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

Research goals

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

Review

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

Review

Using artificial neural network models in stock market index prediction

Using artificial neural network models in stock market index prediction

Erkam Guresen [a], Gulgun Kayakutlu [a,*], Tugrul U. Daim [b]

[a] *Istanbul Technical University, Istanbul, Turkey*
[b] *Portland State University, Portland OR, USA*

A R T I C L E I N F O

Keywords:
Financial time series (FTS) prediction
Recurrent neural networks (RNN)
Dynamic artificial neural networks (DAN2)
Hybrid forecasting models

A B S T R A C T

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.

Figure: Using artificial neural network models in stock market index prediction [5]

Review

Using artificial neural network models in stock market index prediction

- 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

Review

Neural networks performance in exchange rate prediction

Neural networks performance in exchange rate prediction

Svitlana Galeshchuk*

Ternopil National Economic University, 11 Lvivska str., 46000 Ternopil, Ukraine

ARTICLE INFO

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

ABSTRACT

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.

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

Review

Neural networks performance in exchange rate prediction

- 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

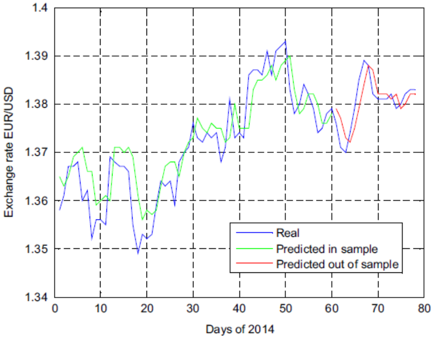


Figure: NN model for EURUSD 1-day ahead prediciton [4]

Review

Evaluating machine learning classification for financial trading: An empirical approach

Evaluating machine learning classification for financial trading: An empirical approach

CrossMark

Eduardo A. Gerlein [a,c,*], Martin McGinnity [a,b], Ammar Belatreche [a], Sonya Coleman [a]

[a] Intelligent Systems Research Centre, University of Ulster, Londonderry, UK
[b] School of Science and Technology, Nottingham Trent University, Nottingham, UK
[c] Electronics Department, Pontificia Universidad Javeriana, Bogotá, Colombia

A R T I C L E   I N F O

Keywords:
Trading
Financial forecasting
Computer intelligence
Data mining
Machine learning
FOREX markets

A B S T R A C T

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 to some extent, new techniques arising from the field of computational intelligence such as machine learning and data mining have emerged to analyse 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 in the FOREX market. It assesses the performance of the models and how particular setups of the models produce systematic and consistent predictions for profitable trading. Due to the inherent complexities of financial time series the role of attribute selection, periodic retraining and training set size are discussed in order to obtain a combination of those parameters not only capable of generating positive cumulative returns for each one of the machine learning models but also to demonstrate how simple algorithms traditionally precluded from financial forecasting for trading applications presents similar performances as their more complex counterparts. The paper discusses how a combination of attributes in addition to technical indicators that has been used as inputs of the machine learning-based predictors such as price related features, seasonality features and lagged values used in classical time series analysis are used to enhance the classification capabilities that impacts directly into the final profitability.

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, EURGPB and EURUSD
- Multiagent system trading (2 years)
- 6 hour time trame
- Binary classification (up/down)
- Test set - 2510 (01.2007 - 06.2009)
- Test set II - 6442 (01.2007 - 04.2013)

Review

Evaluating machine learning classification for financial trading: An empirical approach

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

Ticker	**USDJPY**								
Training Set Size	**5242**								
Attributes	**Attributes : <hour>,<day>,<closing_price>,<ppc>,<lppc>,<lppcma>,<RSI>,<Williams%R>,<class> (9 attributes)**								
Model	**Accuracy (%)**	**DOWN Accuracy (%)**	**UP Accuracy (%)**	**SHORT Accuracy (%)**	**Long Accuracy (%)**	**MAXDD (%)**	**Average Ret/Trade (%)**	**Cummulative Return (%)**	**Trades**
K*	**49.84**	50.72	49.14	49.02	49.14	29.68	-0.0068	**-13.59**	2510
C4.5	**51.08**	53.06	50.20	51.89	50.20	23.32	0.0034	**1.20**	2510
Jrip	**49.76**	50.59	49.02	49.41	49.02	33.10	-0.0090	**-22.08**	2510
NB	**50.60**	51.65	49.83	50.61	49.83	34.30	-0.0132	**-26.45**	2510
LMT	**50.92**	52.96	50.06	52.15	50.06	26.71	-0.0054	**-10.46**	2510
OneR	**51.27**	50.06	52.96	50.98	50.43	13.13	0.0119	**31.96**	2510
Rand	**48.49**	47.58	49.34	48.26	47.58	39.59	-0.0142	**-31.47**	2510

