Robert SZÓSTAKOWSKI[1]

The use of the Hurst exponent to investigate the quality of forecasting methods of ultra-high-frequency data of exchange rates[2]

1. INTRODUCTION

In this article, a review of forecasting method was conducted using historical tick data from the foreign exchange market. Both ask and bid prices from one trading day – from 5 pm 05-03-2012 to 5 pm 06-03-2012 for every exchange rate AUD/CAD, AUD/USD, GBP/JPY, GBP/PLN, GBP/USD, USD/CHF, USD/JPY were analyzed separately as independent data. Forecasting methods used in the article range from simple statistical methods like moving average, linear regression to more advanced like Kalman filter, ARMA, ARIMA models. Finally, machine learning methods like linear discriminant analysis and logistic regression were tested.

In the first part of the research, point forecasts were calculated using statistical models. For them, AMAPE errors were calculated to compare results between methods. In the second part of the research, logistic regression and linear discriminant analysis models were trained on historical data to predict a direction of bid/ask price change (up, down, the same). Additionally, the point forecasts from the previous part of the research were transformed into predictions of a direction of bid/ask price change. For a comparison of results a metric called forecasting accuracy was used (a percent of accurate forecasts).

The models were optimized by selecting the best hyperparameters based on their performance on historical data. The time series were divided into subseries of 10000 values. Every model was trained and optimized on the previous subseries and tested on out-of-sample data (on next 10000 values). These two steps were repeated for all subseries.

Moreover, for every time series, a rolling window of past 150 values was used to calculate values of the Hurst exponent. Every exponent value was assigned to the last 150th bid/ask quote from the rolling window. These results were compared with forecasting accuracy metrics from both parts of the research. At the end based on the forecasting results and time series characteristics, a few discovered dependencies were presented.

The main hypothesis of this research is the fact that the forecasting error decreases and the percentage forecasting accuracy increases when the value of the Hurst exponent increases. It was proven that analyzing the time series characteristics based on the chaos theory like a value of the Hurst exponent can be helpful in achieving better forecasting results. Finally, the average forecasting accuracy was higher for machine learning methods than for statistical methods, regardless of values of the Hurst exponent.

The idea behind this research was to develop a methodology which can be applied to the art of forecasting to increase the performance of a variety of models. The article main contribution to the science is a set of rules how to use the Hurst exponent for developing more accurate forecasting models.

The paper is divided into following parts. The first section is an introduction to the article. It describes the idea behind the research. Moreover, this section explains the main hypothesis and article contribution to the science. The next section describes the current state of the science in the field of forecasting and the Hurst exponent analysis. The third section characterizes the data used in the research. The next section is a detailed introduction to the Hurst exponent analysis with references to its application in other papers. The fifth section describes forecasting methods which were used in the research. The sixth section shows how the research was conducted. It describes an optimization algorithm and explains how the forecasting models were adjusted for obtaining the best results. The next section puts attention to the problems which were encountered by the researcher while forecasting high-frequency data. The eighth section includes a description of benchmarks which were used to compare forecasting models. The ninth section shows the empirical results of forecasting using statistical and machine learning models applied to tick currency data. The article ends with a section which summarizes the research and proposes further extensions.

2. RELATED LITERATURE

Forecasts based on large data sets gained a significant importance in every branch of economics. The necessity of forecasting led to a discovery of a range of methods starting from simple, autoregressive models to complicated nonlinear specifications. In most cases, a level of model complexity does not correlate with the expected results which is shown in the work of Green, Armstrong (2015). Moreover, it is impossible to discover a true model which generates a given time

series data according to the effective market hypothesis developed by Fama (1970) or the fractal market hypothesis discovered by Peters (1994) which says that price changes in financial markets are random or come from the deterministic chaos process.

An attempt to measure long memory effects was done by Hurst (1951) in his work in the area of hydrology. Analyzing the Hurst exponent on financial data might lead to more accurate, competitive out-of-sample forecasts and it is described in the work of Mitra (2012), Castillo, Melin (1996). Their research was conducted on daily stock market returns. According to their conclusions, values of the Hurst exponent which differ from 0.5 (random series) can be used as a predictor for investment strategies. Based on this fact, it was assumed that we can increase our investment results by analyzing values of the Hurst exponent over time. Furthermore, Qian, Rasshed (2004) proved that investigating values of the Hurst exponent calculated from Dow-Jones daily returns can increase the forecasting performance of neural networks. Unfortunately, he proved that analyzing only values of the Hurst exponent bigger than 0.5 leads to more accurate forecasts. In this research, a similar approach was applied to the full order book data of exchange rates. This article shows that the Hurst exponent results calculated from tick data have totally different density plots than in the research done by Qian, Rasshed (2004).

In the art of forecasting, it is almost impossible to repeat the forecasting accuracy in out-of-sample data based on historical results. Usually, models are selected by their performance on historical data what leads to the assumption that the best historical models will perform with at least the same accuracy in the future. A research presented by Aiolfi, Timmermann (2006) shows that a significant persistence in the forecasting performance can be found. The authors used data sets of stock market returns, interest rates and spread from main G7 economies during the 1959 and 1999 year.

In this article, their research is extended by analyzing time series characteristics based on the fractal theory developed by Peters (1994) to measure the accuracy of forecasting methods on high-frequency data of exchange rates. The research of Andersen, Bollerslev (1997) showed the importance of long-memory dependence in financial market volatility. This dependence in market volatility is characterized by slowly mean-reverting fractionally integrated process. Their article proved that forecasts of low-frequency volatility are more precise when based on high-frequency data. It established a link between financial markets microstructure and lower-frequency data relevance. Moreover, Cheung (1993) showed an evidence of long memory in exchange rates data. His research indicates that the empirical evidence of unit roots in exchange rates data is not robust to long memory alternatives. The author mentioned that the dependence is also hard to detect using impulse-response function analysis. The model used in the research – ARFIMA (integrated autoregressive moving average) did not

outperform a random walk in out-of-sample forecasts. Huang, Yang (1999) came to a similar conclusion using high-frequency data – one-minute time intervals in a given trading day for NYSE and NASDAQ. The researchers used the Modified Rescaled Range Analysis to discover a long-term memory in analyzed series. They showed sub periods during trading sessions when the random walk hypothesis was not supported. Their research conclusion is similar to Peters (1994) – local randomness and global determinism can actually coexist.

In the review of the forecasting methods, a wide range of models is used, starting from a simple moving average, ending with machine learning methods like linear discriminant analysis. Selecting two machine learning methods for the article was driven by analyzing the research results obtained in this area, particularly in the research of Ahmed et al. (2010), which shows how accurate predictions can be obtained using machine learning methods. Two models which were not used in the above-mentioned article: linear discriminant analysis and logistic regression were selected for the following research.

