A comparison of Fundamental and Technical Analysis of Linear and Non-Linear models for FOREX prediction

USING AS PREDICTORS: PRICES OF COMMODITIES, BUSINESS INDEXES AND SENTIMENT DERIVED FROM NEWS

ANGELOS STAMATOPOULOS

SNR: 2017855
ANR: 645523

THESIS SUBMITTED IN PARTIAL FULFILMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE IN COMMUNICATION AND INFORMATION SCIENCES,
MASTER TRACK DATA SCIENCE: BUSINESS AND GOVERNANCE,
AT THE SCHOOL OF HUMANITIES AND DIGITAL SCIENCES
OF TILBURG UNIVERSITY

THESIS COMMITTEE:

DR. H.J. (HENRY) BRIGHTON

DR. EE (ELSKE) VAN DER VAART

Table of contents

Acknowledgement

I would like to thank all the people who contributed in some way to the work described in this thesis. First and foremost, I thank my academic advisor, dr. H.J. Brighton of the Department of Cognitive Science and Artificial Intelligence at Tilburg University. His guidance helped in all the time of research and writing this thesis. The door to prof. Brighton office was always open whenever I ran into a trouble spot or had a question about my research.

Besides my advisor, I would also like to thank the second reader, dr.ir. M.van Otterlo, for his encouragement and the insightful comments and hard questions.

Last but not least, I would like to thank my parents whose love is with me in whatever I pursue, my brother who supported me throughout my studies and my loving fiancé Noémie, who provides unending inspiration.

Abstract

Foreign currency exchange market (FOREX) is a highly volatile complex time-series for which predicting the next price is a challenging problem. In this paper, a twofold approach has been implemented; a comparison of the two different methods (technical vs. fundamental analysis) and a comparison between linear and non-linear models. Linear and non-linear univariate models were built for the technical analysis. For the fundamental one, macroeconomic features of commodities, business indexes, and also features derived from news sentiment were used. Deep Learning techniques (LSTM NN) were implemented to build the multivariate models, and three major forex currency pairs were investigated. This research combines the implementation of the LSTM algorithm which has not been used widely for forex prediction and the sentiment derived from general news (GDELT database). It is the first time that the data from the GDELT database are used in a forex prediction task. The results show the power of the non-linear neural network models for the forecasting task, the significant contribution of the sentiment derived from the news as a predictor and the necessity for multiple predictors in forecasting tasks (fundamental analysis), where models predict consecutive prices.

1. INTRODUCTION

1.1 Context

The presence of a primal form of the foreign exchange rate (Forex or FX) can be traced back in time at the beginning of bartering, when humans started to exchange commodities, food, and products for a number of common currency units. Nowadays, by “foreign exchange rate” we refer to the rate at which one country’s currency is exchanged with another country’s currency. Thus, it measures the number of units of one currency required to be exchanged with the number of units of another currency. A stable currency exchange unit has always been of great importance for the traders, as it successfully performs its function, as a means of exchange and as a storage of its value because it preserves its purchasing power stable. In our times, the need of the traders not only to hold stable currencies but also to know the price of currencies in the future is amplified. The increase of international trades and the fact that business operates on a worldwide scale are connected to the importance of foreign exchange. The forex market could be described as the backbone of international trade and investment. The role of forex remains critical in supporting imports and exports, which are necessary to gain access in resources and to create additional demand for goods and services. This is the reason also why an obvious upsurge can be noticed in economical dependencies between countries, organizations, and industries. In April of 2016, FOREX market was characterized as the biggest financial market where on an average, $5.09 trillion is traded daily, and the most commonly traded pairs are the following (Mozetič, Gabrovšek, & Novak, 2018):

United States Dollar and Euro (USD/EURO)
United States Dollar and Great Britain Pound (USD/GBP)
United States Dollar and Japanese Yen (USD/JPY)
United States Dollar and Swiss Franc (USD/CHF)
United States Dollar and Australian Dollar (USD/AUD)
United States Dollar and Canadian Dollar (USD/CAD)
United States Dollar and Brazilian Real (USD/BRL)

The importance of the forecasting could be described as a critical element of financial and managerial decision making, in order to reduce the risk and increase the efficiency of the decisions that are crucial for financial organizations, macroeconomic planning of countries and even for a firm or private investor. Exchange rates forecasting has been proven an extremely difficult task, and in general, it is considered as one of the most challenging modern applications of time series forecasting. Primarily steered by the intensive degree of noise, non-stationarity, deterministically chaotic and also by the semi-strong form of Efficient Market Hypothesis (Malkiel & Fama, 1970). The semi-strong form of Efficient Market Hypothesis (EMH), suggests that prices reflect all the publicly available information and the prices change instantly and accordingly to reflect the newly available information, and states that stock market prices are unpredictable. One of the most famous papers of the field was published in 1983, introducing the Random Walk model and, in which macroeconomic fundamentals were used and it was stated that the particular model was almost impossible to outperform (Meese & Rogoff, 1983). The authors proved that exchange rate forecasts which were based on structural models performed worse than a naive random walk, where the unknown price of the next day follows a random pattern based on today's price. Since then, a lot of different techniques and methods have been implemented in order to solve this forecasting puzzle. All the methods that have been used and developed can be divided into two categories (Montañez & Antonio, 2011):

- Fundamental analysis – in this case, the analysis is based on the exact knowledge of the factors that have an effect on the economy and the relationship between them. The analysis is based profoundly on the financial condition of the country while focusing on the effect of the balance between supply and demand on each currency. Empirical models like the monetary model, the hybrid model, and the currency substitution model are some examples of such analysis.

- Technical analysis – this kind of analysis and its prediction focuses on the assumption of existence and discovery of empirical regularities that emerge from historical data and using various methods, like time series analysis, regression, machine learning, neural networks and more. The predictions produced by technical analysis assume that future movement of prices follows trends that can be forecasted.

The primary goal of this thesis is to examine the difference between technical analysis and fundamental analysis for the prediction of foreign exchange rates of three pairs: USD/JPY (United States Dollar and Japanese Yen), USD/CHF (United States Dollar and Swiss Franc), USD/BRL (United States Dollar and Brazilian Real). For the fundamental analysis, a wide range of features was selected including the sentiment of news from one country towards the other, commodities and business indexes. Also, differences between linear and non-linear models are also going to be examined, by comparing their efficiency in predicting the next price and also testing their ability to forecast on a long-term horizon.

1.2 Research Questions

a. Which forecasting method and what types of models should be chosen to predict the next price? In other words, technical or fundamental analysis, and linear or non-linear models are more efficient for the forex prediction task?

b. Can random walk still be considered invincible?

c. Can commodities indexes, business indexes and sentiment from the news, combined or individually, improve the prediction of Forex rates?

d. Could sentiment derived from news be considered as a good indicator?

e. How do models perform in predicting prices far in the future rather than the next price?

1.3 Structure

This paragraph provides a brief overview of the structure of this research paper. In the first part of the introduction, the research questions were posed. The second part, Literature Review, is composed of sections where fundamental concepts are presented, and they describe methods, models, and findings in the literature regarding financial predictions. In a few words, fundamental analysis, commodities and business indicators, linear/non-linear models, Artificial Neural Networks (ANNs) and Natural Language based Financial Forecasting (NLFF) are decomposed. The third part consists of the design of the experimental method, and the tools that were used for this research are analyzed. Next, in the fourth and fifth part, the models that were built are presented, followed by the sixth part, where the results of the experiments are described. Finally, in the seventh part, the conclusions of this research are discussed, and the thesis concludes with a small note for further research suggestions.

2. LITERATURE REVIEW

2.1 Fundamental analysis

Regarding fundamental analysis, as described in the previous section, the prediction is based on the exact economic, social and political factors that influence the prices of a financial asset. However, do we know which those factors are? The problem with this approach is that the understanding of the rules that govern the forex behavior is not readily available (Lolis, Kodogiannis, 2001). Moreover, economic fundamentals – like trade balances, money supply, national income and other key variables – have not been found useful for the forex prediction. It was justified by the fact that the models are not only influenced by the past values of the fundamentals used, but also by the expectation of the future values of them (J. Wang, 2008). An explanation of this particular behavior could be attributed to the approach that economic models are using, as their coefficients aim to clarify the relationship between the fundamentals and the exchange rate. There had been an attempt to explain and analyze this instability and weakness in the relationship between exchange rate prediction and fundamentals by the proposed scapegoat theory (Bacchetta, & Wincoop, 2004). The authors suggested that although agents have a rather good intuition about the relationship between exchange rate and fundamentals, in the long run, there is evidence for high-level uncertainty over the short to medium term. This entails the risk that when foreign exchange rate movements are inconsistent in comparison with their priors about the underlying structural relationship, then scapegoats are blamed for these inconsistencies. The following quote pictures accurately this phenomenon of participants in the financial market blaming scapegoats (individual fundamentals):

"The FX market sometimes seems like a serial monogamist. It concentrates on one issue at a time, but the issue is replaced frequently. Dollar weakness and US policy have captured its heart. But uncertainties are being resolved … The market may move back to an earlier love …" [Financial Times, November 8, 2010].

2.2 Commodities & Leading Indicators as Predictors

As a potential macroeconomic fundamental, prices of commodities were used for prediction of foreign exchange rates of countries whose exports share was highly based on commodities, such as Australia, New Zealand and Canada (Y. Chen & Rogoff, 2003). The exchange rates are endogenously determined in equilibrium alongside with more macroeconomic variables, and it was considered hard to predict changes of exchange rates based on reduced-form models.

The hypothesis tested, suggested that if it was possible to identify an exogenous shock - in economics, that is a shock positive or negative shock by unexplained factors - due to fluctuations of the prices of the commodities, then they could also predict fluctuations in the actual exchange rates. Therefore, each one of the aforementioned countries would normally experience an increase in their exchange rates appreciations in case of an increase in the price of the commodity they are exporting. Chen and Rogoff proved that commodity price indices could improve the prediction for in-sample and out-of-sample data. On the contrary, in a later research, indexes of commodities were characterized as not significant predictors for out-of-sample data when they were used with quarterly data (Y.C. Chen, Rogoff, & Rossi, 2010). On the other hand, they were found as good predictors in out-of-sample daily data. (Ferraro, Rogoff, & Rossi, 2015).

Leading indicators is another category of macroeconomic fundamentals that are mainly used in predicting trends of financial markets, and in the case of foreign exchange rates, they are widely used for crises predictions by econometricians. In econometric models, it is not uncommon to use twenty or more indicators on their models. Leading indicators can reveal the trend towards the economy is headed and many different participants in the economy are using them to obtain insights. Investors use them to adjust investment strategies to benefit from future conditions of the market, while policymakers use them for considering adjustments to monetary policy. Businesses use them to calculate future revenues, by having insights for future economic conditions. In the existing literature, findings regarding leading indicators are controversial. Famously used indicators, such as Gross Domestic Product (GDP) and Purchasing Power Parities (PPP), were found not that useful in improving the predictive power of models (Kilian & Taylor, 2003); actually, they were affecting the prediction negatively. On the contrary, evidence was found that the Business Cycles Index (BCI) was correlated with stock prices and exchange rates, and also in the same study, the Consumer Confidence Index (CLI) was used without reporting any specific finding for that one (Andersen, Bollerslev, Diebold, & Vega, 2007).

2.3 Comparing linear and non-linear models

In general, financial forecasting and, particularly, the exchange rates prediction remains a subject of interest in both the academic community and the economic community. Even though, as mentioned, they are highly volatile which makes them very difficult to predict over short to medium time frames (Nag & Mitra, 2002).

