doi: 10.18559/ref.2017.2.5

Research Papers in Economics and Finance

RESEARCH PAPERS IN ECONOMICS AND FINANCE

JOURNAL HOMEPAGE: www.ref.ue.poznan.pl

Analysis of the EUR/USD exchange rate in binary-temporal representation

Michał Dominik Stasiak[1]

[1] Poznan University of Economics and Business, Poland

ABSTRACT

An exchange rate between two currencies can be described in a binary representation. The binarization algorithm transforms the exchange rate represented by tick data into a binary string. Each course change equal to a given discretization unit is assigned a binary value indicating the direction of the change. The performed statistical analysis confirms the existence of a correlation between previous course changes and the probability of future direction of the changes. In order to conduct a more detailed analysis of the exchange rate in a binary representation, each shift in the trajectory can be assigned a parameter representing the duration of the change. Depending on the current market dynamics, course trajectory changes may occur at different moments in time. The main goal of the presented research is to verify the existence of any dependences between the duration of a change and the probability of future direction of the change.

Keywords: foreign exchange market, high frequency econometrics, technical analysis, modeling of currency exchange rates.

1. Introduction

Currency exchange rates are characterized by a high variability in time (they can change even every few seconds). Therefore, their values are usually presented by broker platforms in the form of a candlestick chart. This representation depends on established time intervals, i.e. the time based on which a single candlestick is being constructed. Traditional intervals are usually 1, 5 or 30 minutes, 1 or 4 hours and 1 day [Schlossberg 2006]. Each candle is represented by four values: the maximal rate, minimal rate, opening rate and closing rate. This method of visualizing a course trajectory can also be applied in a technical analysis for appointing particular indicators [Murphy 1999; Schlossberg 2006; Yazdi & Lashkari 2013; Valcu 2004; Neely& Weller 2011]. Applying the candlestick representation, in which the candle parameters are dependent on a predetermined time interval, can lead to some significant interpretation difficulties. This kind of representation can lead to a loss of information about the order and number of changes "inside" the candle. In order to solve this problem, the author proposes a binary representation of the exchange rate course.

In this kind of representation, each exchange rate trajectory change equal to a given discretization unit is being assigned a binary value indicating the direction of the change. This approach allows for eliminating from the analysis intervals characterized by lack of variability, that is, for example, nights. On the other hand, the representation still encompasses key information, i.e. the direction and level of rate changes. The achieved course representation (in the form of a binary string) is dependent only on the assumed discretization unit. Yet, one may ask if the duration of a change equal to a given discretization unit in an exchange rate trajectory has a substantial impact on the probability of further change direction? This article answers the abovementioned question by presenting a new representation method for the course – a binary-temporal representation – in which each price change is described by both a binary variable and duration of a rate change, calculated in seconds. In the following chapters, an attempt to evaluate the prognostic value of time in proposed representation model was made. To realize this goal, time distributions for a given change type were analyzed using proper statistical tests.

The presented article is organized in the following way. In the second chapter we describe the general premises and advantages of a binary representation for a currency pair exchange rate. In the third chapter we propose the assumptions of a binary-temporal representation. The next chapter is dedicated to statistical analysis of the time parameter's impact on the direction of a future rate change. The last chapter presents the results of performed research.

2. Binary representation of an exchange rate

As was already mentioned before, currency exchange rates are characterized by high variability in time. It is common knowledge that significant rate changes can happen in an interval of a few or a dozen seconds. At the same time, the rate course can alternate in a highly dynamic way - with changes of a high amplitude – because of, for example, presenting significant macroeconomic information (central banks' decisions regarding the level of interest rates, price indicators, etc.). In some instances, the exchange rate shifts over a few minutes can be more dramatic than changes registered during some long periods of hours or even days. Moreover, some changes are characterized by different dynamics during the day (that is, when stock markets, banks and other financial institutions are actually open and macroeconomic date is being presented) and during the night [Cheung & Chinn 1999; Murphy 1999; Schlossberg 2006; Oberlechner 2005]. This kind of situation leads to high complexity of the models encompassing the time parameter in order to describe the exchange rate course. The second problematic issue lies in the volume of the data. Most of the broker platforms present the exchange rate in the form of a candlestick chart for a given time interval (typically within the range of 1 minute to 1 year). The processed data is then used to calculate the value of market indicators.