3. DATA AND PROGRAMMING FRAMEWORKS

The research was conducted on the ultra-high-frequency data set. The full order book data from one trading day – from 5 pm 05-03-2012 to 5 pm 06-03-2012 for seven major exchange rates was analyzed (AUD/CAD, AUD/USD, GBP/JPY, GBP/PLN, GBP/USD, USD/CHF, USD/JPY). Each exchange rate order book has a bid and ask series and they were both used in the forecasting research. As an example, one trading day from AUD/CAD exchange rate (bid and ask prices) is shown in figure 1.

Figure 1. Bid and ask time series from one trading day, AUD/CAD
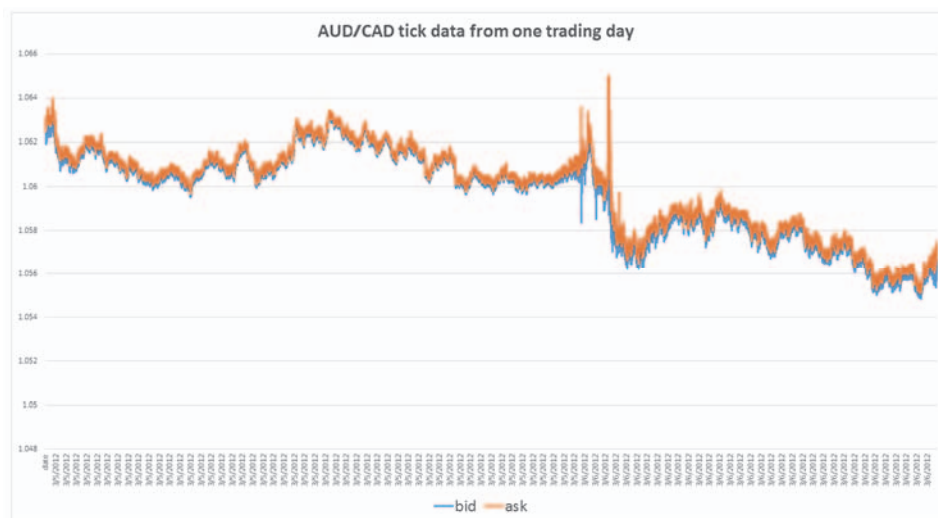
The forecasts of the next value (in this case it was tick – a change in the price of a security from trade to trade) were conducted on both ask and bid prices separately. In summary, fourteen different time series were taken into the analysis, when we consider ask and bid series separately. Moreover, given time series were used to calculate factors necessary for forecasting models like the Hurst exponent, logarithmic returns, which are described in a more detailed oriented manner in the sixth section. The volume of analyzed data can be seen in table 1.

Table 1. NUMBER OF TICKS FOR BID AND ASK PRICES IN ANALYZED EXCHANGE RATES

Number of ticks	AUD/CAD	AUD/USD	GBP/JPY	GBP/PLN	GBP/USD	USD/CHF	USD/JPY
ask	143 395	533 049	359 564	93 417	307 348	441 409	483 167
bid	143 395	533 049	359 564	93 417	307 348	441 409	483 167

In analyzed time series several zeros were removed and replaced by the previous value. This algorithm was chosen as the safest cleansing approach since repeating data in exchange rates time series are very common.

The research was conducted in the framework developed in python 3.5.2. using libraries: scikit-learn 0.16.1, numpy 1.8.2, matplotlib 1.4.2, pandas 0.16.2.

4. HURST EXPONENT

In the article, the Hurst exponent analysis is used to discover subparts of the time series, which have different characteristics like persistency, randomness or anti-persistency. A value of the Hurst exponent is calculated by rescaled range analysis (R/S analysis) which is a statistical measure of the variability. It is calculated by dividing the range of the values exhibited in subseries by the standard deviation of the values over the same subseries. Namely, a series of length N is divided into a number of shorter time series d with a length n, where $d \cdot n = T$. Then for every sub-period with a length n, a rescaled range is calculated:

1. Calculation of the mean value

$$m = \frac{1}{n} \sum_{i=1}^{n} X_i, \tag{1}$$

2. Calculation of the adjusted series Y

$$Y_t = X_t - m, \quad t = 1, 2, \ldots, n, \tag{2}$$

3. Calculation of the cumulative deviation

$$Z_t = \sum_{i=1}^{t} Y_i, \quad t = 1, 2, \ldots, n,$$

(3)

4. Calculation of the range series – R

$$R(n) = max(Z_1, Z_2, \ldots, Z_n) - min(Z_1, Z_2, \ldots, Z_n),$$

(4)

5. Calculation of the standard deviation series – S

$$S(n) = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (X_i - m)^2},$$

(5)

where m is the mean value calculated in point 1.

6. Calculation of the rescaled range series $\frac{R}{S}$

$$\left(\frac{R}{S}\right)_n = \frac{R(n)}{S(n)},$$

(6)

7. Hurst discovered that $\left(\frac{R}{S}\right)_n$ scales by power-law when the time increases which leads to the equation

$$E\left(\frac{R}{S}\right)_n = c\, n^H,$$

(7)

where c is a constant, independent of n and H is called the Hurst exponent.

The procedure described above is repeated for different values of $n = N, \frac{N}{2}, \frac{N}{4}, \frac{N}{8} \ldots$, where the minimum length of eight is usually chosen for the length of the smallest subseries. Finally, to estimate a value of the Hurst exponent, a simple, ordinary least squares regression is calculated on natural logarithms obtained from the equation from the 7th point.

$$lnE\left(\frac{R}{S}\right)_n = \ln c + H \ln n,$$

(8)

where H is the Hurst exponent.

The Hurst exponent divides time series into three groups:
— anti-persistent when $H < 0.5$ – price in time series tends to come back to the long-term mean (weak trends),
— persistent when $0.5 < H < 1$ – this is the opposite of anti-persistent, which means there is a strong trend and long memory effects,
— random when $H = 0.5$.

In the research, the values of the Hurst exponent were calculated for all fourteen bid/ask time series with maximal n which is equal to 150 observations, which is very close to 2^7. That means the Hurst analysis was done for subseries with lengths $2^7, 2^6, 2^5, ..., 2^3$. After the Hurst exponent estimation, every value was classified into ten segments from 0 to 1 with a step 0.1. The candle density plot with the results of the classification can be seen in figure 2.