In recent years, the research mainly focuses on technical analysis, and a variety of models and techniques have been implemented, namely linear models, non-linear, SVM, decision trees, decision forests, Artificial Neural Networks and many hybrid models. In this domain, research is characterized by many complications. Researchers have indeed many decisions to make, and this is the main reason for the variety of studies. A great diversity can be observed in the selection of the currencies, in the predictors selected, the time period examined, the metrics of evaluation reported, and the forecast horizon chosen to predict for each experiment and research (J. Wang, 2008).

The most commonly used linear model in the literature for foreign exchange predictions is Autoregressive Integrated Moving Average (ARIMA). These models combine autoregressive models and moving average models as special cases to deal with non-stationary data. Stationarity implies that the time-series remain at a fairly constant level over time. In economics though, it is common to identify a trend in the data, which declares that the data are non-stationary. The main application of the Autoregressive Integrated Moving Average (ARIMA) models is in the area of short-term forecasting, and it functions better when the data exhibits a consistent pattern over time with a minimum number of outliers. Improvements in performance obtained from using linear and non-linear univariate time-series models are minimal over forecasts generated by a random walk model (Brooks, 1997). Nowadays, ARIMA models are still being used for statistical analysis, as one of their advantages remains their simplicity to use and their broad acceptance in economics and financial studies. Moreover, it is considered one of the standard benchmark models.

Nonlinear models that were used at an early stage showed mixed results and acceptance from the scientific community. The computational power and the models that could be applied until recently were not sufficient to beat linear and random walk models. Boero and Marrocu (2002) published a comparative study comparing the forecasting performance of three traded exchange rates (the French franc, the German mark, and the Japanese yen) against the US dollar. They used three univariate non-linear time-series models, which were compared and contrasted against two linear models, primarily Autoregressive Integrated Moving Average (ARIMA) models, and random walk models. The results show that the advantages of non-linear models over linear ones lie in the criteria used to assess forecast accuracy. The authors conclude by stating that in their analysis non-linear models generate more forecasting gains than linear ones.

2.4 Artificial Neural Networks (ANNs)

Recent research has shown that Artificial Neural Networks (ANNs) are much more efficient in the out of sample predictive abilities comparing to linear time-series models for financial assets. ANNs outperformed random walk and linear models based on a number of recursive out-of-sample forecasts in both error metrics and the percentage of the correctly predicted direction of signs. (Gradojevic & Yang, 2006)

Artificial Neural Networks (ANNs) are well-known for their advantage of being self-adaptive and heavily data-driven, as the amount of data required to be successfully trained is very large in size. As a general conclusion of the literature so far, we could say that ANNs are a valuable tool for cases of nonlinear models, like foreign exchange rates, and especially their nonlinear nature is considered as a significant advantage as there is no need to specify their functional relationship between input and output variables (Nag & Mitra, 2002). Negative side of the neural networks, as mentioned by the majority of the researchers is the fact that it is easy for these models to get overfitted and their functionality usually is described as a black box, wherein some occasions it is quite hard to demystify the accurate way the outputs are produced (Benitez, Castro, & Requena, 1997)

Rudy, Dunis, Giorgioni, & Laws (2010) conducted their research using daily data from 2000 to 2009. Their test out of sample period included the last two years, from 2007 to 2009 (financial crisis period) and the purpose of their work was to examine whether neural network models are still superior to all other models for forecasting EUR/USD exchange rates. Two different types of Artificial Neural Networks (a Multilayer Perceptron and a Recurrent Neural Network) were compared against a random walk and ARIMA models. The result showed that still, neural networks were superior to traditional methods. Recent works in forecasting exchange involve accuracy in prediction, where the accuracy is calculated as a percentage of correctly identified the trend. It is not uncommon also, for a naïve trading algorithm to be provided, made of simple trading rules providing sell or buy signals, based on the value of the prediction. The accuracy of these signals was measured and also a Return on Investment (ROI) was calculated (Hafezi, Shahrabi, & Hadavandi, 2015).

Wong, Xia, & Chu (2010) proposed a novel Adaptive Neural Network (ADNN). The adaptive metrics of inputs and outputs that were developed could avoid overfitting and produce more accurate predictions in periodical time-series with complicated structure. Their model outperformed the AR (Auto-Regression), the ANN (Artificial Neural Network) and the AKN (Adaptive K-nearest neighbors) models with high robustness for both chaotic and real time-series predictions. Sun & Chang (2017) applied a special case of Radial Basis Function Neural Network (RBFNN) to forecast exchange rates of USD/EUR, USD/CHF, and JPY/USD. A radial basis function is an artificial neural network that uses radial basis functions as activation functions. Radial basis function could be described as a function whose values are real numbers and depend on the distance from another point called center; the norm is usually Euclidean distance. The output of the network is a linear combination of radial basis functions of the inputs and neuron parameters. The specific neural network can detect the complex nonlinear relationships between dependent and independent variables. The learning algorithm used in the model uses a diverse dataset for training to adapt itself quickly to the new exchange data rate. The authors compared the performance of MLP and RBF network models where they found that among neural network models RBF do considerably better than MLP.

Fischer, & Krauss (2018) used a special type of Recurrent Neural Networks, the Long Short-Term Memory Neural Network (LSTM) to predict out-of-sample directional movements of the constituent stocks of S&P 500. The LSTM neural network's structure is composed of a cell, an input gate and a forget gate. The cell is responsible for memorizing values over arbitrary time intervals, and the forget gate is responsible for dropping values that are no longer computationally necessary. Each of the three gates can be considered a conventional artificial neuron; they compute an activation function of a weighted sum. The researchers compared their findings with other classification algorithms without memory, such as Logistic Regression Classifier (LOG), Random Forest Classifier (RAF) and Deep Neural Network (DNN). The LSTM Neural Network outperformed all the models and was found a suitable algorithm for noisy financial time-series data due to its ability to capture, memorize and extract meaningful information from patterns. Long Short-Term Memory networks have been found suitable for time series and sequence learning and have become the state-of-the-art models for a wide range of applications in various fields (Greff, Srivastava, Koutník, Steunebrink, & Schmidhuber, 2017).

2.5 Natural Language based Financial Forecasting (NLFF)

Another experimental field that has emerged over the last decade is the use of machine learning techniques to conduct sentiment analysis from the news, and the usage of this information to predict financial assets. It is common knowledge that markets thrive on new incoming information of all kinds, whether it is just rumors, eavesdropping or even scandals. However, financial or political news/articles are considered to report the available information and to be a much more stable and trustworthy source.

One of the first types of research on the specific field was conducted by Wuthrich et al. (1998). They used data mined from the web regarding general news, in order to predict the closing value of major stock market indices in Asia, Europe, and America. They concluded that the information extracted from the news not only increased the quality of the input, but also provided better predictive results. Therefore, their research proved highly promising and was one of the most iconic in the field, which combines Natural Language Processing and Financial Forecasting (NLPFF). Their research showed that news had significant information and revealed the potential for a further future investigation.

The specific field for some years remained inactive, but because of the improvements in computational capacity, the development of techniques and the massive availability **on** public information, research of the specific field has been revived (Zhang, Patuwo, and Hu, 2009). Schumaker, Zhang, Huang, & Chen(2012) used the Arizona Financial Text (AZF in Text) system for their analysis, a financial news article prediction system, and paired it with a sentiment analysis tool. They found that news with a negative tone had a greater impact on the stock market and their prediction of downward trends was more accurate than those of the positive trends. Their study focused on classifying the sentiment of news and then using that information to predict the movement of stocks. In another research, Hagenau, Liebmann, & Neumann(2013) used text mining techniques on financial news, and they showed that a robust feature selection could significantly improve the results of the classification task that they performed. Their model was also found profitable for trading, as their naive trading algorithm could provide high returns based on the highly successful trend capturing model, given that they reported that the accuracy of the model was up to 76%.

Khadjeh Nassirtoussi, Aghabozorgi, Ying Wah, & Ngo(2015), they conducted one of the first exploration efforts of the predictive relationships of news and the FOREX market. Their most significant finding was that they managed to prove such a relationship and they characterized it as a successful feasibility test. Their focus was on predicting the directional movement of a currency pair in the foreign exchange market based on the words used on news headlines using natural language processing and sentiment analysis. As a result, first and foremost, their work showed the predictive power of news on the specific financial asset. They also provided a complete system that could be used by researchers as a framework to build upon and to investors or organizations to comprehend the foreign exchange market better.

3. Experimental Setup (Method)

This chapter describes the data used, the design of the Random Walk models, the process followed, and the design of the Autoregressive Integrated Moving Average (ARIMA) models and the architecture of the Long Short-Term Memory neural network (LSTM NN) models for the exchange rates prediction. In order to answer the research questions regarding the comparison of linear/non-linear models and the contrast between different indicators as predictors, technical and fundamental analysis were employed. For the technical analysis the prices of the exchange rates in correlation with time, also known as univariate time-series, were used. For the fundamental analysis, the Long Short-Term Memory neural network (LSTM NN) was applied, in order to obtain results related to the different indicators and explore their suitability concerning the forex prediction task.

3.1 Data

For this research, multiple resources were used for the collection of data, and in this section, we will explore and examine them thoroughly. Three exchange rates were chosen USD/BRL (United States Dollar - Brazilian Real), USD/CHF (United States Dollar – Swiss Franc) and USD/JPY (United States Dollar – Japanese Yen) for the period 1/1/2005 – 31/12/2015, and the foreign exchange data were retrieved via the yahoo finance website. The exchange rates mentioned above were chosen for two reasons; firstly, all of them have been used in the literature for predictions and secondly, all of these economies are based on exports. Therefore, the commodity prices are expected to influence the fluctuations of their currencies. For the same period (01/01/2005-31/12/2015), commodities indexes and business indicators were downloaded from the OECD.com. The commodity indexes available included:

- the Agricultural Raw Material Index,
- the Beverage Index,
- the Food Index,
- the Beverage and Food Index,
- the Energy Price Index,
- the Non-Fuel Index,

- the Crude Oil Index,
- the Industrial Inputs Price Index,
- the Metal Price Index,
- and the Commodities Price Index.

Following the literature closely, the business indicators chosen were the CLI, BCI and CCI (Berge, 2015).

- The composite leading indicator (CLI) is designed to provide early signals of turning points in business cycles showing the fluctuation of the economic activity around its long-term potential level. CLIs show short-term economic movements in qualitative rather than quantitative terms.

- The consumer confidence index (CCI) is based on household plans for major purchases and their economic situation, both currently and their expectations for the immediate future. Opinions compared to a “normal” state are collected and the difference between positive and negative answers provides a qualitative index on economic conditions.

- The business confidence index (BCI) is based on enterprises' assessment of production, orders and stocks, as well as its current position and expectations for the immediate future. Opinions compared to a “normal” state are collected and the difference between positive and negative answers provides a qualitative index on economic conditions.

Furthermore, data regarding sentiment based on the news were retrieved from Global Database Events Language and Tone (GDELT) via SQL (structured query programming language) for each one of the three pairs of currency examined (USD/BRL, USD/CHF, USD/JPY). As explained in the official webpage of the GDELT Project, the database monitors the world's broadcast, print, and web news from all over the world in over 100 languages and identifies the people, locations, organizations, themes, sources, emotions, counts, quotes, images and events driving our global society every second of every day. Indeed, it is a massive database the archives of which start from 1/1/1979 and date up to now. For the year 2015 alone, the records of the database are estimated to be about three-quarters of a trillion. The database was created to provide a free open platform for computing in the entire world (Leetaru, 2014). More specifically, for each pair of currencies for the examined period (2005-2015), general news reflecting the interaction between the two countries were selected. The columns containing the described information, as recorded in the database's manual (CAMEO) are:

- GoldsteinScale. (floating point) Each CAMEO event code is assigned a numeric score from -10 to +10, capturing the theoretical potential impact that the specific type of event will have on the stability of a country.

