AFM423: Report 1 - Review of Literature on Forecasting Realized Volatility

Zhenning Pei[1], Yasien Sayeed[2], and Zixi Tao[3]

[1] *ID: 20625199 Email: z8pei@uwaterloo.ca*
[2] *ID: 20616222 Email: ygsayeed@edu.uwaterloo.ca*
[3] *ID: 20599273 Email: reginald.tao@uwaterloo.ca*

1 Overview

Modelling the conditional volatility of financial time series has always been one of the major topics in finance. The concept of volatility is widely used in derivatives pricing, portfolio optimization, and risk management tasks. The most well-known model is GARCH proposed by Bollerslev (1986).

However, we note that the squared returns exhibit high excess kurtosis. Return distributions also show a very slow convergence to the normal distribution as time scales increase. Most importantly, the returns, and the realized volatility, display a clear property of long-memory, as the autocorrelations are low but still statistically significant and very slowly decreasing. The standard GARCH model cannot capture all of those features.

The search for an adequate framework has led us to the analysis to the high-frequency intraday trading data. Thanks to the fast-developing technology, those data are readily available on a lot of financial data platforms like Reuters, Bloomberg, and Wind. Merton (1980) first noted that volatility can be estimated accurately as the sum of squared realizations, with high frequency data. Anderson showed in separate papers that even with the distinct features of slowly decreasing autocorrelation functions, using 5-minute returns can accurately measure the *ex post* volatility in the foreign exchange market (Andersen *et al.* 2001; Andersen and Bollerslev 1998).

Long-memory processes, or processes that try to mimic the long-memory property of the realized volatility series, are introduced to tackle the problems presented above. Some of the works include Corsi (2009) on the HAR model, which performs well in terms of short-term prediction and is often used as a baseline to examine and compare more complex models such as LASSO based feature selection models and neural network models.

Hence in this survey, we first introduce the technical setup of realized volatility. We then introduce the HAR model as it's the baseline model for the papers we picked, followed by a literature review of some realized volatility models that may or may not account for the long-memory property. Finally, we formulate a plan for future work on the topic.

2 Realized Volatility

Suppose log-price X_t follows the standard stochastic process

$$dX_t = \mu(t)dt + \sigma(t)dB_t$$

where B_t is the standard Brownian Motion, and $\mu(t)$, $\sigma(t)$ are mean and volatility functions (that are possibly time dependent) independent of B_t. The integrated volatility over one day is then defined as

$$IV_t^{(d)} = \sqrt{\int_{t-1d}^{t} \sigma^2(u)du}$$

Under regularity conditions (Shephard, Barndorff-Nielsen, and Shephard 2002) the unobservable IV is then approximated by realized volatility (RV) as

$$RV_t^{(d)} = \sqrt{\sum_{i=2}^{N} r_{t,i}^2} \tag{1}$$

where N is the number of intraday observations and r the log-return. RV over longer time horizons are simply the average of the daily RV over given time periods.

$$RV_t^{(w)} = \frac{1}{5}\sum_{i=0}^{4} RV_{t-id}^{(d)},$$
$$RV_t^{(m)} = \frac{1}{22}\sum_{i=0}^{22} RV_{t-id}^{(d)}$$

3 HAR Model

Under conditions and definitions above, we state the HAR model proposed in Corsi (2009).

$$RV_{t+1d}^{(d)} = \beta_0 + \beta^{(d)} RV_t^{(d)} + \beta^{(w)} RV_t^{(w)} + \beta^{(m)} RV_t^{(m)} + \omega_{t+1d} \tag{2}$$

where ω_{t+1d} is the error term.

Note the HAR(1,5,22) model can be also specified as a constrained AR(22) model as below

$$RV_{t+1d}^{(d)} = \theta_0 + \sum_{i=1}^{22} \theta_i RV_{t-(i-1)d}^{(d)} + \omega_{t+1d} \tag{3}$$
$$\theta_j = \begin{cases} \beta^{(d)} + \frac{1}{5}\beta^{(w)} + \frac{1}{22}\beta^{(m)} & ,j = 1 \\ \frac{1}{5}\beta^{(w)} + \frac{1}{22}\beta^{(m)} & ,j = 2, \cdots, 5 \\ \frac{1}{22}\beta^{(m)} & ,j = 6, \cdots, 22 \end{cases}$$

4 Flexible HAR Using LASSO

4.1 Summary of the Study

We draw our attention first to *Flexible HAR Model for Realized Volatility* by Audrino, Huang, and Okhrin (2019). The assumption of fixed number of components included in the model is relaxed and therefore a penalized regression method is used to select the active terms in an additive fashion.