Figure 2. Number of ticks classified into the Hurst exponent segments for every exchange rate
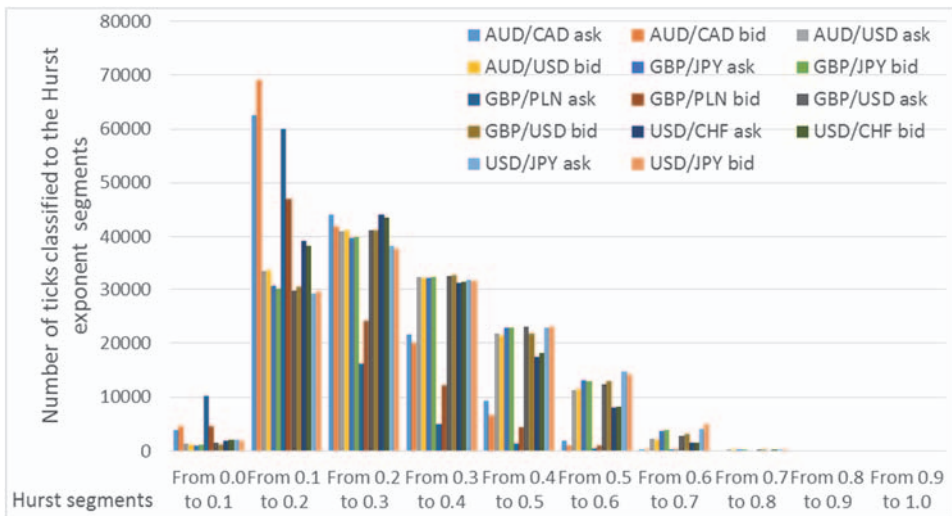
Most of the values seem to be categorized to the anti-persistent group (from 0 to 0.5) which indicates that the time series show a mean-reverting tendency.

5. FORECASTING METHODS USED IN THE RESEARCH

Eight forecasting methods were analyzed in this article. Starting from simple statistical methods like moving average or exponential smoothing which were chosen mostly as benchmarks, going to more advanced, statistical methods like linear regression or Kalman filter. The last two methods were chosen with a hope that more advanced methods should provide better results than simple models.

Moreover, two time series models were chosen – ARMA and ARIMA as a representative of the econometric modeling approach. At the end, the approach to forecasting was changed and two methods based on machine learning were selected: logistic regression and linear quadratic regression.

A forecasted point result obtained for models from paragraphs 5.1 to 5.6 for selected time spans was always taken as a forecast for the next value. For models 5.7 and 5.8 a possible direction of the price change was forecasted. Possible three states were used:

— price increased,
— price decreased,
— price stayed the same.

In the second part of the research, the point forecasts for the models described in 5.1 to 5.6 were transformed to the forecasts of possible direction of bid/ask price change to make a comparison between statistical and machine learning models possible.

5.1. Simple moving average

"Moving average" referring to a type of stochastic process is an abbreviation of Wold's (1939) process of moving average. It is an un-weighted mean of the previous n data which was used as a forecast for the next value. For calculating a simple moving average (SMA or running average) of n observations $(x_t, x_{t-1}, \ldots, x_{t-(n-1)})$ the following formula is used:

$$\text{SMA} = \frac{x_t + x_{t-1} + \ldots + x_{t-(n-1)}}{n} = \frac{1}{n} \sum_{i=0}^{n-1} x_{t-i}. \tag{9}$$

5.2. Exponential smoothing

Historically, the method was independently developed by Robert Goodell Brown and Charles Holt and it was described in "Smoothing, Forecasting and Prediction of Discrete Time Series" written by Brown (2004). The output of the exponential smoothing algorithm is defined as s_t, which can be regarded as the best estimate of what the next value of x will be. Given a sequence of observations which starts at the time $t = 0$ and a smoothing factor α, such as $0 < \alpha < 1$, the exponential smoothing model is given by the formulas:

$$s_0 = x_0, \tag{10}$$

$$s_t = \alpha \, x_t + (1 - \alpha) \, s_{t-1}. \tag{11}$$

5.3. Linear regression

It is an approach for modeling the relationship between a scalar dependent variable y and explanatory variables (or independent variables) denoted x. In linear regression, the relationships are modeled using linear predictor functions whose unknown parameters are estimated from the data.

In practice, there is an approximate linear relationship between the variables rather than exactly linear. This approximation can be represented by adding a non-observable variable ε, fancied as a collection of small errors. The variable is often called "noise" and represents all other factors which influence the dependent variable y_t. Given n observations $(x_t, x_{t-1},\ldots, x_{t-(n-1)})$ called "regressors", "exogenous variables" or "independent variables" the following hypothesis describes the estimation of the future value y_t which is called a "regressand" or "endogenous variable".

$$y_t = \beta_{t-1}x_{t-1}+\ldots+\beta_{t-(n-1)}x_{t-(n-1)} + \varepsilon, \tag{12}$$

where β are regression coefficients and ε is the error term or noise. This value captures all factors which influence the dependent value y_t other than the regressors x_t.

In order to estimate a linear regression model, few assumptions about the error term and data must be met (Greene, 2000):
— there is a random sampling of observations,
— the conditional mean should be zero $- E(\varepsilon \mid x) = 0$,
— there is no multicollinearity,
— the error terms should all have the same, finite variance (homoscedasticity),
— there is no autocorrelation (the error terms of different observation should not be correlated),
— the error term should be normally distributed.

In the research, the linear regression model was estimated few millions of times. Probably, for some estimations, the above-mentioned assumptions were met, but for some were not. The goal was to compare this linear method with other, more advanced methods (as a benchmark). For the same reasons, the linear regression was used in the articles of Altay (2005) and Ahangar et al. (2010).

Moreover, the linear regression was chosen for forecasting because of rapid price changes which can be observed in the analyzed time series. Moreover, this model is widely used in the scientific research, for example by Lin, Tsai (2015). For calculating linear regression forecasts the model from the scikit-learn library was used. It uses Ordinary Least Squares method to estimate the linear regression parameters.

5.4. Kalman Filter

It is the advanced statistical algorithm used widely in physics also known as linear-quadratic estimation (LQE). The model uses a series of measurements observed over time, containing statistical noise and other inaccuracies. It produces estimates of unknown variables that tend to be more precise than those based on a single measurement alone by using Bayesian Inference and by estimating a joint probability distribution over the variables for each time frame.

