

Forecasting Stock Market Realized Variance with Echo State Neural Networks

Milan Fičura^{*}

Abstract:

Echo State Neural Networks (ESN) were applied to forecast the realized variance time series of 19 major stock market indices. Symmetric ESN and asymmetric AESN models were constructed and compared with the benchmark realized variance models HAR and AHAR that approximate the long memory of the realized variance process with a heterogeneous auto-regression. The results show that asymmetric models generally outperform symmetric ones, indicating that a correlation between volatility and returns plays an important role for volatility forecasting. Additionally, models utilizing a logarithmic transformation of the time series achieved generally better results than models applied directly to the realized variance. Echo State Neural Networks outperformed HAR and AHAR models for several important indices (S&P500, DJIA and Nikkei indices), but on average they achieved slightly worse results than the AHAR model. Nevertheless, the results show that Echo State Neural Networks represent an easy-to-use and accurate tool for realized variance forecasting, whose performance may potentially be further improved with meta-parameter optimization.

Key words: Realized variance; HAR model; Echo State Neural Networks.

JEL classification: C45, C53, C58.

1 Introduction

Forecasting the volatility of the stock market plays an important role in many areas of finance including risk management, portfolio construction, derivatives pricing and quantitative trading. Many different types of volatility models have been proposed in the literature. Among the most commonly used are ARCH and GARCH models, modelling volatility as a linear combination of squared past daily returns, and their extensions, utilizing effects such as long memory (FIGARCH model) or negative correlation between stock returns and their volatility (GJR-GARCH or EGARCH models) (see Bollerslev 2008 for a review).

Another popular class of volatility models, especially in the area of option pricing, are stochastic volatility models, modelling the underlying volatility of asset returns as a separate latent state process (Shephard, 2005). These models include the log-variance model, the stochastic-volatility jump-diffusion class of models, or the Markov Switching Multifractal model (Calvet and Fisher, 2001).

^{*} Milan Fičura; University of Economics, Prague, Department of Finance and Accounting, <milan.ficura@vse.cz>.

The article is processed as an output of the research project IGA - IG102027, *Advanced methods of risk management with the use of artificial intelligence*, 59/2017.

Over the past two decades, a new class of so-called realized volatility models emerged, utilizing power-variation estimators to estimate the underlying volatility of asset returns from high-frequency data. Multiple power-variation estimators of volatility have been proposed, including the realized variance (Andersen and Bollerslev, 1998), the bi-power variation, which is robust to jumps, or the realized kernel estimator which is robust to microstructure noise of the high-frequency price process. Volatility estimated with such estimators can then be modelled with standard time series models such as ARIMA, or more commonly, the long-memory ARFIMA model (Pong et al., 2003). A special place among these models holds the HAR model (Corsi, 2004), which is easier to estimate than the ARFIMA model while at the same time successfully capturing the long-memory of the volatility process by using a heterogeneous auto-regression over the realized variances aggregated over different time periods. An asymmetric extension of the HAR model, called the AHAR model, captures in addition to that also the negative correlation between the stock returns and their volatility, by adding past stock market returns into the regression equation.

Finally, machine learning models have recently started to be adopted for realized volatility modelling and forecasting as well (McAleer and Medeiros, 2011, Vortelinos and Dimitrios, 2015), in order to capture the possible nonlinear dependencies in the realized variance time series.

In our study, we apply Echo State Neural Networks (ESN) to forecast the realized variance of 19 stock market indices and we compare their performance with the performance of HAR and AHAR models used as benchmarks.

Echo State Neural Networks are a type of recurrent neural networks (RNN) developed for modelling of temporal phenomena, including the prediction of time series. Earlier recurrent neural networks such as the Elman neural networks (Elman, 1990) proved difficult to train efficiently with gradient based optimization algorithms due to the so-called vanishing gradient problem. To cope with this problem, other types of recurrent neural networks and training approaches have been proposed, including LSTM networks, Evolino neural networks and Echo State Networks (Jaeger and Haas, 2004).

The vanishing gradient problem obstructs training of the recurrent part of neural networks, preventing them from learning long-range dependencies in time series. Echo State Neural Networks solve the vanishing gradient problem by not training the recurrent part of the neural network, but instead randomly generating a large recurrent layer, called reservoir, and then training only the readout from the reservoir, typically with a simple penalized linear regression (Ridge regression or Lasso regression).

