sciendo

Neural Network Approach in Forecasting Realized Variance Using High-Frequency Data

Josip Arnerić
Faculty of Economics and Business, University of Zagreb, Zagreb, Croatia
Tea Poklepović
Faculty of Economics, Business and Tourism, University of Split, Croatia
Juin Wen Teai
National University of Singapore, Singapore

Abstract

Background: Since high-frequency data have become available, an unbiased volatility estimator, i.e. realized variance (RV) can be computed. Commonly used models for RV forecasting suffer from strong persistence with a high sensitivity to the returns distribution assumption and they use only daily returns. **Objectives:** The main objective is measurement and forecasting of RV. Two approaches are compared: Heterogeneous AutoRegressive model (HAR-RV) and Feedforward Neural Networks (FNNs). Even though HAR-RV-type models describe RV stylized facts very well, they ignore its nonlinear behaviour. Therefore, FNN-HAR-type models are developed. **Methods/Approach:** Firstly, an optimal sampling frequency with application to the DAX index is chosen. Secondly, in and out of sample predictions within HAR models and FNNs are compared using RMSE, AIC, the Wald test and the DM test. Weights of FNN-HAR-type models are estimated using the BP algorithm. **Results:** The optimal sampling frequency of RV is 10 minutes. Within HAR-type models, HAR-RV-J has better, but not significant, forecasting performances, while FNN-HAR-J and FNN-LHAR-J have significantly better predictive accuracy in comparison to the FNN-HAR model. **Conclusions:** Compared to the traditional ones, FNN-HAR-type models are better in capturing nonlinear behaviour of RV. FNN-HAR-type models have better accuracy compared to traditional HAR-type models, but only on the sample data, whereas their out-of-sample predictive accuracy is approximately equal.

Introduction

Volatility is most important variable in asset pricing, portfolio management and risk estimation since it is an indicator of financial disturbances. Volatility is usually measured by variance of price returns, i.e. as a constant parameter. Nowadays it is accepted as a time-varying unobservable variable. Therefore, both researchers and practitioners are interested in modelling and forecasting volatility. It poses a challenge to both financial modellers and economists due to financial time series features that volatility models have to take into account: strong persistence, i.e. the autocorrelation of the squared and absolute returns; not normally distributed returns and slowly converging to the normal distribution as the time scale increases (Corsi, 2009); heteroscedasticity; and volatility clustering.

There are five types of volatility estimators (Degiannakis et al., 2015): historical volatility measure (HV); implied volatility (IV); AutoRegressive Conditional Heteroscedasticity (ARCH) model; stochastic volatility (SV) model and realized volatility (RV) which uses intra-daily (high-frequency) data to estimate *ex-post* volatility using different sub-sampling methods. According to the theory of quadratic variation of semi-martingales, the integrated volatility (true but unknown parameter) can be consistently estimated by the realized variance (Barndorff-Nielsen et al., 2002a, b). This is unrealistic, as prices are not continuously observed in practice and leads to the biased estimates of true volatility and to the market microstructure noise, which makes RV less accurate (Degiannakis et al., 2015).

Theoretically speaking, for a very short time interval there is an "infinite" number of intra-daily returns, but high-frequency data is not available for all stocks and markets. Only for most liquid markets, one can assume that the stock prices evolve nearly continuously during the trading day. In practice, RV estimator becomes biased with the increase of sampling frequency (Bandi et al., 2008; 2011). The bias is caused by non-synchronous trading, discrete price observations and bid-ask bounce. Market microstructure noise in high-frequency data exhibits serial correlation (Aït-Sahalia et al., 2005; Hansen et al., 2005; 2006). The broad opinion in practical implications is not to sample too often, i.e. sparsely sampling is recommended from 5 to 30 minutes. Therefore, it is important to choose appropriate sampling frequency, i.e. RV should be adjusted and sampling frequency should be selected based on a trade-off between accuracy and potential biases.

Moreover, RV shows few features (Corsi et al., 2012): long-range dependencies, i.e. hourly, daily, weekly and monthly RV displays significant autocorrelation at very long lags which is known as long-memory data generating process; leverage effects, i.e. returns are negatively correlated with volatility, where volatility bursts are associated with negative past returns; and jumps from previous period that have strong impact on future volatility (Zhang et al., 2005; Bandi et al., 2008). Additionally, empirical studies argue that daily RV estimates based on high-frequency data provide volatility forecasts that are superior to forecasts constructed from daily returns.

Based on daily log-returns, among the others, ARCH model proposed by Engle (1982) and its generalized version (GARCH), proposed by Bollerslev (1986), are widely used in empirical researches. The empirical results strongly suggest that simple models of RV outperform the popular GARCH model in out-of-sample forecasting (Andersen et al., 2003).

Corsi (2003) proposes an additive cascade model of different volatility components. Market participants with different time horizons, i.e. different trading frequencies, react to and cause different types of volatility components: short-term traders with daily or intra-daily trading (such as dealers and intra-day speculators);

the medium-term agents who rebalance their positions weekly; and the long-term investors with less frequent trading, i.e. one or more months. This idea led to a simple AR-type model in the RV which considers volatilities over different time horizons, i.e. Heterogeneous AutoRegressive model of Realized Volatility (HAR-RV). Although rather simple, this model is able to reproduce the dynamics and stylized facts of RV. HAR-RV model used on USD/HF data showed good out-of-sample forecasting performances which outperformed standard models. Therefore, HAR-RV is broadly applied (Andersen et al., 2007a; Corsi, 2003; 2009) due to its simplicity in estimation and prediction. Corsi (2009) uses three data sets of RVs (USD/CHF, S&P500 and T-Bond futures) at the frequency of ten-ticks return. Results show that HAR-RV model has good properties in modelling the long-memory behaviour of volatility and good forecasting performance compared to the short-memory (AR-type) models at all-time horizons. Recent empirical studies suggested the importance of jumps (J) in modelling RV (Andersen et al., 2007a; Corsi et al., 2012; Huang et al., 2013; Clements et al., 2013). Therefore, the extension of HAR-RV model, i.e. HAR-RV-J, is analysed. Andersen et al. (2007a) use a bipower variation measure and jump tests. The analysis uses a decade of five-minute high-frequency returns for the DM/USD spot market, S&P500 and 30-year US Treasury yield futures. They demonstrate that important gains are obtained in volatility forecast accuracy by incorporating jumps into HAR-RV model.

Junior and Pereira (2011) consider HAR-RV and Mixed Data Sampling (MIDAS-RV) model, using intra-daily data for the most liquid stocks on the São Paulo Stock Market. The sample includes 3 years of high-frequency data using 5-minute interval as the best among the others with lower frequency. Firstly, they estimate daily volatilities using different GARCH-type models and RV. Further, they augment the research by filtering the RV to correct for microstructure problems and estimate the two models for time horizons of 1, 5, 10, 15 and 20 days. They compare in- and out-of-sample forecasts and conclude that better results are obtained with MIDAS-RV model for in-the-sample. For out-of-sample forecasts, there is no significant difference between models. They contributed to the literature of RV for emerging countries. Huang et al. (2013) combine HAR model with continuous volatility and jumps (HAR-CJ), and use the influence of momentum (M) to develop a new model called HAR-CJ-M. It is compared to two models, HAR-ARV and HAR-CJ. The application of these models to Chinese stock market index using 5 years of 5-minutes high-frequency data shows that continuous sample path variation, momentum effect and ARV (adjusted RV) have a good forecasting performance on the future ARV. Moreover, the HAR-CJ-M model shows better forecasting performance than the other two models.

