Application of selected methods of statistical analysis and machine learning in predictions of EURUSD, DAX and Ether prices

Mateusz Zaborski
M.Zaborski@mini.pw.edu.pl

Faculty of Mathematics and Information Science
Warsaw University of Technology

05.12.2018
Table of Contest

1. Research goals
2. Review
3. Ether price prediction
   - Ethereum (blockchain) graph
   - Correlations
   - Prediction model - Logistic Regression
   - Prediction model - Neural Network
4. Experiments
   - Moving average crossover
   - Logistic regression
5. Conclusion

Zaborski
Application of selected methods of statistical analysis and machine learning
Research goals

Zaborski

Application of selected methods of statistical analysis and machine learning
Research goals

Prices modeling

- Price modeling and prediction
- Trading strategies
- Application of novel ML and AI models
- Including new data sources in the models
Research goals

Efficient market hypothesis[1]

- Weak-form efficiency
  - Future prices cannot be predicted by analyzing past prices
- Semi-strong-form efficiency
  - Neither fundamental analysis nor technical analysis techniques cannot be used in prediction
- Strong-form efficiency
  - Price reflects also private information
Main prediction techniques

- Based on historical prices
  - AR, MA, ARIMA models
  - Bayesian networks
- Machine Learning
  - Regressions
  - k-NN
  - SVM
- Artificial Intelligence
- Social media (e.g. Twitter, boards)
  - Sentiment analysis

Application of selected methods of statistical analysis and machine learning
Using artificial neural network models in stock market index prediction

Erkam Guresen\textsuperscript{a}, Gulgun Kayakutlu\textsuperscript{a,}\textsuperscript{*}, Tugrul U. Daim\textsuperscript{b}

\textsuperscript{a} Istanbul Technical University, Istanbul, Turkey
\textsuperscript{b} Portland State University, Portland OR, USA

\textbf{ARTICLE INFO}

\textbf{Keywords:}
- Financial time series (FTS) prediction
- Recurrent neural networks (RNN)
- Dynamic artificial neural networks (DAN2)
- Hybrid forecasting models

\textbf{ABSTRACT}

Forecasting stock exchange rates is an important financial problem that is receiving increasing attention. During the last few years, a number of neural network models and hybrid models have been proposed for obtaining accurate prediction results, in an attempt to outperform the traditional linear and nonlinear approaches. This paper evaluates the effectiveness of neural network models which are known to be dynamic and effective in stock-market predictions. The models analysed are multi-layer perceptron (MLP), dynamic artificial neural network (DAN2) and the hybrid neural networks which use generalized autoregressive conditional heteroscedasticity (GARCH) to extract new input variables. The comparison for each model is done in two view points: Mean Square Error (MSE) and Mean Absolute Deviate (MAD) using real exchange daily rate values of NASDAQ Stock Exchange index.

© 2011 Elsevier Ltd. All rights reserved.

\textbf{Figure:} Using artificial neural network models in stock market index prediction [5]
Related researches (goals):
- Comparing methods
- Exchange prediction
- Forecasting improvement
- Comparing ARIMA and ANN
- Crisis and bankruptcy prediction
- Investigate effect of volume on prediction
- Integration of fundamental and technical analysis
- Applying hybrid models
Neural networks performance in exchange rate prediction

Svitlana Galeschuk*
Ternopil National Economic University, 11 Lvivska str., 46000 Ternopil, Ukraine

A R T I C L E  I N F O

Article history:
Received 25 September 2014
Received in revised form
1 February 2015
Accepted 18 March 2015

Keywords:
Neural networks
Multilayer perceptron
Currency exchange-rate changes
Foreign exchange rate prediction

A B S T R A C T

Exploration of ANNs for the economic purposes is described and empirically examined with the foreign exchange market data. For the experiments, panel data of the exchange rates (USD/EUR, JPN/USD, USD/GBP) are examined and optimized to be used for time-series predictions with neural networks. In this stage the input selection, in which the processing steps to prepare the raw data to a suitable input for the models are investigated. The best neural network is found with the best forecasting abilities, based on a certain performance measure. A visual graphs on the experiments data set is presented after processing steps, to illustrate that particular results. The out-of-sample results are compared with training ones.