Course discretization was used for the first time in the early years of the 20th century in order to create and analyze data charts in the so-called Point and Figure method [De Villiers, 1933]. Even though this method is far more precise than others, it was soon replaced by the candlestick chart analysis and was never further developed. In works of [Stasiak 2016], a binary representation which allowed for the conversion of an exchange rate trajectory into a binary string was used. In this approach, the sole parameter is the discretization unit that describes the length of the trajectory change for which the change will actually be noted.

The discretization algorithm works in the following way: the algorithm assigns an upper and lower change limit for the initial exchange rate value. The limit is equal to a positive or negative value shift of one discretization unit. When the exchange rate falls below the lower change limit, the algorithm assigns a binary value of (0) to the respective change, and in a contrary situation, i.e. in the case of exchange rate increasing above the upper change limit, it assigns it the binary value (1). In the next steps, the algorithm calculates further limit values based on the current exchange rate course. As a result of this procedure, the exchange rate course trajectory is presented as a binary string (e.g. 10101……01111). In case of encountering any data gaps for prices, e.g. after weekends, the algorithm checks which value would have been reached and assigns a respective result. The general idea of the discretization algorithm's performance is presented in Figure 1.

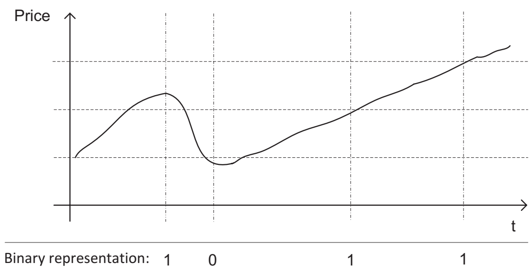

Figure 1: Binarization algorithm results for the considered exchange rate

2.1. Binary representation characteristics

Depending on the established value of the discretization unit, the achieved trajectory of exchange rate changes may have a different character. When appointing the discretization unit, it is crucial to include the characteristics and frequency of the changes for a considered currency pair as well as the spread offered by the brokers, etc. Choosing a high discretization unit can lead to such consequences as ‚losing' a part of significant information. In the case of using small (only a few pips) discretization units, the representation will register some random fluctuations of the so-called ‚noise'. The problem of the level of changes qualified as a noise was widely analyzed in existing literature, e.g. [Lo, Mamaysky & Wang 2000; Logue& Sweeney 1977, Menkhoff & Taylor 2007; Neely & Weller 2011]. Yet, it is difficult to indicate any specific recommendations for a binary analysis of a given currency pair. The minimal discretization unit can be described as optimal if it eliminates the influence of the noise.