- NumMentions. (integer) This is the total number of mentions of this event across all source documents during the 15-minute update in which it was first seen. Multiple references to an event within a single document also contribute to this count. This can be used as a method of assessing the "importance" of an event: the more discussion of that event, the more likely it is to be significant.
- NumSources. (integer) This is the total number of information sources containing one or more mentions of this event during the 15-minute update in which it was first seen. This can be used as a method of assessing the "importance" of an event: the more discussion of that event, the more likely it is to be significant.

- NumArticles. (integer) This is the total number of source documents containing one or more mentions of this event during the 15-minute update in which it was first seen. This can be used as a method of assessing the "importance" of an event: the more discussion of that event, the more likely it is to be significant.

- AvgTone. (numeric) This is the average "tone" of all documents containing one or more mentions of this event during the 15-minute update in which it was first seen. The score ranges from -100 (extremely negative) to +100 (extremely positive). Common values range between -10 and +10, with 0 indicating neutral.

To sum up, data from the sources described were combined to create three datasets, one for every exchange rate pair. For all three datasets, the Date column and the Prices of Commodities Indexes were identical. The three Indicators BCI, CLI and CCI used for each dataset were unique for every country. Therefore, each dataset was composed of six of the indicators, half of them for the first country's currency rate and the other half of its pair country currency rate. Both Commodities Indexes and Business Indicators prices data were monthly. Each foreign exchange rate pair data was daily, and data from GDELT were also daily, but for each day there were multiple events recorded.

Below, table 1, shows the total number of data gathered for this research. In table 2, the different type of features for the prediction of the three exchange rates (USD/BRL, USD/CHF, USD/JPY) are presented. Note that for this research three different datasets were created, one for every exchange currency pair.

Table 1

Size of data used in each forex dataset (USD/BRL, USD/CHF, USD/JPY), for the time period examined 01/01/2005-31/12/2015

	Forex currencies		
	USD/BRL	USD/CHF	USD/JPY
Exchange Rates (Daily prices)	2 875	2875	2875
Monthly Business Indexes (BCI, CCI, CLI)	120	120	120
Weekly Commodities Indexes	520	520	520
GDELT	729 615	162 338	189 156

Table 2

Features of each dataset

Date	Exchange Rate Pair	Agr_Raw_Mat	Bev_Price	Comm.Price	Energy Price
Food_Bev Price	Food_Price	Indus_Input Price	Metal_Price	Non_Fuel Price	Crud_Oil Price
Countr1 BCI	Country1 CCI	Country1 CLI	Country2 BCI	Country2 CCI	Country2 CLI
Country1 Avg_Tone	Country1 Num_Sources	Country1 Num_Articles	Country1 Num_Mentions	Country1 Goldstein_Scale	Country2 Avg_Tone
Country2 Num_Sources	Country2 Num_Articles	Country2 Num_Mentions	Country2 Goldstein_Scale		

Note: In this table, where Country1 is mentioned is always data from the US economy and where Country2 is mentioned, is always data of the other economies (BRL/CHF/JPY). Therefore, 3 datasets were created (USD/BRL, USD/CHF, USD/JPY)

3.2 Preprocessing

The major challenge during the preprocessing phase was how to combine daily financial data with the data gathered from GDELT. As can be observed from Table 1, the data from GDELT exceeds the number of all of the remaining financial data, which means that per day there were approximately 80-100 daily news regarding the examined economies. All that information had to be filtered and compressed in order to have one value per day and to be able to be combined with the rest of the data. The preprocessing phase was implemented using the R language (see Appendix A). Most of the data gathered, were in a relatively proper format by default, as they are financial data.

However, GDELT’s data that were selected required normalization before using it for analysis, as it is advised in the CAMEO manual book that accompanies the database (“The GDELT Project”).

Not only could the size of the data not fit the other datasets from the financial sources, but also it was observed that events could have much publicity with neutral impact, therefore they would have a large number of Number of Mentions, Number of Sources and Number of Articles and zero Average Tone. Alternatively, the other extreme was a minimal Number of Mentions, Number of Sources and Number of Articles accompanied with extreme positive or negative value of Average Tone. In other words, events that had much publicity had non-important tone, and events that had almost no publicity had a very extreme value of tone, either positive or negative. At this point, it is essential to mention that this problem and any suggested solutions were not available in the literature examined.

As can be deducted, the less important and published news with extreme values of tone could potentially add extra noise to the already noisy data. A simple average for the GDELT data could not work, since the problems mentioned above would affect the results and change the actual values that carry the information. To tackle these issues, and since the GDELT dataset had never been used before for exchange rate prediction, at least in the literature research that has been conducted, the data were transformed using the following transformation method:

$$New\_Avg\_Tone = (Num\_Mentions + Num\_Sources) * Num\_Articles * Avg\_Tone \qquad (1)$$

By transforming the data with the previous eq. (1), the Number of Articles made a crucial contribution to the final result. The news that went more popular and according to the sign of their Avg_Tone had a chance to have a higher or lower New_Avg_Tone. Therefore, the news that could cause a real impact was separated from those who could not impact the market. The news with more extreme Tone values, either extremely positive or extremely negative, but their publicity was too shallow to include any influencing power to the foreign exchange rate market.

After the transformation, all the features that were going to be used as input for the models (all except Date and Exchange Rate Pair), were normalized ranging from 1, to -1. Then, the GDELT data were gathered by day and an average per day of New_Avg_Tone and Goldstein_Scale was calculated so that they could correspond with the rest of the financial data. Each one of the three final datasets had the following format, as it is described in table 3. For a better understanding of table 3, the exchange rate pair is one the exchange rates examined (USD/BRL, USD/CHF, USD/JPY) and the commodities prices are in common for each data set. The industrial indexes (BCI, CCI, CLI) are corresponding to each country. Finally, the transformed GDELT's data are describing the new average tone calculated by eq. (1) and the Goldstein scale. Each country's news had different average tone and Goldstein scale towards the other country in each case.

Table 3

Features of each dataset, after the GDELT's data transformation.

Date	Exchange Rate Pair	Agr_Raw_Mat	Bev_Price	Energy_Price	Comm_ Price
Food_Bev_Price	Food_Price	Indus_Input_Price	Metal_Price	Non_fuel_Price	Crud_Oil_ Price
Country_1_ BCI	Country_1_ CCI	Country_1_CLI	Country_2_ BCI	Country_2_ CCI	Country_2_ CLI
Country_1_New Avg_Tone	Country_1_ Goldstein	Country_2_New_ Avg_Tone	Country_2_ Goldstein		

Note: The GDELT's data, Country_1_New_Avg_Tone, and Country1_Goldstein are describing the tone and importance of news from Country1 towards Country2 and vice versa.

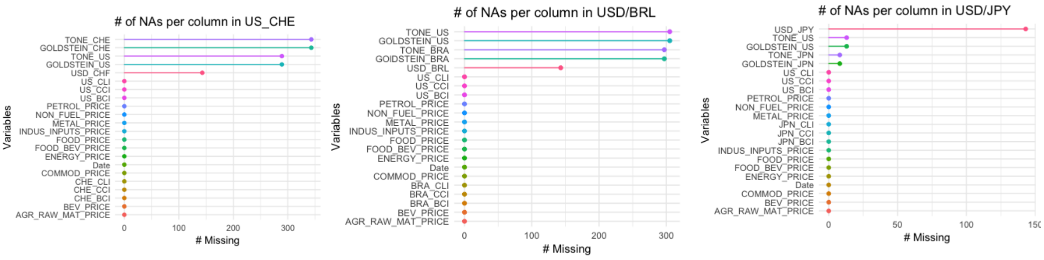
3.3 Missing values

In the full datasets created, missing values were generated since commodities indexes data were of weekly frequency and the Business Indexes were of monthly frequency. The method selected for the imputation is the cubic spline interpolation, as proposed by Ajao et al. (2012). One advantage of the spline basis is that it can handle non-uniform spacing of samples, and the robust results of the procedure were proved to be a powerful data analysis tool. The objective in cubic splines is to derive a third-order polynomial for each interval between knots, as in:

$$f(x_i) = a(x - x_i)^3 + b_i(x - x_i)^2 + c_i(x - x_i) + d_i \qquad (2)$$

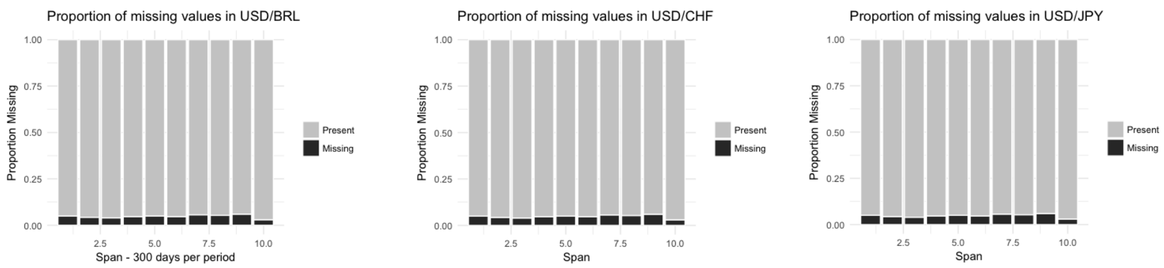
Thus, for n +1 data points (i = 0,1, 2, … n), there are n intervals and, consequently, 4n unknown constants to evaluate. As it can be observed in figure 1, after the imputation of weekly and monthly data to daily, the number of NAs, for the commodities Indexes and the Business Indexes features dropped to 0. Data from GDELT though included missing values as for some days there were not any available data.

Figure 1: Total number of Missing values after the cubic spline interpolation of the financial data. Financial data were imputed to daily.



The remaining missing values, as pictured in figure 1, exist in the Exchange rate pair, which can be explained since in the period of 11 years (2005-2015) the market did not operate during weekends. Some missing values have been identified and in the GDELT data, due to the fact that some days there was no news recorded. The proportion of missing data was distributed equally during the time, as it is evident in figure 2. The timespan of every dataset was split into ten periods (x-axis) and on the y-axis shows the proportion of missing values for each period. The remaining missing values were treated with the exact same method of cubic spline interpolation was used, as implemented for the financial data in eq. (2).

Figure 2: Proportion of missing values in each dataset per period (300 days)



The final three datasets after the preprocessing had no missing values. All the data were imputed to daily, so even for weekends, when markets were closed, prices for exchange rates were generated through imputation**.** The reason that prices of the exchange rates had to be generated for weekends as well was the fact that data from GDELT included weekends. The news is produced on a daily basis, and the information is continuous in comparison with financial markets that remain closed during weekends. In order for the relationship between news, commodities and business indexes regarding their predictive power with the three selected foreign exchange rates to be examined, the datasets were transformed to include daily prices of every feature for every day of the year, for each of the ten years of the sample.

3.4 Input and Output Selection

This research paper could be divided into two sections, the univariate time-series analysis/prediction (technical analysis) and the multivariate one (fundamental analysis). This section aims to clarify which inputs were used in each model -experiment and which steps were followed. For the task of foreign exchange rate prediction, especially when using neural networks, it is a common tactic to perform it with only one input (univariate time-series) without affecting the predictive power of the model (Huang, Lai, Nakamori, & Wang, 2004). On the other hand, when performing fundamental analysis, there is not any specific rule regarding the choices of the variables/features that will be selected, and it entirely depends on the researcher.