© 2015 Elsevier B.V. All rights reserved.

Figure: Neural networks performance in exchange rate prediction [4]
Main assumptions:

- Three-layer perceptron (5-10-1) is used
- EUR/USD, GBP/USD, USD/JPY
- Three steps
  - Daily - 01.01.2014 - 25.04.2014 (83 values)
  - Monthly - 05.2009 - 05.2014 (60 values)
  - Quarterly - 05.1999 - 59.2014 (59 values)
Review
Neural networks performance in exchange rate prediction

**Figure:** NN model for EURUSD 1-day ahead prediction [4]
Evaluating machine learning classification for financial trading: An empirical approach

Eduardo A. Gerlein, Martin McGinnity, Ammar Belatreche, Sonya Coleman

Abstract

Technical and quantitative analysis in financial trading use mathematical and statistical tools to help investors decide on the optimum moment to initiate and close orders. While these traditional approaches have served their purpose, some recent techniques arising from the field of computational intelligence such as machine learning and data mining have emerged to analyze financial information. While the main financial engineering research has focused on complex computational models such as Neural Networks and Support Vector Machines, there are also simpler models that have demonstrated their usefulness in applications other than financial trading, and are worth considering to determine their advantages and inherent limitations when used as trading analysis tools. This paper analyses the role of simple machine learning models to achieve profitable trading through a series of trading simulations on the FOREX market. It assesses the performance of the models and how particular aspects of the models produce systematic and consistent predictions for profitable trading. Due to the inherent complexities of technical price patterns, the role of attribute selection, periodic rotation and training set size are discussed in order to obtain a combination of attributes not only capable of generating positive cumulative returns for each set of the machine learning models but also to demonstrate how simple algorithms, traditionally produced from financial forecasting for trading applications produce similar performances as more complex counterparts. The paper discusses the combination of attributes in addition to technical indicators that have been used as inputs of the machine learning-based predictions such as price-related features, seasonality features and lagged values used in classical time series analysis are used to enhance the classification capabilities that impact directly into the final profitability.

© 2016 Elsevier Ltd. All rights reserved.

Figure: Evaluating machine learning classification for financial trading: An empirical approach [6]
Review
Evaluating machine learning classification for financial trading: An empirical approach

- Low complexity ML models
- USDJPY, EURGBP and EURUSD
- Multiagent system trading (2 years)
- 6 hour time frame
- Binary classification (up/down)
- Test set I - 2510 (01.2007 - 06.2009)
- Test set II - 6442 (01.2007 - 04.2013)
Evaluating machine learning classification for financial trading: An empirical approach

Table 2. Simulation results of the USD/JPY trading agents, Single training at inception.

<table>
<thead>
<tr>
<th>Ticker</th>
<th>USD/JPY Training Set Size</th>
<th>Attributes: &lt;hour&gt;, &lt;day&gt;, &lt;closing_price&gt;, &lt;ppc&gt;, &lt;lpcc&gt;, &lt;lppl&gt;, &lt;RSI&gt;, &lt;Williams%R&gt;, &lt;class&gt; (9 attributes)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>Accuracy (%)</td>
<td>DOWN Accuracy (%)</td>
</tr>
<tr>
<td>--------</td>
<td>---------------</td>
<td>-------------------</td>
</tr>
<tr>
<td>K*</td>
<td>49.84</td>
<td>50.72</td>
</tr>
<tr>
<td>C4.5</td>
<td>51.08</td>
<td>53.06</td>
</tr>
<tr>
<td>Jrip</td>
<td>49.76</td>
<td>50.59</td>
</tr>
<tr>
<td>NB</td>
<td>50.60</td>
<td>51.05</td>
</tr>
<tr>
<td>LMT</td>
<td>50.92</td>
<td>52.96</td>
</tr>
<tr>
<td>OneR</td>
<td>51.27</td>
<td>50.06</td>
</tr>
<tr>
<td>Rand</td>
<td>48.49</td>
<td>47.58</td>
</tr>
</tbody>
</table>