Through the literature, HAR-RV model for RV forecasting is inevitable starting point since it outperforms the short-memory models and is comparable to long-memory models. However, some authors argue that inclusion of jump component into models for RV forecasting significantly improves the predictive accuracy. Therefore, the HAR-RV-J model with and without leverage (L) effects is also considered.

Furthermore, nonlinear dependence of financial time series is usually present. Moreover, empirical research found that dynamics of the volatility is also nonlinear (Corsi et al., 2012; Rosa et al., 2014), however, it is not addressed appropriately. Several authors showed that RVs are subject to structural breaks and regime switches driven by shocks (Corsi et al., 2012). To capture nonlinearity logarithms or the square root of RV are used in most research (Andersen, 2007a; Clements et al., 2013). However, to account for nonlinearity, recently the attention is given to Neural Networks (NNs). NNs are a valuable method for modelling and forecasting time

series in general. Empirical studies show that NNs are appropriate in forecasting extremely volatile time series such as exchange rates (Zhang, 2003; Chaudhuri et al., 2016), interest rates (Dedi et al., 2011; Aljinović et al., 2013), stocks (Zekić-Sušac et al., 2002; Medeiros et al., 2006; Wang et al., 2016), inflation (Teräsvirta et al., 2005; Binner et al., 2006; Choudhary et al., 2008; Al-Maqaleh et al., 2016), GDP growth (Gonzales, 2000; Teräsvirta et al., 2005; Aminian et al., 2006; Teräsvirta et al., 2008) and volatility in general (Bildirici et al., 2009; 2012; 2014; Bektipratiwi et al., 2011; Mantri et al., 2010; 2012; Arnerić et al., 2014; 2016). Unlike traditional models NNs are assumption free, i.e. do not require any distributional assumption or functional form between inputs and outputs. However, they require an enormous amount of data, i.e. sufficiently enough data for training, validation and testing set. These requirements can easily be meet using high-frequency data. The Feedforward Neural Networks (FNNs), i.e. multilayer perceptrons, are most commonly used (Teräsvirta et al., 2008; Bildirici et al., 2009; 2012; 2014; Mantri et al., 2010; 2012; Bektipratiwi et al., 2011). However, there are only few papers dealing with NNs in forecasting RV (Jurkovič, 2013; Rosa et al., 2014; Baruník et al., 2016). Jurkovič (2013) use NNs to forecast daily RV of the EUR/USD, GBP/USD and USD/CHF based on 5-minute data. Their performance is benchmarked against HAR-RV and traditional ARIMA models. A simple enhancement to HAR model, named HARD, is introduced. NNs are used as nonparametric estimation technique where the whole set of NNs is estimated, from the simplest network (28 parameters) to more complex network (600 parameters). He demonstrates that NNs surpass other models in both disciplines, better fit in-the-sample and better forecasting performance. Some of them do so statistically significant, as shown by Diebold–Mariano (DM) test. However, this type of NN modelling through "black-box" is neither suitable nor appropriate from an econometric perspective. Moreover, up to two hidden layers are used, which leads to overfitting. Therefore, only one hidden layer should be considered to benefit from the econometric interpretation of the estimated parameters. Rosa et al. (2014) suggest a hybrid fuzzy NN (eHFN) approach for RV forecasting with jumps, which is nonlinear and time varying. They simultaneously choose the number of neurons and weights. The performance of eHFN is compared to FNN, linear regression and evolving fuzzy models using data for S&P500, Nasdaq, FTSE, DAX, IBEX and IBOVESPA with 4 years of 1-minute quotations. The results show that eHFN describes time varying RV with jumps very well, while FNN performed slightly worse. However, it is not defined how the inputs and hidden neurons are selected. Baruník and Křehlík (2016) compare HAR, ARFIMA and HAR-FNN models for prediction of energy prices volatility. They use different RV estimators and conclude that high-frequency models outperform standard GARCH model. Moreover, FNN model gives less precise short-term forecasts in crisis period, but it reduces overfitting. However, FNN prespecification is not done correctly since they use all 22 lags of volatility as inputs and either 7 or 15 hidden neurons. This paper, as most of the papers regarding NNs, uses them as nonparametric estimation methods, relying on trial-and-error technique in search of the appropriate architecture and neglecting the interpretability of parameters. Therefore, it is extremely important to understand the structure of RV to build a well-designed NN. Although there is no universal input variable selection in NN modelling, this paper relies on input variables that describe stylized facts of RV such as long memory, jumps and leverage effect. The overfitting problem due to the excessive hidden units is also considered and solved in this paper.

The main objective of this paper is RV forecasting using high-frequency data. Because of recent methodological developments, two approaches are compared: HAR-RV and FNN approach. Namely, a parsimonious FNN model as an extension to

HAR-RV model is developed that captures the nonlinear behaviour of RV along with other stylized facts. Few questions arise. What sampling frequency is low enough to avoid the bias when obtaining RV? Which number of hidden units is optimal in RV modelling? Which diagnostic statistics can be used in testing FNN and HAR-type models' goodness of fit? Which model yields the most accurate RV forecasts: FNNs or HAR-type models? In that context, the present paper contributes to the existing literature in a several ways. Firstly, through the selection of the optimal sampling frequency to compute RV, which is free of market microstructure noise. Secondly, finding the appropriate NN model that gives the most accurate RV forecasts of DAX index, where the structure specification of NN models that is comparable to the HAR-RV-type models is given *a priori*. From this perspective, suggestions about NN settings in achieving most accurate RV prediction are given. These settings include selection of input variables and optimal sampling frequency, choosing the number of hidden neurons, type of optimization technique and specifying starting values of the initial weights. Thirdly, this paper contributes by comparing FNNs against HAR-type models in context of in- and out-of-sample forecasting performance.

The remainder of the paper is organized as follows. Section 2 describes the data and methodology. Section 3 presents empirical findings. Discussion is given in Section 4. Concluding remarks and directions for future research are provided in the last section.

Methodology

The dataset of intra-daily price observations of the DAX index for each trading day covers the period from 04.01.2010 to 16.09.2016. It is provided by Thomson Reuters Tick History service at high-frequency of 1-minute interval, containing in total 894647 high-frequency data. The empirical analysis starts with the choice of the appropriate sampling frequency for computing the least biased RV, which should be selected based on a trade-off between accuracy and potential biases due to market microstructure frictions (Degiannakis et al., 2015). Namely, market friction is a source of additional noise in the volatility estimate. A 1-second sampling frequency would result with many gaps in transaction data, lots of zero prices and thus poor quality of intra-daily observations. Contrary, Andersen et al. (2003) prove that the 30-minute returns ensure stable volatility, but this leaves few daily observations at hand. In this paper the optimal sampling frequency is not chosen randomly, as in many researches, but based on minimization of the mean squared error (MSE) of RV estimator considering equidistant sampling interval (Zhang et al., 2005). Many studies propose a sampling frequency of 5 or 30 minutes. However, in this paper the prices observed every 10 minutes are the most appropriate, since as much information as possible is kept and they are free of market microstructure noise. Before further analysis, the stylized facts of DAX indexes' RV based on 10-minute prices are illustrated in Figure 1. The differences between non-robust and robust RV are given in panels in the first row. Robust RV is obtained as a bipower variation (BPV) (Barndorff-Nielsen et al., 2004). BPV is widely applied as an efficient estimator of integrated variance that is robust to jumps and microstructure noise. The difference between RV and BPV is used to perform ABD test for detecting significant jumps (Anderson at al., 2007b). The significant jumps according to the ABD test (left bottom panel) are later used as input variable in extended HAR-RV model, i.e. HAR-RV-J model. Autocorrelation function (ACF) of RV computed on 10-minute returns of DAX index is given on the right bottom panel (thin line). ACF of RV proves the existence of long memory as it takes a long time to die out and decays very slowly. Lagged jumps also seem to have positive effect on RV as all cross-correlation coefficients are positive

even with lower values compared to autocorrelation coefficients (dotted line). Cross-correlation coefficients between lagged negative daily returns and RV (dashed line) suggest the existence of leverage effect, i.e. past negative returns are positively correlated with current RV and highly persistent. This confirms the existence of a heterogeneous structure that can be described by a cascade of heterogeneous volatility components aggregated at different time horizons, i.e. HAR-RV model.