Raw financial inputs and outputs could also be used for the models, especially in the case of neural network algorithms. Further preprocessing of the data, though, has been found to improve the predictive power of the neural network models. The steps for that further processing, as proposed by Huang, Lai, Nakamori, & Wang (2004), and will be investigated further below are the following (every step is analyzed in the following sections 3.4.1 - 3.4.5):

- Sampling
- Transforming
- Normalizing
- Dividing
- Windowing

3.4.1 Sampling

During the sampling phase, sample size and sampling rate have to be decided. Sampling size is an important decision to make, and Zhang & Hu (1998) showed in their research that a larger sample, improves the prediction of the model. In their study for exchange rate prediction, they fed an ANN (Artificial Neural Network) algorithm two samples, a large one of 887 data points, which outperformed a much smaller one of 261 data points. In the literature, there is a variety of different data sampling approaches for the specific task and researchers have used different sizes of datasets ranging from 2 to over 16 years. Walczak (2001), concluded that any period of more than two years of daily forex data, or more than 730 data points, is enough for training and testing models for the prediction.

The sampling rate is another decision that must also be made by the researcher, and it also varies in studies of the field and defines at what rate the raw data is sampled in terms of time. As stated in the 'Data' section, GDELT were daily (weekends incl.), and also the exchange rates were gathered daily (no weekends incl.). Commodities and Business Indexes were available only weekly and monthly, respectively. All the data were imputed to daily (weekends included), according to the cubic spline imputation method described in 'Preprocessing' part. The sampling rate for the data of exchange rates were the closing prices of each working day of the forex market, as it is widely used by many other researchers (Liu & Wang, 2008; Huang et al., 2004).

In this research, 11 years of daily exchange rates were used, namely from 1/1/2005-31/12/2015. Given the amount of data and the amount of noise, it was decided for the models to be built on a weekly basis. Therefore, all the available data were resampled on a weekly basis. Considering the literature above, a final sample of almost 3000 data points is completely reasonable, and the data gathered are more than enough to train and evaluate the models.

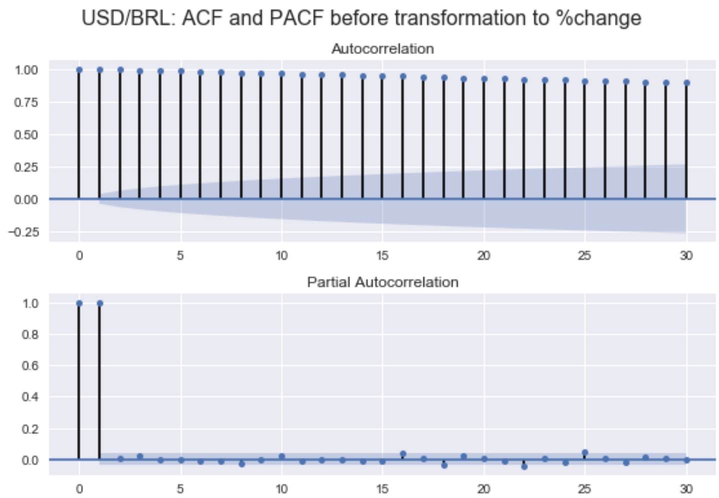
3.4.2 Transforming

Dunis & Williams (2002) argued that transforming the price can upgrade significantly the prediction model's forecasting ability, compared to working directly on the raw price. The nature of foreign exchange prices is found to be nonstationary and random since the early stages of research (Meese & Rogoff, 1983). The main purpose of any transformation is to make the distribution of the transformed data symmetrical and as close as possible to produce a shape of a normal distribution. De Bodt, Rynkiewicz, & Cottrell (2004) concluded that any forecasting or modeling attempt with financial time-series should be based on successive variations of price and not on the prices directly. A popular measure of successive variation method that is often used is referred to as the return (Dunis et al, 2004) and is calculated with the use of the following formula:

$$R_t = \frac{P_{t+1} - P_t}{P_t} \qquad (3)$$

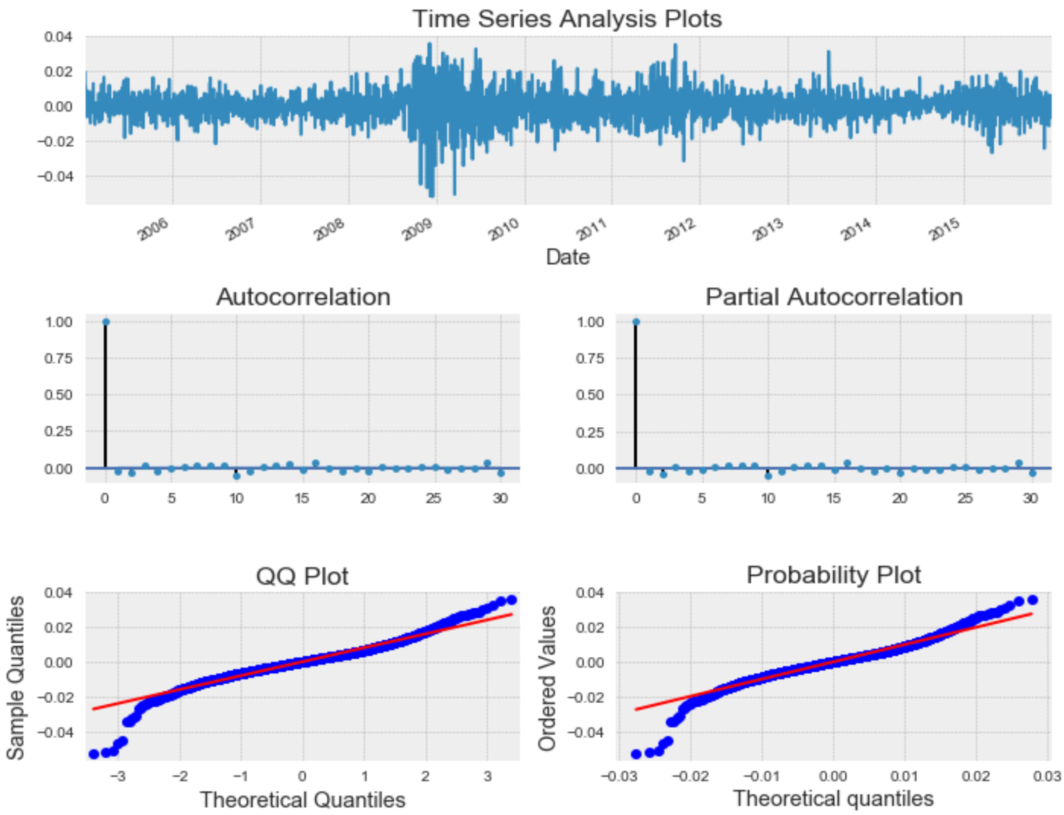
where R_t is the return on time t and P_t is the price on time t.

Figure 3: Autocorrelation and Partial Autocorrelation in USD/BRL exchange rate



Dunis et al., also showed that by transforming the data in returns, the time-series becomes stationary. All three datasets before the transformation of the exchange rate prices appeared to have similar characteristics with high autocorrelation and partial autocorrelation, but after the transformation, the data were ready to be modeled. Below, plots for the exchange pair of USD/BRL are provided indicatively, as the plots for the other two exchange rates are identical. Figure 3 shows the presence of Autocorrelation for the USD/BRL foreign exchange rate throughout the sample and Figure 4 shows the transformation to stationary data. Note that the other two pairs had identical plots for autocorrelation and partial autocorrelation, that was eliminated by the same transformation (as described for the USD/BRL pair).

Figure 4: Autocorrelation and Partial Autocorrelation in USD/BRL foreign exchange after transformation



3.4.3 Normalizing

Normalization is the statistical process of adjusting the values, measured in different scales to a common scale. Most of the studies using advanced machine learning or deep learning algorithms rely on this process for the input data, but not all researchers agree on the necessity of this.

El Shazly et al. (1997), concluded that data normalization did not provide any better results on their neural networks tests for the classification task of foreign exchange rates. Zhang and Hu (1998), also shared the same opinion on this subject, and they did not normalize the data for the regression task of foreign exchange rates. On the contrary, a vast majority of researchers in the field support that normalization, although not mandatory, should become so since it can minimize any potential computational problems. The major advantage of normalized data is that higher values affect small ones less in the training phase, and as a result, the prediction error will be smaller (Shrinivasan et al., 1998).

Notably, for multivariate time-series models, it is mandatory to normalize all the input data in order for different variables tl acquire the appropriate weight for the prediction task. For the normalization task the formula that was used for all the inputs of the models is a common practice for financial time-series (Lachtermacher, 1995):

$$x_n = \frac{x_o}{x_{max}} \qquad (4)$$

where x_n and x_o represent the normalized and the original data respectively. In this research, all the input data in every model are normalized with the described formula.

3.4.4 Dividing

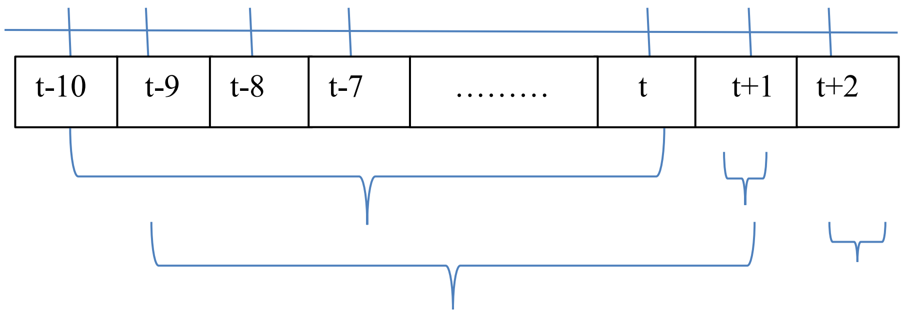
Dividing data into subsets with appropriate sizes for training and testing is one more decision that has to be made by the researcher, as there is no common solution or approach. The primary factor to be considered for that decision is the data size. Yao & Tan, (2000) proposed that the training set should consist 70% of the data, 20% for validation set and finally, 10% for the test set, according to their experience. Other researchers though support that for data samples greater than 1500 data points, 80% of the dataset should be used for training and the rest 20% for validation and testing. Kim (2003) and Tay & Cao (2001), both studies used the 80% - 20% division rule, and this research will also follow this rule. Below, graphs of the discussed division are presented (Figure 5).

Figure 5: The division of data into training and test set

3.4.5 Windowing

Windowing is the process whereby the number of successive data inputs is selected that is used in the model to predict the subsequent value each time. An example of a rolling window process is the following: if a window of size three is chosen, then to predict the price $\hat{p}_t$, the three successive samples of p_{t-1}, p_{t-2} and p_{t-3} will be used as inputs to produce the output of $\hat{p}_t$. In this research, the window of size ten (10) was used in the non-linear models, and more specifically in the LSTM models of this paper, both univariate and multivariate ones. The size of the sliding window was taken arbitrarily, as two more sizes tested, one of 5 and one larger of 15 and both showed poorer predictive potential than the one of size 10.

Figure 6: Illustration of windowing process of size 10



3.5 Performance Measures – Errors

For this research and, more precisely, for the evaluation of the models, four types of error were chosen, the RMSE (Root Mean Squared Error), the MAE (Mean Absolute Error), the MAPE (Mean Absolute Percentage Error) and the R^2 (R squared).

The first measure that quantifies accuracy is the regression error, which illustrates the amount of deviation as an error between the actual value and the predicted value. The standard regression error in statistics and machine/deep learning is the Mean Square Error (MSE) (Z. Wang & Bovik, 2009), defined by:

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (Y_i - \widehat{Y_i})^2 \qquad (5)$$

$$RMSE\ (\hat{\vartheta}) = \sqrt{MSE\ (\hat{\vartheta})} \qquad (6)$$

The MSE is a measure of the quality of an estimator, and if its value is closer to zero, the model described the better is. It is considered the second moment (a specific quantitative measure, used both in statistics and mechanics, of the shape of a set of data points), and it incorporates both the variance of an estimator and its bias.