Figure: Results without retraining [6]
Review
Evaluating machine learning classification for financial trading: An empirical approach

Table 4. Prediction results for USD/JPY. Retraining period = 50, Retraining test size = incremental since inception, 9 attributes

<table>
<thead>
<tr>
<th>Ticker</th>
<th>Retrain Set Size</th>
<th>Retrain Periods</th>
<th>Attributes: &lt;hour&gt;,&lt;day&gt;,&lt;closing_price&gt;,&lt;ppc&gt;,&lt;ppcma&gt;,&lt;RSI&gt;,&lt;Williams%R&gt;,&lt;class&gt; (9 attributes)</th>
<th>USD/JPY Incremental</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>50</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model</td>
<td>Accuracy (%)</td>
<td>DOWN Accuracy (%)</td>
<td>UP Accuracy (%)</td>
<td>SHORT Accuracy (%)</td>
</tr>
<tr>
<td>OneR</td>
<td>49.86</td>
<td>51.19</td>
<td>48.39</td>
<td>48.61</td>
</tr>
<tr>
<td>C4.5</td>
<td>52.25</td>
<td>54.57</td>
<td>50.78</td>
<td>52.59</td>
</tr>
<tr>
<td>Jrip</td>
<td>50.81</td>
<td>52.17</td>
<td>49.54</td>
<td>51.24</td>
</tr>
<tr>
<td>LMT</td>
<td>52.84</td>
<td>54.38</td>
<td>51.41</td>
<td>51.41</td>
</tr>
<tr>
<td>Rstar</td>
<td>50.73</td>
<td>52.05</td>
<td>49.45</td>
<td>50.33</td>
</tr>
<tr>
<td>NaiveBayes</td>
<td>52.59</td>
<td>55.42</td>
<td>51.51</td>
<td>50.98</td>
</tr>
</tbody>
</table>

Figure: Results with retraining [6]
### Review

Evaluating machine learning classification for financial trading: An empirical approach

Table 7. Experiment set ups for (a) maximum cumulative returns for EURGBP, (b) maximum accuracies for EURGBP, (c) maximum cumulative returns for EURUSD and (d) maximum accuracies for EURUSD

<table>
<thead>
<tr>
<th></th>
<th>OneR</th>
<th>C4.5</th>
<th>Jrip</th>
<th>LMT</th>
<th>Kstar</th>
<th>NaiveBayes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Accuracy (%)</strong></td>
<td>58.9003</td>
<td>64.0882</td>
<td>63.7465</td>
<td>64.3678</td>
<td>56.8810</td>
<td>63.8644</td>
</tr>
<tr>
<td><strong>Max Cumulative Return (%)</strong></td>
<td>32.0000</td>
<td>57.1697</td>
<td>95.9409</td>
<td>93.4583</td>
<td>73.1667</td>
<td>22.8258</td>
</tr>
<tr>
<td><strong>Retrain Set Size</strong></td>
<td>500</td>
<td>1000</td>
<td>1000</td>
<td>1000</td>
<td>500</td>
<td>1000</td>
</tr>
<tr>
<td><strong>Retrain Periods</strong></td>
<td>15</td>
<td>10</td>
<td>5</td>
<td>5</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td><strong># of Attributes</strong></td>
<td>9</td>
<td>5</td>
<td>5</td>
<td>9</td>
<td>5</td>
<td>9</td>
</tr>
</tbody>
</table>