The study then states a few points in statistical inference theory, and further specify the differences between the flexible HAR model and its original version. It also proposes a few alternative models such as AR-AIC and AR-LASSO to justify the performance of the flexible model in mind.

Empirical data from 10 individual equities are then used to illustrate the feature selection on the lags. From the illustration we can see clearly that the HAR(22) model largely holds. However, there are small peaks in probability of being selected for lag of 2, 4, 6, and 8 weeks and somewhat differ from the original assumptions.

The study then uses stock indices data to compare the estimation and forecasting accuracy. It finds minor improvement in both in-sample and out-of-sample prediction accuracy by using Flexible HAR over the baseline model, and the proposed model performs the best among all alternative models in terms of RMSE. Later, more formal tests such as predictive ability test (Hansen 2005) and model confidence set (Hansen, Nason, and Lunde 2010) are also used to show that none of the alternative models significantly outperforms the proposed model.

One interesting inclusion of the paper is to dynamically allocate weights on the HAR and the Flexible HAR models in an effort to produce a *super learner*. The flexible model usually gets a larger weight, and more importantly, the weights assigned to the flexible model is higher when there are market downturns. It is intuitive that when the market are volatile, the data-drive model is much more preferred and the assumptions in classical models may not hold during such time periods. The study henceforth arrives at the conclusion that the LASSO based, more flexible counterpart of the HAR model have to be favored, especially during the hard times when we really need the model to work.

4.2 Methodology: LASSO

The Least Absolute Shrinkage and Selection Operator (LASSO) is first introduced in Tibshirani (1996) and uses l_1 norm penalty to restrict the number of features selected in the model. In our case, the estimator is given as

$$\hat{\beta}_\lambda = \arg\min_{\beta} \left\{ \sum_{t=p}^{T} \left(RV_{t+1d}^{(d)} - \beta_0 - \sum_{i=1}^{p} \beta_i \sum_{j=1}^{i} RV_{t-(j-1)d}^{(d)} \right)^2 + \lambda \sum_{i=1}^{p} \lambda_i |\beta_i| \right\} \tag{4}$$

where λ_i is the weight for each coefficient. Ordinary LASSO is a special case of the above model where $\lambda_i = 1, \forall i = 1, \cdots, p$.

LASSO method, by choosing the tuning parameter λ, forces some of the coefficients to become exactly zero and thus reduce the dimensionality and the compexity of the model. It has been widely used in economic and financial applications (Fan, Lv, and Qi 2011). The paper, adopting the model above, reserves the structural property of the HAR models.

4.3 Data

There are 2 parts of the paper that use empirical data. First data from 10 individual stocks on NYSE are used to show feature selection procedure. Later 19 indices data are used, with a focus on estimation and prediction accuracy.

The individual stock data are cleaned through the following steps:

1. Only keep transactions between market open and close, with non-zero prices.
2. Exclude correction trades.
3. If multiple transactions have the same timestamp, replace all with one data point, with the median price.

Realized volatility is then calculated based on returns every 5 minutes.

Data from indices, mostly within the equity space, are used and cleaned in a similar way.

4.4 Significant Findings

One of the many findings of the study is that it confirms the validity of the HAR models. As stated earlier in our review, realized volatility has a long-memory, and the HAR models, by including relatively long lag terms, try to mimic the structure of the true process. It is shown in the paper that most of the lags beyond 22 are not significant. It must be note that, however, there is no strong evidence that the HAR model recovers the true structure.

In terms of prediction accuracy, however, it is really hard to outperform the HAR model. The result is supported by a variety of tests including RMSE, Superior Predictive Ability test, and Model Confidence Set ranking. It partially explains why the HAR model is often used as a baseline for comparison in the literature.

Another very interesting point presented is the *super learner*. Choosing a dynamic, data-driven weight between the HAR and Flexible HAR model shows us that the more flexible counterpart proposed thrives in volatile markets. It could be a great incentive for the academia as well as market participants to embrace the machine learning based models.

