \beta_{0m}\text{RV}_t^{(m)} \\ &+ \sum_{j=1}^q \beta_j \left(1 + \exp\left(-\gamma_{j0} - \gamma_{jd}\text{RV}_t^{(d)} - \gamma_{jw}\text{RV}_t^{(w)} - \gamma_{jm}\text{RV}_t^{(m)}\right)\right)^{-1} + \epsilon_{t+1d}. \end{aligned} \quad (3.2)$$

We will denote this model as HAR-NN. In fact, HAR-NN model is also studied in Hillebrand and Medeiros (2010), McAleer and Medeiros (2011); however, they are more focused on the bagged estimator rather than the HAR-NN model estimates themselves.

Our second mode is an extension of HAR(∞) to the neural network model. However, practically, the HAR(∞) model should be approximated by some large values of p . Since the HAR model is a special case of AR(22), we set $p = 22$. The HAR(∞)-NN model is referred to as the model (3.1) with

$$\mathbf{z}_t = \left(1, \text{RV}_t^{(d)}, \text{RV}_{t-1d}^{(d)}, \dots, \text{RV}_{t-21d}^{(d)}\right)'.$$

For example, the hyperbolic tangent activation function gives

$$\text{RV}_{t+1d}^{(d)} = \beta_{00} + \sum_{i=1}^{22} \beta_{0i}\text{RV}_{t+(1-i)d}^{(d)} + \sum_{j=1}^q \beta_j \tanh\left(\gamma_{j0} + \sum_{i=1}^{22} \gamma_{ji}\text{RV}_{t+(1-i)d}^{(d)}\right) + \epsilon_{t+1d}. \quad (3.3)$$

We may expect a better forecasting performance because the HAR(∞)-NN reflects both non-linearity and long-memory properties with more coefficients. However, the number of parameters used in HAR(∞)-NN is $23 + 24q$, and are almost five times more than that in HAR-NN. This factor may lead to estimation errors, and thereby deteriorate the forecasting performance. Therefore, we also consider a hybrid of two models as:

$$\text{RV}_{t+1d}^{(d)} = \beta_{00} + \sum_{i=1}^{22} \beta_{0i}\text{RV}_{t+(1-i)d}^{(d)} + \sum_{j=1}^q \beta_j H(\mathbf{z}_t) + \epsilon_{t+1d} \quad (3.4)$$

with

$$\mathbf{z}_t = \left(1, \text{RV}_t^{(d)}, \text{RV}_t^{(w)}, \text{RV}_t^{(m)}\right)'.$$

Then, the number of parameters to be estimated is $23 + 5q$, so uses less variables for approximating the non-linear terms. We will denote this model as HAR-AR(22)-NN. We fit the HAR-AR(22)-NN model by iteratively applying the multi-staged fitting. First, we fit AR(22) to the original data, and fit feed-forward NN model (using HAR model's independent variables (\mathbf{z}_t) as input values in the NN) to the residuals. Subsequently, by subtracting the fitted value obtained from the feed-forward NN model from $\text{RV}_{t+1d}^{(d)}$ and by fitting AR(22) model again, we update the coefficients in the HAR-AR(22)-NN model until convergence.

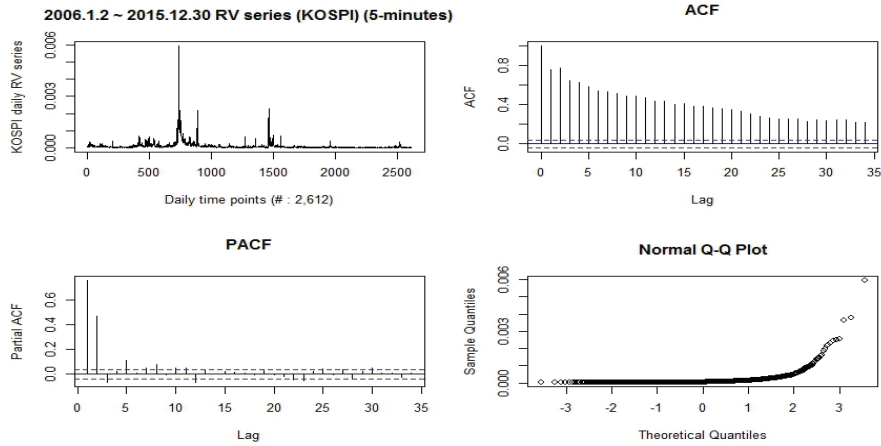


Figure 1: *TS, ACF, PACF, and QQ plots of KOSPI daily RV series. RV = realized volatility.*