(a) EURGBP Maximum Cumulative Return

<table>
<thead>
<tr>
<th></th>
<th>OneR</th>
<th>C4.5</th>
<th>Jrip</th>
<th>LMT</th>
<th>Kstar</th>
<th>NaiveBayes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Accuracy (%)</strong></td>
<td>52.5074</td>
<td>56.4974</td>
<td>56.8079</td>
<td>56.7814</td>
<td>56.9207</td>
<td>53.7184</td>
</tr>
<tr>
<td><strong>Max Cumulative Return (%)</strong></td>
<td>14.5224</td>
<td>33.3007</td>
<td>37.0147</td>
<td>10.3977</td>
<td>92.0795</td>
<td>-9.5866</td>
</tr>
<tr>
<td><strong>Retrain Set Size</strong></td>
<td>500</td>
<td>2000</td>
<td>4000</td>
<td>500</td>
<td>500</td>
<td>500</td>
</tr>
<tr>
<td><strong>Retrain Periods</strong></td>
<td>10</td>
<td>15</td>
<td>20</td>
<td>10</td>
<td>5</td>
<td>15</td>
</tr>
<tr>
<td><strong># of Attributes</strong></td>
<td>3</td>
<td>9</td>
<td>9</td>
<td>5</td>
<td>9</td>
<td>9</td>
</tr>
</tbody>
</table>

(b) EURGBP Maximum Accuracy

<table>
<thead>
<tr>
<th></th>
<th>OneR</th>
<th>C4.5</th>
<th>Jrip</th>
<th>LMT</th>
<th>Kstar</th>
<th>NaiveBayes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Max Accuracy (%)</strong></td>
<td>53.3513</td>
<td>57.3513</td>
<td>56.8079</td>
<td>57.7801</td>
<td>58.7869</td>
<td>57.2116</td>
</tr>
<tr>
<td><strong>Cumulative Return (%)</strong></td>
<td>14.1645</td>
<td>4.5894</td>
<td>37.0147</td>
<td>4.7742</td>
<td>2.8077</td>
<td>-11.8703</td>
</tr>
<tr>
<td><strong>Retrain Set Size</strong></td>
<td>500</td>
<td>1000</td>
<td>4000</td>
<td>2000</td>
<td>2000</td>
<td>4000</td>
</tr>
<tr>
<td><strong>Retrain Periods</strong></td>
<td>15</td>
<td>10</td>
<td>20</td>
<td>15</td>
<td>15</td>
<td>20</td>
</tr>
<tr>
<td><strong># of Attributes</strong></td>
<td>3</td>
<td>9</td>
<td>9</td>
<td>5</td>
<td>9</td>
<td>9</td>
</tr>
</tbody>
</table>

(c) EURUSD Maximum Cumulative Return

(d) EURUSD Maximum Accuracy

**Figure:** Results with retraining for best set up [6]
Price prediction is permanently open problem
New data intervals and sources can be applied
New models (hybrid) can be applied
Market is not constant
Results can be statistically insignificant
Ether price prediction
Ethereum graph

Blockchain parsing

1. Full blockchain download (via Parity)
2. Blockchain parsing to transaction list (via Parity API)
3. Transactions grouped into one-day packages
4. Few per-day measures extracted
   - From 2015-08-08 to 2018-04-29
Application of selected methods of statistical analysis and machine learning

**Figure**: Ether Price
Ethereum graph
Number of nodes

Figure: Number of nodes
Application of selected methods of statistical analysis and machine learning

Ethereum graph

Number of edges

Figure: Number of edges
Research goals
Review
Ethereum (blockchain) graph
Correlations
Prediction model - Logistic Regression
Prediction model - Neural Network

Ethereum graph
Total flow in Wei

Figure: Total flow in Wei

Application of selected methods of statistical analysis and machine learning
Ethereum graph
Total flow in USD
Ether correlations

Time series

Figure: Correlation of raw time series
Ether correlations

Time series of returns

Figure: Correlation of time series of returns
Channel: Correlations
Lagged time series of returns

Figure: Correlation of lagged time series of returns
Ether prediction model - Logistic Regression

Model description

- Logistic Regression
- Differences as features
- Rolling training/test set
  - Training - 333, 498, 663, 828 ..
  - Test - 165
- Results - AUC
  - 0.47
  - 0.55
  - 0.48
  - 0.51
  - 0.46
Ether prediction model - Neural Network

Model description

- Neural Network
  - Activation - ReLu + sigmoid (last layer)
  - Optimizer - Adam
  - Loss - binary cross entropy
  - MinMax scaler - based on train set

- Differences as features

- Rolling training/validation/test set
  - Training - 390, 490, 590, ..
  - Validation - 100
  - Test - 100