In order to verify the achieved binary strings, a statistical analysis aiming to exclude randomness was performed. In the case of any random changes in the exchange rate trajectory represented by the binary method, a further analysis would have been futile. In the performed research, four different statistical tests were used. Their general function is to verify the pseudo-random number generators [Soto 1999, Weigl & Anheier 2003] and they include a frequency test, runs test, non-overlapping template matching test and long run test. Because of the differences in notation and actual algorithms encountered in existing literature, in this article we use algorithms from SP800-22 packet [Rukhin et al. 2010]. The tests find if the analyzed data has a random character, i.e. if there are any occurrences of ensuing changes not correlated with previous trajectory. The simultaneous confirmation of the randomness hypothesis by all those tests allows to qualify the given data as random. In the presented research, this was confirmed for the cases of both high and low discretization units, that is, for the research performed for discretization units equal to 20, 25 and 30 pips respectively. Considering units smaller than 20 pips is not effective because of the level of spreads offered by brokers and because of random fluctuations. On the other hand, appointing discretization units higher than 35 pips can cause registering any small changes and thus, a decrease in the quality of information about the course trajectory. Only the first test (Frequency Test) confirmed the randomness hypothesis for the analyzed binary string. This result can be derived from the fact that in the long periods the summed numbers of zeros and ones in the binary strings were similar [Rukhin et al. 2010; Menezes, Oorschot & Vanstone 1996]. The three remaining tests reject the hypothesis of randomness. This means that there exists a possibility of some regularities connected with different frequencies in which zero-one sequences occur. To summarize, one can state that the statistical test results indicate the possibility of some correlation between the historical trajectory and the direction of future changes [Godbole & Papastavridis 1994; Chung, 2012; Rukhin et al. 2010; Menezes, Oorschot & Vanstone 1996, Weigl& Anheier 2003; Soto 1999].

In the presented research the tick historical data from Ducascopy was used, regarding the period of 5 years (01.01.2010-01.01.2015), for a EUR/USD currency pair. Relatively recently, the access to the currency market was restrained for professional investors. As a result of the development in telecommunication and informatics, in recent years the number of market participators highly increased (development of broker platforms, decrease of required deposit levels). Technological changes allowed for faster transactions, where time is actually measured in microseconds. Moreover, the implementation possibilities of HFT systems (which are becoming more and more popular) making hundreds or thousands of transactions, also have an influence on the market performance and characteristics. Taking into account all of the above premises, considering a five-year period for the analysis seems to be sensible. When using the statistical tests, the level of significance recommended by NIST was used, that is 0.05 [Ruthin et al. 2010]. The research results of the performed analysis suggest that the binary representation can be applied in constructing prediction models, yet, it lacks some information connected with the duration of a single exchange rate change. This raises the question if the knowledge about the duration time increases the prediction capabilities of the binary representation analysis.

2.2. A comparison of Binary and Candle representations of an exchange rate

Let us now consider exchange rate trajectory modelling in order to construct HFT systems. In HFT systems, the algorithm automatically performs hundreds of transactions of a few or a few dozen pips range. Investor's revenue is assured by a statistical majority of successful transactions. In the case of modelling such small changes, candlestick representation leads to a loss of information about the changes "inside" the candle, and in consequence, to unreliable modelling results. Even when using one-minute candlesticks (i.e. the most precise of candlestick charts offered by broker platforms), an investor still loses a lot of important information about respective changes occurring in the periods of high investors' activity shorter than one minute.

In Figure 2, a comparison between time- and binary representation of exchange rate changes is presented. Two exchange rate trajectories in a given time period are shown (Figure 2a and Figure 2b). Both of them can be represented by a candle of the same parameters. The example clearly depicts that, despite the same candle representation, the process of exchange rate trajectory changes can have an entirely different scenario and be actually characterized by different binary strings (101 in Figure 1a and 1 in Figure 1b). Moreover, the abovementioned example shows another disadvantage of the candlestick representation. Let us assume that the HFT system algorithm makes a sale transaction in which parameters SL and TP are placed one discretization unit above and below the price. In the case presented in Figure 2a the investor achieves a profit; however, in the case presented in Figure 2b they will register a loss. It is impossible to differentiate the situations based only on historical candlestick charts.

Binary representation can be used in a simple way to create HFT systems. Let us consider a sale transaction with a TP parameter smaller than the current price by a discretization unit and with an SL parameter higher than the current price by the same discretization unit. The probability of a profit is equal to the probability of the occurrence of the value 0 in the binary representation, and the loss probability is equal to the probability of encountering a 1. An analogous situation takes place in the case of the purchase transaction.