4.5 Possible improvements

We observe that the individual stocks selected are blue-chip companies from different industries. We note that it might be difficult to extract intraday data from some less covered companies due to illquidity, but the focus on those huge corporations may introduce bias and affect the generality of the study on financial time series. Similar problems can be also found in selection of indices with the heavy focus in the equity space.

One other possible improvement is to put more emphasis on the selection of λ, the tuning parameter. One may find better prediction performance with different λ's.

5 Linear HAR, flexible HAR, NN HAR and Bagging

5.1 Summary of the Study

The second paper is *Forecasting Realized Volatility with Linear and Nonlinear Univariate Models* by McAleer and Medeiros (2010). This study aims to compare and contrast the linear HAR model, two nonlinear HAR models inspired by neural networks estimated by least squares (LS) and Bayesian regularization (BR). It is worth noting that the models used in this study are actually quite similar to the flexible HAR described above, but it differ in that it utilizes bagging to improve its models, which is quite rare among the papers that we surveyed. The application of bagging would be our major point of interest in reviewing this paper.

Overall, the paper was able to achieve comparable performance with flexible linear HAR and the nonlinear HAR model estimated with BR, but the performance of the nonlinear HAR with LS is not as good. The author conjecture that without limiting the number of parameters using BR, the effects of bagging is offset by the over-parameterization in the nonlinear HAR model with LS.

5.2 Methodology: Linear HAR

The linear HAR method is based on the work of Corsi (2009) as described above. The paper attempts to extend Corsi's work by considering more lags than the standard $I = (1, 5, 22)$ as well as other factors such as weekdays, macroeconomic announcements and past cumulative returns based on work of Hillebrand and Medeiros (2010). It then uses bagging for feature selection, which is described in a separate section. This is, in a way, similar to the flexible HAR described above as it aims to reduce the complexity of the model.

5.3 Methodology: Neural Networks based HAR model

The neural network approach in this paper can be interpreted as incorporating a smooth transition regression where the transition variable is an unknown linear combination of explanatory variables McAleer and Medeiros (2007). Although not explicitly stated, we suspect that the neural network aspect of this model comes from the usage of the logistic function as an activation function in its curve smoothing terms. The equation is as follows:

$$y_t = \beta_0'\omega_{t-1} + \sum_{i=1}^{m} \beta_i f(\gamma_i'\omega_{t-1}) + \epsilon_t \tag{5}$$

The set up is essentially the HAR model with the addition of the smoothing term (the sum term) where f is the logistic function. Note that the term of $\beta_0'\omega_{t-1}$ is also the most important term in the benchmark

HAR model; this is significant because through preservation of this term, this model preserves the most important quality of long memory as in the HAR model.

5.4 The Bagging Process

The bagging process aims to further reduce overfitting by drawing blocks of samples of size m with replacement, further fine-tuning the model to each sample drawn and then averaging over the results of all the samples chosen. The size m is chosen to capture possible dependence in the error term of the realized volatility series, such as conditional variance ('volatility of volatility') according to McAleer and Medeiros (2010). In the study, the bootstrap and sampling process is the same for models tested with bagging, meaning that they only differ in their feature selection process.

Bagging the linear HAR model uses a diagonal selection matrix S^* where the jth diagonal entry is set to 1 (indicating selected) if the t-value of the corresponding sample term is greater than a pre-specified threshold. In this study, the threshold is set to 1.96 to correspond to a two-sided test at the 96th confidence level.

Bagging the non-linear HAR model uses the same process of feature selection to remove statistically insignificant regressors, but it adds an additional step of randomly choosing the M logistic transitions from a uniform distribution on $[0, 20]$; it then uses the same process to compute and average over all forecasts to obtain the final prediction.

5.5 Data

This study uses high-frequency tick-by-tick data on S&P 500 futures from Jan 1996 to Mar 2007 (2796 observations) and FTSE 100 futures from Jan 1996 to Dec 2007 (3001 observations); realized volatilities are calculated daily using the realized kernel estimator with the modified Turkey-Hanning kernel of BHLS as by Stock, Chan, and Watson (1999).

Out of the data collected, 1000 is used for out-of-sample assessment purposes.