In spite of the simplicity of this approach, Echo State Neural Networks upon their introduction significantly outperformed standard recurrent neural networks in a wide variety of benchmark tasks, especially regarding the prediction of univariate chaotic time series such as the Mackey-Glass oscillator. They proved to be very efficient also in a wide variety of empirical applications, including wind speed forecasting and financial time series prediction (Lukoševičius and Jaeger, 2009).

In spite of these successes, there seems to be no previous study (to the knowledge of the author) that would apply Echo State Networks to the issue of realized variance forecasting, which is why the given topic was chosen for this paper. The rationale is that unlike the HAR model, ESN may be able to capture not only the long memory, but also the nonlinear relationships in the realized variance time series, while at the same time being far easier to train than other types of RNN.

The rest of the paper is organized as follows. In the section two, realized variance and the HAR model are explained. The third section introduces Echo State Neural networks and in the section four are presented results of the empirical study.

2 Realized Variance and the HAR Model

Let us assume that the logarithmic price of an asset follows a generally defined Stochastic-Volatility Jump-Diffusion process expressed as:

$$dp(t) = \mu(t)dt + \sigma(t)dW(t) + j(t)dq(t) \quad (1)$$

where $p(t)$ is the logarithm of the price, $\mu(t)$ is the instantaneous drift rate, $\sigma(t)$ is the instantaneous volatility, $dW(t)$ is a differential of the Wiener process, $j(t)$ is a process determining the size of the price jumps and $dq(t)$ is a differential of the counting process determining the times of jump occurrences.

The total variability of the price process over a period between $t-1$ and t can then be expressed with its *quadratic variation* as follows:

$$QV(t) = \int_{t-1}^t \sigma^2(s)ds + \sum_{t-1 \leq s < t} \kappa^2(s) \quad (2)$$

where $\kappa(t) = j(t)I[q(t)=1]$ and $I(\cdot)$ is the indicator function. The first term in the equation corresponds to the *integrated variance*, representing the continuous component of the price variability, while the second term corresponds to the *jump variance*, representing the discontinuous component of the price variability.

As the quadratic variation is an unobservable quantity of the price process, it has to be estimated. Squared daily returns provide an unbiased estimate of the underlying quadratic variation but they are plagued by a large degree of noise.

Andersen and Bollerslev (1998) proposed a much more accurate estimator of the underlying quadratic variation based on the asymptotic theory of power variations and intraday data. The estimator is called *realized variance* and it can be calculated as the sum of squared high-frequency returns over the given day.

Formally, if we denote $r(t, \Delta)$ as the logarithmic return between $t - \Delta$ and t , we can define the realized variance as follows:

$$RV(t, \Delta) = \sum_{j=1}^{1/\Delta} r^2(t - 1 + j\Delta, \Delta) \quad (3)$$

And it holds that $RV(t, \Delta) \rightarrow QV(t)$ as $\Delta \rightarrow 0$.

Although the realized variance may be plagued by high-frequency microstructure noise and thus provide biased estimates of the quadratic variation on certain occasions (especially if we move to the ultra-high frequencies), it will be used as the variance measure to be predicted in the rest of the study.

In applications, the realized variance is commonly viewed as a de-facto observable measure of the underlying market volatility, which can be used to construct future volatility forecasts using any standardly adopted time series model. Among the plethora of different models, the HAR model (proposed by Corsi 2004) has become the industry standard due to its easy estimation (via simple linear regression) coupled with its ability to approximate the long memory of the volatility process.

HAR stands for a Heterogeneous Autoregressive Model and it predicts the future market volatility (daily realized variance) based on a linear regression on the realized variance calculated over the past day, week and month. The model can be expressed as follows:

$$RV_d(t) = \beta_c + \beta_d RV_d(t-1) + \beta_w RV_w(t-1) + \beta_m RV_m(t-1) + \varepsilon(t) \quad (4)$$

with $RV(t)$ denoting the realized variances at time t (with indices d , w and m , corresponding to the aggregation over the past day, week and month), $\varepsilon(t)$ is a Gaussian white noise and betas are parameters of the model.