4. Forecasting comparison with multinational RVs

In this section, we compare the forecasting performance of the proposed neural network-based HAR models. We use the daily volatility data from the Oxford-Man Institute of Quantitative Finance (available online). We use the 5 minutes high-frequency data to calculate the RV from January 2, 2006 to December 30, 2015 and obtain 2,612 daily RV series. We use 11 stock indices from around the world: Standard & Poor’s (S&P) 500, Financial Times Stock Exchange (FTSE) 100, Russel, Dow Jones Industrial Average (DJIA), Nikkei 225, Hang Seng, KOSPI, Indice Bursatil Espanol (IBEX35), Bovespa, Euro, and Deutscher Aktienindex (DAX). Figure 1 shows the RV time plot, sample autocorrelations, partial autocorrelation, and Normal QQ-plot for the KOSPI index. It is observed that the volatility level changes over time with very strong and slowly decaying autocorrelations, which are possibly heavy-tailed. Therefore, it seems to be reasonable to consider the HAR models to explain such features.

We evaluate the one-step-ahead out-of-sample mean squared prediction error (MSPE) and mean absolute prediction error (MAPE) for the performance measure of forecasting. It is given by

$$MSPE := \frac{1}{T_t} \sum_{t=T}^{T+T_t-1d} \left(\widehat{RV}_{t+1d}^{(d)} - RV_{t+1d}^{(d)} \right)^2, \quad MAPE := \frac{1}{T_t} \sum_{t=T}^{T+T_t-1d} \left| \widehat{RV}_{t+1d}^{(d)} - RV_{t+1d}^{(d)} \right|, \quad (4.1)$$

where the model is fitted from the training data from daily time point 1 to T , and the square and absolute prediction errors are calculated in the test set from daily time point $T + 1$ to $T + T_t$ using rolling-window method. For example for the KOSPI RV, we set $T = 2,512$ and $T_t = 100$. First, the model is built upon the initial training data from 1 to 2,512 and predict 2,513th value. Then, we use data from 1 to 2,513 to reestimate model parameters while all other tuning parameters remain fixed and forecast 2,514th value. By iterating this procedure to all 100 observations in the test set, we obtain 1-step-ahead out-of-sample forecasts. Finally, MSPE and MAPE are calculated for model comparison.

It is important to note that our proposed neural network-based HAR models require selecting an optimal order of approximation q . This must be done by minimizing the pre-specified loss as a function of q . We followed the popular training-validation approach to determine the tuning parameter as

Table 1: The optimal choice of q with the sigmoid activation function for the KOSPI RV series. The criterion is to minimize $MSPE \times 10^5$ and $MAPE \times 10^3$ according to q

	HAR-NN		HAR(∞)-NN		HAR-AR(22)-NN	
	MSPE	MAPE	MSPE	MAPE	MSPE	MAPE
$q = 1$	0.2330	1.140	0.2710	1.157	0.3080	1.485
$q = 5$	0.0005	0.054	0.0028	0.078	0.0098	0.095
$q = 10$	0.0278	0.519	0.0327	0.602	0.0413	0.663
$q = 15$	0.0472	0.698	0.0508	0.728	0.0467	0.687
$q = 20$	0.0367	0.610	0.0501	0.722	0.0752	0.816

HAR = heterogeneous autoregressive model; NN = neural network; AR = autoregressive; MSPE = mean squared prediction error; MAPE = mean absolute prediction error.

Table 2: The optimal choice of q with the tanh activation function for the KOSPI RV series. The criterion is to minimize $MSPE \times 10^5$ and $MAPE \times 10^3$ according to q .

	HAR-NN		HAR(∞)-NN		HAR-AR(22)-NN	
	MSPE	MAPE	MSPE	MAPE	MSPE	MAPE
$q = 1$	0.2300	1.060	0.2450	1.087	0.2570	1.099
$q = 5$	0.0317	0.563	0.0397	0.581	0.0402	0.612
$q = 10$	0.0125	0.204	0.0208	0.278	0.0377	0.579
$q = 15$	0.0382	0.580	0.0424	0.711	0.0411	0.699
$q = 20$	0.0501	0.712	0.0372	0.569	0.0397	0.602

HAR = heterogeneous autoregressive model; NN = neural network; AR = autoregressive; MSPE = mean squared prediction error; MAPE = mean absolute prediction error.