5.6 Significant Findings

From the empirical results, both the flexible HAR model and the NN-HAR model with BR outperformed the benchmark HAR model overall. The NN-HAR model estimated with nonlinear LS was the worst model among all the alternatives considered, with both poor in-sample and out-of-sample performance. The paper conjectures that this might be due to over-parameterization of the NN-HAR model without BR, which is similar to ridge regression in a sense and thus reduces number of parameters overall according to Burden and Winkler (2009).

It is also interesting that the paper's neural network HAR model approach still preserves the most important factor in forecasting realized volatility: the long memory property. Ultimately, the paper's neural network approach is still fundamentally a HAR model. In the next section, we will make the point that the long memory property that HAR captures is crucial for the performance of any model aiming to accurately forecast realized volatility.

5.7 Possible improvements

It is noted that the sample period for this paper ends just before the financial crisis; it would be insightful to see how the same methodologies tested in this paper behave during historically volatile times.

It would also be interesting to highlight the effect of bagging on the alternative models by including a set of results with alternative models applied without bagging. We see that the paper concludes that two of the alternative models are better compared to its benchmark, but we do not know how much of the improvement is attributed to the effects of bagging. Nevertheless, this inclusion provides an important inspiration for future research attempts.

6 Principle Components Combining, Neural Networks and GARCH against HAR

6.1 Summary of the Study

Our third paper is *Forecasting Realized Volatility: HAR Against Principal Components Combining, Neural Networks and GARCH* by Vortelinos (2015). This study aims to determine whether the nonlinear models stated in the title would improve realized volatility forecast accuracy when compared with the traditional HAR method. The paper concludes that because the new models are worse at capturing the important long memory property of realized volatility, they appear to underperform the traditional HAR method.

This study is quite ambitious in scope: it examines data across seven US financial markets from 2002 to 2011: spot equity, spot FX rates, ETFs, equity index futures, US Treasury bonds futures, energy future and commodity options (Vortelinos (2015)). Because of the large scope, we observe slight differences in performance of different models across different markets, but the HAR stands as the best-performing model overall.

The study uses 4 methods to forecast 1-day realized volatility: the principle components combining model, a single-layer feed forward neural network, GARCH, and HAR as benchmark. These models are evaluated using many methods: R^2 of the Mincer-Zarnowitz regression by Mincer and Zarnowitz (1969) which tests for unbiasedness, directional criterion to check whether prediction and the true value have the same sign, average predictive accuracy in the Diebold Mariano test by Diebold and Mariano (1995), Root Mean Squared Error and Mean Absolute Error. While different models' performance differ slightly under different evaluation methods, HAR had the best performance over most criteria, with PCC coming close at second place. It is also worth mentioning that although GARCH is considered industry standard, the empirical results from the paper suggests that its performance is simply not as good as HAR or PCC models; so we will not be focusing our study on GARCH.

Overall, it is interesting to see this paper comparing many different models for realized volatility forecasting. It may be surprising at first that the more flexible models did not perform better than the simpler HAR, but this result precisely highlights the significance of the long memory nature of realized volatility.

6.2 Methodology: PCC

The PCC method is based on the work of Stock, Chan, and Watson (1999) and Stock and Watson (2004); it fits a regression model on the first m principle components of individual forecasts based on this formula

$$y^h_{s+h} = \phi_1 \hat{F}^h_{1,s+h|s} + \dots + \phi_m \hat{F}^h_{m,s+h|s} + v^h_{s+h} \tag{6}$$

where $s = R, \dots, t-h$ with R being the last observation of the in-sample period, h is the forecasting horizon of 1-step; the number of principle components m is chosen empirically based on the IC_{p^3} information criterion developed by Bai and Ng (2001).

6.3 Methodology: Neural Networks

The neural network used by this paper is a single-layer feed forward neural network; it takes all in-sample data series as initial inputs and is refined using AIC. Experimental data for this neural network is split into 3 sets: training (65%), testing (15%) and validation (20%). The paper specified a learning rate of 0.5 without obvious reasons. The neural network can be interpreted as a class of flexible nonlinear function of the form

$$f(X_{t-1}, ..., X_{t-p}) = \beta_0 + \sum_{j=1}^{q} G(\gamma_{0,j} + Y_t^T \gamma_j)\beta_j \tag{7}$$

where $Y_t = (X_{t-1}, ..., X_{t-p})^T$, $X_{t-p} = RV_{t+1|n}$, $\gamma_j = (\gamma_{1,j} + ... + \gamma_{p,j})^T$ and β_i is scalar coefficient for corresponding i. G is the activation function chosen to be the hyperbolic tangent transfer function from Zhang and Qi (2005).

