Real vs Back Testing Data on MetaTrader: Analysis of Impact for Scalper Strategies

Lucas Freitas, Augusto Lima, Maria Lívia Coutinho, Leandro Coelho, Diego Frias

Curso de Bacharelado em Sistemas de Informação – Universidade do Estado da Bahia (UNEB) – Campus de Salvador
41.150-000 – Salvador – BA– Brasil

{lfreitas,alima,marialivia,leandrocoelho,diegofrias}@uneb.br

Abstract. *With the advent of small trades over Foreign Exchange market, operating through platforms such as MetaTrader® and the possibility of using automatic strategies, a lot of traders have bet on this model to profit easily. Although MetaTrader (MT) allows simple and easy forms for building and testing an automatic strategy, several simulations indicate positive results that are not possible to replicate on real world data. In this paper we provide an analysis of back testing automatic strategies using MT Strategy Tester. We use two data sources for testing automatic strategies: real data from broker's server and MetaTrader's historical center from the same period. The results show the risk of back testing using MT simulation tool for scalper strategies.*

1. Introduction

The Foreign Exchange market (FOREX) is considered the major and most intricate financial market in the world and has been growing in the past few years achieving 3.98 trillion US dollars daily turnover [Alrefaie 2013].

The FOREX, is the market in which traders are able to buy, sell, exchange and speculate on different currencies. Foreign exchange markets are composed by a mix of investors such as banks, commercial companies, hedge funds, and retail FOREX brokers [Alrefaie 2013]. Most users interact with the Market using platforms supplied by the brokers, that allows trading at home, just having stable internet connection [Sim 2010]. Actually small traders with low balance can still trade in such market through broker companies that provides all support (capital and applications) to buy or sell orders and manage their money [Montgomery 2015]. But how can broker companies allow small traders to deal over FOREX?

The answer for this question is quite simple! The brokers use margin control on a leverage system that allows small traders to operate in the market. Meanwhile the leverage system is extremely efficient; it can lead traders to have considerable loss if their strategy is not well applied. Thus, an accurate testing is needed to avoid wrong analysis or even biased results.

Many programs of technical analysis were developed to allow testing of trading strategies based on real history data. For example, MetaTrader 4 (MT4), developed by MetaQuotes Software Corp., includes a strategy tester (back testing tool) interface that allows simulations over broker's historical data. Back testing can be performed using data constructed with three different models: 1) Every Tick; 2) Control points; 3) Open prices. Depending on the strategy used, the divergence of results can be drastic when

choosing different data generation models. Although the first model generates data with the highest resolution in time (smaller time scale) it is the less accurate. It is due to a fractal interpolation method used to generate random ticks values expected to approach the non-recorded real data variation within a minute.

To the best of our knowledge, we found no references issuing the addressed problem. [Galati 2000] provides a good overview of the volume and spread on such market, and, although lagged in time, it has enough current concepts. [Sim 2010] studies the problems found in technical indicator computations but he does not discuss the strategy tester of MT4. [Pinto 2012] presents an Expert Analyzer (EA) for automatic trading on FOREX market based on a trading strategy. He uses the MetaTrader 5 to test and optimize the strategy, but does not refer to the quality of the simulation, nor to the performance of the optimized EA in actual operations, as we addressed in this work.

In this paper, we revisit this problem by using a scalper EA based on a simple static stop-loss and take profit strategy for stopping operations and a regression trend predictor for determining when opening a promissory buy or selling operation. As scalpers works in very short time intervals, targeting also small profits, most operations often open and close in a few seconds. In these conditions, the EA need to be tested using data modeled by the Every Tick model. The motivation of this study was the negative performance of the EA in the market, in spite of being optimized using the MT4 Back testing framework over the last 6 months. While the profit increased regularly with an attractive rate when using the data generated by the MT4 Strategy Tester, in real life the loss increased drastically. Reviewing the literature we found several articles of EA programmers and trade experts noticing the same fact with different strategies, but most identify several procedural and conceptual human errors, as being the major cause of back testing-optimized EAs to fail in real market conditions [Daniel, 2011, Rolf and Moritz, 2015, Desroches, V. 2015, Balch, T. 2015, Greenberg, M. 2010]. However, none of these studies reported as probably cause some limitation of the strategy tester. Therefore, we decided to collect data in real time from the MT4 server to compare with the data generated by the MT4 Strategy Tester. We also developed a platform to perform back testing with the recorded data that is not possible with the MT4 platform.