The RMSE is a measure frequently used. RMSE describes the differences between values predicted by a model or an estimator (sample or population values) and the values observed. It represents the square root of the differences between predicted values and the observed ones, and it serves to aggregate the magnitudes of the errors in predictions for various times into a single measure of predictive power. These individual differences are called *prediction errors* when computed out-of-sample. RMSE is a measure of accuracy to compare forecasting errors of different models for a particular dataset and not between datasets, as it is scale-dependent. (Willmott & Matsuura, 2005)

The second measure selected is the MAPE (Mean Absolute Percentage Error), usually expressing the accuracy of the prediction as a percentage, and it is defined by the following formula:

$$MAPE = \frac{100\%}{n} \sum_{t=1}^{n} \left| \frac{A_t - F_t}{A_t} \right| \qquad (7)$$

where A_t is the actual, n is the size of the sample and F_t is the forecasted one. The difference between the two values, A_t and F_t, is divided by the actual value A_t again. The absolute value is summed for every forecasted point in time and divided by the number of fitted points n. Multiplying by 100%, we get the mean absolute percentage error. However, while MAPE is a very straightforward measure of error and can be considered easy to understand and convincing, still, drawbacks of it have to be mentioned. First of all, it cannot be used when in the data set zero values exist, which is not a problem in our case, as the datasets used had no zero values. Second, for forecasts that are too high, the measure can exceed 100%, because there is no upper limit to the percentage error. Third, if MAPE is used to compare the accuracy of prediction models, then it is biased, thus it will systematically select a method whose forecasts are too low. In our case, we can overcome this issue, as MAPE is not the only measure to decide about the suitability of the models (Everitt, 2002).

The third measure selected is the MAE (Mean Absolute Error), which as the name suggests, is an average of the absolute errors $|e_i| = |y_i - x_i|$, where y_i is the prediction and x_i is the true value and is given by the formula:

$$MAE = \frac{\sum_{i=1}^{n}|y_i - x_i|}{n} \qquad (8)$$

MAE is an easy error to evaluate and understand and has a clear interpretation as it measures the average absolute difference between y_i and x_i. Another valuable property of the MAE error is that each of the errors contributes to it in proportion to the absolute value of the error, which is not true for other errors such as RMSE.

The last error that was used in this research is the R^2. In statistics, the coefficient of determination, denoted as R^2, is the proportion of the variance in the dependent variable that is predictable from the independent variable(s).

$$R^2 = 1 - \frac{SSR}{SST} \qquad (9)$$

where SSR (Residual Sum of Squares) and SST (Total Sum of Squares) are given by the formulas below:

$$SSR = \sum_i (y_i - f_i)^2 = \sum_i e_i^2 \qquad (10)$$

$$SST = \sum_i (y_i - \bar{y})^2 \qquad (11)$$

In simple linear regression, where only one intercept is included, then r^2 is simply the square of the sample correlation coefficient between the observed outcomes and the observed predictor values. If additional regressors are included, R^2 is the square of the coefficient of multiple correlations. In both such cases, the coefficient of determination ranges from 0 to 1. In cases where negative values are produced, it is an alert interpreted as the mean of the data provides a better fit to the outcomes than do the fitted function values. In regression, the R^2 coefficient of determination is a statistical measure of how well the regression predictions approximate the real data points. An R^2 of 1 indicates that the regression predictions fit the data perfectly. In a non-simple linear model, R^2 is a measure of the global fit of the model and is often interpreted as the proportion of response variation "explained" by the regressors in the model.

4 Univariate Linear Models

4.1 Random Walk Models

The random walk models in this paper were used as benchmark models. The purpose of this paper is twofold, namely, first to explore the contradictive findings in the literature between linear and non-linear models, and second to explore the contradictive findings in the comparison between technical and fundamental analysis.

In general, the random walk is a mathematical model, also known as a stochastic or random process, that describes movements and it consists of a succession of a number of random steps on some mathematical space such as integers. An elementary example of a random walk is the random walk on the integer number line, which starts at 0 and at each step it moves by 1 or -1 with equal probability. These types of models have been applied to numerous applications in economics (fluctuating stock price), game theory (financial status of a gambler), or even biology (the search path of a foraging animal). In this research two types of random walk were calculated, the first one is a random walk derived by a normal distribution (Dupernex, 2007), and the second is, the model used in the case of Meese & Rogoff (1983), a random walk which is a special case of ARIMA models the ARIMA (1,0,1), where tomorrow's price is equal to today's price.

For the first model, random values come from a Normal probability distribution with mean, μ, and standard deviation, σ, as N (μ, σ).

$$PredPrice_t = PredPrice_{t-1} * \varepsilon, \;\; where\ \varepsilon \sim N(\mu, \sigma) \qquad (12)$$

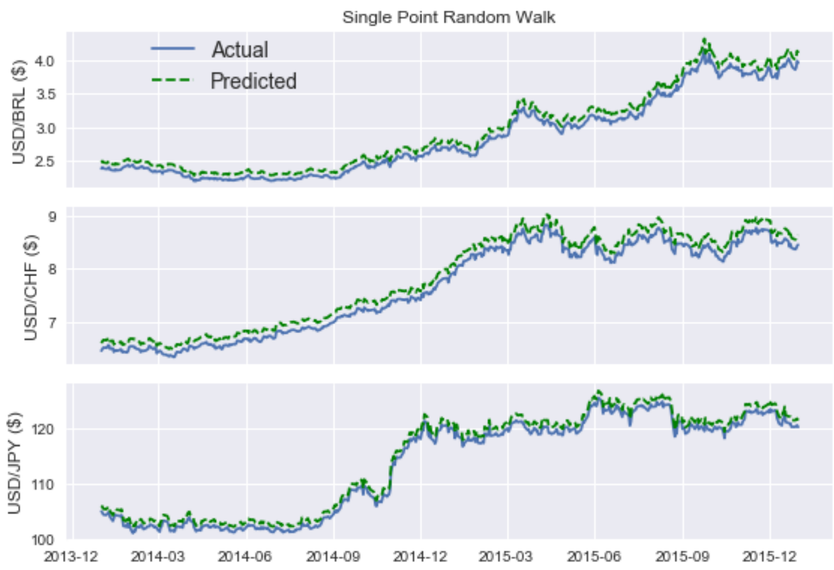


Figure 7: Random walk models, with random values derived from a normal distribution with mean μ and standard deviation σ for the three exchange rates (USD/BRL, USD/CHF, USD/JPY). The previous value is used to predict the next one

The graphs provided show the implementation of the previously described method. The single point prediction graphs show the random steps derived from a normal distribution, where mean = μ of training data, and standard deviation = σ of the training data. The previous value (t_{-1}) used the random steps to predict the t value (fig. 7).

In another experiment with the same method, there was an attempt for the random walk to produce the prices of the full interval of the test set, and the model was unable to capture the long-term real values for the out-of-sample predictions of each dataset. During the training period, the mean and the standard deviation were calculated, and the values were used to produce random steps for the total of the two-year test period (Figure 8). Although this experiment was conducted to investigate the abilities of a random walk model to predict in once the prices for the next two years, useful insights could be noticed. The random walk model fails to capture the trend in more extended period predictions, and it cannot be used for more than t+1 period prediction.

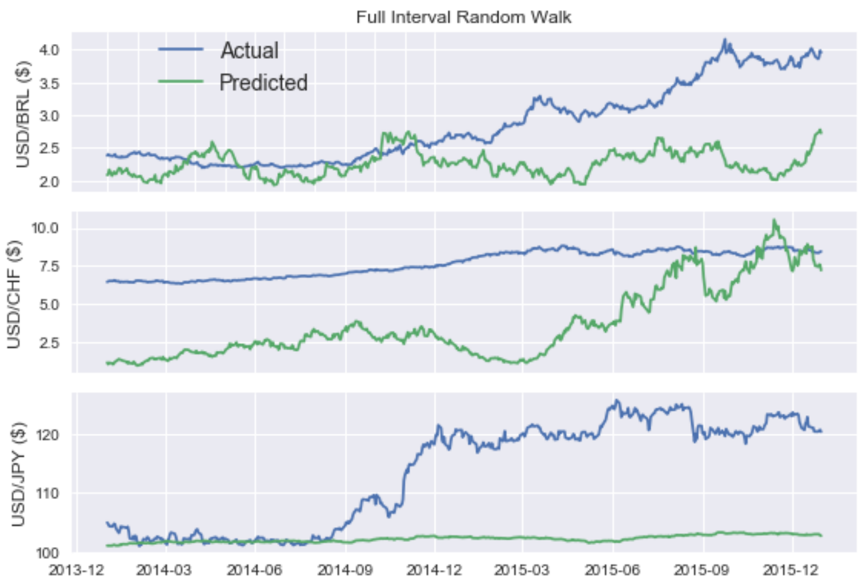


Figure 8: Random walk models, with random values derived from a normal distribution with mean μ and standard deviation σ for the three exchange rates (USD/BRL, USD/CHF, USD/JPY). `The previous value is used to predict the total number of the test set values.

4.2 ARIMA models

In academic research, the Box-Jenkins approach is widely acknowledged as a benchmark technique for univariate methods because of its structured modeling basis and acceptable forecasting performance (G. E. P. Box, Jenkins, Reinsel, & Ljung, 2015). Box and Jenkins provided an analytical description for their method of Autoregressive Integrated Moving Average (ARIMA) models, and this is one of the reasons that ARIMA models are so notorious in the scientific community (Box, 2013). The Box Jenkins approach was followed in this paper to build the ARIMA models, and since the data were already transformed to stationary (a necessary step for the Box and Jenkins approach), in the preprocessing part (see 3.4.2), ARIMA models can be used. The stationarity was checked with graphs (figure 4) and also the Augmented Dickey-Fuller test was implemented in the data of each dataset (Dickey, & Fuller, 1981). The augmented Dickey-Fuller (ADF) statistic, used in the test, is a negative number. The more negative it is, the stronger the rejection of the hypothesis that there is a unit root (a feature that can cause problems in time-series models) at some level of confidence.

In order to construct the best ARIMA (p,d,q) model for the exchange rate time-series datasets, the autoregressive (*p*) and moving average (*q*) parameters are to be identified for an effective model. The time-series has been transformed to stationary, as shown in the preprocessing part (see 3.4.2), by lag differencing. The next step in fitting an ARIMA model is to determine whether AR or MA terms are needed to correct any autocorrelation that remains in the differenced series. Autocorrelation (ACF) and partial autocorrelation (PACF) correlograms were used to identify autoregressive term and moving average term. Since the ACF and PACF coefficients are not significant, the best model based on Bayesian Information Criterion (BIC) for various orders of autoregressive (*p*) and moving average (*q*) terms keeping integrated term (*d*).