it becomes standard method in machine learning literature. We further divide the training set into two sets; the first 2,312 observations compose the training set for the tuning parameter selection, and the next 100 observations make up the validation set. The loss function is the MSPE- and MAPE-based one-step-ahead out-of-sample forecast given as (4.1). For example, Tables 1–2 show the $MSPE \times 10^5$ and $MAPE \times 10^3$ as a function of q when they are evaluated from the validation set (in the KOSPI RV series set). With the sigmoid activation function, the minimum is achieved when $q = 5$ for all models and both performance measures. However, with the tangent hyperbolic activation function, the minimum is achieved with $q = 10$. This indicates that a detailed order of approximation is needed with the tanh activation function. Once the optimal q is selected for each RV series and model, we merge the validation set into the training set to re-estimate the parameters to obtain better forecasts. We implemented above algorithm with R and posted them for readers convenience to open repository GitHub as

<https://github.com/yools56/Neural-Network-based-HAR-models>

Tables 3–4 show the one-step-ahead MSPE for multinational stock indexes when the sigmoid and tanh activation functions are used for the neural network models, respectively. We also compare the traditional HAR model and the ARFIMA model as a reference for long memory model. The order of ARFIMA model is selected through the Bayesian information criterion (BIC). The optimal order of ARFIMA model was chosen by having the smallest BIC value. For example, the optimal order of ARFIMA model of KOSPI RV series is $(1, d, 1)$.

When the sigmoid function is used for the activation function, it is observed that the HAR-NN model achieves the minimum MSPE for most of the stock indexes, such as S&P 500, FTSE 100, Russel, DJIA, Nikkei 225, Hang Seng, KOSPI, and Euro. The HAR(∞)-NN model performs best for Bovespa, and the ARFIMA model performs best for IBEX 35 and DAX. However, when it comes to the tanh activation function, the HAR-NN model uniformly achieves the minimum for all stock

Table 3: MSPE $\times 10^5$ with optimal q and sigmoid activation function

Stock	Model				
	HAR	HAR-NN	HAR(∞)-NN	HAR-AR(22)-NN	ARFIMA
S&P 500	3.429	3.313	3.425	3.648	3.405
FTSE 100	0.999	0.928	1.155	1.199	0.990
Russel	0.642	0.605	0.638	0.870	0.632
DJIA	5.635	5.556	6.117	6.046	5.585
Nikkei 225	2.242	2.056	2.916	2.636	2.287
Hang Seng	0.765	0.749	0.843	0.965	0.795
KOSPI	0.315	0.294	0.301	0.426	0.317
IBEX 35	1.491	1.595	1.509	2.070	1.474
Bovespa	0.6985	0.712	0.692	1.296	0.6986
Euro	2.119	2.085	2.364	2.880	2.100
DAX	1.680	1.676	1.781	2.409	1.671
The average	1.819	1.779	1.976	2.222	1.806

MSPE = mean squared prediction; HAR = heterogeneous autoregressive model; NN = neural network; AR = autoregressive; ARFIMA = autoregressive fractionally integrated moving average.

Table 4: MSPE $\times 10^5$ with optimal q and tanh activation function

Stock	Model				
	HAR	HAR-NN	HAR(∞)-NN	HAR-AR(22)-NN	ARFIMA
S&P 500	3.429	3.393	3.442	3.666	3.405
FTSE 100	0.999	0.988	1.019	1.192	0.990
Russel	0.642	0.638	0.667	0.864	0.632
DJIA	5.635	5.574	5.708	5.927	5.585
Nikkei 225	2.242	2.171	2.348	2.705	2.287
Hang Seng	0.765	0.699	0.858	0.994	0.795
KOSPI	0.315	0.287	0.321	0.370	0.317
IBEX 35	1.491	1.457	1.559	1.901	1.474
Bovespa	0.6985	0.672	0.695	1.288	0.6986
Euro	2.119	2.095	2.287	2.828	2.100
DAX	1.680	1.670	1.724	2.303	1.671
The average	1.820	1.786	1.875	2.185	1.814

MSPE = mean squared prediction; HAR = heterogeneous autoregressive model; NN = neural network; AR = autoregressive; ARFIMA = autoregressive fractionally integrated moving average.

indexes. It is interesting to observe that higher-order HAR models, such as the HAR(∞)-NN or HAR-AR(22)-NN models, perform worse than the HAR-NN model. We believe that this is because of the estimation error emerging as a result of estimating too many parameters relative to the sample size. Out-of-sample forecasting does not necessarily improve by having a model with smaller in-sample errors. This becomes more evident in Figures 2–3 wherein the out-of-sample forecasts are overlaid in one figure. The purple line presents the HAR-AR(22)-NN model, and they are clearly away from other forecasts.