- Epochs from 50 to 2000 (500 when 20 inputs)
**Table: Model variants**

<table>
<thead>
<tr>
<th>Model</th>
<th>ethusd lag</th>
<th>btcusd lag</th>
<th>graph lag</th>
<th>Neurons in HL</th>
<th>Scaler</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>N</td>
</tr>
<tr>
<td>Model 2</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>N</td>
</tr>
<tr>
<td>Model 3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>Y</td>
</tr>
<tr>
<td>Model 4</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>5</td>
<td>Y</td>
</tr>
<tr>
<td>Model 5</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>N</td>
</tr>
<tr>
<td>Model 6</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>0</td>
<td>N</td>
</tr>
<tr>
<td>Model 7</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>N</td>
</tr>
</tbody>
</table>
**Ether prediction model - Neural Network**

### Results

<table>
<thead>
<tr>
<th>Model</th>
<th>AUC.1</th>
<th>AUC.2</th>
<th>AUC.3</th>
<th>AUC.4</th>
<th>AUC.5</th>
<th>AUC.6</th>
<th>AUC.7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1</td>
<td>0.62</td>
<td>0.55</td>
<td>0.51</td>
<td>0.54</td>
<td>0.46</td>
<td>0.51</td>
<td>0.47</td>
</tr>
<tr>
<td>Model 2</td>
<td>0.61</td>
<td>0.49</td>
<td>0.52</td>
<td>0.54</td>
<td>0.55</td>
<td>0.52</td>
<td>0.46</td>
</tr>
<tr>
<td>Model 3</td>
<td>0.50</td>
<td>0.50</td>
<td>0.48</td>
<td>0.37</td>
<td>0.46</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Model 4</td>
<td>0.44</td>
<td>0.44</td>
<td>0.52</td>
<td>0.43</td>
<td>0.55</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Model 5</td>
<td>0.55</td>
<td>0.53</td>
<td>0.51</td>
<td>0.47</td>
<td>0.46</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Model 6</td>
<td>0.52</td>
<td>0.52</td>
<td>0.55</td>
<td>0.49</td>
<td>0.42</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Model 7</td>
<td>0.53</td>
<td>0.52</td>
<td>0.50</td>
<td>0.48</td>
<td>0.58</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

---

**Zaborski**

Application of selected methods of statistical analysis and machine learning
Experiments
Moving average crossover

- Strategy based on 2 moving average cross - fast and slow

Figure: Moving average crossover [2]
Moving average crossover

- 5min interval
- 01.01.2016 - 01.09.2018 - divided into half
- Two indexes
  - EURUSD
  - DAX
- Long MA (SMA) lengths: 50, 100, ... 500
- Short MA (EMA) lengths: 5, 10, ... 40
Moving average crossover

EURUSD results

**Figure:** Mean return for long positions
Resea rch goals
Review
Ether price prediction
Experiments
Conclusion

Moving average crossover
EURUSD results

Figure: Mean return for short positions
Moving average crossover

DAX results

Figure: Mean return for long positions
Figure: Mean return for short positions
Logistic regression

- 5min interval
- Two indexes
  - EURUSD
  - DAX
- 01.01.2016 - 01.09.2018
  - 75%/25% - train/test split
- Prediction direction (up/down) 30min ahead
Logistic regression

Features
- Returns - $r_t, r_{t-1}, ..., r_{t-5}$
- Volume
- Log(Volume)
- RSI (n=14)
- WPR (n=14)
- MFI (n=14)
- 3 SMA deviation
  - 5
  - 20
  - 100
Logistic regression

EURUSD results

Figure: ROC (AUC = 0.536) for EURUSD
Logistic regression
EURUSD results

Figure: Mean return for EURUSD test set
Logistic regression
DAX results

Figure: ROC (AUC = 0.512) for DAX
Logistic regression

DAX results

Figure: Mean return for DAX test set
Application of selected methods of statistical analysis and machine learning in price prediction of EURUSD, DAX and Ether.
Conclusion

- Reliable price prediction is challenging problem
- Ether price
  - Relationships between price and blockchain structure exists
  - Using them in price prediction is still open question
Bibliography