

# A Survey on Computer Science Techniques in the FOREX Market: Models and Applications

Diego Aguilar, Ildar Batyrshin, Oleksiy Pogrebnyak

Centro de Investigación en Computación, Instituto Politécnico Nacional,  
CDMX, México

diego.aguilar.m15@gmail.com, batyr1@cic.ipn.mx, olek@cic.ipn.mx

**Abstract.** In this paper a survey on computer science techniques such as natural language processing, machine learning, data mining, association rules, etc. which are applied on the FOREX market is presented. The main objective of investing in FOREX market is to maximize the profits given a certain period of time. This survey can be useful to researchers in this area, in order to compare and extend previous works.

**Keywords:** time series, Forex, data mining, natural language processing.

## 1 Introduction

### 1.1 Foreign Exchange Market

The Foreign Exchange (FOREX)<sup>1</sup> market is one of the biggest and most liquid<sup>2</sup> markets for investors, trading in average \$5.1 trillion dollars per day in April 2016 [1]. It trades in pairs of currencies, the first coin of the pair is called base currency and is the one that an investor can sell or buy. The second coin is called reference currency and it means the money that an investor is willing to pay or to receive for the base currency.

FOREX is a decentralized market, also know as an Over the Counter (OTC) market. OTC markets are not controlled by any government nor have a physical location, all transactions are made by computer, phone and email orders. OTC trades a great number of listed securities, such as currencies, bonds, and shares of companies that does not meet the requirements to be listed on exchanges such as the New York Stock Exchange (NYSE) or the Bolsa Mexicana de Valores (BMV) exchange in Mexico. There are mainly two forms of trades, one where a dealer trade with institutions or corporations and other intermediate trades between different dealers.

Modeling the economic markets and currency pair market is a difficult task as stated by Rey [2], who published a paper in which the role of a vehicle currency

<sup>1</sup> Some sources also use the FX abbreviation.

<sup>2</sup> A liquid asset is described as one that can be easily exchanged for money, i.e. currencies, shares, bonds, etc.

is studied in order to explore the rise and fall of currencies through the evolution of trades patterns between countries.

## 1.2 The Risk on OTC Markets

OTC markets such as FOREX are considered to have more risk implied on the transactions, due to little or non existent regulations, which make OTC markets less transparent; for this reason many advisors do not recommend this type of market for small and unexperienced investors.

## 1.3 Mexico in the FOREX Market

According to the results given by Bank of Mexico (Banxico), in 2016 the Mexican peso (MXN) was the tenth currency with most trade in the world, and the second one in the list of emerging countries<sup>3</sup>, only behind the Chinese renminbi.

It is recognizing patters using fuzzy logical also important to notice that most of the transactions made with the MXN currency are realized outside of Mexico, besides the US dollar (USD) is the main currency that is traded against the MXN [3].

## 1.4 Trading in FOREX

To make transactions in the FOREX market it is necessary to have an account with a brokerage company; as the OTC market works on networks, it is possible to open an account from other country, but it is recommended to use the most trusty broker one can find, i.e. a broker with the biggest capital liquidity.

## 1.5 Spread

The spread is one of the main subjects in FOREX trading, it is described as the difference between the bid and ask price of a currency<sup>4</sup>, is common to find it in practice as the Bid-Ask Spread. Currency pairs usually have the lowest spread, movements around \$0.0001<sup>5</sup> also known as Pip<sup>6</sup>.

Because there are no commissions on the FOREX transactions, brokers use the spread as a way of making profit, this means the tighter a spread is, the better profit an investor can make. The competition between FOREX brokers is fierce, it is recommendable to shop around for the best prices in currencies<sup>7</sup>.

<sup>3</sup> An emerging market economy is described as one who is progressing into becoming an advanced economy, i.e: Mexico, Brazil, Russia, India, Turkey and China.

<sup>4</sup> The ask price is the money that a seller is asking to receive for a security, conversely the bid price is the money a buyer is willing to pay for a security.