The Kalman filter can be written as a single equation but usually, it is described as two distinct equations which symbolize two phases: "predict" and "update". The predict phase uses the state estimate from the previous time step to produce an estimate of the state at the current time step, which is also called "a priori" state. The update phase combines the current "a priori" prediction with the current observation information in order to refine the state estimate. Described phases are formulated by following equations:

Predict:

Predicted (a priori) state estimate

$$\hat{x}_{t|t-1} = F_{t-1}\, \hat{x}_{t-1|t-1} + B_t u_t, \tag{13}$$

Predicted (a priori) estimate covariance

$$P_{t|t-1} = F_t\, P_{t-1|t-1} F_t^T + Q_t, \tag{14}$$

Update:

Updated (a posteriori) state estimate

$$\hat{x}_{t|t} = \hat{x}_{t|t-1}\, K_t\big(y_t - H_t \hat{x}_{t|t-1}\big), \tag{15}$$

Optimal Kalman gain

$$K_t = P_{t|t-1}\, H_t^T \big(H_t P_{t|t-1} H_t^T + R_t\big), \tag{16}$$

Updated (a posteriori) estimate covariance

$$P_{t|t} = P_{t|t-1}\, (I - K_t H_t), \tag{17}$$

where:

$\hat{x}$ – estimated state,	F – state transition matrix (i.e., transition between states),
u – control variables,	B – control matrix (i.e., mapping control to state variables),
P – state variance matrix,	Q – process variance matrix (i.e., error due to process),
y – measurement variables,	H – measurement matrix (i.e., mapping measurements),
K – Kalman gain,	R – measurement variance matrix.

The complexity of the filter was explained in the paper "Understanding the Kalman Filter" by Meinhold, Singpurwalla (1983).

5.5. ARMA – autoregressive–moving-average model

The model consists of autoregressive part $AR(p)$ where p means the order, and of moving-average model $MA(q)$ where q also means the order. The autoregressive part $AR(p)$ uses a linear combination of its own lagged values. It is described using the following equation:

$$x_t = c + \sum_{i=1}^{p} \varphi_i x_{t-i} + \varepsilon_t, \tag{18}$$

where: φ_i are parameters of the linear model on lagged values, ε_t are errors of the model which are assumed to be independent identically distributed random variables (i.i.d.) sampled from a normal distribution with zero mean: $e_t \sim N(0, \sigma^2)$ where σ^2 is the variance.

The moving average part $MA(q)$ models the error as a linear combination of error terms occurring contemporaneously in the past. It is described using the following equation:

$$x_t = \mu + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} + \varepsilon_t, \tag{19}$$

where: θ_i are parameters of the linear model on past error terms, μ is the expectation of x_t.

Finally, the ARMA model consists two models described above $\mathrm{AR}(p)$ and $\mathrm{MA}(q)$ and has the following formula:

$$x_t = c + \varepsilon_t + \sum_{i=1}^{p} \varphi_i x_{t-i} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i}, \tag{20}$$

where: φ_i are parameters of the autoregressive part of the model, θ_i are parameters of the moving average.

A more detailed description of the model can be found in the work of Tsay (2002).

5.6. ARIMA model – autoregressive integrated moving average

This model is a generalization of the autoregressive moving average (ARMA) model. Non-seasonal ARIMA models are generally denoted $\mathrm{ARMA}(p, d, q)$ where parameters p, d, and q are non-negative integers and p is the order of the auto regression. This model is usually applied to the data which show evidence of non-stationarity. In such case, a differencing step can be applied to reduce the non-stationarity in the time series. The ”I” part of the model stands for ”Integrated” which means that the data were reduced by replacing non-stationary series by the difference between their values and the previous values. They can be estimated using the Box-Jenkins approach. A more detailed description of the model can be found in the research of Tsay (2002).

5.7. Logistic regression

This model measures the relationship between the categorical dependent variable and independent variables by estimating probabilities using a logistic function. It is used to estimate the probability of a binary response based on one or more predictors (or independent) variables (features). It is not a classification method, but in terms of economics, it is called qualitative response/discrete choice model. It can be seen as a special case of the generalized linear model which is analogous to linear regression, but it is based on different assumptions. The conditional distribution is a Bernoulli distribution rather than a Gaussian distribution because the dependent variable is binary. This method was developed by statistician David Cox in 1958. A detailed explanation of the logistic regression was described in the article by Yu et al. (2010).

5.8. Linear discriminant analysis

The model is a generalization of Fisher's linear discriminant, which is used to find a linear combination of features that characterizes two or more classes of

objects. The resulting combination may be used as a linear classifier or for dimensionality reduction before conducting a classification. The method is related to ANOVA and to regression analysis, but it has continuous independent variables and a categorical dependent variable when ANOVA has categorical independent variables and a continuous dependent variable. It is widely used for face recognition problems, bankruptcy prediction and in marketing. In the research, the model LDA from the python machine learning library scikit-learn was used. It was implemented based on the description in the work of Hastie et al. (2008).

6. RECALCULATION OF MODELS AND OPTIMIZATION

6.1. Models calibration

In the article, eight forecasting methods were optimized by selecting the best hyperparameters based on their performance on historical data. Every time series out of fourteen was divided into subseries of 10000 values. After selecting the best model for particular subseries, the hyperparameters were used to conduct forecasts on the next subseries (10000 values, out-of-sample data). These two steps were repeated for the whole time series. The same procedure was applied to every time series. A model optimization for each subseries was conducted based on criteria's described below:
— the lowest AMAPE error – for models like moving average, exponential smoothing, linear regression, Kalman filter models, described in paragraphs 5.1 to 5.6,
— the highest accuracy of forecasting a direction of bid/ask price change, which is a percentage of accurate forecasts of possible up, down movements or without changes divided by the number of all forecasts. These criteria were used for logistic regression and linear discriminant analysis models described in paragraphs 5.7 and 5.8.

This approach for model calibration was chosen based on the research done by Katz, McCormick (2000).

6.2. Optimization

For the model optimization, a simple brute force algorithm was used – this approach checks all possible variants of the parameter from the set of possible combinations. For instance, for moving average parameter – window length, we start the optimization from taking 3 last prices, including the current one and then we check the following lengths: 6, 9, 12 with the step 3 until the length is smaller than 20. For every window length forecasting metrics are calculated (described in 4.1) and the best model (the optimal window length) is taken to forecast next 10000 values.

An explanation what parameters were optimized for every model can be found below:

1) moving average was trained on bid or ask prices. Only one parameter was optimized "window length" – indicating how many quotations were used to calculate the forecasts of the model. The optimal value was selected from the set of values (3, 6, 9, 12, 15, 18).

2) linear regression model was also trained on bid or ask prices. It was assumed that endogenous variable is a predicted, next bid or ask price and exogenous prices are last n bid or ask prices, and the number of exogenous prices – n is chosen using the brute force optimization from the set of values (9, 12, 15, 18).