Figure: Results without retraining [6]

Review

Evaluating machine learning classification for financial trading: An empirical approach

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

Ticker	**USDJPY**
Retrain Set Size	**Incremental**
Retrain Periods	**50**
Attributes	**Attributes : <hour>,<day>,<closing_price>,<ppc>,<lppc>,<lppcma>,<RSI>,<Williams %R>,<class> (9 attributes)**

Model	**Accuracy (%)**	**DOWN Accuracy (%)**	**UP Accuracy (%)**	**SHORT Accuracy (%)**	**Long Accuracy (%)**	**MAXDD (%)**	**Average Ret/Trade (%)**	**Cummulative Return (%)**	**Long Average Ret/Trade (%)**	**Short Average Ret/Trade (%)**	**Trades**
OneR	**49.86**	51.19	48.59	48.61	48.71	31.59	-0.0005	**-7.22**	-0.0032	0.1560	6441
C4.5	**52.25**	54.57	50.78	52.39	50.74	20.28	0.0104	**87.50**	0.0065	0.3000	6441
Jrip	**50.81**	52.17	49.54	51.34	51.05	23.13	0.0042	**25.81**	0.0015	-0.0239	6441
LMT	**52.84**	54.58	51.41	51.41	52.34	22.42	0.0048	**30.98**	0.0019	0.0213	6441
Kstar	**50.73**	52.05	49.45	50.33	49.88	37.21	-0.0002	**-5.17**	-0.0028	-0.1077	6441
NaiveBayes	**52.59**	53.42	51.51	50.98	51.26	41.07	-0.0034	**-23.13**	-0.0070	-0.1716	6441

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 EURGPB, (b) maximum accuracies for EURGPB, (c) maximum cumulative returns for EURUSD and (d) maximum accuracies for EURUSD

	OneR	C4.5	Jrip	LMT	Kstar	NaiveBayes
Accuracy (%)	58.9003	64.0882	63.7465	64.3678	56.8810	63.6844
Max Cumulative Return (%	32.0090	57.1697	95.9409	39.4583	73.1667	22.6258
Retrain Set Size	*500*	*1000*	*1000*	*1000*	500	*1000*
Retrain Periods	*15*	*10*	*5*	*5*	15	*15*
# of Attributes	*9*	*5*	*5*	*9*	9	*5*

(a) EURGPB Maximum Cumulative Return

	OneR	C4.5	Jrip	LMT	Kstar	NaiveBayes
Accuracy (%)	58.9003	64.2280	63.7465	64.5076	61.4477	63.6844
Cumulative Return (%	32.0090	27.2639	95.9409	43.4592	21.8619	22.6258
Retrain Set Size	*500*	*2000*	*1000*	*1000*	2000	*1000*
Retrain Periods	*15*	*15*	*5*	*15*	15	*15*
# of Attributes	*9*	*5*	*5*	*9*	5	*5*

(b) EURGPB Maximum Accuracy

	OneR	C4.5	Jrip	LMT	Kstar	NaiveBayes
Accuracy (%)	52.5074	56.4974	56.8079	56.7614	56.9207	53.7184
Max Cumulative Return (%	14.5224	33.3007	37.0147	10.3977	92.0795	-9.5586
Retrain Set Size	*500*	*2000*	*4000*	*500*	500	*500*
Retrain Periods	*10*	*15*	*20*	*10*	5	*15*
# of Attributes	*5*	*9*	*9*	*5*	5	*9*