It is worth noting that this model only differs one term when compared with the neural network HAR model in the previous paper. This model is missing the linear term directly computed from different lags. As we see in further analysis of findings, this is an extremely important point.

6.4 Data

The experimental data from different US financial markets are obtained from the following sources:

1. spot equity market: Dow Jones Industrial Average: INDU
2. spot FX rates: EUR/USD exchange rate
3. ETF: PowerShare QQQ: QQQ
4. Equity Index Futures: E-Mini Dow futures continuous contract: YM
5. Energy futures: Crude oil miNY futures continuous contract: QM
6. Commodities options: CBOE gold index options: GOX

The prices were sampled every minute and realized volatility series are estimated daily. Data is cleaned to exclude weekends, holidays and days with too many missing entries. Out of the data collected, 30% is used for assessment purposes.

6.5 Significant Findings

The most important takeaway from this paper is that long memory is extremely important for forecasting realized volatility and failure to capture it cannot be made up by simply increasing model complexity. The HAR model is the best model in four of the seven markets examined, and its RMSE and MAE are considerably lower than that of the neural network and GARCH.

We see the paper uses a feed forward neural network which is more sophisticated than the HAR model, but it came in third under RMSE and MAE and last under R^2 of the Mincer-Zarnowitz regression which may suggest overfitting. As stated before, this model only differs 1 term from the much better performing neural network HAR model in the previous paper, which actually beat the same benchmark of standard HAR. This again highlights the important of including the long memory property of realized volatility in model selection and design.

It is also difficult to interpret the neural network's fitted model with properties of realized volatility since the feature selection in the hidden layer is non-transparent; this is also a main problem of the PCC model despite its reasonable performance.

6.6 Possible improvements

From this paper, it is clear that we can improve on their neural network approach by incorporating LSTM to adjust for the long memory property of realized volatility. By having an recurrent neural network with LSTM instead of a feed forward one, it is possible that we can significantly improve the model by dynamically choosing the best lag period for forecasting realized volatility.

We also note that the parameter tuning for their neural network also has room for improvement; there is no obvious explanation for their choice of hyperbolic tangent transfer function as their activation function, and the choice of 0.5 as learning rate seems quite arbitrary.

Finally, we could also incorporate cross validation into their assessment strategy so we would not only have more data to train on but also further reduce the likelihood of overfitting.

7 Recurrent Neural Networks

7.1 Summary of the Study

The most recent study we looked at on this topic, *Realized Volatility Forecasting with Neural Networks* (Bucci 2019), builds upon the previous discoveries regarding the HAR model, and tries to improve upon it using non-parametric models, focusing on neural networks. Unlike the previous experiments, this one tried different recurrent neural networks (RNNs) in order to try to estimate the realized volatility. It found that NARX and LSTM networks were very successful, outperforming the statistical models and being more effective in predicting realized volatility.

This study looks at the (log-transformed) monthly realized variance for the S&P 500 index between February 1950 and December 2017, using various macroeconomic and financial ratios as the features. From the initial feature set, it reduced the features to a smaller subset using LASSO regression, and determined that macroeconomic variables such as the inflation rate were generally not relevant. The experiment tried using various different neural network architectures, which were a feed-forward neural network (FNN), Elman neural network (ENN), Jordan neural network (JNN), long-short term memory network (LSTM), and a NARX neural network (NARX). All of these were set to have one input layer, along with one hidden layer and then one output layer. Each of these neural networks were compared to baseline ARFIMA and LSTAR models. Performance was measured by MSE, finding that the best results were found from the NARX and LSTM models, although all the other neural networks did better than the baseline models.

7.2 Methodology: RNN

Recurrent neural networks are a specific type of neural network where outputs from a previous time are fed back into the inputs for the next data point. In this way, it is able to 'look back' at the past, where the value in future outputs is dependent on previous results. This follows to replicate the long-memory property found in the HAR model, by feeding back the past outputs into the network instead.