3) exponential smoothing model was trained on bid or ask prices. Only one parameter was optimized – a smoothing parameter called "alpha". The optimal value was selected from the set of values (0.1, 0.2, 0.3, 0.4, 0.5, 0.6),

4) Kalman filter was also trained on bid or ask prices. Two internal model parameters "Q" and "R" were optimized. Furthermore, for Kalman filter, two initial values for every calculation were taken arbitrary $P_o = 0$, $x_0 = 0$. These two parameters were selected from the set of values (0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9),

5) ARMA and ARIMA models were calculated from logarithmic changes of bid/ask prices. Both models were chosen to the research due to the fact that the exchange rate time series tend to have stationary and non-stationary periods. The optimization of ARMA and ARIMA models was performed by checking the following combinations of autoregressive parts and moving average parts:
   — ARMA(p,q) – with the following autoregressive part p and moving average part q – (1, 0), (1,1), (1,2), (0,1) (2,1), (2,2), (2,0), (0,2),
   — ARIMA(p,1, q) – with the following autoregressive part p and moving average part q – (1,1,0), (1,1,1), (1,1,2), (0,1,1) (2,1,1), (2,1,2), (2,1,0), (0,1,2).

6) logistic regression and linear discriminant analysis were trained on logarithmic changes of bid/ask prices. These two machine learning models were trained by providing an input vector of logarithmic price changes with a labeled outcome based on the next bid/ask price:
   — 1 when the next bid/ask price was higher than the previous one,
   — 0 when the next bid/ask price was equal to the previous one,
   — 1 when the next bid/ask price was lower than the previous one.

The length of input vectors was optimized. The best model was selected after checking the historical performance of models trained on following input vectors lengths (3, 6, 9, 12, 15, 18). Given such training, these models were taught how to predict a direction of bid/ask price change.

Table 2 summarizes the optimization phase.

Table 2. OPTIMIZATION PARAMETERS WITH POSSIBLE VALUES

Method	Parameter	From	To	Step
Moving average	window length	3	20	3
Exponential smoothing	alpha	0.1	0.6	0.1
Linear regression	window length	9	20	3
Kalman filter	Q	0.1	0.9	0.1
	R	0.1	0.9	0.1
Logistic regression	values used for training	3	20	3
Linear discriminant analysis	values used for training	3	20	3

Surprisingly, after the optimization phase, the best forecasting results for many subseries were achieved, using the same internal model parameters like:
— for Kalman filter alpha=0.1 beta=0.1,
— for exponential smoothing alpha=0.5,
— for linear regression the minimal window length 9 observations,
— for moving average the minimal value equal 3 observations,
— for machine learning methods usually 3 last logarithmic returns were taken.

That leads to the conclusion that the last prices which are taken to calculate a forecast, tend to have the highest significance of all of the observations. Particularly, we can see that analyzing the exponential smoothing model. If it could have been possible for the alpha to take a value 1, it would have become a naive forecasting method, due to the process of the optimization. That is why the maximum value was restricted to 0.6.

7. PROBLEMS WITH FORECASTING HIGH-FREQUENCY DATA AND SOLUTION

During the research process, few problems appeared. First of them was connected with the repeated bid or ask values in the analyzed time series. It was caused by adjusting only one value of the tick while keeping the second value at the same level (i.e. ask value was changing when bid stayed the same for several ticks). This behavior can be seen in almost every analyzed time series. In such case, the data were not interrupted manually (for instance by cleansing), because of the bid and ask prices are unbreakable part of the tick, but even though it might have caused many problems during forecasting and less accurate predictions.

A solution for modeling price changes when the price is the same for several values, is developing models which can predict the fact that the price will not change in the next tick. Most often the subparts of the time series with the same prices are encountered when there is a lack of volatility, especially during the night. A predicting model which can detect such subparts of the time series can probably outperform described models.

That is the reason why it was worth to check the performance of the models forecasting a direction of bid/ask price change. In this case, the accuracy means how many times the method was correct forecasting that the price will go up, down or it will stay the same, divided by the overall number of forecasts. Given these three states (up, down, the same) it was possible to introduce machine learning algorithms from the family of supervised learning models which can be taught how to predict price changes.

In the work of Moody, Wu (1995), the authors found a relationship between the forecasting accuracy and bid/ask spreads on tick-by-tick data. Accordingly, this research was focused on finding a possible relationship between values of the Hurst exponent and the forecasting accuracy of statistical models and machine learning models. It is worth to check in further extensions of the research if there is a seasonality in intraday values of the Hurst exponent and how much it affects the forecasting results. The research conducted by Bayraktar et al. (2003) shows that a seasonality in intraday financial data (S&P 500) can affect the Hurst ratio estimation. In order to achieve the robustness of the Hurst exponent estimation, the authors suggest using wavelets with at least two vanishing moments.

8. BENCHMARKS AND FORECASTING ACCURACY MEASURES

Every forecasting technique must be evaluated before a possible application in solving a real economic problem. It is good to have some benchmarks which can be used for comparisons. Gately (1995) has shown that machine learning methods tend to have a better performance when predicting price changes rather than making point forecasts. Due to that fact, two machine learning models chosen in this research (described in paragraphs 5.7 and 5.8) were configured to predict price changes (increased, decreased, without change). This solution is recommended by the literature (i.e. Gately, 1995).

In the first research, the methods from the paragraphs 5.1 to 5.6 were analyzed and following accuracy measures MAE, MSE, RMSD, MAPE and AMAPE were calculated for every one of them[3]. The AMAPE (adjusted mean absolute percentage error) measure was used as a tool for comparing results. For comparing different currency data sets (varying descriptive statistics) AMAPE error, from all measures used to estimate how close forecasts or predictions are to the eventual outcomes, is the most accurate. It makes possible to compare currency data sets with each other in a fair manner which is described by (Doman, Doman, 2009). The metric is defined by the formula:

---

[3] The MAE, MSE, RMSD, MAPE results are available on public repository for every exchange rate in unstructured output from the scripts on: https://github.com/rszostakowski/phd/tree/master/1Article/resultsForEveryExchangeRate.

$$AMAPE = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{y_{t+i} - \hat{y}_{t+i}}{y_{t+i} + \hat{y}_{t+i}}\right|, \tag{21}$$

where:

N – number of forecasts, $\hat{y}_{t+i}$ – a point forecast in the future,
i – the length of the forecast period, t – number of observations in the
                                                                              testing data set.

Moreover, in this part of the research, a naive forecasting was used as a benchmark. In this model, the current price is a forecast, without adjusting them or attempting to establish causal factors:

$$\hat{y}_{t+h|t} = y_t. \tag{22}$$

In the second research, the highest accuracy of forecasting a direction of bid/ask price change was used to compare methods between each other. Machine learning methods were trained to forecast a possible up, down price movements or without changes. For making it comparable to other methods, point forecasts obtained using methods from 5.1 to 5.6 were also classified into these three states, i.e. if exponential smoothing forecast indicates that the price will go up we have a classification to the "up" state. Owing to that it was possible to calculate the accuracy of every described method in the article. For benchmarking two artificial time series were created:
— "Random up and down" – a realization of the Markov process with two possible states (price went up or down) with the equal probability 0.5 of appearance,
— "Random up and down or the same" – a realization of the Markov process with three possible states (up or down, stayed the same) with the equal probability $\left(\frac{1}{3}\right)$ of appearance.