(c) EURUSD Maximum Cumulative Return

	OneR	C4.5	Jrip	LMT	Kstar	NaiveBayes
MaX Accuracy (%)	53.3613	57.3513	56.8079	57.7861	58.7869	57.2116
Cumulative Return (%	14.1643	4.5894	37.0147	4.7742	2.8077	-11.8703
Retrain Set Size	*500*	*1000*	*4000*	*2000*	2000	*4000*
Retrain Periods	*15*	*10*	*20*	*15*	15	*20*
# of Attributes	*9*	*5*	*9*	*9*	5	*9*

(d) EURUSD Maximum Accuracy

Figure: Results with retraining for best set up [6]

Review
Conclusion

- 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

Ethereum graph
Ether price (in USD)

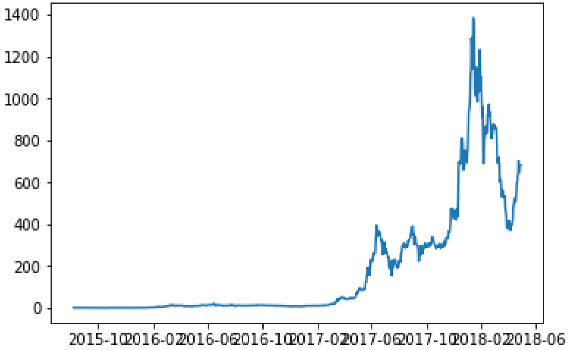


Figure: Ether Price

Ethereum graph

Number of nodes

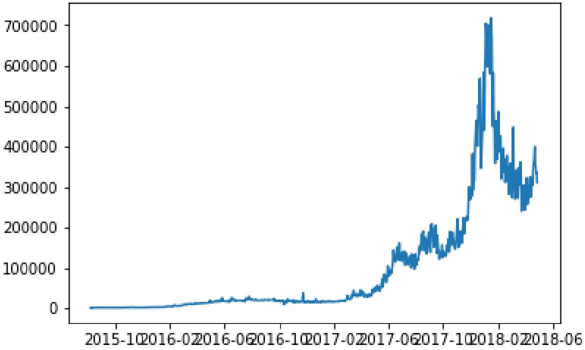


Figure: Number of nodes

Ethereum graph
Number of edges

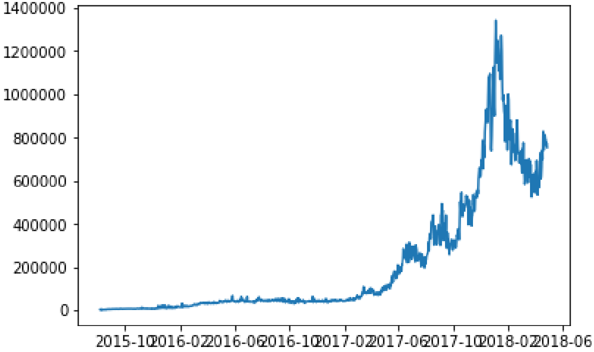


Figure: Number of edges

Ethereum graph
Total flow in Wei

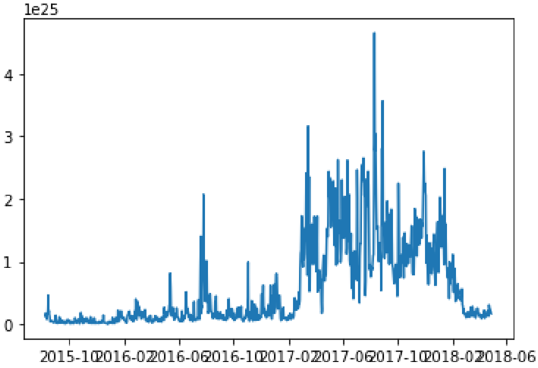
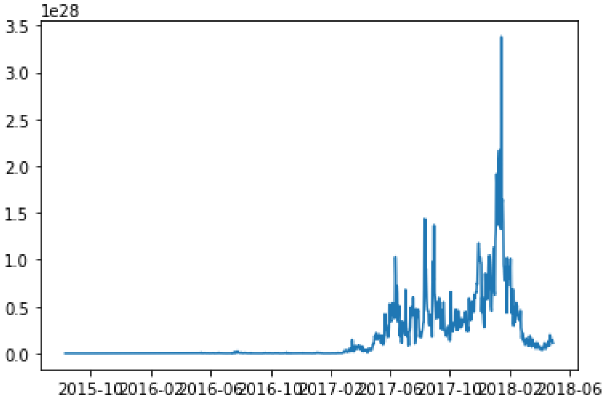


