Big Data Automatic System of Analysis and Trading on Financial Markets

Serhii A. Rybalchenko
Faculty of Economics, Department of Economic Cybernetics
Taras Shevchenko National University of Kyiv
Kyiv, Ukraine
sergiy.rybalchenko@gmail.com

Abstract—The paper considers the main tendencies of the international currency markets and opportunities for individual investments. The methods of decision support of assets purchasing were provided. A mechanism of intellectual trade with the use of a significant data flow is proposed. Open and closed types of information products were combined to automate the system. Guided by the effectiveness of trade signals assessment, recommendations were made for the use of methods of artificial neural networks or logistic regressions. A set of criteria has been formed to support lifecycle of the automated trading system.

Keywords—automated trading system, artificial neural networks, logistic regressions, international currency markets.

I. INTRODUCTION

In modern conditions, the international financial market serves as the main indicator of changes in economic environment and economic development of certain regions and industries. All economic projects need financing. Most projects require cheap international investment. In turn, foreign investments come into efficient and safe environment where the balance between risk and yield is maintained.

The modern foreign exchange market began forming during the 1970s. At that time, the main participants of international currency transactions were: the International Monetary Fund (IMF), International Bank for Reconstruction and Development, and the General Agreement on Tariffs and Trade (GATT) [1, p. 169]. With the widespread adoption of the Internet in the 1990s, banks and small companies created online networks to produce automated quotes and allowed for instantaneous trading [2]. Over time Forex became available to individual investors as well. According to the Bank for International Settlements, the preliminary global results from the 2016 Triennial Central Bank Survey of Foreign Exchange and OTC Derivatives Markets Activity show that trading in foreign exchange markets averaged $5.09 trillion per day in April 2016. This is down from $5.4 trillion in April 2013 but up from $4.0 trillion in April 2010. Measured by value, foreign exchange swaps were traded more than any other instrument in April 2016, at $2.4 trillion per day, followed by spot trading at $1.7 trillion [3]. These data allow us to perceive Forex as the main indicator and instrument of the international financial market.

The leading status of the market entices significant scientific efforts to develop methods of analysis and forecasting of exchange rates. In addition to some developments carried out or supported by commercial organizations, a lot of material is being developed by educational institutions. This is due to easy access to the market, historical data and volume of information. The data of the international currency market reflect the real processes of the economy, while at the same time are suitable for student learning proven and the newest methods. Accordingly, students at the relevant departments annually conduct 10-15 courseworks and diploma projects on application of economics and mathematical methods for analyzing and forecasting the exchange rates of the international exchange market. Within this publication, the main directions of scientific research on optimization and automation of trading and other systems are singled out.

The object of study – automated trading and analysis systems. Subject of research – analysis and synthesis method, quantitative method of measuring and comparing, historical and simulation methods. The aim – to create effective automatic system of analysis and trading, based on artificial intelligent and simulation methods. The goal will be revealed through the following objectives:

• Determine the conditions and trends of the international currency markets;
• Outline the main opportunities for individual investments;
• Explain weaknesses of decision support methods;
• Summarize the experience of forming of analysis and trading systems;
• Show effective methods of Big data analysis with open-source statistical applications;
• Develop automatic system of analysis and trading, recommendations for further lifecycle system management.