9. EMPIRICAL STUDY

9.1. Statistical models forecasts

The Hurst exponent analysis was performed for every rolling window of 150 values and the result was assigned to the last tick from these 150 values. It was conducted to check, if subseries with a higher or lower value of the Hurst exponent than 0.5 (random time series) can be used to make more ac-

curate forecasts, than for the series with a value of the Hurst exponent oscillating around 0.5. Thereafter, for every tick, the assigned value of the Hurst exponent was classified into the segments from 0 to 1 with a step 0.1. The results of the Hurst exponent classification can be seen in figure 2. Surprisingly, the values of the Hurst exponent bigger than 0.8 were not obtained. They are typical for time series with a strong trend. The majority of Hurst exponent ratios was classified into the first two segments (from 0 to 0.1 and from 0.1 to 0.2).

Figure 3. The Hurst exponent segment-based classification of forecasting accuracy measured by AMAPE, historical, tick-by-tick foreign exchange rates data from one trading day – 5pm 05-03-2012 to 5pm 06-03-2012
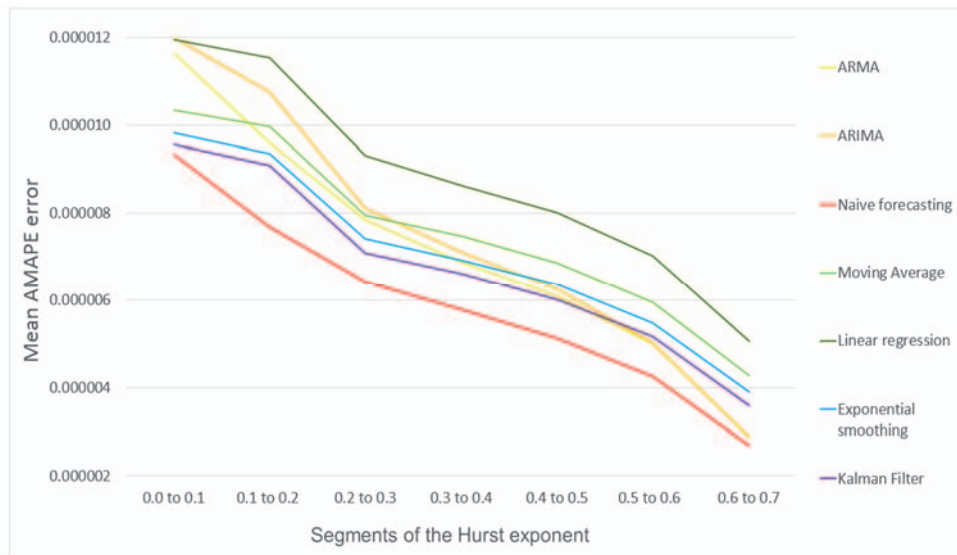
For every statistical model, an average adjusted mean absolute percentage error (AMAPE) was calculated to show how close forecasts of these models are to the eventual outcomes[4]. It was done to show a general performance of the forecasting methods, according to the Hurst exponent classification. The results can be seen in figure 3.

Unfortunately, average AMAPE for statistical models seems to be higher than AMAPE for naive forecasting. It proves, that the forecasting using sophisticated statistical methods on the high-frequency data cannot outperform the results of

---

[4] The partial results of AMAPE measures for every exchange rate are available on github repository: https://github.com/rszostakowski/phd/blob/master/1Article/ResultsWithHurstExponent.xlsx.

naive forecasting. The same conclusion can be found in the work of Cheung (1993), Green, Armstrong (2015) and Kilian, Tylor (2001). The closest method to achieving this goal was Kalman Filter. Surprisingly, the average performance of ARMA and ARIMA methods was similar for most of the exchange rates. Moreover, their forecasts were more inaccurate than the predictions of trivial methods like exponential smoothing for the Hurst ratio between 0 and 0.3 but for the Hurst exponent segments above 0.4 they increased their performance and were more accurate than any other method except for the naive forecasting. Unsatisfactory results were obtained due to a large number of data, which were repeated (bid and ask prices). The worst method for all Hurst exponent segments was the linear regression. Even strong persistence observed for the highest values of the Hurst exponent did not improve significantly the results of the linear regression model.

The most important discovery found in this research is the fact that the forecasting errors tend to decrease for every forecasting method when the value of the Hurst exponent increases. This observation is persistent for all forecasting models. Even for the Hurst exponent ratio close to 0.5, which indicates a random time series, better forecasts were obtained than for values of the Hurst exponent close to 0 (a mean-reverting time series). This fact proves the hypothesis of the research that forecasting methods are more accurate when the value of the Hurst exponent increases.

9.2. Forecasting a direction of bid/ask price change

In the second part of the research, an alternative approach to forecasting a point value was introduced. Forecasting a direction of bid/ask price change was chosen to overcome some of the drawbacks of the previous approach. For every method, an average accuracy of forecasting a direction of bid/ask price change was calculated and classified to the Hurst exponent ranges[5]. It is worth to mention that for time series without many repeated values a benchmark process "random up and down" should have an average accuracy of forecasting close to 0.5. As we can see in figure 4 a higher mean forecasting accuracy was obtained for the "random up and down or the same" process which indicates that analyzed times series have many repeated bid/ask prices. It is approximately the one-third of ticks. The second process slightly improves its performance for values of the Hurst exponent above 0.6 due to the fact that the statistical sample for these segments is too small (figure 2).

---

[5] The accuracy of forecasting a direction of bid/ask price change for every exchange rate are available on github repository: https://github.com/rszostakowski/phd/blob/master/1Article/ResultsWithHurstExponent.xlsx.