In order to reflect the negative correlation between the stock market returns and their volatility (the so-called leverage effect), the HAR model can be extended into the so-called AHAR model (Asymmetric HAR). The AHAR model is defined as:

$$RV_d(t) = \beta_c + \beta_d RV_d(t-1) + \beta_w RV_w(t-1) + \beta_m RV_m(t-1) + \gamma_d Ret_d(t-1) + \gamma_w Ret_w(t-1) + \gamma_m Ret_m(t-1) + \varepsilon(t) \quad (5)$$

with $Ret(t)$ corresponding to the past asset price returns over the last day, week and month (indices d , w and m) and gammas to the parameters associated with these returns. Other variables remain the same as in equation (4).

While HAR and AHAR models are able to approximate the long memory of the volatility process, they are basically linear models in a sense that they calculate their volatility forecasts based on a linear combination of the past daily, weekly and monthly volatility (and respectively also the returns in the case of the AHAR model).

As the realized variance process may contain non-linear dependencies, we propose to use Echo State Neural Networks for its prediction which, as universal function approximators, have the potential to capture any non-linear relationships between the past realized variances and the future ones.

The standard Echo State Neural Network model can be expressed with the following 3 equations:

$$Rez_t^* = f(W_{In}[1; X_t] + W_{Rez} Rez_{t-1}), \quad (6)$$

$$Rez_t = (1 - \alpha) Rez_{t-1} + \alpha Rez_t^*, \quad (7)$$

$$y_t = W_{Out}[1; Rez_t], \quad (8)$$

with X_t denoting the vector of normalized explanatory variables at time t , Rez_t the vector of the reservoir at time t , $f(\cdot)$ the activation function of the neural network (logistic function in our case), and W_{In} , W_{Rez} and W_{Out} the weight matrices. The variable α is a smoothing parameter that in our case will be set to 1 (no smoothing). W_{In} and W_{Rez} are set randomly, by sampling the values of the parameters from the uniform distribution with bounds of $[-1, 1]$, with the W_{Rez} matrix subsequently made sparse (by replacing 50% of its values by zeros) and its spectral radius normalized to the value of one. The W_{Out} vector is found with a ridge regression with the penalization parameter set to 1 and the size of the reservoir (i.e. number of neurons) set to 50.

Meta parameters of the Echo State Neural Network (spectral radius of W_{Out} , scaling of W_{In} , value of α , sparsity of W_{Rez} , number of units in the reservoir Rez_t ,

and penalization of the ridge regression used for the estimation of W_{Out}) may further be optimized. However, this will not be performed in our study due to the relatively short length of the time series that are further plagued by extreme events such as the financial crisis, potentially leading to an overfitting of the data should the meta-optimization be carried out.

The model is applied in a symmetric (ESN) and asymmetric (AESN) versions. In the symmetric version, the realized variance $RV(t-1)$ is the only explanatory variable X_t of the model and the realized variance $RV(t)$ is the target variable y_t . In the asymmetric version (AESN), analogically to the AHAR model, the return $Rev(t-1)$ is added to X_t as an additional explanatory variable. Unlike HAR and AHAR models, only the realized variances and returns at $t-1$ are used (i.e. only daily variances and returns), as the model should be able to capture the long memory of the realized variance by itself, so it is not necessary to include past weekly and monthly realized variances as additional inputs.