II. OPPORTUNITIES FOR INDIVIDUAL INVESTORS AND RESEARCHERS

Prior to the propagation of information technologies to the household level, individual investors could only use Forex forwards and futures contracts. With the proliferation of computers, there was an expansion of opportunities for investing. A person with a laptop can easily connect to the international financial system by one of many broker options. In turn, brokers interact with banks and other financial institutions where they can carry out foreign exchange operations at their discretion and at customer's request. If we look towards the implementation of customer contracts, brokers can usually use the mechanism of "clearing". For the same currency pair, there will always be several different
offers from clients: someone wants to buy the first currency of the pair, someone the second one. Even by volume and opposite by direction customer orders are offset and the broker will not have to carry out any currency exchange transactions on the interbank market, and will not pay therefore any the commission. When executing customer transactions, the broker will copy off the loss of the first client to the win of the second, and write down some of the funds on his own account as a commission. If all transactions will not be offset with the opposite ones, the lack of demanded currency will be purchased on the interbank market. Obviously, with the "clearing" system the volume of real transactions can be less than 20%. Such a system could be neutral for the client, if not for abuse by dis-honest brokers. With a neutral system customer does not care about the mechanism for the implementation of transactions, he is set to receive speculative profits. But when the deal is not backed by real market performance, improbable temporary events in the fluctuation of the asset rate may occur. For example, drop by 30% for 1 second with a subsequent sharp return to the normal level. After adoption of automatic trading systems by large companies, such cases became a reality [4]. But exceptional and possible only when the system crashes for significant part of market participants. Such market participants are in high-level jurisdictions and customers are protected by the law and reputation of the company. Derived problems arise. Small common brokerage companies purchase informational blocks of currency quotations from banks or other financial institutions. With the proper manipulation of the courses in the "clearing" system brokers shift responsibility on the supplier of quotations. The client will have to enter into a legal dispute. In such circumstances, jurisdiction is important. Such models were common in the 2000s. Although the international currency market does not have a single regulator and is by all the general good. At this stage, thriving brokerage companies have full execution of customer agreements, i.e. without internal "clearing". Most companies have a level of jurisdiction not lower than "C".

Let's consider the factors of income formation for a client of brokerage company. The basic law of commerce here as well is to buy cheaper and to sell with profit later. Therefore, speculative profit:

\[
Income_i = (Price_i - Price_{i-4}) \cdot Volume_i ,
\]

where \( Income_i \) is income in time period \( i \), \( Price_i \) is currency price in time period \( i \), \( k \) is periods number of order (between open and close order), \( Volume_i \) is assets (currencies) volume in time period \( t \).

Under such formal conditions, the break-even point of a speculative investor is in equality:

\[
\sum_{j=1}^{n} Price_i^j = \sum_{j=1}^{n} Price_{i-4}^j ,
\]

where \( j \) is order index. So,

\[
E(Price) = E(Price_{i-4}) .
\]

The interpretation of formulas (1) - (3) consists in the need to correctly guess or predict the direction of price movement in 50% of cases under neutral conditions \((Price_{const}, Volume_i = const)\). This is an acceptable task for both scientific methods and for any coin (situation with coin flip). We call such a situation a point of zero informational advantage. That is, when we know nothing about the history and prospects of the market. All that is needed is the current asset price and the ability to fulfill the condition. It is possible to draw the equivalence between the available information and the probability of correctly predicting the direction of the trend of the asset price.

\[
P_{0.5} = I_0.
\]

In reality, of course, this is not enough. When executing customer order, and closing the order at the same price, the client receives a loss from the executed order. The nature of the negative difference at an even price lies in the broker commission, currency spread and other payments.

Brokerage commission is the main source of income for financial intermediation. Clearly exhibited commission rates have been the subject of fierce competition between brokers. And 0% became minimum completion of lowering of explicit commission for a client. Most FOREX brokers do not demand commission fees for individual investors with small volumes. Trading in amounts greater than $100,000 is possible with fixed fees for each transaction, relative to the amount of the transaction, periodic as well for the implementation of individual services. For investors with small amounts, the source of the broker's profit is transferred to the spread. The spread is the difference in the buy and sell price of any asset or currency pair [5]. Buying an asset at price \( P_1 \) we can sell an asset in the same moment at lower price \( P_2 \). The difference between the prices of \( P_1 \) and \( P_2 \) is the spread of the broker. It can be fixed or relative, depending on the time of day and market volatility.

Brokers measure spread in pips. Pip is the minimal change in the internationally accepted format of currency pairs. For example, for EUR/USD pip is 0.0001, for USD/JPY - 0.01. For most major currencies for brokers, the relative spread ranges from 0.01% to 0.1% (Fig. 1.), but such a low spread-commission is misleading. Since it is impossible to execute an order on the international interbank market for a client with a small amount, the broker provides the client with a financial leverage (1:50-500). That is, to buy a minimum USD lot on the market ($ 100,000), at the most popular leverage 1:100 client needs to spend $ 1,000. Recently, brokers commonly provide clients with opportunities to enter into deals for a part of the minimum market lot. All losses or profits due to the rate change of purchased USD (asset) are not shared in proportion to the invested funds, but are fully transferred to the client.