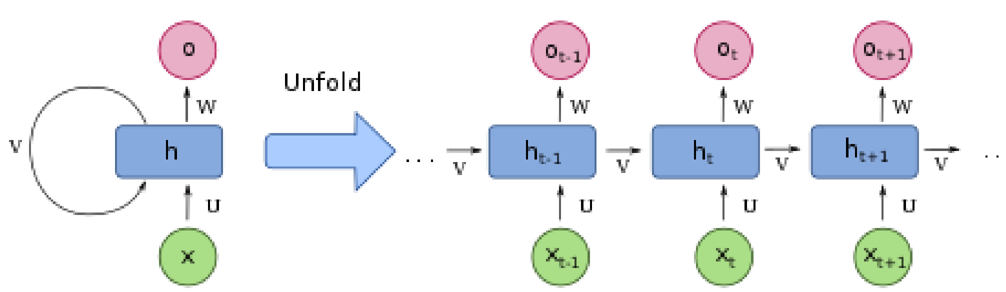

Figure 1. A simplified RNN (Commons 2017)

The specific structure of each different type of RNN varies depending upon the architecture of the RNN, but in all cases, it involves some form of reusing the previous hidden states or outputs and consuming them as another input in a future state, similar to the diagram above, which shows a simplified RNN which takes 1 input and the data from the previous state. This ability to reuse previous data makes them well-suited to model time series data or other sequential inputs.

Although the models are similar, there are some differences between each of them. ENNs are a type of neural network in which the value of each hidden unit at time t is passed into the input for time $t + 1$. Similarly, JNNs work in the exact same way, except that they pass the value of the outputs back into the input, instead of the hidden unit. NARX works similarly to JNNs, although there are 2 differences, where rather than taking the predicted output from the data, it uses the actually observed output at time t as an input for time $t + 1$ instead, and it also allows for more time lags to occur, where data at some time t will take the values for times $t - 1$, $t - 2$, $t - 3\ldots$ as well. The LSTM model takes a different approach from these, augmenting each node in it with a memory block and allowing it to determine how to transform the data it receives based on past results, and has the ability to selectively forget what data it processes to make use of what is most important and remove noise.

One common problem that RNNs have is the vanishing gradient problem, which has to do with the backpropogation algorithm for updating the weights for the RNNs after each training input. After each training point is computed, the RNN must calculate the error and update the weights of each point so that they reduce the error of the neural network. However, since most activation functions have a derivative that has an absolute value less than or equal to 1, the results tend to vanish towards 0, leading to very small or insignificant changes to the weights, meaning that it requires additional time for training, otherwise the weights will not be set properly (Pascanu, Mikolov, and Bengio 2012). This prevents the model from learning dependencies over a longer period of time, as the model cannot tell if there is a dependency between the current time and a previous state or if it simply has the wrong parameters. The LSTM and NARX models both take steps in an attempt to prevent the gradient vanishing problem, although the other neural network architectures used in this experiment do not.

7.3 Data

This experiment looked at several different financial and macroeconomic ratios initially, and only selected a smaller subset of them for the final model, which was done by using LASSO regression to find only the features that were most relevant. It found that these tended to be financial ratios, whereas the economic metrics were far less relevant for making predictions most of the time. The response variable is the monthly realized volatility for the S&P 500 for every month between February 1950 and December 2017, which has been log-transformed. The features considered are displayed below, along with those which were significant enough to be chosen in the end.

Feature	Definition	Included in model
Dividend Yield Ratio	Dividends over the past year divided by current price for stocks in the S&P 500	True
Earning/Price Ratio	Earnings over the past year divided by current price for stocks in the S&P 500	False
Market Excess Return	Return of the S&P 500 minus the 1-month T-bill rate	True
Value Factor	Average returns on Value stocks minus Average returns on growth stocks	False
Size Premium Factor	Average returns on small-cap stocks minus average returns on large-cap stocks	False
Short Term Reversal Factor	Average returns on stocks that did poorly in the previous period minus average returns on stocks that did well	True
T-Bill rate	3-month T-bill rate	False
Term Spread	Difference between long term bond yield and T-bill rate	False
Default Spread	Difference in bond yields between Baa and Aaa bonds	True
Monthly Inflation	US inflation rate	False
Monthly Industrial Production Growth rate	US industrial production growth	False

Table 1. Features used in RNN

7.4 Significant Findings

From this experiment, there are 3 key findings.