2. Description of the MT4 Strategy Tester tool

There are many different types of trading strategies for the FOREX market and there are different ways to test those strategies [John 2007] [Li 2010]. Some strategies are based on opening and closing values for a bar of given time; others are based on values provided through tick by tick and others by high/low points of past bars and so on. Regardless of the strategy used, the most popular retail trading interface is the MetaTrader4 (MT4) [Korczak 2013]. In this case the MT4 offers a strategy tester module for performing EAs evaluations. The problem with such module is that it has characteristics that are not clearly described for dealers.

Most platforms, including MT4, allow analysis on different time frames, somehow standardized.The lowest time frame is M1, which comprises every minute data needed to draw a candle in the chart, that is: Open, Close, High and Low prices within the minute, and optionally the volume of transactions. A candle, is a bar extended

from Low to High prices with a filled rectangle between Open and Close prices. The filling color changes accordingly the price decreased (black) or increased (white) on the time interval.

The same occurs for larger time frames which are constructed combining the candles of the nearest lower time frame. Figure 1 shows the time frame candles of 30 minutes collected by the MT4 historical data center from 01:00 pm. until 01:30 pm. of January, 15th 2016. As it can be observed there is an open price (1.29315), high price (1.29350), low price (1.29290) and close price (1.29340) for the 30 minutes bar (figure 1a). Figures 1b, 1c, and 1d shows the candles for lower time frames over the same time interval.

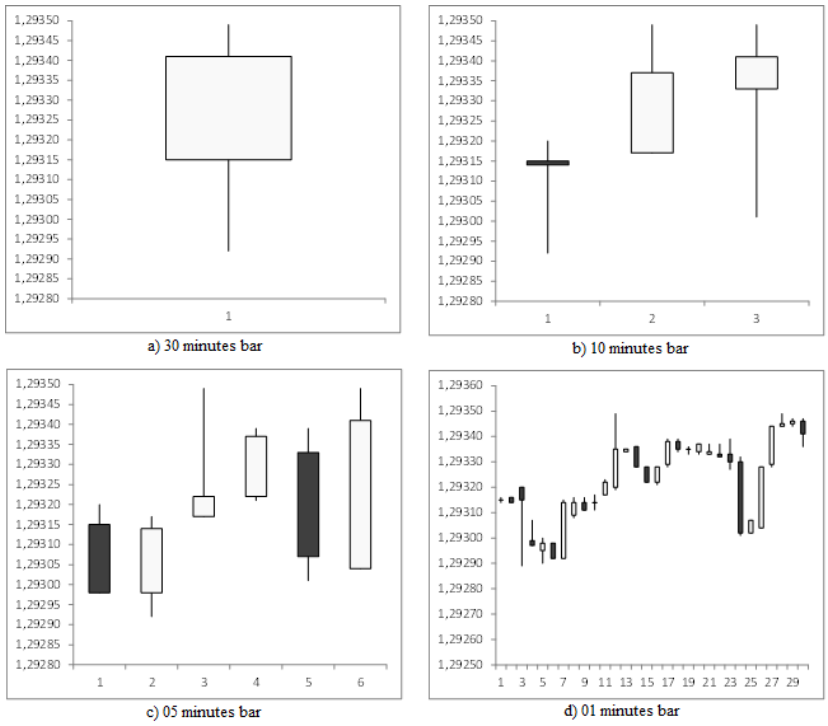


Figure 1. Time frame candles: The same data represented by different timeframes.