To summarise, binary representation allows for more accurate modelling of the course trajectory as compared to candlestick representation and thus can be used in constructing HFT systems.

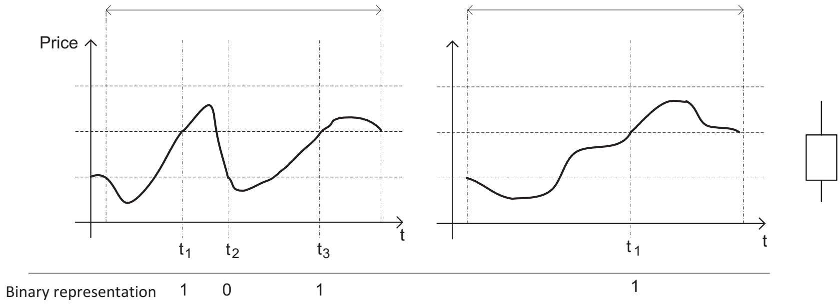

Figure 2: Two different trajectories of course changes given by the same candle but two different binary representations

3. Binary-temporal Representation

As indicated in the research results described in the previous chapter, in order to create a valid prediction model, binary representation can be applied to the data. In the case of small discretization units (a few or a few dozen pips), the duration of the change is not significant from the investor's (or HFT automaton's) point of view. On the other hand, the parameter can have an additional information value when used in the prediction of future exchange rate changes. For example, let us consider a discretization unit of 20 pips. If an increase in the exchange rate occurred in a short time, let us say, 5 seconds, then the probability of the further increase is higher than in the case of the same change lasting, for example, two hours.

In order to perceive the duration time of a change as a parameter of binary representation, in the presented article we propose binary-temporal representation. The main idea of this method's algorithm can be described as follows. The procedure indicates the upper and lower change limit for an initial exchange rate value. The limits are equal to the positive and negative shift in the exchange rate with a length of one discretization unit. The algorithm also registers the initial time of the change occurrence. If the exchange rate falls below the lower limit, the algorithm assigns two values to the registered change – a binary value of (0) and the duration of the change given in seconds (t). In the case of an increase in the price being higher than the upper limit, the algorithm assigns the binary value of (1), along with the change duration time (t). In the next steps, the algorithm indicates next limits calculated using the current exchange rate and the time when the previous course change ended. As a result of the algorithm operation, the exchange rate trajectory is presented by a string (e.g. 1(5), 0(450), 1(100), 0(345)). Figure 3 presents exemplary algorithm results for creating the binary-temporal exchange rate representation.

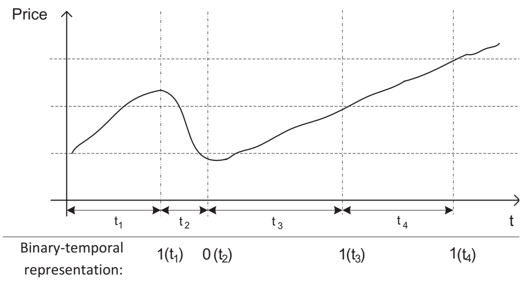

Figure 3: Binary-temporal exchange rate representation algorithm results

4. Statistical analysis of a EUR/USD exchange rate in binary-temporal representation

In order to confirm the possibility of using time as a parameter of binary representation, a statistical analysis was performed for the change duration distribution times, calculated for the given discretization unit of a currency pair. Because of the difficulties in indicating potential correlations between the change duration and the direction of future changes, an analysis was performed in order to verify if the registered time distribution has a memory-loss character (that is, if there are no influences of previous events on the current event). The registered time distribution was then compared with an exponential distribution, which is characterized by a memory-loss. The main idea was to research the similarity exchange rate trajectory change duration time distribution to the exponential distribution. In order to do so, the Kołmogorow-Smirnow Test and Anderson--Darling were conducted.