The first is that RNN models that attempt to handle the vanishing gradient problem (LSTM and NARX) outperformed the ones that didn't handle this issue (JNN and ENN). These models are better able to handle the long-term dependence that is not present in the others, while the others cannot identify it. As a result, any further improvements to this problem with RNNs should involve a model that is able to avoid this problem and not ignore it.

The second discovery is that in general, the RNN models used here have shown an improvement in accuracy compared to the previous baseline models, especially with LSTM and NARX. All of the models used outperformed the ARIMA and LSTAR models used as the baseline. The experiment found that LSTM and NARX also outperformed the other models in times where the volatility was very high, such as the 2008 financial crisis.

Finally, the last key discovery is that most macroeconomic features considered here are not very important for forecasting realized volatility, and it was more dependent on the financial ratios, as well as volatility for previous periods.

7.5 Possible improvements

This experiment only attempts to predict monthly volatility, however, it does not examine if RNNs can be used for smaller predictions, such as ones on a weekly or daily basis. Attempting to fit a model to see how well it can predict on that time frame would potentially be more useful for investing compared to only predicting the monthly volatility.

An additional improvement would be to attempt to increase the number of hidden layers and make deeper neural networks, as the author simply decided to use 1 hidden layer in each of the neural networks, and see if this change will make the models better at predicting results.

8 Connection between Papers

All of the papers examined indicate the importance of the long-memory property of realized volatility. Starting with the first HAR(22) model, the rest of the papers went on to show similar findings, that predicting the realized volatility was relatively accurate when looking at the results for about 22 trading days (or 1 month). Both the initial autoregressive models and the later neural networks did well when they made use of previous predictions, and future improvements would want to attempt to build on these rather than move away, as the existing data for alternative models did not show signs of improvement.

Most of the models did not use a large number of features, usually only basing the predictions on the returns of the assets, and not looking at other indicators. Models that did look at additional features found that most of them were not very significant to predicting the realized volatility, eventually excluding most of them.

While the studies were largely in agreement on most topics, most of them focused solely on equities or equity futures, and there was not as much focus on predicting the realized volatility for other asset classes, so there is less indication if these same types of models will hold there, or if the effectiveness will decrease among different asset classes.

9 Future Plans

We have decided to restrict our analysis to the realized volatility of indices (equity, currency, bond, gold commodities); however, we may not be able to obtain high-frequency for indices in every field, so we will be looking at stocks instead as a backup if intra-day trading data is not available.

We will mainly be using intra-day index/stock prices (every 5 min or so) and macroeconomic market conditions as input; depending on data availability, if we decide to restrict our analysis to stocks only, we could add more stock-specific features like dividend yield.

We plan to integrate some of the improvements on the models proposed by reviewed literature by comparing and contrasting the following models:

1. Recurrent Neural Network with Long Short-Term Memory trained with high-frequency intra-day returns (instead of daily returns which the last paper uses)
2. Elastic net so we could examine a compromised effect of L1 and L2 penalties on overall fit and model complexity (as an attempt to improve upon the flexible HAR proposed in the first paper)
3. HAR(22) with bagging to see the effect of cross validation on forecasting realized volatility (in response to the interesting question raised in the second paper)
4. HAR(22) or GARCH, which would be used as our benchmark since both are widely accepted in the industry

References

Andersen, T., and T. Bollerslev. 1998. "Answering the Skeptics: Yes, Standard Volatility Models Do Provide Accurate Forecasts." *International Economic Review* 39 (April). doi:10.2307/2527343.

Andersen, T., T. Bollerslev, F. Diebold, and H. Ebens. 2001. "The Distribution of Stock Return Volatility." *Journal of Financial Economics* 61 (July): 43–76. doi:10.1016/S0304-405X(01)00055-1.

Audrino, F., C. Huang, and O. Okhrin. 2019. "Flexible HAR Model for Realized Volatility." *Studies in Nonlinear Dynamics Econometrics* 23 (August). doi:10.1515/snde-2017-0080.

Bai, J., and S. Ng. 2001. "Determining The Number Of Factors In Approximate Factor Models." *Econometrica* 70 (January). doi:10.1111/1468-0262.00273.

Bollerslev, T. 1986. "Generalized autoregressive conditional heteroskedasticity." *Journal of Econometrics* 31:307–327. doi:10.1016/0304-4076(86)90063-1. https://www.sciencedirect.com/science/article/pii/0304407686900631.