Figure 4. The Hurst exponent segment-based classification of mean accuracy of forecasting a direction of price change, historical, tick-by-tick foreign exchange rates data from one trading day – 5pm 05-03-2012 to 5pm 06-03-2012
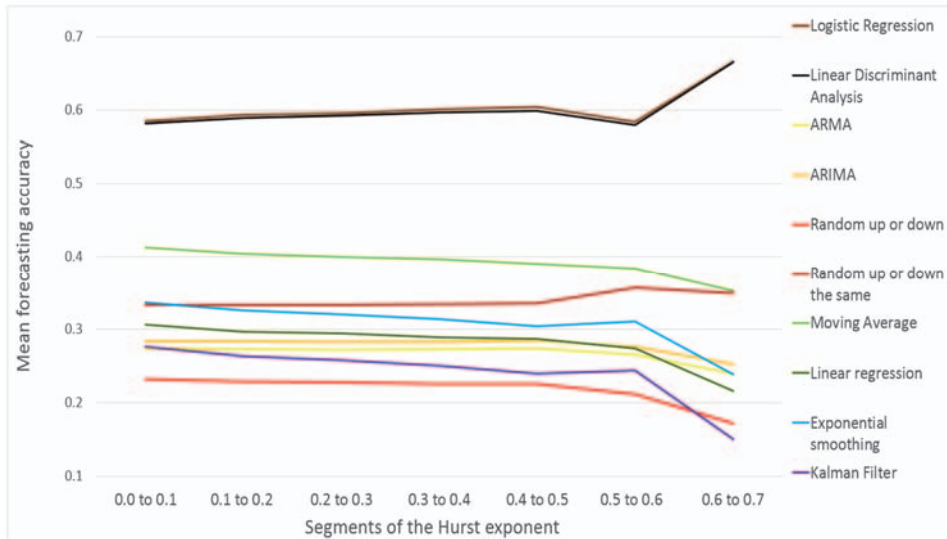
In this part of the research, the most accurate forecasting results were obtained by models based on machine learning techniques – the logistic regression and the linear discriminant analysis. For both of them, the average accuracy was close to 0.6, which means that in average the machine learning models were accurate in predicting 60% price directions. Such good results indicate that this approach can be successfully applied in liquidity forecasting or in active investment. Moreover, in most cases, the moving average model outperformed the random processes and other statistical models like ARMA and ARIMA. The most sophisticated statistical method like Kalman Filter appeared to be inaccurate. It might have been caused by too narrow model calibration.

Interestingly, for GBP/PLN ask and bid series the average forecasting accuracies were higher than for any other currency. Moreover, for these two series even models like Kalman filter, moving average and exponential smoothing were more accurate than the naïve forecasting. It might have been caused by a lower liquidity observed on GBP/PLN exchange rate. It means that almost every new tick in the series was changing both of the prices (bid and ask) simultaneously.

This part of the research has revealed that the statistical models tend to have a contrary tendency to the one, discovered in the first part of the research. For them, an average accuracy of forecasting a direction of bid/ask price change decreases when the value of the Hurst exponent increases. Fortunately, this tendency is not observed in the forecasting results of machine learning methods. These methods seem to benefit from the increase of the Hurst exponent – the

bigger the value of the Hurst exponent the higher their average performance. That indicates that a further research into statistical learning methods should be pursued.

10. CONCLUSIONS AND FURTHER RESEARCH

In this paper, seven different exchange rates series were tested using the Hurst exponent analysis (full order book, bid and ask prices). For most of the data, the values of the Hurst exponent oscillated around 0.2. These results differ from the study of Mitra (2012), who has shown that the Hurst exponent calculated from daily returns of stock market indices oscillates around 0.5.

Unfortunately, the first part of the research revealed that analyzed statistical methods tend to be more inaccurate than the naive forecasting on high--frequency-data. Moreover, their average AMAPE was higher than the average adjusted mean absolute percentage error of naïve forecasting for all Hurst exponent segments. From the other side, the results from the second part of the research were far more satisfying. The machine learning methods used for predicting a direction of bid/ask price change were more precise than the random process and the statistical methods. Finally, none of the methods were able to show a significant increase in their performance for mean reverting parts of the time series.

The most important research discovery is the fact that the average forecasting errors tend to decrease with an increase of the Hurst exponent for statistical and machine learning methods. Moreover, for the values of the Hurst exponent close to 0.5 forecasts from all analyzed models, seem to be more accurate than for the values which indicate a mean reverting time series (from 0.1 to 0.2). A positive effect of analyzing the Hurst exponent in forecasting was observed in the research paper of Mitra (2012) for time series with the Hurst exponent bigger than 0.5. Unfortunately, the author did not analyze the effect of the Hurst exponent below 0.5 on forecasting, which was done in this article.

Furthermore, for machine learning methods the accuracy of forecasting a direction of price change tends to increase when the value of the Hurst exponent increases. According to the author's best knowledge, this approach has not been analyzed in any other research paper. Finally, only machine learning methods discovered nonlinear structures hidden in the data and proved to be successful in predicting price changes. Their accuracy usually oscillated around 60%.

Forecasting high-frequency data of exchange rates seems to be very complex, due to the currency market characteristics (volume, number of investors). The research has shown several areas for an improvement in the art of forecasting which tend to follow in three directions.

First of them, is the path of analyzing market characteristics. It is worth to try creating time series regimes when it is highly probable to outperform the naive forecasting based on other metrics than the Hurst exponent. A second path leads to developing better forecasting models. Vengertsev (2014) and Ghahramani (2001) have shown that more advanced machine learning methods like support vector machine or deep learning models which can be successfully applied in the forecasting. Moreover, based on the papers written by Menkhoff, Taylor (2007) and Osler (2003) it is worth to apply technical analysis indicators as input vectors to machine learning models.

Finally, it is worth to try to develop a statistical model which would detect the fact, that a bid or ask price can be unchanged in the future tick. It can increase forecasting performance and lead to outperforming the naive forecasting model.

REFERENCES

Ahangar R. G., Yahyazadehfar M., Pournaghshband H., (2010), The Comparison of Methods Artificial Neural Network with Linear Regression Using Specific Variables for Prediction Stock Price in Tehran Stock Exchange, *arXiv preprint arXiv*:1003.1457.

Ahmed N. K., Atiya A. F., El Gayar N., El-Shishiny H., (2010), An Empirical Comparison of Machine Learning Models for Time Series Forecasting, *Econometric Reviews*, 29 (5), 594–621.

Aiolfi M., Timmermann A., (2006), Persistence in Forecasting Performance and Conditional Combination Strategies, *Journal of Econometrics*, 1–2, 31–53.

Altay E., Satman M. H., (2005), Stock Market Forecasting: Artificial Neural Network and Linear Regression Comparison in an Emerging Market, *Journal of Financial Management and Analysis* 18 (2), 18–33.

Andersen T. G., Bollerslev T., (1997), Heterogeneous Information Arrivals and Return Volatility Dynamics: Uncovering the Long-Run in High Frequency Returns, *The Journal of Finance,* 52 (3), 975–1005.

Bayraktar E., Poor H. V., Sircar K. R., (2003), Efficient Estimation of the Hurst Parameter in High Frequency Financial Data with Seasonalities using Wavelets, *Computational Intelligence for Financial Engineering, Proceedings of the IEEE 2003 International Conference*, 309–316.