Figure 1
Stylized Facts of DAX Indexes' RV
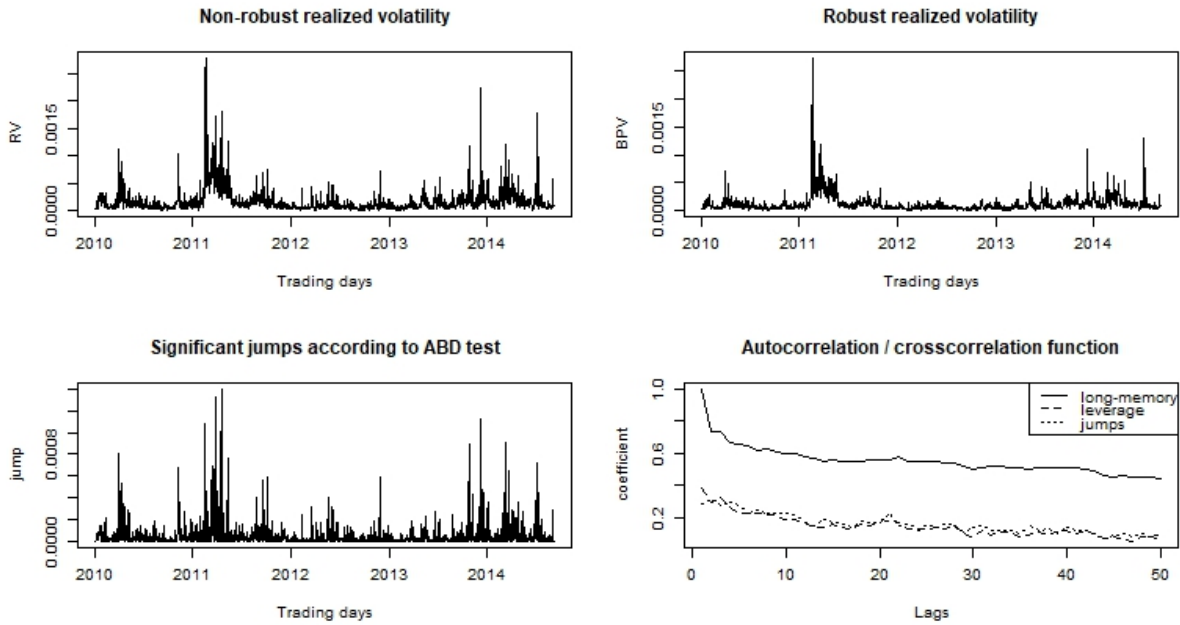
Source: Authors' work

After choosing the optimal sampling frequency, the sample is divided into two subsamples, i.e. from 04.01.2010 to 30.12.2015 and from 04.01.2016 to 16.09.2016. as in and out of sample periods respectively. The number of 10-minute intra-daily returns in-the-sample is 80698 which correspond to the first 1504 trading days. Out-of-sample data set (9640 intra-daily returns, i.e. the last 181 trading days) is kept for forecasting accuracy comparisons between HAR-type and FNN models.

The main advantage of the HAR-RV model is its simplicity and easiness to estimate by least squares as a special type of linear model. The standard HAR model proposed by Corsi (2009) is as a linear function of the lagged RV over daily, weekly and monthly horizons, since the agents with different time horizons perceive different types of volatility components. Three components are usually used with length $h_1 = 1$ (daily), $h_2 = 5$ (weekly), $h_3 = 22$ (monthly). Thus HAR-RV model for $t = 1, 2, ..., T$ trading days is:

$$RV_t = \beta_0 + \beta_D RV_{t-1} + \beta_W RV_{t-5} + \beta_M RV_{t-22} + \varepsilon_t, \text{ where } \varepsilon_t \square i.i.d.(0,1)_.$$ (1)

HAR-RV-J model that explicitly considers jumps is extension of standard HAR-RV model, aggregated at the same horizons is:

$$RV_t = \beta_0 + \beta_D RV_{t-1} + \beta_W RV_{t-5} + \beta_M RV_{t-22} + \beta_D^J J_{t-1} + \beta_W^J J_{t-5} + \beta_M^J J_{t-22} + \varepsilon_t. \tag{2}$$

Variables that describe the jumps are estimated by $J_t = \max(RV_t - BPV_t, 0)$. Only significant jumps are retained as inputs according to the ABD z-statistic (Andersen et al., 2007a). Otherwise, the value of the jump component is replaced with zero, as small and non-significant jumps cause measurement errors.

HAR-RV-J model is extended with leverage (L) components aggregated on daily, weekly and monthly basis. It is well known that volatility increases more after a negative shock than after a positive shock of the same magnitude. Therefore, it is assumed that RV reacts asymmetrically not only to previous daily returns but also to previous weekly and monthly returns:

$$RV_t = \beta_0 + \beta_D RV_{t-1} + \beta_W RV_{t-5} + \beta_M RV_{t-22} + \beta_D^J J_{t-1} + \beta_W^J J_{t-5} + \beta_M^J J_{t-22} + $$
$$+ \beta_D^L L_{t-1} + \beta_W^L L_{t-5} + \beta_M^L L_{t-22} + \varepsilon_t \tag{3}$$

Variables that describe the leverage effects in equation (3) are estimated by $L_t = \min(r_t, 0)$ where r_t is aggregated daily return on certain horizon. The model (3) is labelled as LHAR-RV-J.

To account for possible nonlinearity in the RV dynamics, variables are usually transformed into logs or taking the square root. An alternative solution is to use NNs. Specifically in this paper, an extension of HAR-RV, HAR-RV-J and LHAR-RV-J models within NN architecture is proposed. The most commonly used type of NN is feed-forward neural network (FNN).

The FNN consists of an input layer, one (or more) hidden layers and an output layer that are connected by weights and activation functions. The method used for weight estimation in a FNN is called backpropagation (BP) learning algorithm. It is based on a gradient descent approach, which iteratively updates the network weights with the aim to minimize the objective function. Mean squared error (MSE), defined as mean of squared differences between the observed and fitted values of time series, is most commonly used objective function in a FNN. During learning, input patterns are fed forward through the network, layer by layer, until the expected (fitted) value of output is calculated. Expected output is then compared to the observed (or target) output based on which the error value is calculated. The errors are used as inputs in a reverse propagation where the weights are corrected backward layer by layer in iterative manner. This process is repeated until the total error converges to a minimum or until the maximum number of iterations is reached. Three-layer FNN in most general form is (Adhikari et al., 2013):

$$y_t = f\left(\phi_{co} + \sum_{h=1}^{q} \phi_{ho} g\left(\phi_{ch} + \sum_{i=1}^{p} \phi_{ih} x_{t,i}\right)\right) + \varepsilon_t, \tag{4}$$

where y_t is the output vector of a time series presenting dependent variable, x_t is the input matrix with p time series as input variables $i = 1, 2, \ldots, p$, while $f(\cdot)$ and $g(\cdot)$ are linear and logistic activation functions. Weights ϕ_{co} and ϕ_{ch} are the constant terms of an output and hidden neurons (where $h = 1, 2, \ldots, q$) in one hidden layer

respectively. Weight ϕ_{ih} is the connection between *p*-inputs and *h*-hidden neurons, ϕ_{ho} denotes the connection between *h*-hidden neurons and one output, while ε_t is an error term. FNN containing *p* inputs, *h* neurons in one hidden layer and 1 output is denoted as FNN(*p,h,*1).