The MT4 strategy tester uses a database known as historical data center and allows traders to choose time intervals for testing their strategies. However, the backtest data can be generated using three different models for processing database data. Depending on which model is utilized the results for the same strategy can be extremely different from each other. Next we describe the three simulation models offered by MT4 platform:

a) Open prices: This model uses only the open bar price value for perform ticks over the market, that is, sends a single price every bar. This way allows fast simulations because the system avoids additional market price values. This model cannot be used for strategies that explicitly use bar sizes, maximum or minimum values, or differences between opening and closing prices.

b) Control Point: This model estimates representative prices within a bar. This method demands historical data of the nearest lower time scale. For this reason it is only recommended for strategies working on time-frames larger than 1 minute.

c) Every Tick: Is the model with the highest time resolution sending a price at several ticks within a 1 minute bar. However, as the real tick number is not recorded in the database, ticks are generated according with a probability distribution depending on the volume of transactions and other candle values. The prices at such ticks are also generated using a fractal model that satisfies four control values: open, close, high and low prices. The main problem of this model is the uncertainty about the inaccuracy of such within bar estimates. It is worth noting that imprecision at this time scale mainly impacts scalper strategies.

2.1. Every Tick data generation model

In MT4 Strategy Tester framework when Every Tick model is chosen for data generation, a fractal interpolation is used to generate tick data subject to some constraints regarding the saved data for each 1 minute candle: Open, Close, High, Low and Volume. This is done for any chosen time frame.

In order to compare real-time data with back testing data generated by MT4 Strategy Tester using the Every Tick model, real time tick data was collected from broker's server during ~69.5 hours (from October 21, 2015 at 00:21 until October 23, 2015 at 20:50). This was done using a properly designed robot running on the MT4 platform with time frame M1. Posteriorly, we used the same robot, but running in tester mode on the same platform, over the same time interval with the same time frame, to collect generated back testing tick data. We should note that in tester mode the spread must be fixed. The spread is the difference between the buy (also called Bid) price and the sell (also called Ask) price. Two prices are given for a currency pair. The spread represents the difference between what the broker gives to buy from a trader, and what the broker takes to sell to a trader. The spread we set on tester mode was the average spread calculated previously from the saved real time data.

Comparing tick by tick Bid and Ask prices between the two data sets over the entire period we calculated the integral error. Also the average tick number by candle was calculated for each data set. Several qualitative analyses were also done for selecting time intervals for illustrating our conclusions.

3. Expert Advisor

To perform the study we developed an EA based on the scalping philosophy. The Scalper strategy is based on performing various orders in a very short time, taking advantage of the small amplitude but high frequency oscillation of the market, achieving small profits with each order. The EA implements a predictive method based on spline regression over a selected number of ticks before the current tick in order to predict the next price change. A free end condition was applied in the last tick to allow a more consistent extrapolation to the future, to estimate an expected future value at some prefixed instant ahead. When the predicted future value differs from the current value in

some prefixed amount in pips an operation is initiated by the EA, that is an order of buy or sell is opened for increasing or decreasing trend, respectively. A pip is the smallest price change that a given exchange rate can make, also called basis point. Most major currency pairs are priced to four decimal places, however, the EUR/USD currency pair, chosen for this study, is priced to five decimal places. The amount in pips to open an operation is calculated as a function of the desired profit, the current spread and the commission of the broker. Once an order is opened the EA controls the profit or loss, closing the order when the profit reaches the desired profit (the order price equals the take profit price - TP) or the loss reaches the specified accepted loss (the order price equals the stop loss price - SL).