Figure: Total flow in Wei

Ethereum graph

Total flow in USD



Ether correlations

Time series

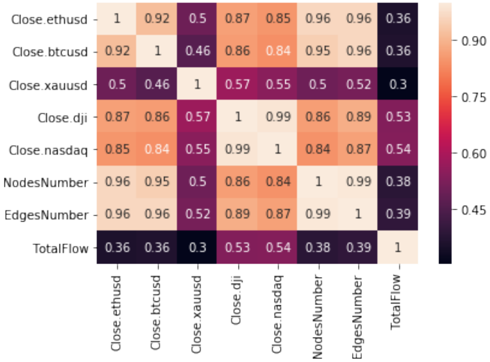


Figure: Correlation of raw time series

Ether correlations

Time series of returns

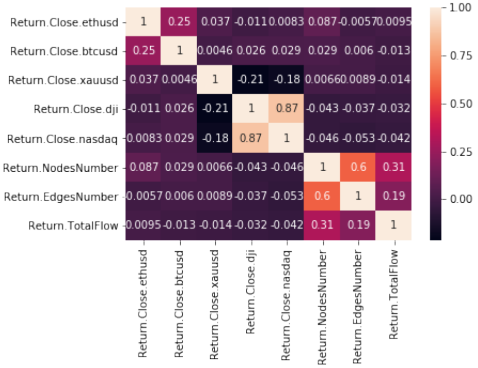


Figure: Correlation of time series of returns

Ether correlations

Lagged time series of returns

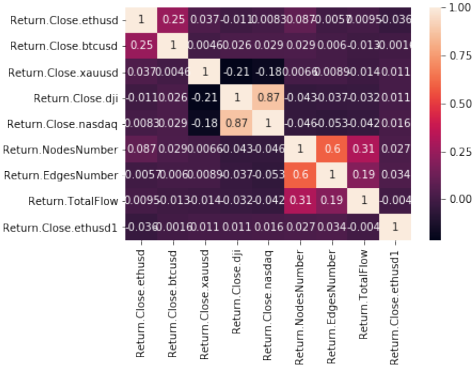


Figure: Correlation of lagged time series of returns

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)

Ether prediction model - Neural Network

Model variants

Table: Model variants

Model	ethusd lag	btcusd lag	graph lag	Neurons in HL	Scaler
Model 1	4	0	0	0	N
Model 2	4	0	0	2	N
Model 3	2	2	2	0	Y
Model 4	2	2	2	5	Y
Model 5	2	2	2	0	N
Model 6	4	4	4	0	N
Model 7	4	4	4	5	N

Ether prediction model - Neural Network

Results

Table: Results

Model	AUC.1	AUC.2	AUC.3	AUC.4	AUC.5	AUC.6	AUC.7
Model 1	0.62	0.55	0.51	0.54	0.46	0.51	0.47
Model 2	0.61	0.49	0.52	0.54	0.55	0.52	0.46
Model 3	0.50	0.50	0.48	0.37	0.46	-	-
Model 4	0.44	0.44	0.52	0.43	0.55	-	-
Model 5	0.55	0.53	0.51	0.47	0.46	-	-
Model 6	0.52	0.52	0.55	0.49	0.42	-	-
Model 7	0.53	0.52	0.50	0.48	0.58	-	-