<sup>5</sup> Some currencies trade at movements of \$0.01

<sup>6</sup> A Pip is the equivalent of 1/100 of 1%.

<sup>7</sup> Many variables can affect the spread of a currency pair, such as volatility, high or low demand and also affects the political issues on the base currency country.

## **1.6 Making a Profit on the FOREX Market**

### **1.7 Two Strategies for Trading**

There are mainly two strategies for trading in currencies, shares, bonds, etc. The long and the short position.

#### **Long Position**

The long position is the most common among investors in markets, it follows the principle “Buy low, sell high”, in the FOREX market is the same as buying a currency which an investor thinks is going to appreciate in the future (in FOREX changes may occur in relatively small intervals of time, such as minutes), if the currency price rise then the investor will sell it obtaining a profit from the operation. On the other hand if the currency loss its value the investor could sell it at lower price taking losses.

#### **Short Selling**

The short selling is a strategy in which an investor predicts the value of an asset, such as currency, will drop in the future. In practice an investor could “borrow currencies” from a margin account<sup>8</sup>, the short investor exchange the currency pair, if the prediction is true and the value of the base currency drops such investor could buy the same currency at a lower cost, finally returning it to broker in the margin account.

Short selling is seen as a highly risky strategy, because the maximum profit can be achieved when a base currency drops down to near zero; however, the maximum losses tend to infinity since there is no limit for a currency to appreciate. Thus this strategy is recommended only for experienced investor. The main idea behind short selling is to bet against a security.

## **2 Previous Related Surveys**

In this section, previous survey related to FOREX and financial markets using computer science techniques are described.

Most recently, Kumar et al. [4] performed a survey on applications of text mining in financial domain; it categorized the applications into FOREX rate prediction, stock market prediction, customer relationship management (CRM) and cyber security which were published in the period of 2000-2016. They divide the text mining task into four categories, namely classification, clustering, summarization, topic identification and association rule mining. The survey concluded that more efficient techniques are needed to handle and predict a significant amount of data, also the most frequently used text mining task is classification followed by prediction: one of the purposes is to standardize the performance metrics to compare various works.

<sup>8</sup> Brokerage firms offer margin accounts in which they lend money to investors in order to purchase securities.

Batyrshin et al. [5] presented a survey paper on perception based patterns in time series. They discussed the importance of recognizing patterns using fuzzy logic on time series in order to support decision making processes. In this study they found that most of the decision making procedures in problems on financial and economics are based on human decisions supported by statistics, data mining or data processing software.

The main contributions are the applications that uses fuzzy logic for the description of patterns such as sign derivatives, scaling trends, linguistic interpretations of patterns as result of clustering, relationship between patterns, etc.

This survey includes studies which use financial data, i.e stock prices, news, indicators as well as non-financial data, i.e climate conditions. They conclude that time series data mining along with natural language computation and computing with word and perceptions is an opportunity field for researchers to explore new method and publish new techniques.

Soft computing techniques applied on the stock market where reviewed by Atsalakis et al. [6]. In this research they stated that soft computing techniques are candidate to capture stock market nonlinear relations returning significant forecasting results, the methods included in the survey are mainly neural networks and neuro-fuzzy models with several types of training. The results were compared against conventional models such as artificial neural networks, linear and multi-linear regression, ARIMA models, genetic algorithms, random walk, buy and hold, etc.

Data mining techniques with applications in the stock markets were surveyed by Hajizadeh et al. [7], they stated that data mining techniques in the financial domain are gaining interest from the researchers. One of the most important problems is finding efficient ways to summarize and visualize the stock market data in order to give useful information for investment decisions.

The survey categorized six data mining techniques and its applications on the stock market, namely decision tree, neural networks, clustering, association rules, factor analysis and time series. They conclude that the use of data mining techniques in the financial domains has reached good results and they expect the number of studies in this field increase.

### **3 Review of Computer Science Techniques for the FOREX market**

This survey highlights some applications of computer science which have an application in the FOREX market. These applications are categorized mainly into FOREX rate forecast, up-down movements, trend analysis.