The results for the one-step-ahead out-of-sample MAPE is presented in Tables 5–6. The results are more delicate here. The HAR-NN model performs the best among the HAR type of models; however, the ARFIMA model also works well when compared to the HAR-based model. This finding is slightly contrary to expectations because the number of parameters used in the ARFIMA model is smaller than NN-based HAR models. Similarly to the MSPE, too many parameter estimations may worsen the forecasting performance.

Since the stock market shows higher volatility, point estimates may not successfully evaluate the forecasting performance. Hence, we also consider the comparison of the one-step-ahead prediction

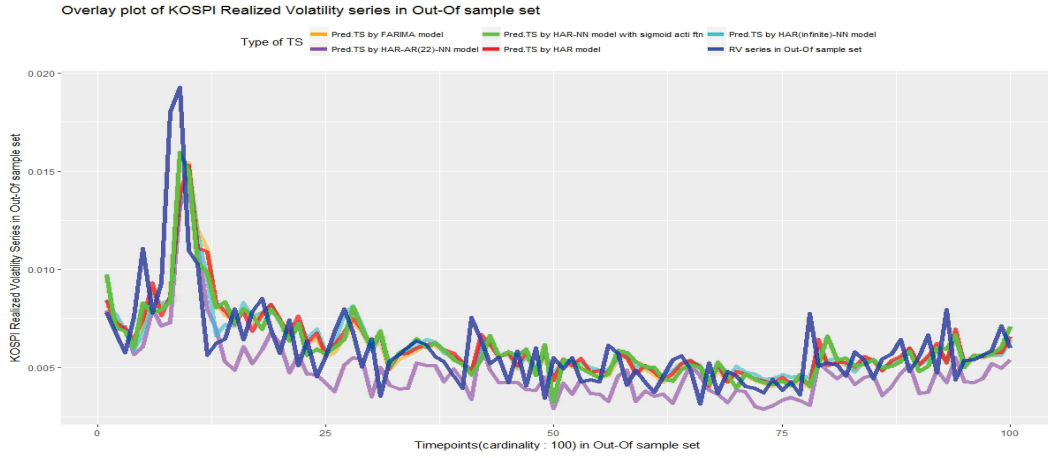


Figure 2: Forecasting overlay plot for the KOSPI RV series in the out-of-sample set with sigmoid activation function.

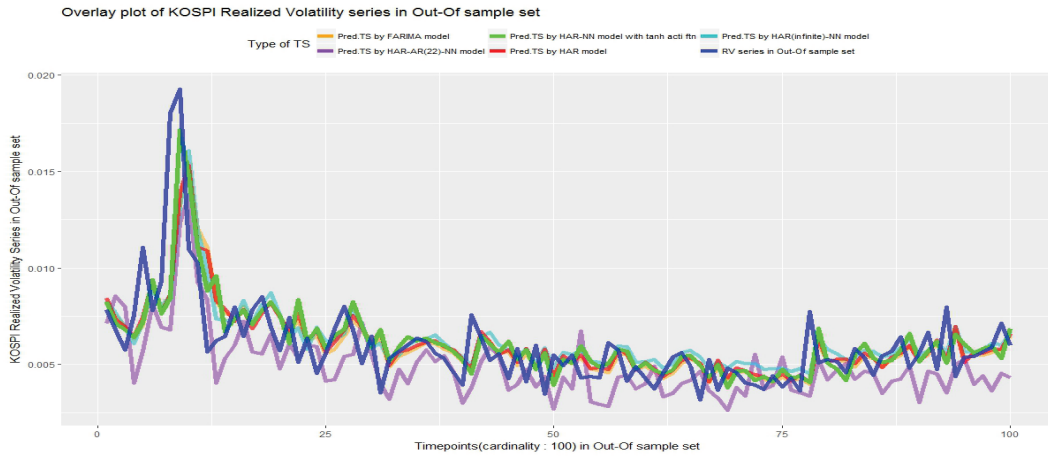


Figure 3: Forecasting overlay plot for the KOSPI RV series in the out-of-sample set with tanh activation function.