To exemplify just suppose a buy order opened at time t_0 and an open Bid price value O_p. Let SL_p be the stop loss gap and TP_p the take profit gap, both in pips. Let also MT_o, Spr_p and Com_p be the maximum time for each order, the current spread in pips and the broker's commission in pips, respectively. Denoting by C_p the current Bid price the EA monitors at every tick the departure from the opening condition, calculating $H_p = (CP_v - O_p)*1e+5$. If $H_p >= TP_p + Spr_p + Com_p$, or even if $H_p <= -SL_p$ the order is closed and the account balance is updated with the net profit or loss. If successful the net profit of the order is calculated as $P_p = H_p – (Spr_p + Com_p) > 0$, and in the opposite case the order loss is calculated as $Lp = H_p - (Spr_p + Com_p) <0$.

4. The new platform EMtrader

The platform EMTrader was developed in the project and consists of: (a) a database that stores the date and time of each transaction, the buy (Bid) and sell (Ask) values, and the volume of the transactions every minute, (b) a robot that runs on MT4 platform and collects data in real time on the broker's server by saving them in the database, (c) a library that contains a C version of all the native functions of MQL4, used in robots developed by the group, and others that simulate some basic operations of the broker server. For example, this library includes functions for performing simulated buy and sale transactions, to monitor the price in order to close the operations when the condition of Stop Loss or Take Profit occurs, for the calculation of the balance of each operation and the total balance, among others, (d) an environment to edit, compile and run robots that read values from the database, which allows us to test the robots using real data.

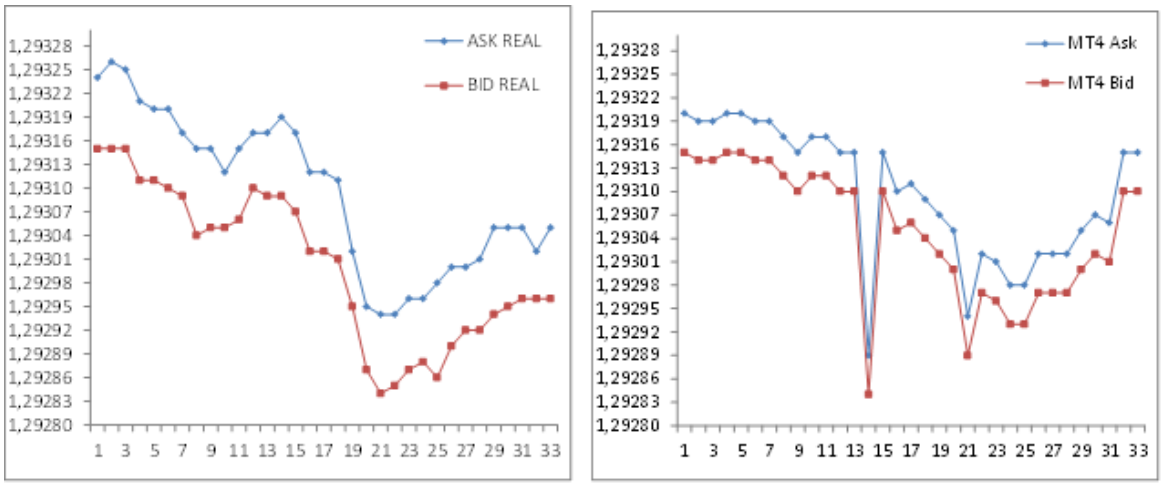
Each platform component was tested separately, concentrating on the computational accuracy and stability using a representative set of entries according to the type of tasks. The expected results for each entry were obtained by performing a similar robot in MetaTrader saving the input and corresponding output to a file. Integration testing was subsequently performed for checking the stability of communication between platform components.

5. Experimental Results

5.1 Comparison of data sets.

Taking into account that the spread for generated data in MT4 back testing tool is a constant to be chosen, we calculated the average spread of the real data over the interval used for comparison (69.5 hours). The average spread was ~3.9 pips, having peaks up to

40 pips. The spread in pips for the EUR/USD currency pair was calculated as the difference between the Ask and Bid prices multiplied by 1e+5. Doing this, we noticed that: (1) the fractal model generates only the Bid price and the Ask price is calculated by adding the prefixed spread value. Figure 2, shows the real time (left) and generated tick data (right) during a selected minute. Notice the similarity of the initial (opening) Bid (red) prices (~1.29315) and the dissimilarity of the Ask (blue) prices, (2) The fractal model satisfies the Bid opening, high and low prices but not the closing price, as can be also observed on Figure 2.



a) Real data from market

b) Historical data from MT4

Figure 2. Every Tick data for a 1 minute candle comprising 33 ticks in both real case and back testing, for comparison: (a) Real data; (b) MT4 back testing data.