In the context of nonlinear HAR-RV model, FNN-HAR model is proposed by replacing the output vector y_t with daily RV based on 10-minute intra-daily returns of DAX index and replacing input matrix x_t by RVs aggregated on daily, weekly and monthly basis. When a set of input variables includes RVs with the lags $t-1$, $t-5$ and $t-22$ then FNN(3,*h*,1) is applied. The number of hidden neurons *h* should be determined in a single hidden layer to avoid the overfitting. Reducing the number of hidden layers, in general, as well as neurons of hidden layers helps in avoiding local minima. Overfitting occurs due to the inclusion of excessive hidden layers and/or neurons, which, with existing number of inputs, increases the number of parameters to estimate (Franses et al., 2003). In majority of empirical research, the number of hidden neurons is determined by trial and errors. However, there are different methods developed for hidden neurons selection (Angus, 1991; Zapranis et al., 1999; Kaashoek et al., 2001; Franses et al., 2003; Medeiros et al., 2006). The first approach, backward or pruning, in which a model with a large number of hidden units is estimated first, and the size of the model is subsequently reduced by a cross-validation which gradually eliminate neurons that do not significantly contribute to the NN (Zhang et al.,1998; Kaashoek et al., 2001). The second approach, forward or cascading, starts from the smallest NN to larger ones, which implies adding neurons until the improvement of information criteria is achieved (Franses et al., 2003; Medeiros et al., 2006). In this paper, the second approach is used as discussed in Li et al. (1995) and since Medeiros et al. (2006) argue that suitable statistical conclusions cannot exist if the choice of NN architecture goes from larger to smaller models, i.e. from general to specific. Moreover, some research recommended as a rule of thumb in a single hidden layer to use the number of neurons equal or less than the number of inputs (Hagiwara, 1994; Keeni et al., 1999; Moshiri et al., 2000; Doukim et al., 2010). This recommendation is incorporated in this paper along with cascading approach.

Results

Firstly, the optimal sampling frequency for RV modelling is selected. It is not chosen randomly but based on minimization of the MSE of RV estimator considering equidistant sampling interval. The optimal sampling frequency is 10-minutes. After the sampling frequency is selected, the estimation and comparison of HAR-type and FNN models is performed. Three sets of inputs are considered in both types of models. First set of inputs describes long-memory (RVs aggregated on the daily, weekly and monthly basis), the second set of inputs is extended for the jumps (J), and the third set of inputs includes leverage (L) effects (all aggregated at the same horizon levels). According to three sets of inputs, HAR-type models given by equations (1)-(3) are estimated using least squares method. The same data in-the-sample is used as in FNNs' training and validation set, while the out-of-sample data is used for forecasting accuracy comparison.

As well-designed NN depends on many configuration settings, more attention is given to NN modelling. Specifically, three types of FNNs are estimated, i.e. FNN-HAR (Feedforward Neural Network-Heterogeneous AutoRegressive), FNN-HAR-J (FNN-HAR-Jump) and FNN-LHAR-J (FNN-Leverage HAR-Jump) model, using BP algorithm

with small learning rate (0.001) and large number of iterations (1000). BP is comparable to the gradient descent technique as a special case of the Newton-Raphson method where the Hessian matrix is replaced by learning rate. Learning rate is the step size that controls the length of the search while the negative of the gradient determines the direction of the search. Too high learning rate will cause divergence. Therefore, smaller learning rate is applied with large number of iterations at the cost of slow convergence. Furthermore, output and input variables for training and validation are normalized to assure convergence in the iterative optimization process. As recommended in the literature 70% of in-the-sample data is split for training and 30% for validation. Even there is no general rule for specifying starting values for initial weights, interval [0,1] is selected due to variance positivity. Logistic and linear activation functions are applied to input-to-hidden and hidden-to-output layer respectively. Logistic activation function seems convenient when transforming normalized inputs to a [0,1] interval. FNN are criticized in the literature for large number of parameters to estimate and their sensitivity to overfitting. Hence, in this paper most settings of FNNs are defined in advance to benefit from parsimonious model with sufficiently small number of hidden neurons in order to avoid overfitting and at the same time to obtain most accurate RV forecasts. All results are obtained using the "R" package.

Figure 2
FNN-type models architecture
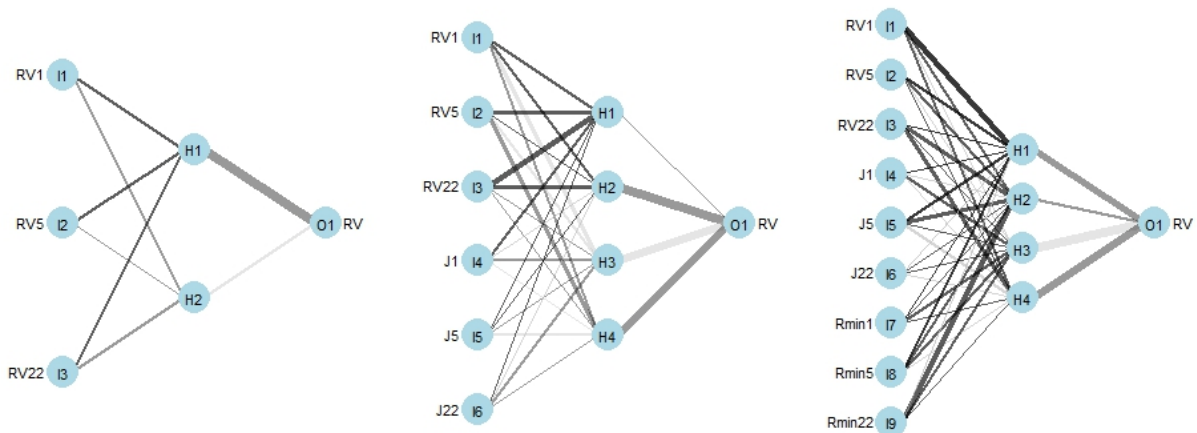
Source: Authors' work

Adopting the cascading approach 15 FNNs are estimated for three different sets of inputs, based on a rule of thumb recommended in the literature, each time adding an additional neuron in a single hidden layer and computing RMSE in-the-sample until the improvement in RMSE is achieved. Therefore, the optimal number of hidden units is defined for all models with the minimum RMSE in-the-sample. The results showed that 2 neurons for FNN-HAR, 4 for FNN-HAR-J and 4 for FNN-LHAR-J is the optimal number of hidden neurons. Thus, it is confirmed that the number of hidden neurons in a three-layered FNN with one output is less than the number of input variables. This is in accordance with the parsimony principle as expected due to the valid selection of input variables. A small number of hidden neurons reduce dimensionality and overfitting. Architectures of FNN-type models are presented in Figure 2. Line thickness of connections is proportional to the magnitude of each weight. Nodes of inputs, hidden neurons and output are labelled with capital I, H and O respectively. Bias nodes connected to hidden and output neurons are

omitted due to reduction of the number of parameters. The optimal results with the in-the-sample diagnostics for both HAR- and FNN-type models are given in Table 1.