interval for $RV_{T+1d}^{(d)}$. We use the confidence interval formula for neural network models given by Allende *et al.* (2002). We rewrite the neural network-based HAR model as

$$RV_{t+1d}^{(d)} = g(\mathbf{z}_t) + \epsilon_{t+1d}, \quad \epsilon_{t+1d} \sim \mathcal{WN}(0, \sigma^2), \quad (4.2)$$

where $g(\cdot)$ is a non-linear function that combines HAR and NN. Subsequently, Allende *et al.* (2002) proposed the $100(1 - \alpha)\%$ asymptotic prediction interval as

$$\hat{g}(\mathbf{z}_t) \pm t_{1-\frac{\alpha}{2}}(df) \hat{\sigma} \sqrt{1 + \hat{S}}, \quad (4.3)$$

where $\hat{g}(\mathbf{z}_t)$ is a neural network estimate, k is the number of input variables in the neural network, $t_{(1-\alpha/2)}(df)$ is the $(1 - \alpha/2)^{th}$ quantile of student t distribution with the degree of freedom $df = T - (k +$

Table 5: MAPE $\times 10^3$ with optimal q and sigmoid activation function

Stock	Model				
	HAR	HAR-NN	HAR(∞)-NN	HAR-AR(22)-NN	ARFIMA
S&P 500	2.771	2.653	2.778	2.936	2.750
FTSE 100	1.697	1.650	1.817	2.009	1.653
Russel	1.662	1.633	1.663	1.946	1.615
DJIA	3.413	3.341	3.560	3.575	3.379
Nikkei 225	2.804	2.764	3.152	3.023	2.818
Hang Seng	1.865	1.896	1.931	2.099	1.882
KOSPI	1.174	1.155	1.125	1.433	1.154
IBEX 35	2.519	2.564	2.562	3.010	2.474
Bovespa	2.000	2.009	1.959	2.520	1.970
Euro	2.735	2.789	2.850	3.353	2.698
DAX	2.468	2.436	2.569	3.208	2.438
The average	2.283	2.262	2.360	2.647	2.257

MAPE = mean absolute prediction error; HAR = heterogeneous autoregressive model; NN = neural network; AR = autoregressive; ARFIMA = autoregressive fractionally integrated moving average.

Table 6: MAPE $\times 10^3$ with optimal q and tanh activation function

Stock	Model				
	HAR	HAR-NN	HAR(∞)-NN	HAR-AR(22)-NN	ARFIMA
S&P 500	2.771	2.709	2.745	2.930	2.750
FTSE 100	1.697	1.659	1.694	1.917	1.653
Russel 2000	1.662	1.679	1.728	1.868	1.615
DJIA	3.413	3.399	3.411	3.504	3.379
Nikkei 225	2.804	2.830	2.890	3.031	2.818
Hang Seng	1.865	1.847	1.931	2.102	1.882
KOSPI	1.174	1.143	1.204	1.398	1.154
IBEX 35	2.519	2.531	2.587	2.877	2.474
Bovespa	2.000	1.963	1.989	2.627	1.970
Euro	2.735	2.727	2.947	3.344	2.698
DAX	2.468	2.464	2.487	3.100	2.438
The average	2.283	2.268	2.328	2.609	2.257

MAPE = mean absolute prediction error; HAR = heterogeneous autoregressive model; NN = neural network; AR = autoregressive; ARFIMA = autoregressive fractionally integrated moving average.

2) $q - 1 - k$, and

$$\hat{\sigma}^2 = \frac{1}{df} \sum_{t=1}^T (\text{RV}_{t+1d}^{(d)} - \hat{g}(\mathbf{z}_t))^2.$$

The $\hat{\Sigma}$ is an asymptotic variance of $g(\mathbf{z}_t)$ that is calculated from the continuity of the non-linear function g , based on the delta method; however, this method is not detailed here for brevity (see, equation 30 in Allende *et al.* (2002)).

Figures 4–5 show the one-step-ahead 95% asymptotic prediction intervals for the HAR-NN model with the sigmoid and tanh activation functions, respectively. The HAR-NN provides a slightly wider interval than the traditional HAR, and shows more peaks and valleys on turning points. It means that the HAR-NN model can capture sharper changes due to a higher volatility than the HAR model. Hence, it is recommended to use the HAR-NN model for prediction interval to account for higher stock market volatility.