3 Results and Discussion

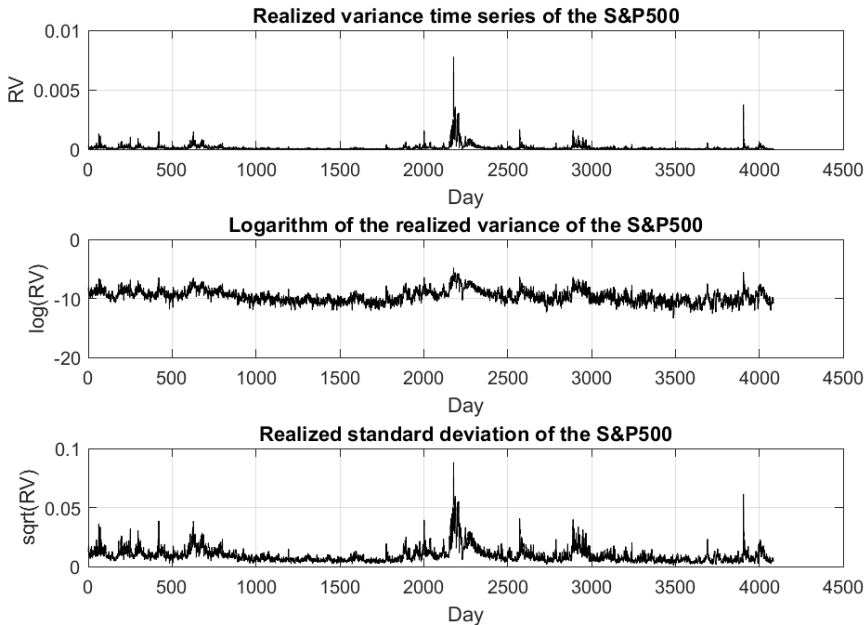
All of the models (HAR, AHAR, ESN and AESN) were applied to 19 stock market indices realized variance time series downloaded from the Oxford Man Institute realized volatility library. The dataset contains daily realized variances and daily returns of the stock market indices over the period of 4,274 days, ranging from January 3, 2000 to May 11, 2016. The realized variances were computed using 5-minute high-frequency returns.

For testing of the models, the first 3,000 days were set as the in-sample period (ranging from January 3, 2000 to June 29, 2011), to estimate the model parameters and the last 1,274 days as the out-sample period (ranging from June 30, 2011 to May 11, 2016), to evaluate their predictive power, based on the R-Squared criterion. All the models were implemented using the Matlab software.

In addition to the direct application of the models to the realized variances, they were further applied to the log-transformed realized variances and to the square roots of the realized variances (i.e. to the realized standard deviations). These transformations are sometimes performed so that the models' residuals correspond more closely to the normal distribution assumed by the linear regression and thus increase the model performance. The final R-Squared values were in all the cases (including transformations) calculated for the one period ahead predictions of the realized variance so that the results are comparable across the models.

Figure 1 shows the realized variance time series for the stock index S&P 500 and the time series of logarithmic and square root transformations of the realized variance.

Fig. 1 Realized variance time series for the S&P 500 and its transformations



Source: Authorial computation.

Tables 1 and 2 show the out-sample results (R-Squared criterions) for all of the models, transformations and realized variance time series.

From the results it is apparent that asymmetric models (AHAR and AESN) provide on average more accurate volatility forecasts than symmetric models, indicating that the correlation between volatility and returns plays an important role in stock market volatility forecasting.

We can also see that the best results were on average achieved for the logarithmic transformation of the realized variance time series, which exhibits a closer to normal distribution than the original realized variance time series or the square root transformation.

The proposed neural network based AESN model outperformed the AHAR model on some of the time series, most importantly on the S&P 500 stock index, the DJIA index and the Nikkei 225 index. Nevertheless, when an average R-Squared value for all of the time series is calculated, we can see that the benchmark AHAR model slightly outperformed the AESN model. This indicates that nonlinearities in

the realized variance time series may not be large enough in order to exploit the advantages of neural networks for modelling of their behaviour.

Nevertheless, Echo State Neural Network models proved to be a viable alternative for realized variance modelling, achieving a competitive performance with the industry benchmark HAR and AHAR models, despite the fact that they used only the last day realized variance (and return in the case of the AESN model) as the predictor, indicating that they were able to partially approximate the long memory dynamics of the volatility process with their recurrent layer.