4.1. Exponential distribution

Exponential distribution can be described in terms of its density function [Bobrowski 1986]:

$$f(t) = \begin{cases} 0 & for \quad t < 0 \\ \lambda e^{-\lambda t} & for \quad t \geq 0 \end{cases}, \tag{1}$$

where:
f(t) – a density function
λ – the intensity of events
The cumulative distribution function for exponential distribution is given by the s formula:

$$F(x) = 1 - e^{-\lambda t}, \qquad \text{for } t \geq 0 \tag{2}$$

By using the cumulative distribution of time between occurring events we can determine the unconditional probability value $P(T>t)$, i.e. the probability that the time between given events is longer than T:

$$P(T > t) = 1 - F_0(t) = e^{-\lambda t}. \tag{3}$$

Simultaneously, we can calculate the probability values P(T>t+τ), i.e. the probability of the time between events being at least t+τ long:

$$P(T > t) = e^{-\lambda(t+\tau)} \tag{4}$$

The above stated probability can be rewritten as:

$$P(T > t + \tau) = P(T > \tau | \mathrm{T} > t)\mathrm{P}(\mathrm{T} > t), \tag{5}$$

where $P(T>\tau \mid T>t)$ is the conditional probability stating that the time between events will last for at least another period of τ. Based on (3)-(5) we can conclude that:

$$P(T > \tau | \mathrm{T} > t) = \mathrm{P}(T > \tau). \tag{6}$$

The conditional P(T>τ│T>t) and unconditional P(T>τ) probabilities are identical, which means that knowledge about the past does not have any influence over the variable characterized by exponential distribution [Stasiak et al. 2010]. In consequence, changes characterized by exponential distribution cannot be used in the analysis aiming to predict future events.

4.2. Distribution testing

In order to verify the similarity of the variable distribution to a given theoretical distribution we can use the statistical tests. The Kołmogorow-Smirnow Test (often called the Kołmogorow Test) verifies if the registered distribution of a variable differs from a given theoretical distribution, when only a finite number of variable observations is known to the researcher (i.e. statistical sample). The test compares the cumulative distributions of the theoretical distribution (F0) and the empirical distribution (Fn) obtained from the analyzed data [Bobrowski & Łybacka 2006; Gibbons & Chakraborti 2011;Górecki 2011]:

$$D = \sup_{x \epsilon R} |F_n(x) - F_0(x)|. \tag{7}$$

where D is the test's statistic.

The verification of the coincidence between the empirical and theoretical distribution can be also performed with the use of the Anderson-Darling Test [Anderson & Darling, 1952]. This test is characterized by stronger statistical power than the Kolmogorov-Smirnov test [Razali & Wah, 2011], yet it also compares the differences between the empirical and theoretical cumulative distribution. Statistic A of this test can be calculated based on the following formula:

$$A^2 = n \int_{-\infty}^{+\infty} \left[\frac{F_n(x) - F_0(x)}{F_0(x) - F_0(x)^2}\right] dF_0(x). \tag{8}$$

As a result of the applied tests, we can confirm or reject with a given statistical significance, the hypothesis of coincidence between the theoretical and empirical distribution.

4.3. Time parameter analysis of a EUR/USD exchange rate in binary-temporal representation

The distribution of change duration times was researched for a 5-year period (0.1.01.2010-01.01.2015) of tick data observations for EUR/USD exchange rates. The tick historical data was sampled from the Ducascopy broker. Based on the MQL4 software written by the author in order to perform described research, binary-temporal exchange rate representation using a 20-pips discretization unit was achieved for a given period. Taking into consideration such parameters as noise, the spread value and change dynamics, the author deemed the 20-pips discretization unit as ‚optimal' for application in HFT systems (that is, the value allows the representation not to include too much of random noise). Next, by using a statistical packet R with the implemented Kołmogorow-Smirnow and Anderson-Darling Test [Górecki 2011], the hypothesis of the coincidence between the duration distribution of falls or increases of an exchange rate with a theoretical (exponential) distribution was verified. The tests results confirm the lack of coincidence with the exponential distribution for the established 0.05 significance level. In Picture 3 cumulative empirical distribution achieved from the researched data and theoretical cumulative distribution of the exponential distribution for the duration of increases (Picture 3a) and the duration of falls (Picture 3b) are presented.