Table 1
HAR-RV- and FNN-type Models Performance and Goodness-of-fit in-the sample

a) HAR-RV-type models			
	HAR-RV	**HAR-RV-J**	**LHAR-RV-J**
Inputs	3	6	9
RMSE	0.000164	0.000159	0.000152[a]
AIC	-26198.67	-26289.06	-26430.66[a]
Wald test	HAR-RV	33.027***	43.856***
	HAR-RV-J	-	51.352***
b) FNN-type models			
	FNN-HAR	**FNN-HAR-J**	**FNN-LHAR-J**
Neurons	2	4	4
Weights	8	28	40
RMSE	0.000147[a]	0.000149	0.000148
AIC	-26512.5[a]	-26446.36	-26434.69
Wald test	FNN-HAR	1271	0.418
	FNN-HAR-J	-	1004

Note: [a] represent the smallest RMSE and AIC within each type of model. *, **, *** represent significance at the 10%, 5% and 1% level of Wald test.
Source: Authors' work

Since each diagnostic test has its unique properties, it is necessary to use more than one measure in comparing different types of models. The most commonly used measure for model comparison is MSE or RMSE (Gonzales, 2000; Moshiri et al., 2000; Zhang, 2003; Teräsvirta et al., 2005; Binner et al., 2006; Medeiros et al., 2006; Choudhary et al., 2008; Bildirici et al., 2012; 2014; Chaudhuri et al., 2016; Al-Maqaleh et al., 2016; Wang et al., 2016). It measures the mean squared deviations of the observed from the expected (or predicted) values. In this case, errors of the opposite sign are not cancelled and therefore (R)MSE provides a good measure of estimated error. However, it does not take into account the principle of parsimony. In order to select a suitable model, it is necessary to use an indicator that takes into account the number of parameters, MSE and sample size, i.e. measure as Akaike information criteria – AIC. This paper also uses a Wald test (Buse, 1982) to test the zero restriction on weights for comparison purpose. It is appropriate diagnostic statistics in testing goodness-of-fit of the models. In addition to the selected indicators that can be calculated in-the-sample and out-of-sample, to test the predictive accuracy of HAR-type, as well as FNN-type models, Diebold-Mariano (DM) test (Diebold et al., 1995) is used. The models are tested in pairs. DM test is adopted to MSE loss function. It is assumed by the null hypothesis that unconditional expectation of loss function differential between two models is zero. If null hypothesis is rejected in favour of one sided alternative it is concluded that model j has significantly better predictive accuracy than model i. DM test is computed between models with the lowest value of MSE against other models using one sided alternative. Thus, a negative DM statistics can be shown in Table 2. Performance and goodness-of-fit measures out-of-sample are computed, along with HAR-type models in Table 2.

Table 2
HAR-RV- and FNN-type Models Performance and Goodness-of-fit out-of-sample

	a) HAR-RV-type models		
	HAR-RV	**HAR-RV-J**	**LHAR-RV-J**
RMSE	0.000193	0.000188[a]	0.000189
AIC	-3088.40	-3091.19[a]	-3083.92
DM test	HAR-RV	-0.729	-0.705
	HAR-RV-J	-	-0.099
	b) FNN-type models		
	FNN-HAR	**FNN-HAR-J**	**FNN-LHAR-J**
RMSE	0.000203	0.000192[a]	0.000194
AIC	-3062.58[a]	-3041.81	-3014.60
DM test	FNN-HAR	-2.7150***	-1.365*
	FNN-HAR-J	-	-0.442

Note: [a] represent the smallest RMSE and AIC within each type of model. *, **, *** represent
significance at the 10%, 5% and 1% level of Diebold-Mariano (DM) test.
Source: Authors' work

Out-of-sample forecasting results are compared and presented via plots of fitted
values of each model alongside the RV in Figure 3 for HAR and FNN pure models, in
Figure 4 for HAR and FNN models with jumps and in Figure 5 for HAR and FNN models
with jumps and leverage effects. Forecasting horizon includes 181 trading days, from
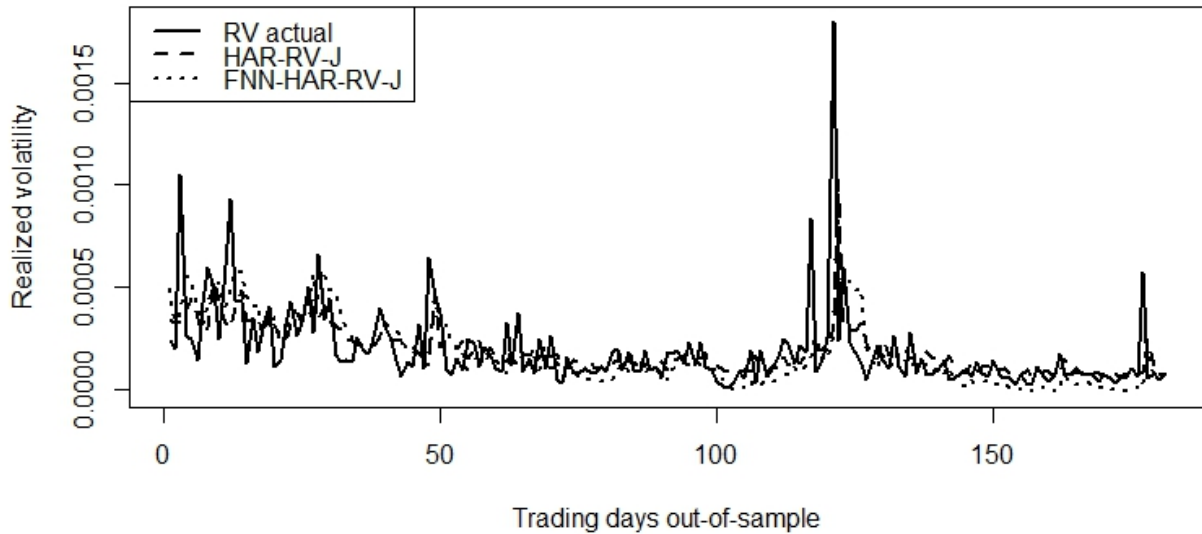04.01.2016 to 16.09.2016.