The FOREX rate predictions attempts to give an accurate price of a currency pair in the future, using available data, such as past values of bid/ask quotes.

Due to space limitations most of the mathematical framework of the studies will not be presented.

Nassirtoussi et al. [8] published a research in 2015, in which they classified the future of a currency pair; a binary classification was used, namely POS if a

currency pair will appreciate or NEG if the bid/ask price will drop. The study included a novel system that used text mining, natural language processing, statistical pattern recognition as well as sentiment analysis that predicts the directional movement on currency pairs in the FOREX market on the words used in news headlines in previous hours. The forecast horizon for this study is 1-2 hours as they were working with intra-day trading data, news headlines were obtained from a financial news web page via web service and data from EUR/USD currency pair.

Before the research of Nassirtoussi the idea of forecasting the FOREX rates from news headlines was revised by Peramunetilleke and Wong [9], in their study they proposed that the information contained in news was better in contrast to numeric time series, because they not only contain the effect, but also the possible cause of the event that affects the FOREX rates.

They implemented a new technique that relies on natural language processing, which automatically evaluated the news. The system receives as an input news headlines from Reuters and the output is a categorical forecast divided in three: the dollar moves up, remains steady or goes down within the next one, two or three hours.

The system generates weights for keywords on headlines, then it creates rules that can handle continuous attributes, rather than boolean expressions, to create the rules they use two methods, namely *TFxIDF* which consist in term frequency, discrimination factor and normalization and the *TFxCDF* method which uses category frequencies.

The results they obtained showed that the *TFxCDF* method reaches better results than other methods, it was also compared to a conventional time series analysis tool and two different neural network approaches and outperformed them.

Idvall et al. [10] in their master thesis investigated the use of Markov chains in hidden Markov models with two implemented extensions: a Gaussian mixture model (GMM) and an exponentially weighted expectation maximization (EWEM) in order to create a new way of algorithmic trading which main objective is to increase the profit of the alpha returns on intra-day trading.

They concluded that the use of hidden Markov models on the FOREX markets is still an area of opportunity for researchers. Their model in some settings was able to obtain significance profit, whereas other settings the predictor looked close to a random generator, this was attributed to the great quantity of factors that have to be considered when setting up a model.

They propose that a different machine learning method and improving the initial estimates of the parameters would produce better results. As for the model itself it could be expanded in order to diversify the portfolio of HMM, including different currency pairs, since each pair might depend on different causes thus have different patterns, also the usage of leverage could be advantageous to enhance the profits.

Regarding the machine learning (ML) area, Baasher and Fakhr [11] published an article in which they used ML techniques to forecast the High exchange rate

daily trend; They represented the forecast problem as a binary classification task with the two possible outcomes being uptrend and downtrend. This article used four major currencies, namely USD/YEN, USD/EGP, EUR/EGP, EUR/SAR.

The ML algorithms used for classification were radial basis function (RBF), multilayer perceptron neural network (MLP) and support vector machines (SVM). The results for forecasting of the generalized autoregressive conditional heteroscedasticity (GARCH) were used as part of the feature set, along with other classical technical analysis techniques for time series i.e momentum. Feature selection and extraction method were applied in order to optimize the results given by the ML algorithms. For the comparison analysis, they used the percentage classification performance and a novel function called percentage normalized profit (PNP), which represents the ratio between accumulated profits using the predicted trends versus the accumulated profit using perfect predictions during a test period.

The study concludes that a good feature selection along with regularization on the ML techniques can provide good results, making a profit in the FOREX market.

Karuzzman and Sarker [12,13] published a research on predictions of foreign currencies rates using neural networks with three different learning algorithms, namely Standard Backpropagation (SBP), Scaled Conjugate Gradient (SCG) and Backpropagation with Bayesian Regulation (BPR). They concluded that the SCG learning algorithm outperformed the other models.

In their tests they used moving averaged values to feed the neural network to predict the following week, comparing six currencies against the Australian dollar (AUD), in terms of performance measure the SCG model achieves very close prediction in terms of NMSE and MAE metrics. The Figure 1 shows the results obtained.