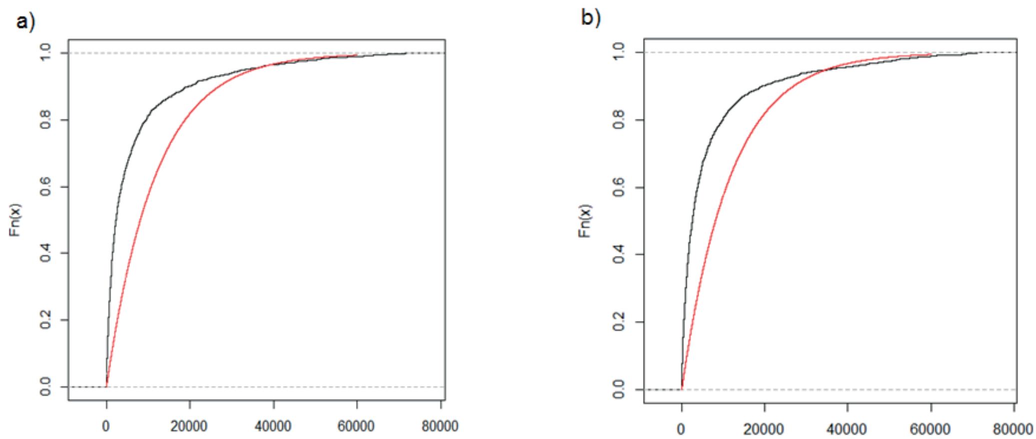

Figure 4: Theoretical cumulative distribution (black) and empirical cumulative distribution (red) of time distribution for a) falls and b) increases

The majority of changes lasted about 5000 seconds (the median is 2548 seconds for falls and 2277 seconds for increases). Thus, in further research it can be beneficial to focus on the classification of the changes in a few-minute-long intervals. An unexpectedly long change time, for example two hours, can be treated as an anomaly, which also has high forecast significance.

The statistical analysis mentioned above was performed for higher discretization units as well, that is, respectively 25, 30 and 35 pips. In all cases the results were analogous (the hypothesis of the empirical distribution being coincidental with the exponential distribution was rejected on the significance level of 0.05).

5. Summary

In the paper, a new binary-temporal representation of an exchange rate trajectory was presented. This kind of representation is an alternative to course trajectory visualization as compared to candlestick charts. Its main advantage lies in encompassing more information about the course variability. Binary-temporal representation can be used in creating HFT systems (the probability of the future direction of a change is equal to the probability of an investor's profit/loss). The EUR/USD analysis performed by statistical tests and presented in the article suggests relations between the direction and order of historical data and the direction of a future change. A statistical analysis of dependences between the duration of the previous changes and the direction of a future change was also performed by the means of the Kołmogorow-Smirnow and Anderson-Darling Test. The research results confirmed the existence of such relationships.

Research was conducted for different discretization units and based on the historical data of the EUR/USD course. All tests used a 0.05 statistical significance level and were applied to a few-thousand-element data set. Research results confirm the possibility of applying the binary-temporal representation to precise modelling of an exchange rate trajectory and for constructing prediction algorithms for HFT algorithms.

References

Anderson, T. W., Darling, D. A. 1952, *Asymptotic theory of certain 'goodness of fit' criteria based on stochastic processes*, Ann. Math. Stat. 23, pp. 193–212.

Bobrowski, D., 1986, *Probabilistyka w zastosowaniach technicznych*, Wydawnictwa Naukowo-Techniczne.