Another significant difference between the two time series compared was the number of ticks. The amount of ticks on test data is greater than the data in real time. During the study period, there was 27.9% more ticks on test data than the data in real time.

This tick number difference prevented the point-wise comparison of the two time series, requiring the use of another metric. We choose the average integral error by minute expressed in pips, between the two Bid time series. To do this we calculated the area below both Bid curves by numerical integration, that is, by summation of the products of the mean Bid price between two ticks given in pips (multiplied by 1e5), times the difference in time between the two ticks given in minutes (divided by 60). Then, we calculated the difference between the two areas and divided the result by the number of minutes in the interval (5670).

We calculated an integral error (in pips per minute), relatively small, between the two Ask prices of 14 pips per minute, but a significantly higher error for the Bid prices of 84 pips per minute.

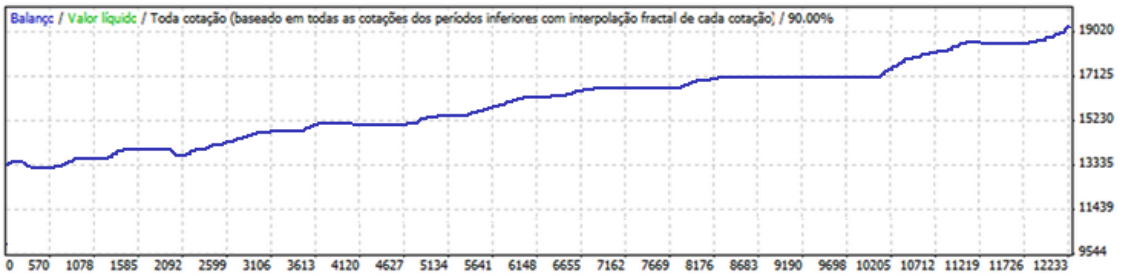
5.2 Comparison of EA performance.

We run the EA described in section 3 in the EMTrader platform with the parameters shown on Table 1.

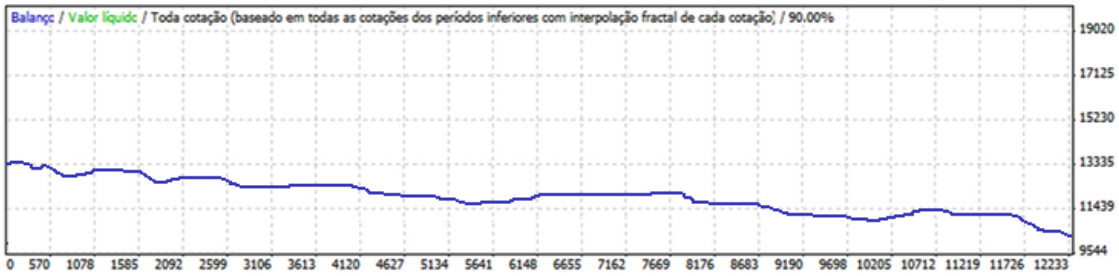
Table 1. EA parameters sets

Parameter	Value
Take Profit	7 pips
Stop Loss	15 pips
Period	From November 21th until November 23th 2015.