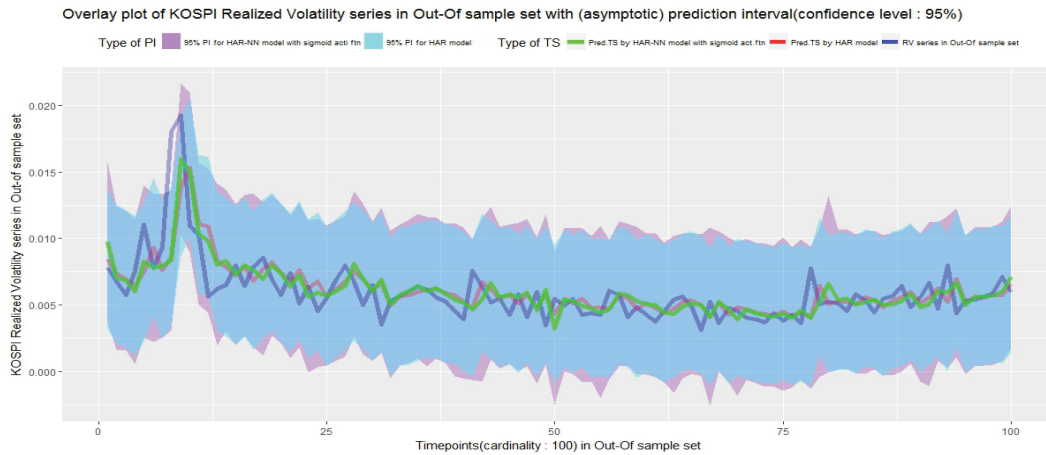


Figure 4: 95% prediction intervals for HAR-NN model with sigmoid activation function. HAR = heterogeneous autoregressive model; NN = neural network.

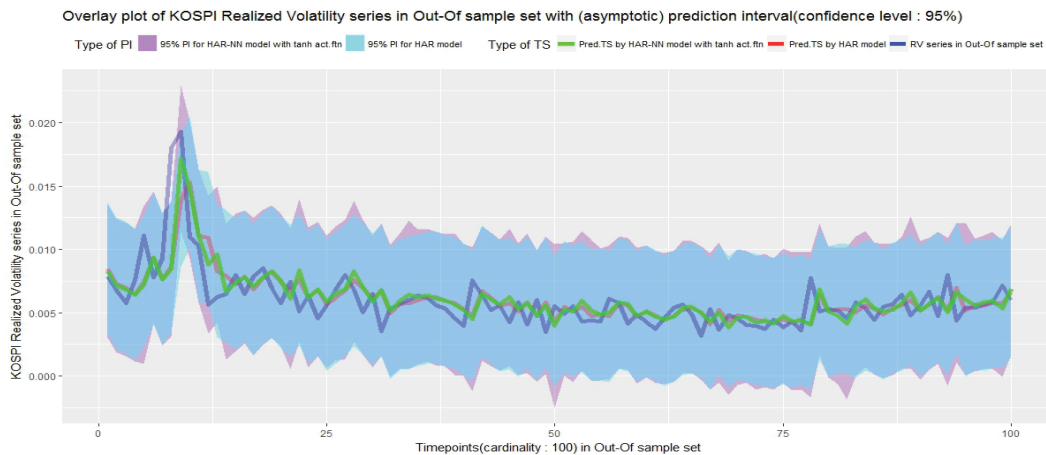


Figure 5: 95% prediction intervals for HAR-NN model with tanh activation function. HAR = heterogeneous autoregressive model; NN = neural network.

5. Residual analysis

Here, we check model adequacy by residual analysis. Figure 6 shows some diagnostic plots from the residuals after fitting HAR-NN model for the KOSPI RV with a sigmoid activation function. They are residuals time plot, sample autocorrelations plot, partial autocorrelations plot, sample autocorrelations plot from the squared series and normal quantile-quantile plot. They show no clear evidence of dependency, remaining trend and unequal variances. The portmanteau test such as Ljung-Box test with 20 lags gives p -value of 0.7707, and Engle's test with lag 1 to determine ARCH effect gives p -value of 0.7985. Therefore, we conclude that there is no strong evidence against white noise assumption and no ARCH effect for the HAR-NN model. Similar conclusions are also drawn for other models. We only report diagnostic plots in Figures 7–8, HAR(∞)-NN and HAR-AR(22)-NN models,