Figure 3
Out-of-sample comparison of actual RV against RV forecasts with pure HAR and FNN
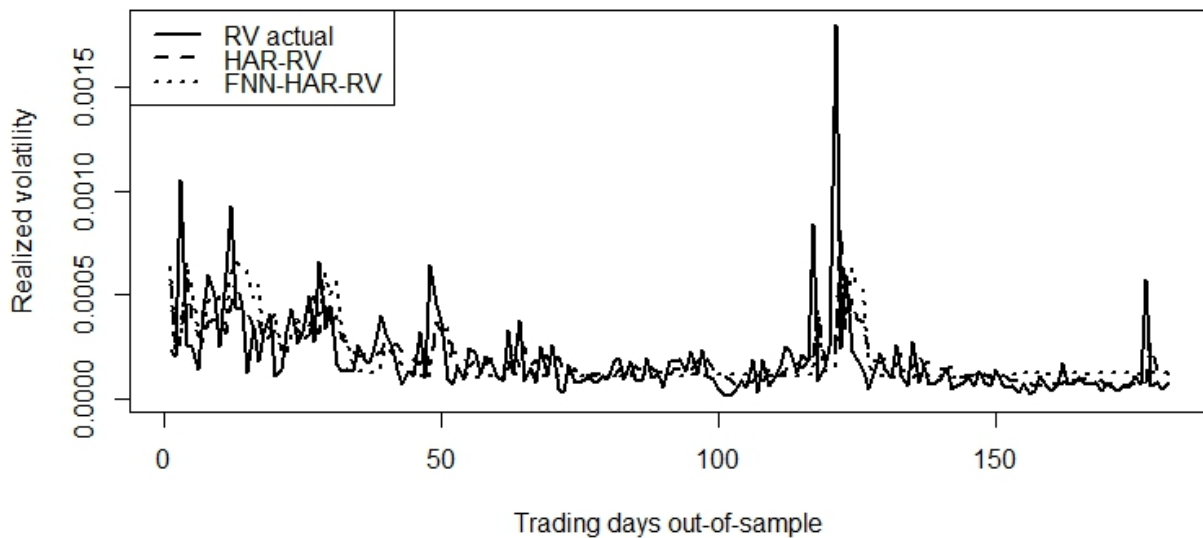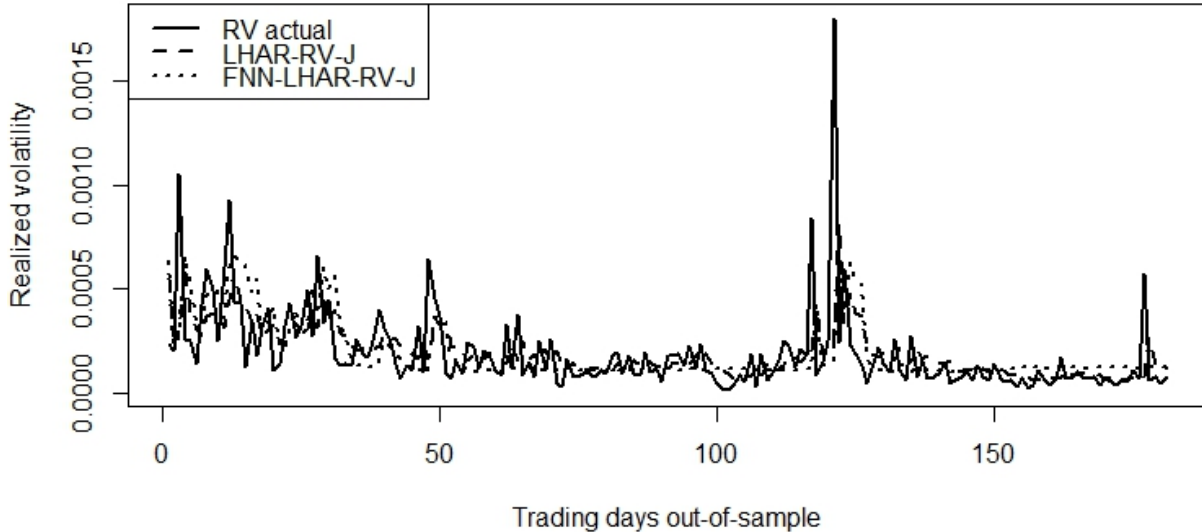

Source: Authors' work

Figure 4
Out-of-sample comparison of actual RV against RV forecasts with HAR and FNN including jumps

Source: Authors' work

Figure 5
Out-of-sample comparison of actual RV against RV forecasts with HAR and FNN including jumps and leverage effects

Source: Authors' work

Discussion

Within HAR-type models evaluation of RMSE in-the-sample shows significant superiority of HAR-RV model that includes both jump and leverage effects. However, the out-of-sample forecasts show better, but insignificant, performances of HAR-RV-J compared to HAR-RV and LHAR-RV-J. This is also confirmed via Akaike information criteria. The null hypothesis of DM test cannot be rejected when comparing HAR-type models, i.e. they have similar predictive accuracy. Therefore, within HAR-type

models it is important to include either jump or both jump and leverage component in modelling and forecasting RV.

Nevertheless, when nonlinearity is included for modelling and forecasting RV using one-layered FNN, the best in-the-sample fit is obtained with more parsimonious model, i.e. FNN-HAR model. However, Wald test shows that there is no significant difference in in-the sample performances between FNN-type models. When comparing forecasting performances of FNN-type models the null hypothesis can be rejected, i.e. according to DM test FNN-HAR-J and FNN-LHAR-J models have significantly lower MSE than FNN-HAR model at 1% and 10% significance level respectively. However, FNN-HAR-J and FNN-LHAR-J models do not differ significantly in their forecasting performances. Therefore, if more parsimonious model is to be selected, that simultaneously has good in-the-sample properties and good out-of-sample forecasting performances, which captures nonlinearity and does not lead to overfitting, than the best-fit model for RV is the FNN-HAR-J model.

Conclusion

As high-frequency data becomes available big amount of information can be used to estimate true unknown volatility more precisely. Traditional approaches to modelling and forecasting volatility mostly rely on close-to-close returns, i.e. close price observed at given trading day. Less traditional models use not only close prices, but open, high and low prices. Regardless, commonly used models for measuring volatility are parametric and therefore suffer from misspecification problem as well as unfulfilled assumptions. Unlike traditional models NNs are assumption free, i.e. do not require any distributional assumption or functional form between inputs and outputs. NN requires an enormous amount of data, i.e. sufficiently enough data for train, validation and test set, which can easily be meet with high-frequency data. Using high-frequency data an unbiased estimator of true volatility can be found, i.e. RV which is free of market microstructure noise. In that context sampling frequency should be selected based on a trade-off between accuracy and potential biases. In this paper the optimal sampling frequency is not chosen randomly, but based on minimization of the MSE of RV estimator considering equidistant sampling interval. The optimal sampling frequency is 10-minutes. After choosing the optimal sampling frequency the RV is computed with its components aggregated on the daily, weekly and monthly basis. A cascade of heterogeneous RV components aggregated at different time horizons can be successfully described by HAR-RV model. HAR-type extensions are used to take into account not only long-memory, but also jumps and leverage effect. Therefore, the focus of the paper is the RV forecasting using traditional HAR-type as well as FNN-type models. More attention is given to development of a parsimonious FNN to capture the main stylized facts of RV including nonlinear dependence. From this perspective, suggestions about FNN settings in achieving most accurate RV prediction are given. These settings include selection of input variables and optimal sampling frequency, the number of hidden neurons, type of optimization technique and specifying starting values of the initial weights. RMSE is minimized to find appropriate number of hidden neurons within every FNN. It is confirmed that the number of hidden neurons in a three-layered FNN with one output is less than the number of input variables. This is in accordance with the parsimony principle as expected due to the valid selection of input variables. Moreover, a relatively small number of hidden neurons reduce dimensionality problem and overfitting. Therefore, the appropriate FNN-HAR-type models for modelling and forecasting of RV are determined. Finally, out-of-sample forecasts between HAR-RV- and FNN-type models are compared to determine their predictive