The EA was executed both over real data and MT4 back testing data, but when run with real data, the prediction function was replaced by a function that read from a file the sequence of times and opened orders generated while run with back testing data. This way we ensure that the EA performed the same operations with real data than that executed with back testing data. In both cases the EA executed 12282 operations, 43% buy and 57% sell. This yields a scalper frequency of 2.16 operations by minute. Figure 3 shows the simulation performance with each data set. The results for MT4 back testing data (Figure 3a) shows a profit of approximately 47% of the initial deposit (USD 13.000) meanwhile the results for real data (Figure 3b) shows a loss of about 21%, yielding a performance discrepancy of 68%.



a) MT4 Historical data Simulation



b)Real data Simulation

Figure 3. Simulations results: (a) Real data result (b), Mt4 Historical center.

6. Discussion

The results suggest that the poor approximation of real data by the fractal model in the Every Tick data generation mode of the MT4 Strategy Tester is the main cause of the performance difference of scalper strategies when executed in the back test mode and real time conditions. Such poor approximation has three main components, organized according to the impact that we assume each has: (1) constant vs variable spread, (2) large local price differences and (3) differences in the number of ticks per unit of time.

In order to illustrate the impact over de trading strategy performance of the second component we isolated an operation that occurs in the time interval shown on Figure 2, more specifically on the 14th tick. Figure 4 shows a Buy order occurring with real data (left) and back testing data (right). Notice that in the real data case, the Bid price decreases 26 pips after tick 14 which exceeds the stop loss value of 15 pips, causing a loss of 15 pips plus the spread of about 7 pips. Meanwhile in the back test case, the Bid price increases after tick 14, reaching up to 32 pips at tick 15. As the take profit value was set at 7 pips, the order closes at tick 15 with profit of 7 pips minus 5 pips of fixed spread, yielding a net profit of 2 pips.

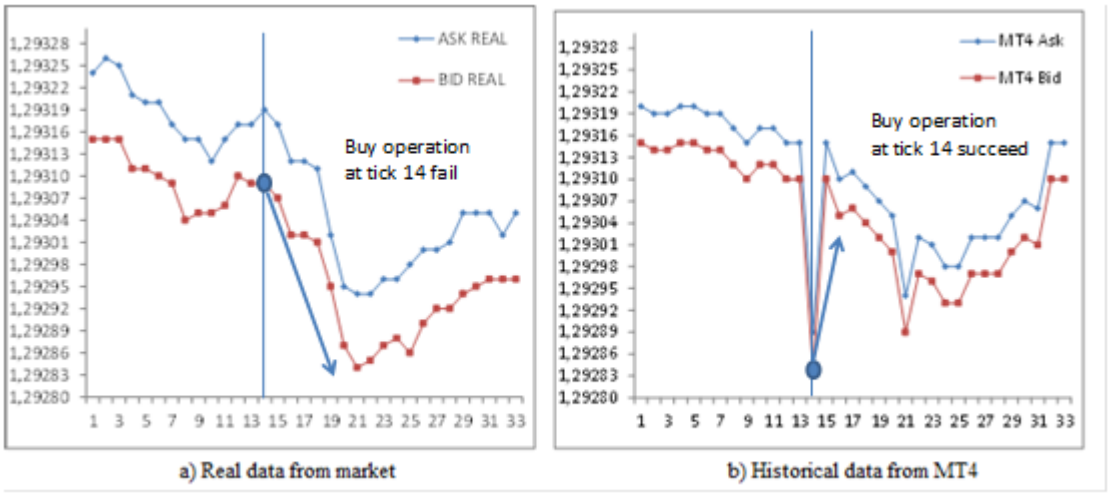


Figure 4. Simulations results - opening a buy order at tick 14: (a) Real data cause loss while (b), Mt4 Historical center data yields a fake profit.

7. Conclusions

We report the investigation of the causes for the usual difference in performance of scalper expert analyzers (AEs) in the FOREX market, when they are executed with back test data generated by the MT4 Strategy Tester tool with the Every Tick model and with data collected from real-time market using the same MT4 trading platform.