Bucci, A. 2019. "Realized Volatility Forecasting with Neural Networks." *Munich Personal RePEc Archive* (August). https://mpra.ub.uni-muenchen.de/95443/1/MPRA_paper_95443.pdf.

Burden, F., and D. Winkler. 2009. "Bayesian Regularization of Neural Networks." *Methods in molecular biology (Clifton, N.J.)* 458 (January): 23–42. doi:10.1007/978-1-60327-101-1_3.

Commons, W. 2017. "Recurrent neural network unfold." https://upload.wikimedia.org/wikipedia/commons/b/b5/Recurrent_neural_network_unfold.svg.

Corsi, F. 2009. "A Simple Long Memory Model of Realized Volatility." *Journal of Financial Econometrics* 7 (February): 174–196. doi:10.1093/jjfinec/nbp001.

Diebold, F., and R. Mariano. 1995. "Comparing predictive accuracy[J]." *J. Bus. Econ. Statist.* 13 (January): 134–144.

Fan, J., J. Lv, and L. Qi. 2011. "Sparse High-Dimensional Models in Economics." *Annual review of economics* 3 (September): 291–317. doi:10.1146/annurev-economics-061109-080451.

Hansen, P. 2005. "A Test for Superior Predictive Ability." *Journal of Business Economic Statistics* 23 (February): 365–380. doi:10.2139/ssrn.264569.

Hansen, P., J. Nason, and A. Lunde. 2010. "The Model Confidence Set." *Econometrica* 79 (March): 453–497. doi:10.2139/ssrn.522382.

Hillebrand, E., and M. Medeiros. 2010. "The Benefits of Bagging for Forecast Models of Realized Volatility." *Econometric Reviews* 29 (August): 571–593. doi:10.1080/07474938.2010.481554.

McAleer, M., and M. Medeiros. 2007. "A Multiple Regime Smooth Transition Heterogeneous Autoregressive Model for Long Memory and Asymmetries." *Journal of Econometrics* 147 (May): 104–119. doi:10.1016/j.jeconom.2008.09.032.

———. 2010. "Forecasting Realized Volatility with Linear and Nonlinear Univariate Models." *University of Canterbury, Department of Economics and Finance, Working Papers in Economics* 25 (January). doi:10.1111/j.1467-6419.2010.00640.x.

Merton, R. C. 1980. "On Estimating the Expected Return on the Market: An Exploratory Investigation." *Journal of Financial Economics* 8:323–361. doi:10.3386/w0444.

Mincer, J., and V. Zarnowitz. 1969. "The evaluation of economic forecasts." *Economic Forecasts and Expectations: Analysis of Forecasting Behavior and Performance* (January): 1–46.

Pascanu, R., T. Mikolov, and Y. Bengio. 2012. "Understanding the exploding gradient problem." *CoRR* abs/1211.5063. arXiv: 1211.5063. http://arxiv.org/abs/1211.5063.

Shephard, N., O. Barndorff-Nielsen, and Shephard. 2002. "Econometric Analysis of Realised Volatility and Its Use in Estimating Stochastic Volatility Models." *Journal of the Royal Statistical Society Series B* 64 (February): 253–280. doi:10.1111/1467-9868.00336.

Stock, J., Y. Chan, and M. Watson. 1999. "A Dynamic Factor Model Framework for Forecast Combination." *Spanish Economic Review* 1 (September): 91–121. doi:10.1007/s101080050005.

Stock, J., and M. Watson. 2004. "Combination Forecasts of Output Growth in a Seven-Country Data Set." *Journal of Forecasting* 23 (September): 405–430. doi:10.1002/for.928.

Tibshirani, R. 1996. "Regression Shrinkage and Selection Via the Lasso." *Journal of the Royal Statistical Society: Series B (Methodological)* 58 (January): 267–288. doi:10.1111/j.2517-6161.1996.tb02080.x.

Vortelinos, D. 2015. "Forecasting Realized Volatility: HAR against Principal Components Combining, Neural Networks and GARCH." *Research in International Business and Finance* (February). doi:10.1016/j.ribaf.2015.01.004.

Zhang, P., and M. Qi. 2005. "Neural network forecasting for seasonal and trend time series." *European Journal of Operational Research* 160 (February): 501–514. doi:10.1016/j.ejor.2003.08.037.