The grid search method was used, and for each dataset, the model with the smallest AICc and BIC values was returned. A range between 0-11 for values of p and q and range between 0-7 for values of d were used to determine the best model. The AICc and BIC criteria were calculated for each combination of the p,d,q values and the model with the smallest AICc and BIC values were chosen. Below the mathematical terms for both criteria are provided.

$$\text{Akaike info criterion (AIC)} = -2 \log(L) + 2\,(p + q + k) \qquad (13)$$

where L is the likelihood of the data, p is the order of the autoregressive part and q is the order of the moving average part. The k represents the intercept of the ARIMA model. For AIC, if $k = 1$ then there is an intercept in the ARIMA model ($c \neq 0$) and if $k = 0$ then there is no intercept in the ARIMA model ($c = 0$). The corrected AIC for ARIMA models can be written as:

$$AICc = AIC + (2(p + q + k)(p + q + k + 1)) / (T - p - q - k - 1) \qquad (14)$$

$$\text{Bayes info criterion (BIC)} = AIC + (\log(T) - 2)(p + q + k) \qquad (15)$$

The ARIMA models can be mathematically described in their general forecasting form. Here, the moving average parameters (θ's) are defined so that their signs are negative in the equation, following the convention introduced by Box and Jenkins (Box, 2013):

$$\hat{y}_t = \mu + \varphi_1 y_{t-1} + \ldots + \varphi_p y_{t-p} - \theta_1 e_{t-1} - \ldots - \theta_q e_{t-q} \qquad (16)$$

In addition to the best ARIMA models that were identified and implemented using a grid search approach, another special case of ARIMA model was built. The ARIMA (0,1,0), or l (1) model, which is a random walk model and is given by the formula:

$$\hat{Y}_t = \mu + Y_{t-1} \qquad (17)$$

The constant term is the average period-to-period change (i.e., the long-term drift) in Y. This model could fit as a no-intercept regression model in which the first difference of Y is the dependent variable. Since it includes only a non-seasonal difference and a constant term, it is classified as an "ARIMA (0,1,0) model with constant." The random-walk-without-drift model would be an ARIMA (0,1,0) model without constant. The ARIMA random walk model that was used in this research was with a constant. Below the plots of the described method, with both the ARIMA (0,1,0) and the best ARIMA chosen model are provided.

Figure 9: Comparison of ARIMA (0,1,0) and Best ARIMA (6,0,1) for the USD/JPY foreign exchange rate

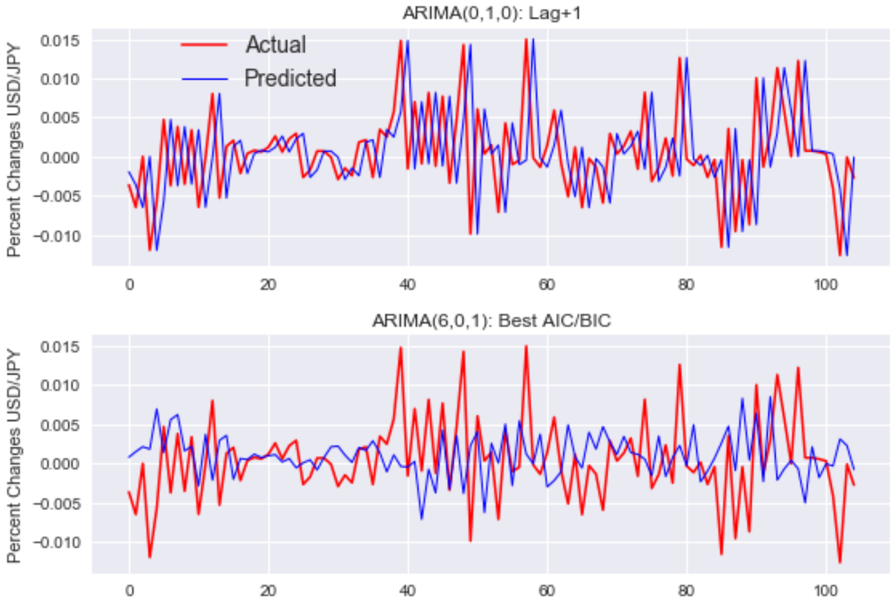


Figure 10: Comparison of ARIMA (0,1,0) and Best ARIMA (7,0,1) for the USD/BRL foreign exchange rate

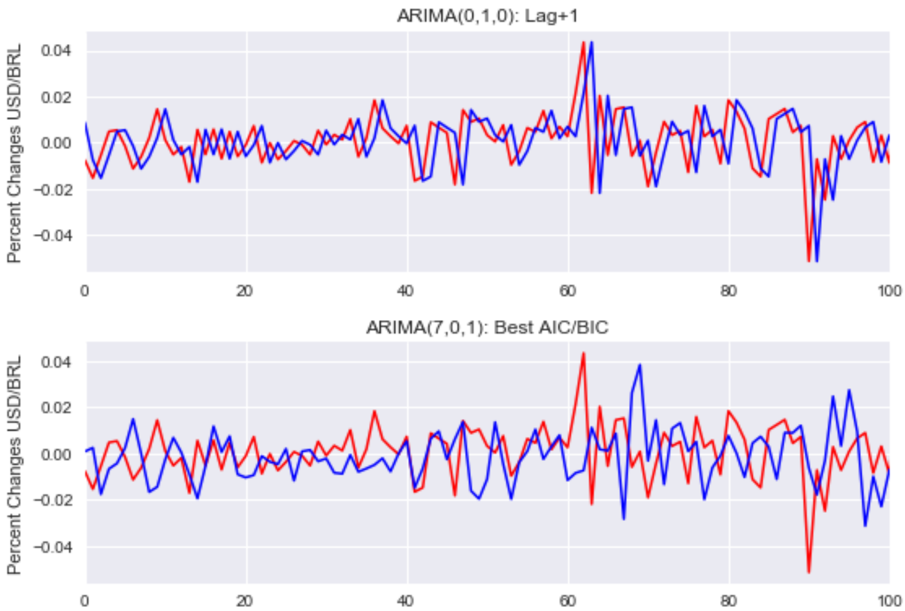


Figure 11: Comparison of ARIMA (0,1,0) and Best ARIMA (6,0,1) for the USD/CHF foreign exchange rate

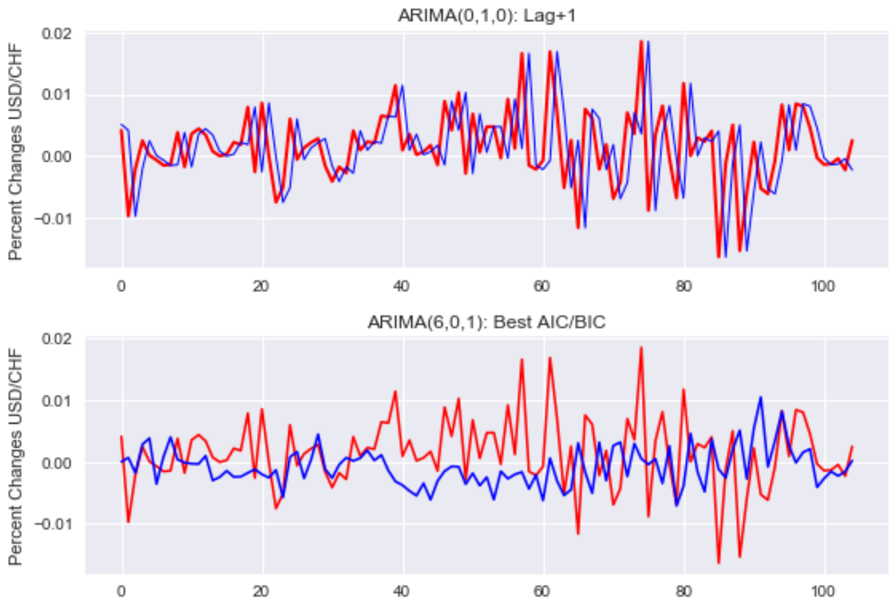
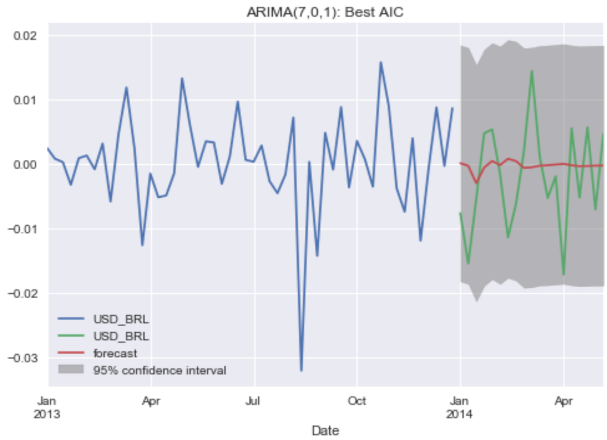


Figure 12: Long-term (t+12) predictions using the Best ARIMA models for each foreign exchange rate pair (USD/JPY, USD/CHF, USD/BRL)

Another experiment was conducted using the best ARIMA models, that were identified in the previous section for each currencies pair, to predict out-of-sample prices for long-term predictions. On the left side, the plots of the experiments show that although ARIMA models fail to capture the exact future price, they can still show the trend which will be followed in the near future. In the graphs, it is observed that each model tries to predict the out-of-sample period of 4 months in weekly prices. The models successfully show the potential trends that will occur in the future, but after a point, the trend line of the model tends to approach 0, and the model loses any ability to predict further in the future.



As it is observed from the figures, the models did not have the same power to predict and the same performance in each case. Mainly, in the case of the exchange rate USD/JPY, the model produced an almost flat line, being unable to capture the trends.

In general, we can observe a well-described trend for the future trend in prices for the pairs of USD/BRL and USD/CHF. These prediction graphs could be characterized as informative, as they give a notion of the future trend but also, they cannot always be trusted, as shown in the USD/JPY case.

5 Non-Linear Models – LSTM NN

In recent years, much research has been done using Neural Networks to predict foreign exchange rates with mostly good results (Gradojevic & Yang, 2006). Long Short-Term Memory Neural Network (LSTM), is a special case of Recurrent Neural Network (RNN), and in fact, it is an RNN composed of LSTM units. A common LSTM unit is composed of a cell, an input gate, an output gate and a forget gate. The cell is responsible to "remember" the values over arbitrary time intervals, and each of the three gates can be thought as a conventional artificial neuron, just like in a simple neural network of a multilayer perceptron (Hochreiter & Schmidhuber, 2006). Short-term memory is the ability to hold, without manipulating a small amount of information in memory in an active, readily available state for a short period of time. Long short-term memory characterizes the described ability, but since it is long, it can last for an extended period. LSTMs have been proven to be a well-suited solution to classify, process and predict time series, given time lags of unknown size and duration between important events.

In fact, LSTMs were built to overcome problems such as the vanishing gradient when training traditional RNNs or RNNs with hidden Markov chain models and as well as other sequence learning applications. LSTM NN since its introduction has been used in a variety of applications and in a variety of field of research that includes time-series analysis and prediction. LSTM algorithms have been implemented successfully for image processing, recognition, and generation. According to their findings, the recurrent long-term models that were built, were directly connected to state-of-the-art visual convolutional network models and could be trained, updating temporal dynamics and convolutional perceptual representations simultaneously. (Donahue et al., 2015). Speech recognition is another field that LSTMs have proved valuable for improving the accuracy of the models used before. One of the most famous works for speech recognition that was one of the first in the field using LSTM algorithms to beat previous RNN algorithms was the one of Graves, Mohamed, and Hinton (2013). Amazon's Alexa, Apple's Siri are also some of the most recent examples of an industrialized product that uses LSTM algorithms to achieve the speech recognition task.

The financial field also used LSTM algorithms to implement market predictions. Recent studies show that stock market predictions can benefit from the properties of LSTM algorithms. In the case of the research of Bao, Yue, & Rao (2017), their findings showed that their proposed model outperformed other similar models, in both predictive accuracy and profitability performance for the task of forecasting the next day's closing price. Similar findings, regarding the predictability power of Long Short-term Memory models (LSTM), were also achieved by K. Chen, Zhou, and Dai (2015). In their research, their focus was on the returns of China's stock market, and their model outperformed the benchmark of a random walk model that they set. Ghosal, Bhatnagar, Akhtar, Ekbal, & Bhattacharyya (2017) used both LSTM and a sentiment analysis algorithm to predict the movements of financial data, and they performed the specific task with success.