Batyrshin et al. [14] proposed a novel definition for a non-statistical measure of time series comovement called “local trend association measure” (*lta*), which is based on an early work the “moving approximation transform” (*mat*) [15]. The *lta* measures positive and negative shape associations between time series.

They conclude that the correlation coefficient is not an useful tool in the analysis of time series shape similarity and shape associations. They realized a comparative analysis between exchange rates with data from Google Finance [16], Figure 2 shows the initial points where the local trends change sign and this can be associated with business events. Also using the approach of “pair of stocks” (or currencies), one can associate pairs which move asynchronous (positively associated) or with inverse movements.

García et al. [17] proposed a method to choose a better sliding window in the moving approximation transform (MAT) [15]. This new method uses the length of a pattern instead of the *cos* function to measure the positive or negative similarity between time series; they also presented a new method to visualize the associations between time series which is easier to understand than the one used in [15,14].

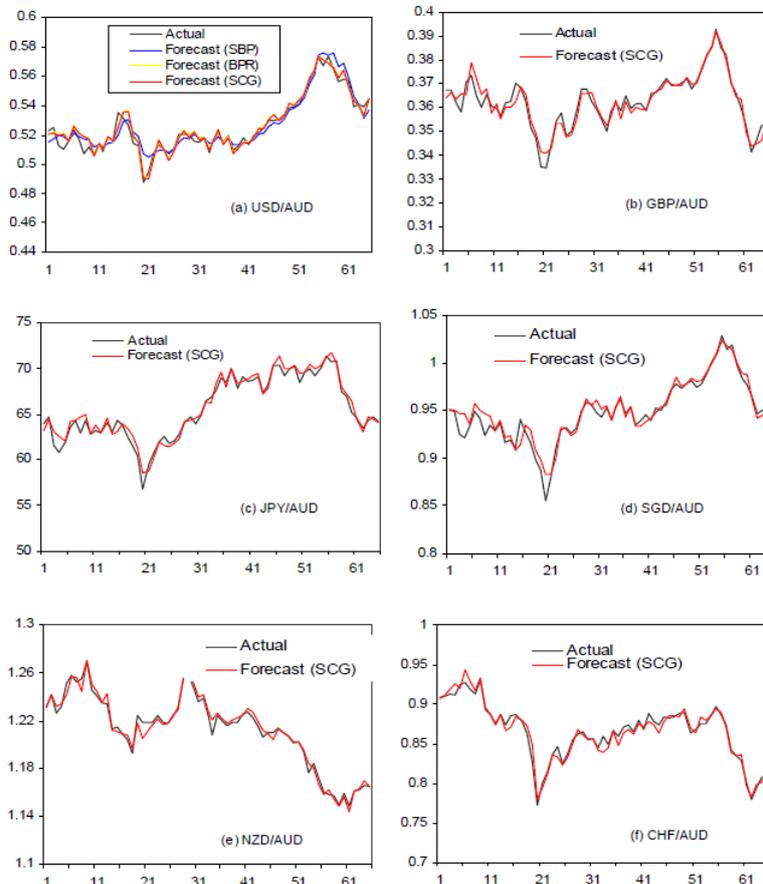


Fig. 1: Forecasting of different currencies by SCG based neural network model over 65 weeks.

Kayacan et al. [18] used gray systems theory-based approaches to predict FOREX rates, because gray systems are more robust with respect to noise which make them ideal to treat the non stationary, and nonlinear time series of FOREX rates.

They used six models in this study, namely GM(1,1), modified GM(1,1) model using modeling errors and Fourier series (EFGM), modified GM(1,1) model at time domain using Fourier series (TFGM), Grey Verhulst model using modeling errors and Fourier series (EFGVM) and Grey Verhulst model at time domain using Fourier series (TFGVM).