Tab. 1 Out-Sample R-Squared values of the one-day realized variance forecasts

	<i>HAR</i>			<i>AHAR</i>		
	σ^2	$\log(\sigma^2)$	σ	σ^2	$\log(\sigma^2)$	σ
S&P 500	0.085	0.132	0.116	0.123	0.152	0.16
FTSE 100	0.235	0.29	0.285	0.115	0.312	0.315
Nikkei 225	0.263	0.273	0.277	0.282	0.302	0.299
DAX	0.27	0.323	0.293	0.168	0.336	0.315
Russel 2000	0.332	0.366	0.375	0.136	0.415	0.429
DJIA	0.006	0.075	0.05	0.049	0.086	0.084
Nasdaq 100	0.178	0.236	0.217	0.198	0.273	0.294
CAC 40	0.32	0.345	0.327	0.273	0.357	0.349
Hang Seng	0.263	0.256	0.276	0.263	0.262	0.284
KOSPI Composite	-0.078	0.296	0.272	-0.374	0.31	0.296
AEX Index	0.278	0.327	0.303	0.274	0.341	0.332
Swiss Market Index	-0.051	0.055	-0.009	-0.105	-0.05	-0.054
IBEX 35	0.306	0.329	0.312	0.281	0.328	0.32
S&P CNX Nifty	-0.169	0.064	0.038	-0.966	0.084	0.078
IPC Mexico	-0.039	0.011	-0.009	-0.014	0.017	0.007
Bovespa Index	0.147	0.177	0.152	0.116	0.183	0.171
Euro STOXX 50	0.289	0.301	0.294	0.208	0.318	0.323
FT Straits Times	0.165	0.289	0.27	0.224	0.305	0.301
FTSE MIB	0.299	0.326	0.307	0.289	0.357	0.347
Average	0.163	0.235	0.218	0.081	0.247	0.245

Source: Authorial computation.

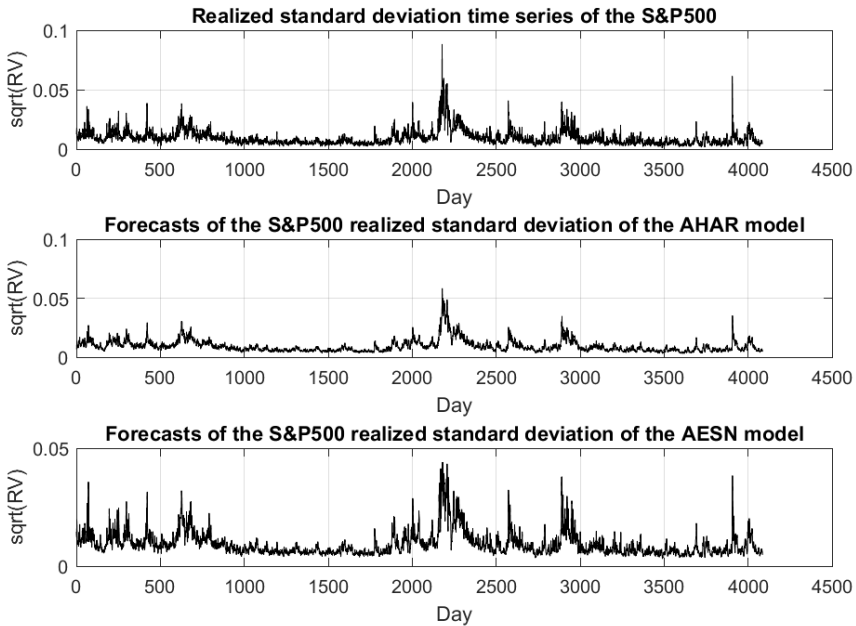
Tab. 2 Out-Sample R-Squared values of the one-day realized variance forecasts - continuation

	<i>ESN</i>			<i>AESN</i>		
	σ^2	$\log(\sigma^2)$	σ	σ^2	$\log(\sigma^2)$	σ
S&P 500	0.05	0.158	0.111	0.141	0.198	0.175
FTSE 100	0.111	0.275	0.226	-0.015	0.294	0.201
Nikkei 225	0.137	0.262	0.233	0.181	0.327	0.289
DAX	0.07	0.266	0.192	0.043	0.277	0.189
Russel 2000	0.241	0.381	0.338	0.249	0.408	0.363
DJIA	0.035	0.101	0.07	0.095	0.133	0.12
Nasdaq 100	0.172	0.251	0.218	0.266	0.302	0.297
CAC 40	0.158	0.323	0.261	0.122	0.341	0.251
Hang Seng	0.196	0.296	0.275	0.16	0.309	0.289
KOSPI Composite	-0.389	0.172	0.036	-0.648	0.165	-0.001
AEX Index	0.215	0.322	0.276	0.21	0.337	0.283
Swiss Market Index	0.06	0.077	0.066	0.032	0.04	0.029
IBEX 35	0.236	0.323	0.298	0.227	0.334	0.296
S&P CNX Nifty	-0.309	0.03	-0.05	-0.188	0.107	0.048
IPC Mexico	-0.003	0.026	0.02	0.007	0.035	0.029
Bovespa Index	-0.092	0.118	0.03	-0.114	0.134	0.053
Euro STOXX 50	0.105	0.271	0.212	0.026	0.283	0.192
FT Straits Times	-0.08	0.189	0.114	-0.141	0.232	0.106
FTSE MIB	0.175	0.322	0.27	0.136	0.356	0.279
Average	0.057	0.219	0.168	0.042	0.243	0.184