accuracy with application to DAX index. According to pairwise DM test, there is no statistically significant difference in predictive accuracy between HAR-type models. Nevertheless, the null hypothesis of DM test can be rejected when FNN-type models are compared, i.e. FNN-HAR-J and FNN-LHAR-J models have significantly lower MSE than FNN-HAR model at 1% and 10% significance level respectively. However, FNN-HAR-J and FNN-LHAR-J models do not differ significantly in their forecasting performances. According to AIC information criteria, FNN model with less hidden neurons is preferred in-the-sample. However, Wald test shows that there is no significant difference in performance between FNN-type models in-the sample. Compared to the traditional ones, FNN-HAR-type models are better in capturing nonlinear behaviour of RV. In-the-sample, FNN-HAR-type models fit better against traditional HAR-type models, whereas their out-of-sample predictive accuracy is approximately the same. Results of this paper do not provide clear superiority of FNN versus HAR-type models. They provide better in-the-sample results, but not out-of-sample performances, although forecasting performances are as good as those of HAR-type models are. FNN models are selected based on the best model only according to in-the-sample performances, which is the limitation of this study. Namely, other approach, which consists of finding the most appropriate NN architecture, based on train and validation set within in-the-sample diagnostics should be used in further research. If more parsimonious model is to be selected, that simultaneously has good in-the-sample properties and good out-of-sample forecasting performances, that captures nonlinearity and does not lead to overfitting, then the best-fit model for RV is the FNN-HAR-J model. This way the investors can benefit from applying the proposed FNN model to accurately model and predict the RV for their potential investments. However, the improvement in performances of NNs in forecasting RV and time series in general should be further investigated. Specifically, implementation of different algorithms and/or other NN architectures in this context opens space for future research. Directions for further studies include modelling and forecasting RV using recurrent NNs to investigate their forecasting performances compared to the FNNs and HAR-type models. Namely, other types of recurrent NNs are developed, whose results are insufficiently analysed in economic time series modelling, such as deep NNs, Long Short Term Memory and Convolutional NNs with a large number of hidden layers, which are proven to be more efficient than classical NNs. The proposed methodology should be additionally tested in this field by conducting the research for different markets to verify the robustness of the obtained results.

References

1. Adhikari, R., Agrawal, R. K. (2013), "An Introductory Study on Time Series Modeling and Forecasting", LAP Lambert Academic Publishing, Germany.
2. Aït-Sahalia, Y., Mykland, P. A., Zhang, L. (2005), "How often to sample a continuous-time process in the presence of market microstructure noise", Review of Financial Studies, Vol. 18, No. 2, pp. 351-416.
3. Aljinović, Z., Poklepović, T. (2013), "Neural networks and vector autoregressive model in forecasting yield curve", The 6th International Conference on Information Technology (ICIT), Al-Zaytoonah University of Jordan, Amman, Vol. 1, pp. 1-8.
4. Al-Maqaleh, B. M., Al-Mansoub, A. A., Al-Badani, F. N. (2016), "Forecasting using Artificial Neural Network and Statistics Models", International Journal of Education and Management Engineering, Vol. 6, No. 3, pp. 20-32.
5. Aminian, F., Suarez, E. D., Aminian, M., Walz, D. T. (2006), "Forecasting Economic Data with Neural Networks", Computational Economics, Vol. 28, No. 1, pp. 71-88.
6. Andersen, T. G., Bollerslev, T., Diebold, F. X. (2007a), "Roughing it up: including jump

components in the measurement, modelling and forecasting of return volatility", The Review of Economics and Statistics, Vol. 89, No. 4, pp. 701-720.

7. Andersen, T. G., Bollerslev, T., Diebold, F. X., Labys, P. (2003), "Modeling and Forecasting Realized Volatility", Econometrica, Vol. 71, No. 2, pp. 579-625.

8. Andersen, T., Bollerslev, T., Dobrev, D. (2007b), "No-arbitrage semi-martingale restrictions for continuous time volatility models subject to leverage effects, jumps and i.i.d. noise: theory and testable distributional implications", Journal of Econometrics, Vol. 138, No. 1, pp.125-180.

9. Angus, J. E. (1991), "Criteria For Choosing The Best Neural Network: Part I", Report No. 91-16.

10. Arnerić, J., Poklepović, T. (2016), "Nonlinear Extension of Asymmetric GARCH Model within Neural Network Framework", in Zizka, J., Nagamalai, D. (eds.). Computer Science & Information Technology, AIRCC Publishing Corporation, Chennai, India, pp. 101-111.

11. Arnerić, J., Poklepović, T., Aljinović, Z. (2014), "GARCH based artificial neural networks in forecasting conditional variance of stock returns", Croatian Operational Research Review, Vol. 5, No. 2, pp. 329-343.

12. Bandi, F. M., Russell, J. R. (2008), "Microstructure Noise, Realized Variance, and Optimal Sampling", The Review of Economic Studies, Vol. 75, No. 1, pp. 339-369.

13. Bandi, F. M., Russell, J. R. (2011), "Market Microstructure Noise, Integrated Variance Estimators, and the Accuracy of Asymptotic Approximations", Journal of Econometrics, Vol. 160, No. 1, pp. 145-159.

14. Barndorff-Nielsen, O. E., Shephard, N. (2002a), "Econometric Analysis of Realised Volatility and its use in Estimating Stochastic Volatility Models", Journal of the Royal Statistical Society, Vol. 64, No. 2, pp. 253-280.

15. Barndorff-Nielsen, O. E., Shephard, N. (2002b), "Estimating Quadratic Variation Using Realized Variance", Journal of Applied Econometrics, Vol. 17, No. 5, pp. 457-478.

16. Barndorff-Nielsen, O. E., Shephard, N. (2004), "Econometric analysis of realised covariation: high frequency covariance, regression and correlation in financial economics", Econometrica, Vol. 72, No. 3, pp. 885-925.

17. Baruník, J., Křehlík, T. (2016), "Combining high frequency data with non-linear models for forecasting energy market volatility", Expert Systems With Applications, Vol. 55, pp. 222-242.

18. Bektipratiwi, A., Irawan, M. I. (2011), "A RBF-EGARCH neural network model for time series forecasting", Proceedings of The IceMATH 2011, pp. 1-8.

19. Bildirici, M., Ersin, Ö. Ö. (2009), "Improving forecasts of GARCH family models with the artificial neural networks: An application to the daily returns in Istanbul Stock Exchange", Expert Systems with Applications, Vol. 36, No. 4, pp. 7355-7362.

20. Bildirici, M., Ersin, Ö. Ö. (2012), "Nonlinear volatility models in economics: smooth transition and neural network augmented GARCH, APGARCH, FIGARCH and FIAPGARCH models", MPRA Paper No. 40330.

21. Bildirici, M., Ersin, Ö. Ö. (2014), "Modelling Markov Switching ARMA-GARCH Neural Networks Models and an Application to Forecasting Stock Returns", Hindawi Publishing Corporation: The Scientific World Journal, Article ID 497941, pp. 1-21.

22. Binner, J. M., Elger, C. T., Nilsson, B., Tepper, J. A. (2006), "Predictable non-linearities in U.S. inflation", Economics Letters, Vol. 93, No. 3, pp. 323-328.

23. Bollerslev, T. (1986), "Generalized autoregressive Conditional Heteroskedasticity", Journal of Econometrics, Vol. 31, No. 3, pp. 307-327.

24. Buse, A. (1982), "The Likelihood Ratio, Wald, and Lagrange Multiplier Tests: An Expository Note", The American Statistician, Vol. 36, No. 3, pp. 153-157.

25. Chaudhuri, T. D, Ghosh, I. (2016), „Artificial Neural Network and Time Series Modeling Based Approach to Forecasting the Exchange Rate in a Multivariate Framework", Journal of Insurance and Financial Management, Vol. 1, No. 5, pp. 92-123.

26. Choudhary, A., Haider, A. (2008), "Neural network models for inflation forecasting: an appraisal", Discussion Papers in Economics, University of Surrey, UK, DP 08/08.

27. Clements, A. E., Liao, Y. (2013), "Modeling and forecasting realized volatility: getting the most out of the jump component", NCER Working Paper Series, Working Paper #93,

August.