The principle of transferring the result entirely to the client is the reason for the high volatility of client accounts. With the growth of USD only by 1%, client funds invested in 1 lot increase by 100%. This is an attractive situation, but it works both ways. That is, for the loss of all funds fluctuation of the rate by 1% is sufficient. Risk-driven investors can be attracted by such a system, but it is often overlooked that commission is paid exclusively by client funds. Accordingly, the range of relative spreads is also multiplied by the financial leverage, and for the standard leverage 1:100 it will be respectively 1% -10%. With the use of coin flip to open orders for EUR/USD, the client's funds will transfer to the broker on average after 100 transactions executed. It is worth
noting that the financial leverage leads not only to the increase of risk and relative spread, but also to the possibility of diversification.

To measure how much really useful information the trading system provides, it's enough to get a share of profitable transactions in the system and divide it by 0.5 (P_0.5).

III. MODERN DECISION SUPPORT METHODS ON FX MARKET

Traditionally, methods for analyzing a currency market and decision making are divided into fundamental and technical ones. Forex fundamental traders evaluate currencies, and their countries, like companies and use economic announcements to gain an idea of the currency's true value [7]. For making decisions in accordance with the methodology of fundamental analysis, macroeconomic indicators, that give an equivalent valuation of the currency, and the price dynamics of the currency them-selves are important. The reaction of the currency market to changes in the macroeconomic environment shows up in the medium and long-term trends. The percents of swap become the source of profit of investor in such situations. In a competitive international market, low percentage rates are offset by fluctuations in the value of the currency itself. When macroeconomic indicators are changed not by objective laws, but subjective decisions of public authorities, the market takes into account the change instantaneously (within a few minutes). Therefore, it is not enough to rely solely on information of the fundamental analysis in modern conditions. The maximally near placing of client servers to the servers of exchanges becomes the determinative factor of objective laws, but subjective decisions of public authorities, the market takes into account the change instantaneously (within a few minutes). Therefore, it is not enough to rely solely on information of the fundamental analysis in modern conditions. The maximally near placing of client servers to the servers of exchanges becomes the determinative factor of fundamental analysis, to have time to carry out agreements before "window of possibilities" closed (a few minutes taking subjective change of macroindexes into account by a market). For individual investors with small sums this is not relevant. Especially in speculative trading system, swap income in the long run will be mutually compensated.

Technical analysis is a trading tool employed to evaluate assets and attempt to forecast their future movement by analyzing statistics gathered from trading activity, such as price movement and volume [8]. The basis of technical analysis is the efficient-market hypothesis [9]. A current competitive market price effectively takes into account all available information about the asset. That is, observing fluctuations in prices can get a sense of and knowledge of all the necessary information regarding the financial instrument.

The main tools of technical analysis traditionally include historical price charts, indicators, financial ratios. Recently, automated systems using artificial intelligence replacing traditional technical indicators. Deutsche Börse AG have announced the use of a global index based on the selection of financial instruments to it by artificial intelligence methods [10]. Use of artificial neural networks does not guarantee success, but it becomes necessary for a successful trading system. These methods allow to process large amounts of information quickly and get the desirable information advantage, and in consequence reduce the risk.

We will focus solely on price indices and volume transaction characteristics. An information advantage means getting more qualitative data, available in all market history, about patterns in the past, that have already begun to operate in the present and will continue for some time in the future.

The general trend is the dominance of international banks, but at the same time increasing the role of brokers who use high-tech information systems investment (XTX Markets) [11]. All top companies can afford expensive systems using artificial intelligence, but their disadvantage is...
the scale and long implementation of the latest methods. But even with the rapid ability to implement analytical systems, which use artificial intelligence, by individual investors, top 10 traders will always be better informed about their own future actions, due to asymmetry. That is, individual investors, using the Pareto rule, should focus on information systems that can reach the probability of successful transactions 0.8. So provide information advantage over 80% of the market participants, which may indirectly influence the market.