Source: Authorial computation.

Figure 2 shows development of the realized standard deviation of the S&P500 index and its 1-day leading forecasts constructed with AHAR and AESN models for illustration. Realized standard deviation is used in the figure instead of the realized variance as differences between the series are more visible for this transformation. The out-sample period starts at the 3,000th day of the series.

Fig. 2 Realized standard deviation of the S&P 500 index and its forecasts



Source: Authorial computation.

4 Conclusion

Echo State Neural Network model (ESN) and Asymmetric Echo State Neural Network model (AESN) were adopted to forecast the realized variance of 19 major stock market indices. The results indicate that Echo State Neural Networks provide a promising alternative to the commonly used HAR and AHAR models of realized volatility, with the AESN model achieving similar performance to the AHAR model. The fact that the nonlinear AESN model did not significantly outperform the AHAR model points towards the possible lack of complex nonlinear relationships in the realized variance time series. Nevertheless, the close average performance of the two models indicates that potential improvements of ESN and AESN models may outperform HAR and AHAR models, currently adopted as industry benchmarks. The presented study also did not apply any optimization to the meta-parameters of the neural networks, which may possibly increase the performance of Echo State models even further. The conclusion is that Echo State Neural Networks represent a promising new method for realized volatility modelling.

References

- Andersen, T. G. and Bollerslev, T., 1998. Answering the Skeptics: Yes, Standard Volatility Models Do Provide Accurate Forecasts. *International Economic Review* 4, 885-905. DOI: 10.2307/2527343.
- Bollerslev, T. G., 2008. Glossary to ARCH (GARCH). *SSRN Electronic Journal*. DOI: 10.2139/ssrn.1263250.
- Calvet, L., Fisher, A., 2001. Forecasting Multifractal Volatility. *Journal of Econometrics* 1, 27-58. DOI: 10.1016/s0304-4076(01)00069-0.
- Corsi, F., 2004. A Simple Long Memory Model of Realized Volatility. *SSRN Electronic Journal*. DOI: 10.2139/ssrn.626064.
- Elman, J. L., 1990. Finding structure in time. *Cognitive Science* 2, 179-211. DOI: 10.1207/s15516709cog1402_1.
- Jaeger, H., Haas, H., 2004. Harnessing Nonlinearity: Predicting Chaotic Systems and Saving Energy in Wireless Communication. *Science* 5667, 78-80. DOI: 10.1126/science.1091277.
- Lukoševičius, M., Jaeger, H., 2009. Reservoir Computing Approaches to Recurrent Neural Network Training. *Computer Science Review* 3, 127-149. DOI: 10.1016/j.cosrev.2009.03.005.
- McAleer, M., Medeiros, M. C., 2011. Forecasting Realized Volatility with Linear and Nonlinear Univariate Models. *Journal of Economic Surveys* 1, 6-18. DOI: 10.1111/j.1467-6419.2010.00640.x.
- Pong, S., Shackleton, M. B., Taylor S. J., Xu X., 2003. Forecasting Currency Volatility: A Comparison of Implied Volatilities and AR(FI)MA Models. *Journal of Banking and Finance* 10, 2541-2563. DOI: 10.1016/j.jbankfin.2003.10.015.
- Shephard, N., 2005. *Stochastic Volatility – Selected Readings*. Oxford University Press, New York.
- Vortelinos, D. I., 2015. Forecasting realized volatility: HAR against Principal Components Combining, neural networks and GARCH. *Research in International Business and Finance* 39, 824-839. DOI: 10.1016/j.ribaf.2015.01.004.