The results they obtained were that GM(1,1) models are better on both interpolation and extrapolation when compared to GVM model without using a filter. On the other hand, using a moving average filter (MAF), the performances

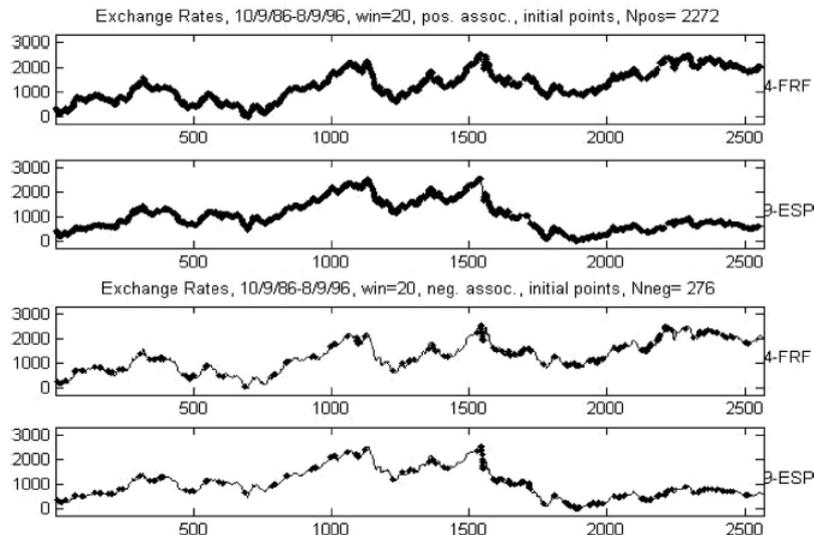


Fig.2: The initial points of positively associated (at the top) and negatively associated (on the bottom) of moving approximations (shown by bold) in sliding window of size 20 of time series 4-FRF and 9-ESP.

of GVM models were increased and the performance of the GM(1,1) model was decreased.

They conclude that the modified GM(1,1) using Fourier series in time domain is the best model fitting and forecasting.

Nieh and Lee [19] revised dynamic relationship between stock prices and exchange rates of the G7 countries. They found that there is no long-run significant relationship between them, on the short-run significant relationship was found for one day in some of G7 countries.

In their research they found that some studies have encountered limited success in identifying a significant correlation between stock prices and currency's fluctuations. The authors implemented various test to explore the dynamic relationships, such as Engle-Granger (EG) and Johansen maximum likelihood cointegration; they applied the vector error correction model (VECM) as an appropriate framework to assess both the short-run inter-temporal co-movement and their long-run equilibrium relationship.

The data they used consisted in stock prices and foreign exchange rates for the G7 countries, namely Canada, France, Germany, Italy, Japan, UK and the US. They concluded that these two financial variables do not have predictive capabilities for more than two consecutive trading days; they stated that from a practical view the investors believed that stock prices and exchange rates can serve as instruments to predict the future of each other, but the VECM only shows that one day prediction power of these two financial assets for certain

G7 countries. Moreover, the significant long-run outlook implies that these two assets share no common trends in their economy system.

A study in 2011 by Li, Bin et al. [20] showed the existence of seasonality in the FOREX market, they studied eight currency, namely the Canadian dollar, EUR, Japanese yen, Swiss franc, Swedish krona, British pound, Australian dollar and New Zealand dollar; which are traded against the USD. They used statically test to measure the returns month by month of the currencies in the years from 1972 to 2010.

The researchers found that five currencies exhibit significantly higher returns in the month of December and a significant reversal in January, this is maybe due to the effect of the “January effect”; this is a commonly know effect in the stock market but this study also detects its presence in the FOREX market.

The study only used hard data and statistics to obtain these results, as for the source of this seasonality effect in the months of December and January is remarked as an open opportunity for other researchers to provide new findings.