5.1 Artificial Neural Network Architecture, Hyperparameter tuning & Evaluation criterion

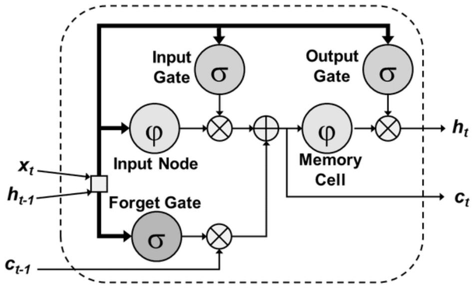
An LSTM model was used for the univariate non-linear model and, also the same was used for the multivariate models as well. The architecture of the model which can be described as consisting of 4 layers, was designed as follows: an input layer, two hidden layers, and one output layer. The input layer consisted of one input neuron in the case of univariate time-series and five input neurons in the case of multivariate, hidden layer was consisted of 50 neurons each, and the output layer of a single neuron. The model was trained using LSTM Backpropagation Through Time algorithm using the hard-sigmoid optimization function and the logistic function (tanh) for the Forward propagation. The past ten instances were fed as inputs to the network and it would forecast the value of the next instance.

$$\tanh x = \frac{\sinh x}{\cosh x} = \frac{e^{x} + e^{-x}}{e^{x} - e^{-x}} = \frac{e^{2x} + 1}{e^{2x} - 1} = \frac{1 + e^{-2x}}{1 - e^{-2x}}, \quad where\ x \neq 0 \qquad (18)$$

The proposed model was trained to minimize the MAE loss-function, as it is one of the most straightforward metrics for the task of foreign exchange prediction.

$$MAE = \frac{\sum_{i=1}^{n} |y_i - x_i|}{n} \qquad (19)$$

Figure 13: A single LSTM memory block. The three gates (Input, Output, and Forget) control the Input Node and Memory Cell to allow long-term memory. The function ϕ is the tanh function and the function σ is the hard-sigmoid function.



One of the fundamental properties of ANNs is their black-box nature (Benitez et al., 1997). This property refers to the non-comprehensive understanding of the researchers about the often incomprehensible internal processes and the internal results produced inside an Artificial Neural Network (ANN) before giving us the output, which is described as highly accurate. Artificial Neural Networks (ANNs) when dealing with multiple features, an internal feature selection process is happening. ANNs tend to use the most informative features and automatically adjust the weights to each feature to learn the requested model by minimizing the error rate. This property will be exploited in this research, and the model described above, will be fed multiple times with each dataset in order to test the behavior of the predictive model in all three currency exchange rates. The first model will be a univariate time-series model (technical analysis), where the LSTM algorithm will predict t+1 out of sample prices.

For the multivariate models (fundamental analysis), the LSTM algorithm will be fed with the commodities indexes, the commodities indexes, and the business indexes, the sentiment analysis GDELT data and for the final model, the complete dataset will be fed to the algorithm.

By implementing the described experiments, and by exploiting the ANNs black box nature will be able to answer to questions regarding what is working and what is not as a good predictor for the specific forex currencies of the research.

Finally, the same model, as described above, was used to perform a t+4 prediction, in each case, where the model and the same size of inputs, were fed to the model the ten previous instances, and the forecast was the next four instances. The chosen parameters of the models were chosen after trial and error experiments and the best parameters were chosen to be applied. For an intuition of how well the LSTM algorithm did perform, the t+1 and the t+4 plots of the univariate experiment will be provided. For the rest of the experiments, the errors will be reported in the results section. All plots below are predictions on unseen data of the test set.

Figure 14: On the left, t+1 LSTM Predictions and on the right, t+4 LSTM Predictions for the USD/JPY foreign exchange rate

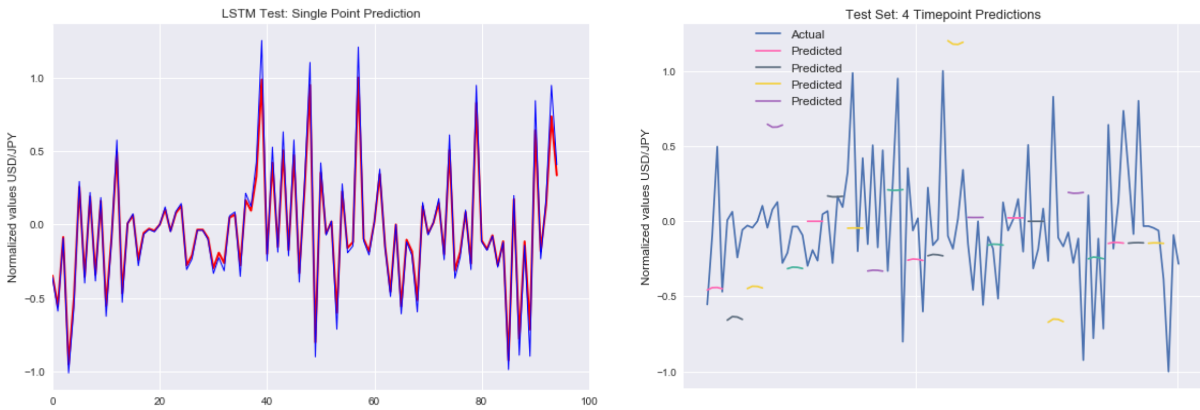


Figure 15: On the left, t+1 LSTM Predictions and on the right, t+4 LSTM Predictions for the USD/CHF foreign exchange rate

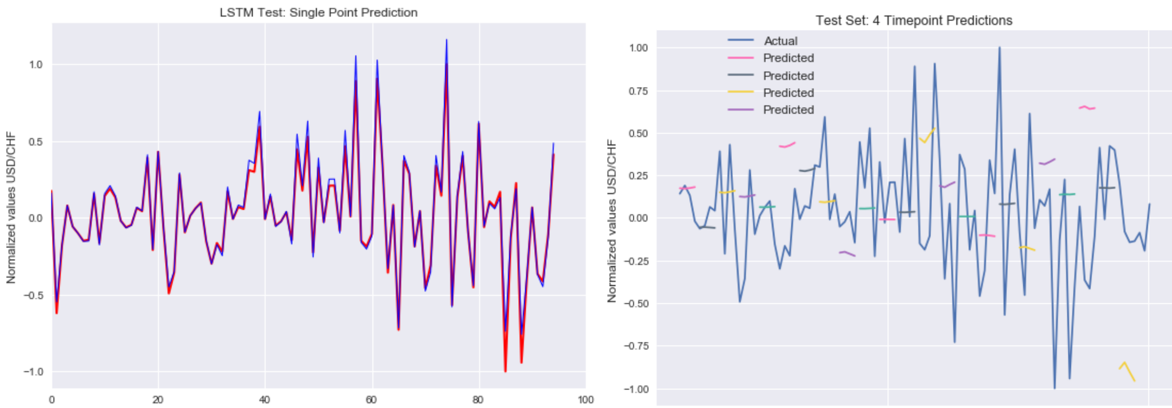
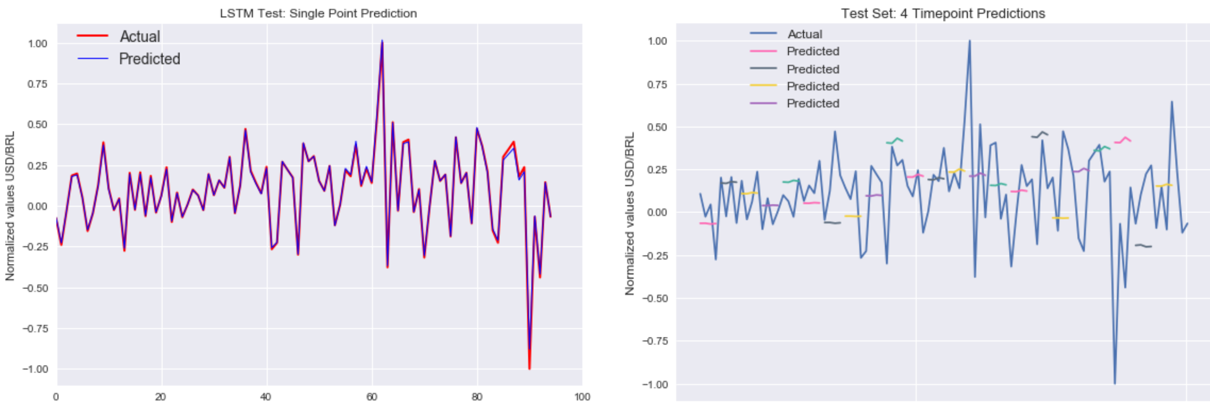


Figure 16: On the left, t+1 LSTM Predictions and on the right, t+4 LSTM Predictions for the USD/BRL foreign exchange rate



6 Results

In this section, the results of the experiments described in the previous sections will be reported. For reasons of consistency, the results will be presented in two subsections, following the same logic as in the experimental parts (3,4), in the first part (6.1) tables will report the errors of the linear models for each currency pair separately and in the second one (6.2) the errors of the non-linear.

6.1 Results of Linear Models Univariate Time-series

Regarding the linear univariate models, by analyzing the results reported in tables 4, 5, and 6 Random Walk models are the best performing models for the next price prediction task. In every case (USD/BRL, USD/CHF, and USD/JPY) the Random Walk models appear to have the lower Mean Absolute Percentage Error (MAPE) and the highest R^2; therefore, their performance could be described as better than the best ARIMA models of each currency pair for the specific task.

The full interval Random Walk model though (fig.8) can be considered an insufficient model for the predictions of consecutive prices. Comparing the efficiency of linear models for forecasts longer than t+1, ARIMA models could be considered a more informative model as they can reveal the future trend of the time-series investigated (fig.12).

Table 4: Metrics of the univariate Linear models for the USD/BRL foreign exchange rate prediction (Technical Analysis)

	RMSE	MAE	MAPE	R^2
Random Walk	0.110	0.1081	3.7987	0.9645
ARIMA (1,0,1)	0.0168	0.0125	315.0351	0.0125
Best ARIMA	0.0121	0.0091	116.2423	-0.0431
Full Interval Random Walk	0.8769	0.6702	20.8746	-1.2466

Table 5: Metrics of the univariate Linear models for the USD/CHF foreign exchange rate prediction (Technical Analysis)

	RMSE	MAE	MAPE	R^2
Random Walk	0.4001	0.3977	5.2004	0.7749
ARIMA (1,0,1)	0.0084	0.0062	878.4059	-1.0711
Best ARIMA	0.0061	0.0045	229.7932	-0.0078
Full Interval Random Walk	5.4695	4.5082	56.7326	-34.4408

Table 6: Metrics of the univariate Linear models for the USD/JPY foreign exchange rate prediction (Technical Analysis)

	RMSE	MAE	MAPE	R^2
Random Walk	1.1401	1.1366	1.0019	0.9831
ARIMA (1,0,1)	0.0080	0.0063	773.8712	-1.2502
Best ARIMA	0.0054	0.0038	179.3645	-0.0271
Full Interval Random Walk	14.3906	11.5768	9.6708	-1.7048

6.2 Results of Non-Linear Models

The results of the non-linear univariate and multivariate models (Long Short-Term Memory Neural Networks) of this research are reported in the tables 7, 8, and 9. The left column of the errors reported (T+1) describe the errors for the next price's prediction task while the right column (T+4) reports the errors for the prediction of four consecutive prices prediction task.

In the T+1 prediction task, the Univariate models of USD/BRL (table 7) and USD/JPY (table 9), had the lowest errors. It is noteworthy that the Full Dataset (multivariate time-series; predictors including commodities prices, business indexes and sentiment data derived from news), in the USD/BRL (table 7) dataset could be considered almost equivalent to its well-performed univariate model.

The most interesting behavior though could be noticed in the USD/CHF dataset (table 8). The particular foreign exchange pair proved hard to be predicted from the linear models, as it showed high errors and low R^2 score (table 5). The linear models, both ARIMA and Random Walk models, failed to predict the USD/CHF currency pair and performed poorly. On the other hand, although the univariate Long Short-Term Memory Neural Network models improved the prediction, the addition of sentiment derived from news drastically boosted the prediction. According to the results, the multivariate model with the data from the GDELT database is the most appropriate for the task of the T+1 prediction of the USD/CHF exchange rate pair.