28. Corsi, F. (2003), "A Simple Long Memory Model of Realized Volatility", manuscript, University of Southern Switzerland.
29. Corsi, F. (2009), "A simple approximate long memory model of realized volatility", Journal of Financial Econometrics, Vol. 7, No. 2, pp. 174-196.
30. Corsi, F., Audrino, F., Reno, R. (2012), "HAR Modeling for Realized Volatility Forecasting", in Handbook of Volatility Models and Their Applications, John Wiley & Sons, New Jersey, USA, pp. 363-382.
31. Dedi, R., Yoga, A. N., Rahmawati, K. D. (2011), "Forecasting the Indonesian Government Securities Yield Curve Using Neural Networks and Vector Autoregressive Model", ISI 58th Session, Dublin, August 21-26, 2011.
32. Degiannakis, S., Floros, C. (2015), "Modelling and Forecasting High Frequency Financial Data", Palgrave Macmillan.
33. Diebold, F. X., Mariano, R. S. (1995), "Comparing Predictive Accuracy", Journal of Business and Economic Statistics, Vol. 13, No. 3, pp. 253-263.
34. Doukim, C. A., Dargham, J. A., Chekima, A. (2010), "Finding the number of hidden neurons for an MLP neural network using coarse to fine search technique", in Proceedings of the 10th International Conference on Information Sciences, Signal Processing and Their Applications (ISSPA '10), May, pp. 606-609.
35. Engle, R. F. (1982), "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation", Econometrica, Vol. 50, No. 4, pp. 987-1007.
36. Franses, P. H., van Dijk, D. (2003), "Nonlinear Time Series Models in Empirical Finance", Cambridge University Press.
37. Gonzales, S. (2000), "Neural Networks for Macroeconomic Forecasting: A Complementary Approach to Linear Regression Models", Working Paper 2000-07.
38. Hagiwara, M. (1994), "A simple and effective method for removal of hidden units and weights", Neurocomputing, Vol. 6, No. 2, pp. 207-218.
39. Hansen, P. R., Lunde, A. (2005), "A realized variance for the whole day based on intermittent high-frequency data", Journal of Financial Econometrics, Vol. 3, No. 4, pp. 525-554.
40. Hansen, P. R., Lunde, A. (2006), "Realized variance and market microstructure noise", Journal of Business and Economic Statistics, Vol. 24, No. 2, pp. 127-161.
41. Huang, C., Gong, X., Chen, X., Wen, F. (2013), "Measuring and Forecasting Volatility in Chinese Stock Market Using HAR-CJ-M Model", Research Article, Hindawi Publishing Corporation Abstract and Applied Analysis, Article ID 143194, pp. 1-13.
42. Junior, M. V. W., Pereira, P. L. V. (2011), "Modeling and Forecasting Realized Volatility: Evidence from Brazil", Brazilian Review of Econometrics, Vol. 31, No. 2, pp. 315-337.
43. Jurkovič, J. (2013), "Forecasting Realized Volatility Using Neural Networks", Master thesis, Charles University in Prague Faculty of Social Sciences
44. Kaashoek, J. F., van Dijk, H. K. (2001), "Neural networks as econometric tool", Econometric Institute Report, EI 2001-05, pp. 1-32.
45. Keeni, K., Nakayama, K., Shimodaira, H. (1999), "Estimation of initial weights and hidden units for fast learning of multi-layer neural networks for pattern classification", in Proceedings of the International Joint Conference on Neural Networks (IJCNN '99), Vol. 3, IEEE, July, pp. 1652-1656.
46. Li, J. Y., Chow, T. W. S., Yu, Y. L. (1995), "Estimation theory and optimization algorithm for the number of hidden units in the higher-order feedforward neural network", in Proceedings of the IEEE International Conference on Neural Networks, Vol. 3, December, pp. 1229-1233.
47. Mantri, J. K., Gahan, P., Nayak, B. B. (2010), "Artificial Neural Networks – An Application to Stock Market Volatility", International Journal of Engineering Science and Technology, Vol. 2, No. 5, pp. 1451-1460.
48. Mantri, J. K., Mohanty, D., Nayak, B.B. (2012), "Design Neural Network for Stock Market Volatility: Accuracy Measurement", International Journal on Computer Technology and Applications, Vol. 3, No. 1, pp. 242-250.
49. Medeiros, M. C., Teräsvirta, T., Rech, G. (2006), "Building Neural Network Models for Time

Series: A Statistical Approach", Journal od Forecasting, No. 25, No. 1, pp. 49-75.

50. Moshiri, S., Cameron, N. (2000), "Neural Networks Versus Econometric Models in Forecasting Inflation", Journal of Forecasting, Vol. 19, No. 3, pp. 201-217.

51. Rosa, R., Maciel, L., Gomide, F., Ballini, R. (2014), "Evolving Hybrid Neural Fuzzy Network for Realized Volatility Forecasting with Jumps", IEEE Conference on Computational Intelligence for Financial Engineering & Economics (CIFEr), London, pp. 481-488.

52. Teräsvirta, T., Tjøstheim, D., Granger, C. W. J. (2008), "Modelling nonlinear economic time series", Oxford University Press, New York,.

53. Teräsvirta, T., van Dijk, D., Medeiros, M. C. (2005), "Linear models, smooth transition autoregressions, and neural networks for forecasting macroeconomic time series: A re-examination", International Journal of Forecasting, Vol. 21, No. 4, pp. 755-774.

54. Wang, J., Wang, J., Fang, W., Niu, H. (2016), "Financial Time Series Prediction Using Elman Recurrent Random Neural Networks", Hindawi Publishing Corporation Computational Intelligence and Neuroscience, Article ID 4742515, pp. 1-14.

55. Zapranis, A., Refenes, A.-P. (1999), "Neural model identification, variable selection and model adequacy", Journal of Forecasting, Vol. 18, No. 5, pp. 299-332.

56. Zekić-Sušac, M., Kliček, B. (2002), "A Nonlinear Strategy of Selecting NN Architectures for Stock Return Predictions", Finance, Proceedings from the 50th Anniversary Financial Conference Svishtov, Bulgaria, 11-12 April, Svishtov, Veliko Tarnovo, ABAGAR, pp. 325-355.

57. Zhang, G. P. (2003), "Time series forecasting using hybrid ARIMA and neural network model", Neurocomputing, Vol. 50, pp. 159-175.

58. Zhang, J., Morris, A. J. (1998), "A sequential learning approach for single hidden layer neural networks", Neural Networks, Vol. 11, No. 1, pp. 65-80.

59. Zhang, L., Aït-Sahalia, Y., Mykland, P. A. (2005), "A Tale of Two Time Scales: Determining Integrated Volatility with Noisy High-Frequency Data", Journal of the American Statistical Association, Vol. 100, No. 472, pp. 1394-1411.

About the authors

Josip Arnerić, Ph.D. is an Associate Professor at the University of Zagreb, Faculty of Economics and Business, Department of Statistics. His main research interests are econometric methods and models, financial time series and volatility, VAR models and cointegration, GARCH and MGARCH models, stochastic processes and risk management. The author can be contacted at **jarneric@efzg.hr**.

Tea Poklepović, PhD is a Teaching assistant at the University of Split, Faculty of Economics, Business and Tourism, Department of Quantitative methods. Her main research interests are statistics and econometrics in business, finance and macroeconomics, especially econometric methods and models, time series and neural networks. The author can be contacted at **tpoklepo@efst.hr**.

Juin Wen Teai is a student at the National University of Singapore, Singapore. His main research interests are econometric methods and models in finance, actuary and investment. The author can be contacted at **juinwen.teai@gmail.com**.