IV. FORMING OF ANALYSIS AND TRADING SYSTEM WITH BIGDATA FLOWS

The first important step in the formation of price dynamics analysis system is a in-formation base. In addition to traditional historical data on price and volume of transactions, the database should include all possible technical indicators, derivative data and proportional, regression levels. A reference point is the information used by other participants in the currency market. The principle of quantitative advantage in information is provided by wider accumulation of all possible data streams. The principle of qualitative advantage in information consists in the effective selection of useful information from a vast set to the next stage of the analysis. Historical asset price data can be found on many online resources. But if the analysis is not static, and includes dynamic principle, we should use one of the online terminals. The most popular is MetaTrader4. Its advantage is the extensive use by brokers and their clients. Therefore, when we install it, we will have the opportunity to immediately open a demo-educational or valid cash account. The terminal includes the program language MQL4. The main problem is the closed structure and the impossibility of online data export to other software applications in a client free version. There is a simple solution to this limitation. The online community provides many useful solutions in the form of open source. One such example is the library mt4R.mqh by Bernd Kreuss (2010) [12], which covers the rights of free use under the terms of the GNU General Public License. By integrating this library into the MetaTrader4 terminal, we will be able to send all the necessary data in the R environment and return the result back to the MetaTrader4 terminal. The main purpose of the R environment is to provide a convenient environment for forming the core of online analysis systems: information filtering, classification of information, model construction, forecasting and signal calculations.

In order to get possibility to exchange data between MetaTrader4 and R, we should add the following at the beginning of the expert advisor .mq4 code:

```plaintext
#define RPATH "C:\Program Files\R\R-3.4.1\bin\Rterm.exe --no-save"
#define RDEBUG 2
#include <mt4R.mqh>
```

In the above lines in order: connect the environment R, indicating the location on the computer of the installed folder with R; set the level of display of events in the register of the operating system; connection of program commands from the library mt4R. At the initial tuning of code, it is recommended that the RDEBUG parameter be equal to 2 and track all events, for example, through the DebugView application. When the automated system working stable without any interruptions, the parameter can be reduced in the value to the minimum.

The main task of the program expert on the MetaTrader4 side will be to send initial online information and trade orders when receiving processed trade signals back from the R script. As part of the MT4 procedure onOpen() - executes each time when a new time period starts (minute, hour, day ...); we can generate a code:

```plaintext
Rm("rHigh", High[], bars, 1);
```

In the above line we used a program command “Rm” from the library mt4R, forming a matrix in R of dimension bars * 1 with the name in " " - rHigh, rewriting data from a one-dimensional array High[] of current currency. Having formed such 5 blocks, we transfer all the basic information – prices on open, close, high, low and volume of trade – to the environment of the R-script. It is possible to calculate all known technical indicators in the environment of R, using specific libraries. But there are differences in the formulas for calculating many indicators, so it is recommended to use the same data that appear on the screens of traders around the world. That is, the calculated indicators in MT4 should also be transferred to R by inserting similar to the previous program blocks and using the indicators iMA (...), iMFI (...), iSAR (...), customized ZigZag iCustom (...), ZigZag, ...) and others. Overall moved an additional 171 technical indicator with averaging for 2-3 periods of 10 years of daily information. It is important to expand the information by all possible methods, especially by adding derivative indicators used by individual traders. Popular is the swing theory of trend definition derived by Gann (1936) [13]. We can calculate the swing in R or by running from R the macro in Excel. Having a swing identification of a short motion and a trend in general, at each moment of time allows us to calculate even the derivative indicators: the duration of the current short and long-term movement in days / minutes; speed of change in prices in the short and long-term. We get the same structure information from the ZigZag indicator. Some traders use Gann theory, others - ZigZag indicator. Our goal is to aggregate each indicators information in the database. After defining all types of wave-like movements (swings), we calculate all possible levels of the Fibonacci. As a result, received 8 basic prices, time, volume streams and 256 streams of indicators.