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

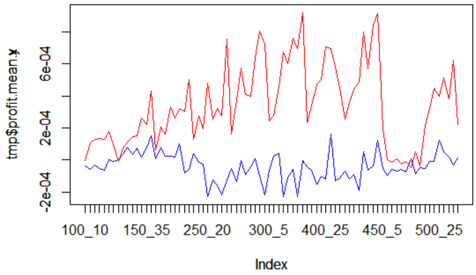


Figure: Mean return for long positions

Moving average crossover
EURUSD results

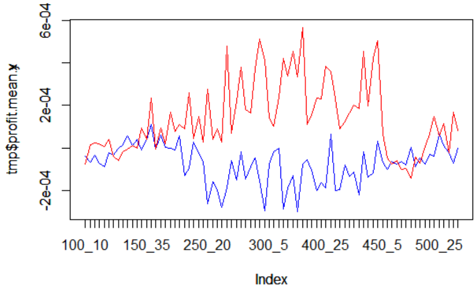


Figure: Mean return for short positions

Moving average crossover
DAX results

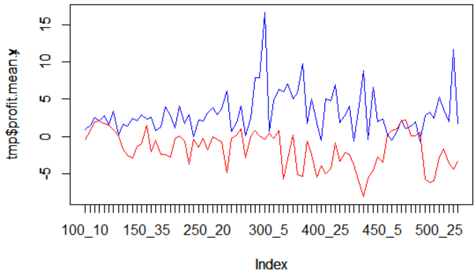


Figure: Mean return for long positions

Moving average crossover
DAX results

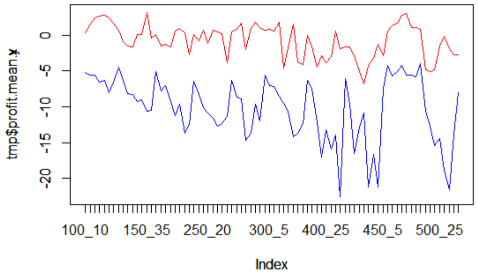


Figure: Mean return for short positions

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 - $r_t, r_{t-1}, \ldots r_{t-5}$
  - Volume
  - Log(Volume)
  - RSI (n=14)
  - WPR (n=14)
  - MFI (n=14)
  - 3 SMA deviation
    - 5
    - 20
    - 100

Logistic regression
EURUSD results

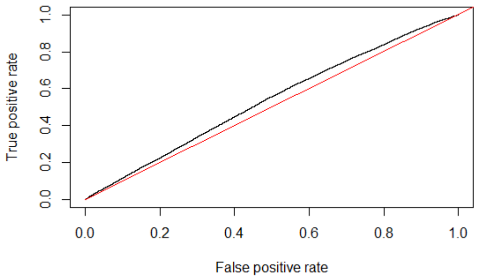


Figure: ROC (AUC = 0.536) for EURUSD

Logistic regression
EURUSD results

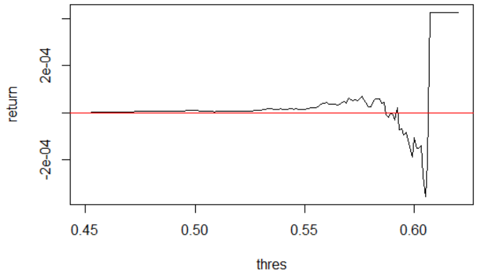


Figure: Mean return for EURUSD test set

Logistic regression
DAX results

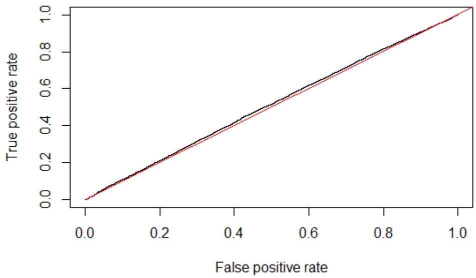


Figure: ROC (AUC = 0.512) for DAX

Logistic regression
DAX results

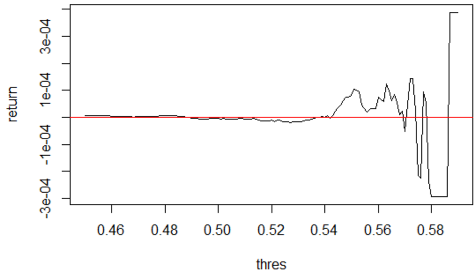


Figure: Mean return for DAX test set

Conclusion

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

[1] E. E. Peters „Teoria Chaosu a rynku kapitałowe", 2005

[2] https://www.investopedia.com/university/movingaverage/movingaverages4.asp, 04.12.2018

[3] Thien Hai Nguyen et al. „Sentiment analysis on social media for stock movement prediction", 2015

[4] S. Galeshchuk „Neural networks performance in exchange rate prediction", Neurocomputing (2015)

[5] E.Guresen et al. „Using artificial neural network models in stock market index prediction", 2011

[6] E. A. Gerlein et al. „Evaluating machine learning classification for financial trading: An empirical approach", 2016