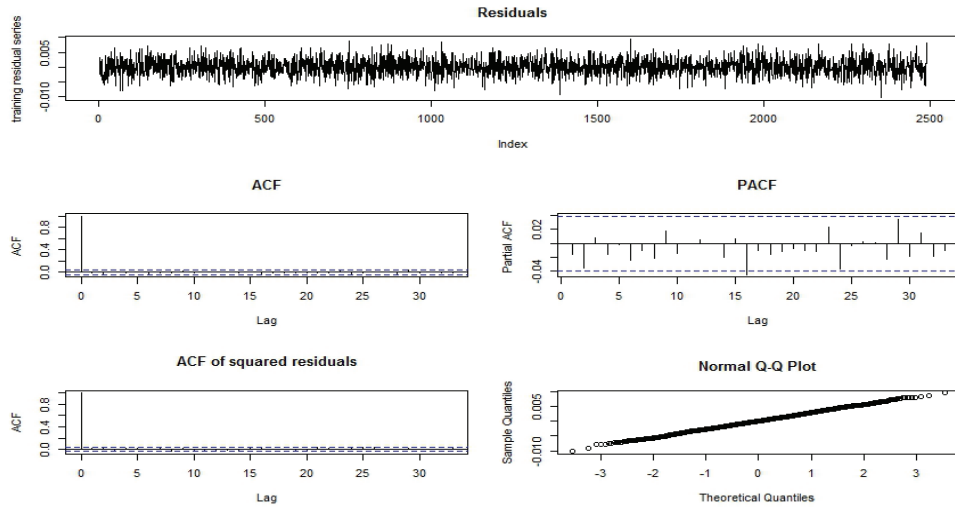


Figure 6: Diagnostic plots from HAR-NN model for the KOSPI RV. HAR = heterogeneous autoregressive model; NN = neural network, RV = realized volatility.

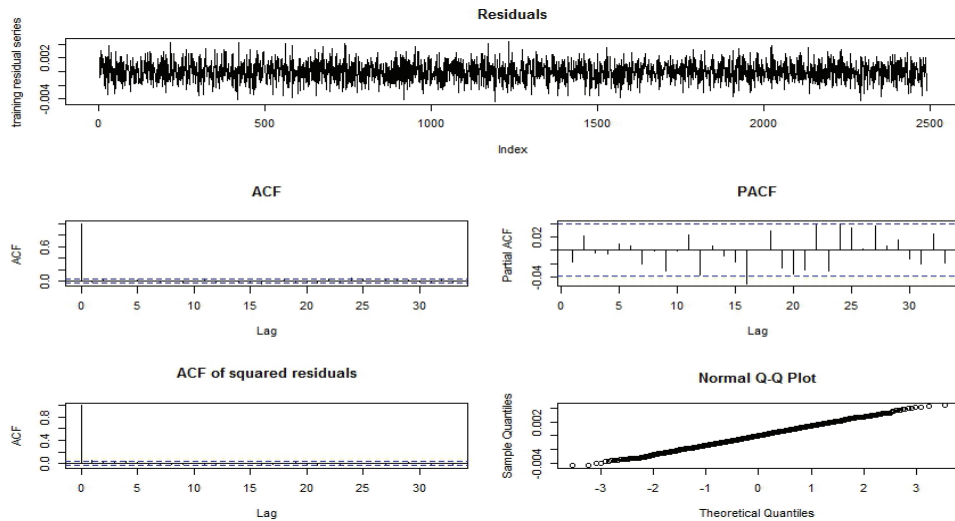


Figure 7: Diagnostic plots from HAR(∞)-NN model for the KOSPI RV. HAR = heterogeneous autoregressive model; NN = neural network, RV = realized volatility.

respectively, for concise presentation. Indeed, we do not observe any significant evidence against white noise assumption and no ARCH effect is observed.

6. Conclusion

In this study, we consider three-neural-network-based HAR models. The traditional HAR model takes daily, weekly, and monthly average volatility, and it is naturally extended to HAR-NN by incorporat-