Brown R. G., (2004), *Smoothing, Forecasting and Prediction of Discrete Time Series*, Dover Publications, Dover Phoenix Ed edition.

Castillo O., Melin P., (1996), Automated Mathematical Modelling for Financial Time Series Prediction Using Fuzzy Logic, Dynamical Systems and Fractal Theory, *Computational Intelligence for Financial Engineering, Proceedings of the IEEE/IAFE* 1996 *International Conference*.

Cheung Y.-W., (1993), Long Memory in Foreign-Exchange Rates, *Journal of Business & Economic Statistics* 11 (1), 93–101.

Doman M., Doman R., (2009), *Modelowanie zmienności i ryzyka*, Wolters Kluwer Polska Sp. z.o.o, Warszawa.

Fama E., (1970), Efficient Capital Markets: a Review of Theory and Empirical Work. *Journal of Finance*, 25, 2, 383–417

Gately E., (1995), *Neural Networks for Financial Forecasting*, John Wiley & Sons Inc.

Ghahramani Z., (2001), An Introduction to Hidden Markov Models and Bayesian Networks, *International Journal of Pattern Recognition and Artificial intelligence*, 15 (1), 9–42.

Green K. C., Armstrong J. S., (2015), Simple Versus Complex Forecasting: The Evidence, *Journal of Business Research*, 68, 8, 1678–1685.

Greene W. H., (2000), *Econometric Analysis,* International edition.

Hastie T., Tibshirani R., Friedman J., (2008), *The Elements of Statistical Learning*, Section 4.3, 106–119.

Huang B.-N., and Yang C. W., (1999), An Examination of Long-Term Memory Using the Intraday Stock Returns, *Technical Report* 99–03, Clarion University of Pennsylvania, Clarion.

Hurst H. E., (1951), Long-Term Storage Capacity of Reservoirs, *Transactions of American Society of Civil Engineers,* 116, 770–799.

Katz J. O., McCormick D., (2000), *The Encyclopedia of Trading Strategies,* McGraw-Hill.

Kilian L., Taylor M., (2001), Why Is It so Difficult to Beat the Random Walk Forecast of Exchange Rates, Europan Central Bank, *Working paper Series, 88*.

Lin T., Tsai C. C., (2015), A Simple Linear Regression Approach to Modeling and Forecasting Mortality Rates, *Journal of Forecasting,* 34, 543–559.

Meinhold R. J., Singpurwalla N. D., (1983), Understanding the Kalman Filter, *The American Statistician*, 37, 2, 123–127.

Menkhoff L., Taylor M., (2007), The Obstinate Passion of Foreign Exchange Professionals: Technical Analysis, *Journal of Economic Literature,* 45 (4), 936–972.

Mitra S. K., (2012), Is Hurst Exponent Value Useful in Forecasting Financial Time Series, *Asian Social Science,* 8 (8), 111–120.

Moody J., Wu L., (1995), Price Behavior and Hurst Exponents of Tick-By-Tick Interbank Foreign Exchange Rates, *Computational Intelligence for Financial Engineering*, Proceedings of the IEEE/IAFE 1995.

Osler C., (2003), Currency Orders and Exchange Rate Dynamics: An Explanation for the Predictive Success of Technical Analysis, *The Journal of Finance*, 7 (5), 37–58.

Peters E., (1991), *Chaos and Order in the Capital Markets*, John Wiley & Sons Inc.

Peters E., (1994), *Fractal Market Analysis: Applying Chaos Theory to Investment and Economics*, John Wiley & Sons Inc.

Qian B., Rasheed K., (2004), Hurst Exponent and Financial Market Predictability, *Proceedings of the IASTED conference on Financial Engineering and Applications* (FEA '04), 203–209.

Tsay R. S., (2002), *Analysis of Financial Time Series Financial Econometrics*, John Wiley & Sons Inc.

Vengertsev D., (2014), Deep Learning Architecture for Univariate Time Series Forecasting, CS229 Technical Report, 3–7.

Wold H., (1939), A Study in the Analysis of Stationary Time Series, *Journal of the Royal Statistical Society*, 102 (2), 295–298.

Yu H. F., Huang F. L., Lin C. J., (2010), Dual Coordinate Descent Methods for Logistic Regression and Maximum Entropy Models, *Machine Learning Journal,* 85, 1, 41–75.

JAKOŚĆ PROGNOZOWANIA CEN W ZALEŻNOŚCI OD WYKŁADNIKA HURSTA PRZY WYKORZYSTANIU DANYCH WYSOKIEJ CZĘSTOTLIWOŚCI Z RYNKU WALUTOWEGO

Streszczenie

Na przestrzeni ostatniego wieku przeprowadzono wiele badań na temat użyteczności metod statystycznych w prognozowaniu cen na rynkach finansowych. Niniejszy artykuł wyjaśnia, dlaczego większość z nich zawiodła, bazując na teorii rynku fraktalnego oraz na podstawie badań przeprowadzonych przy użyciu da-

nych wysokich częstotliwości z głównych par walutowych. Dla wykorzystanych modeli statystycznych i metod uczenia maszynowego zostały policzone miary takie jak AMAPE oraz trafność prognozowania kierunku zmian cen w zależności od wykładnika Hursta. Artykuł pokazuje, że średni błąd prognozowania zmniejsza się wraz ze wzrostem wartości wykładnika Hursta dla zastosowanych modeli prognostycznych. Zaprezentowana w artykule metodyka prognozowania może być skutecznie wykorzystana do podejmowania trafniejszych decyzji inwestycyjnych oraz do budowy automatycznych systemów decyzyjnych.

Słowa kluczowe: dane wysokich częstotliwości, prognozowanie, uczenie maszynowe, metody statystyczne, mikrostruktura rynku, wykładnik Hursta

THE USE OF THE HURST EXPONENT TO INVESTIGATE THE QUALITY OF FORECASTING METHODS OF ULTRA-HIGH-FREQUENCY DATA OF EXCHANGE RATES

Abstract

Over the last century a variety of methods have been used for forecasting financial time data series with different results. This article explains why most of them failed to provide reasonable results based on fractal theory using one day tick data series from the foreign exchange market. Forecasting AMAPE errors and forecasting accuracy ratios were calculated for statistical and machine learning methods for currency time series which were divided into sub-segments according to Hurst ratio. This research proves that the forecasting error decreases and the forecasting accuracy increases for all of the forecasting methods when the Hurt ratio increases. The approach which was used in the article can be successfully applied to time series forecasting by indicating periods with the optimal values of the Hurst exponent.

Keywords: high-frequency data, forecasting, machine learning, statistical models, microstructure, Hurst exponent