Bobrowski, D., Łybacka,, K., 2006, *Wybrane metody wnioskowania statystycznego*, Wydawnictwo Politechniki Poznańskiej.

Cheung, Y. W., Chinn, M. D., 1999, *Macroeconomic implications of the beliefs and behavior of foreign exchange traders*, No. w7417, National Bureau of Economic Research.

Chung, K. L., 2012. *Elementary probability theory with stochastic processes*. Springer.

De Villiers, V., 1933, *The Point and Figure Method of Anticipating Stock Price Movements: Complete Theory and Practice*, Stock Market Publications.

Gibbons, J. D., Chakraborti, S., 2011, *Nonparametric statistical inference* Springer Berlin Heidelberg, pp. 977-979.

Godbole, A. P., Papastavridis, S. G., 1994, *Runs and patterns in probability: Selected papers*, Vol. 283, Springer Science & Business Media.

Górecki, T., 2011, *Podstawy statystyki z przykładami* in: R. Wydawnictwo BTC.

Lo, A. W., Mamaysky, H., Wang, J., 2000, *Foundations of technical analysis: Computational algorithms, statistical inference, and empirical implementation*. The Journal of Finance, Vol. 55(4), pp. 1705-1770.

Logue, D. E., Sweeney, R. J., 1977, *White noise in imperfect markets: the case of the franc/dollar exchange rates*, The Journal of Finance, Vol. 32(3), pp. 761-768.

Menezes, A. J., Van Oorschot, P. C., Vanstone, S. A., 1996, *Handbook of applied cryptography*. CRC press.

Menkhoff, L., Taylor, M. P., 2007, *The obstinate passion of foreign exchange professionals: technical analysis*, Journal of Economic Literature, pp. 936-972.

Murphy, J. J., 1999, *Analiza techniczna rynków finansowych*. WIG-Press, Warszawa.

Neely, C. J., Weller, P. A., 2011, *Technical analysis in the foreign exchange market*, Federal Reserve Bank of St. Louis Working Paper.

Oberlechner, T., 2005, *The psychology of the foreign exchange market*, Vol. 506, John Wiley & Sons.

Razali, N. M., Wah, Y. B., 2011, *Power comparisons of shapiro-wilk, kolmogorov-smirnov, lilliefors and anderson-darling tests*, Journal of statistical modeling and analytics, Vol. 2(1), pp. 21-33.

Rukhin, A., Soto, J., Nechvatal, J., Barker, E., Leigh, S., Levenson, M., Smid, M., 2010, *Statistical test suite for random and pseudorandom number generators for cryptographic applications*, NIST special publication.

Schlossberg, B., 2006, *Technical Analysis of the Currency Market*. John Wiley & Sons.

Soto, J. ,1999, *Statistical testing of random number generators*, In Proceedings of the 22nd National Information Systems Security Conference, Vol. 10, No. 99, p. 12.

Stasiak, M., Glabowski, M., Wisniewski, A., Zwierzykowski, P., 2010, *Modelling and dimensioning of mobile wireless networks: from GSM to LTE*. John Wiley & Sons.

Stasiak, M. D., 2016, *Modelling of Currency Exchange Rates Using a Binary Representation*, Information Systems Architecture and Technology: Proceedings of 37th International Conference on Information Systems Architecture and Technology – ISAT 2016, Springer.

Valcu, D., 2004. *Using The Heikin-Ashi Technique*. Technical Analysis of Stock and Commodities Magazine, Vol. 22(2), pp. 16-29.

Weigl, A., Anheier, W., 2003, *Hardware comparison of seven random number generator tests for smart cards*, Testmethoden und Zuverlässigkeit von Schaltungen und Systemen, p. 55.

Yazdi, S. H. M., Lashkari, Z. H., 2013, *Technical analysis of Forex by MACD Indicator*, International Journal of Humanities and Management Sciences (IJHMS), Vol. 1(2), pp. 159-165.