## 4 Conclusions

This survey presents a review of computer science techniques for the FOREX market from the year 2000 up to date. The presented applications are mainly focused to rates forecast using a diverse variety of techniques, from natural language processing and statistical models to machine learning algorithms. In our study we found that:

- Since 2000 many researchers have been gaining interest in analyzing the FOREX market in order to obtain better profits.
- Due to the FOREX market being an OTC market the studies we found uses different databases, which makes difficult to compare results between them, so we propose to standardize the data used for benchmark, or even create a public repository of FOREX rates and news in order to have a comparison of results.
- The foreign exchanges rates are moistly noisy, non-stationary and chaotic, for this reason many researchers have choose to use machine learning and data mining techniques.
- The main task in FOREX market is to forecast the exchange values in order to generate a bigger profit, but since there are no official benchmarks is difficult to compare results between studies.

## References

1. Bank of International Settlement: Triennial Central Bank Survey - Foreign exchange turnover in April 2016, Monetary and Economic Department (2016)
2. Rey, H.: International trade and currency exchange. *The Review of Economic Studies* 68(2):443–464 (2001)

3. Banco de México: Resultados de la encuesta realizada por el Banco de Pagos Internacionales (BIS) sobre volúmenes de operación en los mercados cambiarios y de derivados durante abril de 2016 (2016)
4. Shravan, K., Ravi, V.: A survey of the applications of text mining in financial domain. *Knowledge-Based Systems* 114. pp. 128–147 (2016)
5. Batyrshin, I. Z., Sheremetov, L., Herrera-Avelar, R.: Perception based patterns in time series data mining. *Perception-based Data Mining and Decision Making in Economics and Finance*. pp. 85–118 (2007)
6. Atsalakis, G.S., Kimon P. V.: Surveying stock market forecasting techniques Part II: Soft computing methods. *Expert Systems with Applications* 36(3):5932–5941 (2009)
7. Hajizadeh, E., Ardakani, H. D., Shahrabi, J.: Application of data mining techniques in stock markets: A survey. *Journal of Economics and International Finance* 2(7) (2010)
8. Nassirtoussi, A. K.: Text mining of news-headlines for FOREX market prediction: A Multi-layer Dimension Reduction Algorithm with semantics and sentiment. *Expert Systems with Applications* 42(1):306–324 (2015)
9. Peramunetilleke, D., Wong. R.: Currency exchange rate forecasting from news headlines. *Australian Computer Science Communications* 24(2):131–139 (2002)
10. Idvall, P., Jonsson, C.: Algorithmic trading: Hidden Markov models on foreign exchange data. Master Thesis, Department of Mathematics, Linkopings Universitet (2008)
11. Baasher, A., Fakhr, M. W.: Forex daily trend prediction using machine learning techniques. In: 11th WSEAS Int. Conf. on Applied Computer Science (2011)
12. Kamruzzaman, J., Sarker, R.A.: ANN-based forecasting of foreign currency exchange rates. *Neural Information Processing-Letters and Reviews* 3(2):49–58 (2004)
13. Kamruzzaman, J., Sarker R. A.: Forecasting of currency exchange rates using ANN: A case study. In: *Neural Networks and Signal Processing, Proceedings of the 2003 International Conference on*. Vol. 1. IEEE (2003)
14. Batyrshin, I. Z., Solovyev V., Ivanov V.: Time series shape association measures and local trend association patterns. *Neurocomputing* 175, pp. 924–934 (2016)
15. Batyrshin, I. Z.: Moving approximation transform and local trend associations in time series data bases. In: *Perception-based Data Mining and Decision Making in Economics and Finance*. Springer Berlin Heidelberg, pp. 55–83 (2007)
16. <https://www.google.com/finance>
17. Garcia-Lopez, F. J., Batyrshin I. Z., Gelbukh, A. F.: Similitud de series de tiempo basada en longitud de patrones de la transformada por aproximación móvil. *Research in Computing Science* 115. pp. 79–92 (2016)
18. Kayacan, E., Ulutas, B., Kaynak, O.: Grey system theory-based models in time series prediction. *Expert systems with applications* 37(2):1784–1789 (2010)
19. Nieh, C., Lee, C.: Dynamic relationship between stock prices and exchange rates for G-7 countries. *The Quarterly Review of Economics and Finance* 41(4):477–490 (2002)
20. Li, B.: Monthly seasonality in currency returns. *JASSA* 3, pp. 1972–2010 (2011)