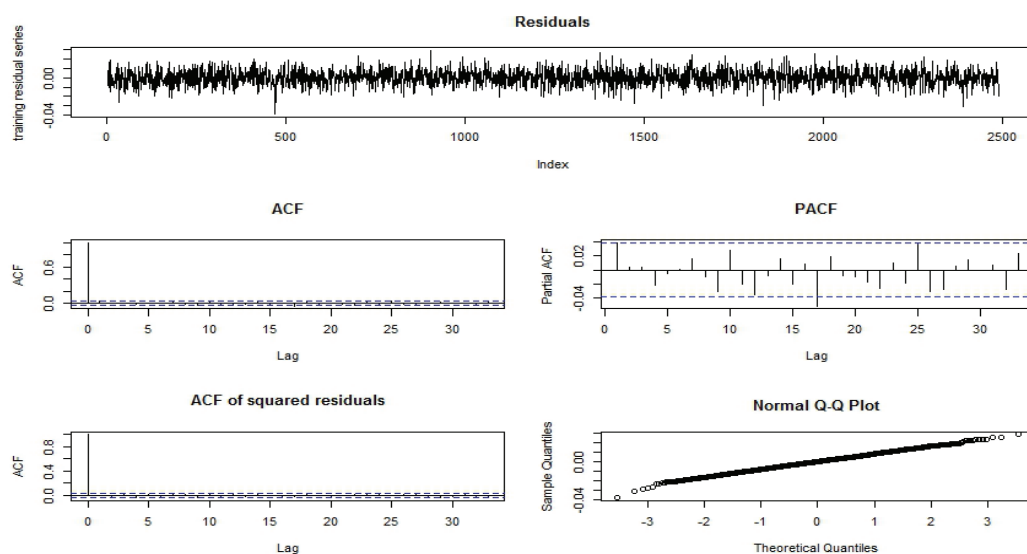


Figure 8: Diagnostic plots from HAR-AR(22)-NN model for the KOSPI RV. HAR = heterogeneous autoregressive model; AR = autoregressive; NN = neural network, RV = realized volatility.

ing three terms in the neural network model. We also consider the extension of the HAR(∞) model to HAR(∞)-NN by taking the order of $p = 22$ as a practical consideration. The hybrid of the two models, HAR-AR(22)-NN model, is also considered for estimating a moderate number of parameters. The optimal tuning parameter (the number of hidden units in the neural network) is chosen data adaptively. Therefore, the training, validation, and testing procedure is applied.

We evaluated the model performance by comparing the forecasting error on multinational RVs. The results are mixed and dependent on the RVs; however, we observed the general tendency that HAR-NN model performs better than traditional HAR. Hence, we confirm that the addition of a non-linear term in the HAR model improves forecasting. However, the addition of too many terms, such as HAR(∞)-NN or HAR-AR(22)-NN, does not necessarily outperform HAR or HAR-NN. This may be attributed to the fact that estimating too many parameters relative to the sample size may worsen the out-of-sample forecasting. We also compared prediction intervals between HAR and HAR-NN models and observed that HAR-NN provides a more reliable prediction interval on peaks and valleys on turning points, and therefore may be effective in capturing the higher volatility of RVs.

Acknowledgements

This work was supported in part by the Basic Science Research Program from the National Research Foundation of Korea (NRF), funded by the Ministry of Science, ICT & Future Planning (NRF-2017R1A1A1A05000831).

References

Allende H, Moraga C, and Salas R (2002). Artificial neural networks in time series forecasting: A comparative analysis, *Kybernetika*, **38**, 685–707.

- Andersen TG, Bollerslev T, Diebold FX, and Ebens H (2001). The distribution of realized stock return volatility, *Journal of Financial Economics*, **61**, 43–76.
- Andersen TG, Bollerslev T, Diebold FX, and Labys P (2001). The distribution of exchange rate volatility, *Journal of the American Statistical Association*, **96**, 42–55.
- Andersen TG, Bollerslev T, Diebold FX, and Labys P (2003). Modeling and forecasting realized volatility, *Econometrica*, **71**, 579–625.
- Barndorff-Nielsen O and Shephard N (2002a). Econometric analysis of realized volatility and its use in estimating stochastic volatility models, *Journal of the Royal Statistical Society, Series B*, **64**, 253–280.
- Barndorff-Nielsen O and Shephard N (2002b). Estimating quadratic variation using realized variance, *Journal of Applied Econometrics*, **17**, 457–477.
- Corsi F (2009). A simple approximate long-memory model of realized volatility, *Journal of Financial Econometrics*, **7**, 174–196.
- Hillebrand E and Medeiros MC (2010). The benefits of bagging for forecast models of realized volatility, *Econometric Reviews*, **29**, 571–593.
- Hwang E and Shin DW (2014). Infinite-order, long-memory heterogeneous autoregressive models, *Computational Statistics & Data Analysis*, **76**, 339–358.
- Kock BA and Teräsvirta T (2011). Forecasting macroeconomic variables using neural network models and three automated model selection techniques, (No. 2011-27) Department of Economics and Business Economics, Aarhus University.
- McAleer M and Medeiros MC (2011). Forecasting realized volatility with linear and nonlinear univariate models, *Journal of Economic Surveys*, **25**, 6–18.

Received July 29, 2018; Revised September 17, 2018; Accepted September 20, 2018