V. AUTOMATIC SYSTEM OF ANALYSIS AND TRADING

We will continue to process existing information flows in R, taking into account the repeatability of the script execution in each new time period: day, hour, minute. Since the swing calculation of swings lasts on average 3 minutes, then it is not recommended to use a system with a complete set of data for minute time intervals. Trading systems within the minute time interval are classified as high-frequency. For such systems it is recommended to include only the basic 8 variables in the dataflow. Based on the received data, we can execute a regression forecast for price changes and accordingly decide on the opening deal. We can also evaluate the price movement indicator (up = 1, down = -1) by a neural network classifier or logit regression. This brief analysis will require an average 15-20 seconds and can be successfully used in high-frequency trading. Going further, we will concentrate on the positional trading systems of hour and day time intervals, with the ability to analyze the full data set in order to obtain information advantage.
All 264 variables should be checked for stationarity by adf or kpss tests (library tsieries). To avoid in-process failures of the automated system, we recommend to use the try () function. Verify the result of the function try(kpss (...)). If the result is not numeric, then repeat the procedure as long as not obtain a numerical result. It is enough to use the while {} cycle with the maximum limit of the number of passes equal to 100. When identifying a non-stationary series, you must use the differences method to transform it into a stationary data. The next step is a factor selection to predict the direction of price movement. There are sufficient differences methods to transform it into a stationary data. The next step is a factor selection to predict the direction of price movement. There are sufficient differences methods to transform it into a stationary data. The next step is a factor selection to predict the direction of price movement. There are sufficient differences methods to transform it into a stationary data. The next step is a factor selection to predict the direction of price movement.

The following code will generate a library "Rgretl" and its function "grmod" integrate those (library stats) , but have chosen function "omit" from other years. Unacceptable option. There is also a function "step" the choice of factors from a large-scale database will last for years. Unacceptable option. There is also a function "step" (library stats) , but have chosen function "omit" from other open-source statistical package - Gretl. A well-established library "Rgretl" and its function "grmod" integrate those statistical packages [14]. The following code will generate a list of optimal factors for the selected price stream:

```r
try(grmod("ols y const X* omit --auto=0.01", data = dataGr, output="$xlist"))
```

dataGr – pre-prepared matrix with data.

In addition, in order to predict future swings and trends in the price of short and long periods, it is necessary to choose the optimal set of factors as well:

```r
try(grmod("logit y const X* omit --auto=0.01",data=dataGr,output="$xlist"))
```

The choice of factors is carried out, but in the classification model - logit. This solves the fundamental problem of the uncertainty of the current price movement. Now, from the time point of the beginning of the uncertainty of the movement direction , we predict the probability of up / down swing in R on the received sets of factors. We have received a completely filled database in any situation, at each moment. In the internal environment R, we forecast price levels using dynamic ARIMA-factors-lagged regressions, neural networks, or other models. We compare the received future prices with the current and get not only information about a possible direction of movement, but also about the horizon and the limits of this movement. If we say in traders' words, we got the type of buy/ sell transaction, the entry point (at which price will open the deal) and the exit point (at which price the transaction will be closed). Results are transmitted from R back to MT4.

```r
NormalizeDouble(Rgd("Hp"),5)
```

Function Rgd () sends the numeric value of the variable Hp (High price forecasted in R) from R to MT4. The average duration of a single run the entire program takes 8-9 minutes on an ordinary computer. With the integration of all safety devices against failures of the automated work, the average duration of work is 25-40 minutes.

VI. CONCLUSIONS

The international currency market is developing towards individual investors. Conditions for entry into the market, even with insignificant capital, have been created. At the same time, investors should pay close attention to trading conditions provided by brokers. The relationship between the success of the trading system, investors' decisions and information flows is shown. Available fundamental and technical indicators do not meet the information needs of the full extent of investors with little capital.

Modern systems should use a broad stream of online data. Within an integrated data stream, a classification, cluster, and factor analysis should be performed in order to obtain new useful information. Thus, the information advantage of an investor with a small capital can be formed.

The current automated system was activated during 01.03.16-10.11.16 on actual market EUR / USD and as a result made 80 daily deals. The probability of profitability of the transaction in the system was 51.25%, the total profit was $ 219.51 at the initial balance of $ 96.4. Profitability is 127.7%, which is far more than the average Ukrainian deposit rate – 14%. The results indicate the achievement of the information advantage by the system, during working in real conditions and taking into account all commissions.

Strengthening the system is possible at choosing the optimal set of prediction factors, applying evolutionary methods to determine the optimal parameters of orders, classification market conditions.

REFERENCES