More specifically, our goal was to provide a research-based answer to the frequently asked question "Why Back Test with MT4 Strategy Tester is not working." While developers of trading platforms and brokers that distribute these platforms, promote and recommend the use of back testing tools to adjust the EAs before using them in actual operations, many developers of EAs and market experts criticize these tools. The main deficiency pointed out is that often EAs optimized using back testing

tools have a very good performance with back test data, but are regularly very bad when operating with real data.

This study, focused on scalper strategy, compared the tick-by-tick data (intra-minute data) generated by the MetaTrader Strategy Tester tool with the data collected in real time form the market using the same platform. The results showed significant differences between the two data sets. We also shown with an example, as a point-wise difference can change the outcome of a transaction, giving profit during training, but loss when operating on real data.

According to the authors' knowledge, this is the first time that the discrepancies in a very short time scale (tick scale) between the data used for testing strategies and the actual data, is identified as a major cause of drastic reduction in the performance of scalpers optimized by back testing when operating in real time.

The result suggests the use of MT4 Strategy Tester just to test the functionality of the algorithm but not to assess the expected performance of any scalper strategy in real applications. The great graphical visualization capability of the MT4 platform is very useful for understanding the way that isolated transactions are opened and managed by an EA. Only forward (real-time) tests should be used to optimize scalping EAs.

References

Alrefaie, M.T., Hamouda, A.-A. and Ramadan, R. “A smart agent to trade and predict foreign exchange market.” IEEE Symposium on Computational Intelligence for Engineering Solutions (CIES), page(s): 141 – 148. 2013.

Sim, K.Y. ; Low, C.S. ; Kuo, F.-C. Detecting faults in technical indicator computations for financial market analysis. 2nd International Conference on Information Science and Engineering (ICISE), page(s): 2749 – 2754. 2010.

MetaQuotes Software Corp. Testing of Trading Strategies in MetaTrader 4. 2000-2011. Available online at http://www.strategytester.com/ . Accessed on January 2015.

Galati, G. “FOREX trading volumes, volatility and spreads in emerging market countries”, BIS Quarterly Review, November, pp 49–51. 2000.

John, L. Person. “Candlestick and Pivot Point Trading Triggers: Setups for Stock, FOREX, and Future Markets”. Jhon Willey & Sons, Inc., Hoboken, New Jersey, pages 279-304. 2007.

Pinto, R.M.C and Silva, J.C.M. “Strategic methods for automated trading in FOREX”. 12th International Conference on Intelligent Systems Design and Applications (ISDA), page(s): 34 – 39. 2012.

Montgomery, D. C.; Jennings, C. L.; Kulahci, M. “Introduction to timeseries analysis and forecasting.” [S.l.]: John Wiley & Sons, 2015.

Li, Y.; Ma, W. “Applications of artificial neural networks in financial economics: a survey”. In: IEEE. Computational Intelligence and Design (ISCID), 2010 International Symposium on. [S.l.], 2010.

Korczak, J.; Hernes, M.; Bac, M. “Risk avoiding strategy in multi-agent trading system”. In: IEEE. Computer Science and Information Systems (FedCSIS), 2013 Federated Conference on. [S.l.], 2013.

Daniel, 2011. “DOES BACK TESTING WORK?” - https://www.danielstrading.com/strategies/2011/10/20/does-back-testing-work/

Rolf and Moritz, 2015. “6 Reasons Why Backtesting Is Not Working For You” - http://www.tradeciety.com/6-reasons-why-backtesting-is-not-working-for-you/

Desroches, V. 2015. “Back Testing in Forex: Why It Doesn’t Work” - http://www.moneyshow.com/articles.asp?aid=currency-19374

Balch, T. 2015. “9 Mistakes Quants Make that Cause Backtests to Lie” by Tucker Balch, Ph.D. - http://blog.quantopian.com/9-mistakes-quants-make-that-cause-backtests-to-lie-by-tucker-balch-ph-d/

Greenberg M. 2010. “Testing Backtesting: Why it doesn’t work” - http://www.financemagnates.com/forex/brokers/testing-backtesting-why-it-doesnt-work/