Another worthwhile observation is the fact that the USD/CHF currency pair's model is the only one that was improving by the addition of more predictors. On the other hand, the USD/BRL and the USD/JPY pairs did not appear to have the same impressive findings. These two pairs did not improve by the addition of more predictors, and their univariate models appeared to have the smallest errors.

For the T+4 task it is important to note that, although the prediction is not as accurate as in the T+1 task, the full dataset with all the predictors had the best results in every pair of currencies. This information reveals that not only the predictors chosen include information for a further and deeper in time forecast, but also indicates the necessity for the addition of more predictors in order to complete this particular task.

Table 7: Metrics of the univariate and multivariate Long Short-Term Memory Neural Network models for the USD/BRL foreign exchange rate prediction (Technical and Fundamental Analysis). T+1 column, describes the errors for the next price prediction task, while T+4 for the 4 consecutive prices prediction task.

LSTM Models	T+1				T+4			
	RMSE	MAE	MAPE	R^2	RMSE	MAE	MAPE	R^2
Univariate	0.0167	0.0090	3.5180	0.9956	0.3003	0.2243	208.0498	-0.3978
Commodities	0.0319	0.0226	10.0243	0.9839	0.2970	0.2221	204.4764	-0.3671
Gdelt	0.0262	0.0202	9.6607	0.9891	0.3059	0.2282	217.4485	-0.4500
Commod + Business Index	0.0342	0.0209	8.3656	0.9815	0.2946	0.2200	201.0305	-0.3453
Full dataframe	0.0213	0.0144	6.4268	0.9928	0.2873	0.2142	185.2608	-0.2791

Table 8: Metrics of the univariate and multivariate Long Short-Term Memory Neural Network models for the USD/CHF foreign exchange rate prediction (Technical and Fundamental Analysis). T+1 column, describes the errors for the next price prediction task, while T+4 for the 4 consecutive prices prediction task.

LSTM Models	T+1				T+4			
	RMSE	MAE	MAPE	R^2	RMSE	MAE	MAPE	R^2
Univariate	0.0445	0.0250	9.0104	0.9832	0.4478	0.3491	187.9372	-0.7329
Commodities	0.0166	0.0111	4.0195	0.9977	0.4501	0.3514	385.9056	-0.7506
Gdelt	0.0027	0.0017	0.5436	0.9999	0.4585	0.3558	384.6239	-0.8167
Commod + Business Index	0.0348	0.0229	8.5778	0.9897	0.4463	0.3464	369.5843	-0.7210
Full dataframe	0.0143	0.0097	3.9985	0.9983	0.4615	0.3556	383.7183	-0.8404

Table 9: Metrics of the univariate and multivariate Long Short-Term Memory Neural Network models for the USD/JPY foreign exchange rate prediction (Technical and Fundamental Analysis). T+1 column, describes the errors for the next price prediction task, while T+4 for the 4 consecutive prices prediction task.

LSTM Models	T+1				T+4			
	RMSE	MAE	MAPE	R^2	RMSE	MAE	MAPE	R^2
Univariate	0.0631	0.0411	14.2024	0.9724	0.5593	0.4296	6953.3975	-1.1926
Commodities	0.1013	0.0738	25.1527	0.9290	0.5552	0.4266	6891.2645	-1.1604
Gdelt	0.0610	0.0451	15.7125	0.9743	0.5703	0.4371	6614.1162	-1.2793
Commod + Business Index	0.0974	0.0723	25.5251	0.9344	0.5454	0.4196	6475.0092	-1.0845
Full dataframe	0.0834	0.0585	19.5499	0.9519	0.5262	0.4036	6272.0233	-0.9405

7. Conclusions

In this section, the results of the experiments will be analyzed in order to report the findings of this research and provide answers to the research questions stated at the beginning of this paper.

1. Which forecasting method and what types of models should be chosen to predict the next price? In other words, technical or fundamental analysis and linear or non-linear models are more efficient for the forex prediction task?

As described in the experimental section, Random Walk models, ARIMA models based on best AIC/BIC and LSTM were implemented in this paper. To begin with, the non-linear LSTM models proved to be the most suitable for the task of the forex prediction. Their ability to learn non-linear time-series and predict was a significant advantage comparing to the linear models, as they even beat the Random Walk based on mean and standard deviation for the prediction of the next day's price.

As a general conclusion regarding the technical or fundamental analysis (univariate vs. multivariate) and the different indicators that were used, it could be said that there is no rule of thumb. Different conclusions can be made depending on the different forex pairs and also the different prediction timespan that was used.

2. Can random walk still be considered invincible?

At this point, it would also be useful to mention that indeed Random Walk models, like the one that was built, are difficult to beat especially with linear models. That explains the massive literature that supports still random walk models, as it is a well-behaved and straightforward model to build, for forecasting next day's price. Nevertheless, as it was shown for long-term prediction, such models fail to capture even the trend. ARIMA, on the other hand, appeared to be vulnerable in its linear approximations that are using, but it can be considered extremely useful as a model to capture the general trend for long-term predictions, as shown in graphs. ARIMA model could clearly show a trend capturing ability for the period of the next three months ahead, in unseen data. As a result, although ARIMA models cannot predict the next day's price accurately, they can be considered as an informative model for someone to identify the future trend of the market.

3. Can commodities indexes, business indexes and sentiment from the news, combined or individually, improve the prediction of Forex rates?

4. Could sentiment derived from news be considered as a good indicator?

5. How do models perform in predicting prices far in the future rather than the next price?

These three research questions will be answered separately for each foreign exchange rate pair (USD/BRL, USD/CHF, and USD/JPY), as for their case there is not a one-fits-all answer.

Focusing on the LSTM models, which were proven as a reliable model with predictive power for the next day's price. Also, based on the different LSTM models built on this paper, we can extract some conclusions about the technical versus the fundamental analysis and, especially the features that were used as predictors for the forex prediction process that was followed in this paper.

This paragraph will focus on both the t+1 and the t+4 timespan for the USD/BRL forex pair. Analyzing the results of the USD/BRL foreign exchange rate, the univariate LSTM model was shown as the most appropriate to predict next day's price. The next best model for the same t+1 prediction was the full dataset. That indicates that all the variables combined for the USD/BRL prediction task were informative and could predict the next day's price adequately, but still even all the extra information provided by the full dataset could not outperform the univariate time series LSTM model. For the t+4 prediction task though, the full dataset outperformed all the other experiments, with each indicator's data set being used separately. Although it appears to show the best metric for the t+4 prediction, the reality is that the prediction is not accurate, which means that the weekly data that were used are not able to predict four continuous prices further in future, based on the prices of the past, regardless the indicators that were used. This conclusion is a general one and applies for every examined currency pair in this research.

For the next currency pair, USD/CHF it is an impressive finding that GDELT's news sentiment data improved the prediction of the pair and outperformed all the other indicators, even the univariate model for the t+1 prediction task. Another interesting finding regarding this particular pair is that the univariate time-series model had the "worst" performance compared to the other models that were tested. It should be clarified that by referring as the "worst" performance, the predictive power of the "worst" model is still high and moreover it outperforms the random walk models and the ARIMA models. For the t+4 prediction task, the indicators that provided the best result was the combination of commodities indexes with the business indexes.

For the last currency pair USD/JPY, the multivariate model with GDELT's data and the univariate time-series proved to be the most efficient models in the t+1 prediction task. The errors reported for these two models showed mixed results to determine which one between the two, as it can be observed in the table USD/JPY (T+1) provided in the 5.2 part. The univariate model had higher RMSE, while MAE and MAPE were lower. R^2 was higher in favor of the multivariate GDELT data frame experiment. For the t+4 prediction task, the full dataset was once again the most appropriate model, without it qualifying as reliable, but this could provide a potential of improving the predictions further in the future while using more informative features-indicators.

To sum up, this research paper proves that the notorious short-term predictions of forex rates for the examined pairs of USD/BRL, USD/CHF, and USD/JPY are predictable and the random walk models can be defeated from non-linear models, such as LSTM Neural Network. LSTM algorithm, although it has been used for financial stock market predictions, is not an algorithm widely used in the forex prediction problems. In this research, its high suitability for the foreign exchange rate riddle is shown. The use of commodities and business indicators can improve the quality of the forecast especially in the longer than t+1 in the future ones. Regarding the news sentiment, can be concluded that they have great potential as shown in the pairs of the USD/CHF and USD/JPY. At this point, it is useful to be mentioned that in comparison with other studies that focused only in use of sentiment derived by the financial news, this study used the sentiment of the general political news, as recorded by the GDELT database and can also be considered a vital indicator. Another important fact is that GDELT's database has never been used before for financial exchange rates prediction, that I know of, following my research. Also, it is shown that fundamental analysis for more than t+1 period is a necessity, and for this task to be implemented successfully, a wide range of indicators should be used.

7.1 Future work

Implications of this research would be the focus on different algorithms based on long short-term memory (LSTM) approach or more complex deep learning algorithms to be tested. Delving deeper in the analysis of the neural network approaches, while using even more indicators for prediction tasks of longer periods, which according to literature, remains a hard task to be accomplished.

BIBLIOGRAPHY

Ajao, I. O., Ibraheem, A. G., & Ayoola, F. J. (2012). Cubic Spline Interpolation: A Robust Method of Disaggregating Annual Data to Quarterly Series. *Journal of Physical Sciences and Environmental Safety, 2*(1), 1-8.

Andersen, T. G., Bollerslev, T., Diebold, F. X., & Vega, C. (2007). Real-time price discovery in global stock, bond and foreign exchange markets. *Journal of International Economics*, *73*(2), 251–277. https://doi.org/10.1016/j.jinteco.2007.02.004

Bacchetta, P., & Wincoop, E. V. (2004). A Scapegoat Model of Exchange-Rate Fluctuations, *94*(2), 5.

Bao, W., Yue, J., & Rao, Y. (2017). A deep learning framework for financial time series using stacked autoencoders and long-short term memory. *PLOS ONE*, *12*(7), e0180944. https://doi.org/10.1371/journal.pone.0180944

Benitez, J. M., Castro, J. L., & Requena, I. (1997). Are artificial neural networks black boxes? *IEEE Transactions on Neural Networks*, *8*(5), 1156–1164. https://doi.org/10.1109/72.623216

Berge, T. J. (2015). Predicting Recessions with Leading Indicators: Model Averaging and Selection over the Business Cycle. *Journal of Forecasting*, *34*(6), 455–471. https://doi.org/10.1002/for.2345

Box, G. (2013). Box and Jenkins: Time Series Analysis, Forecasting and Control. In *A Very British Affair* (pp. 161–215). Palgrave Macmillan, London. https://doi.org/10.1057/9781137291264_6

Brooks, C. (1997). Linear and Non-linear (Non-)Forecastability of High-frequency Exchange Rates. *Journal of Forecasting*, *16*(2), 125–145. https://doi.org/10.1002/(SICI)1099-131X(199703)16:2

Chen, K., Zhou, Y., & Dai, F. (2015). A LSTM-based method for stock returns prediction: A case study of China stock market. In *2015 IEEE International Conference on Big Data (Big Data)* (pp. 2823–2824). https://doi.org/10.1109/BigData.2015.7364089

Chen, Y., & Rogoff, K. (2003). Commodity currencies. *Journal of International Economics*, *60*(1), 133–160. https://doi.org/10.1016/S0022-1996(02)00072-7

Chen, Y.-C., Rogoff, K. S., & Rossi, B. (2010). Can Exchange Rates Forecast Commodity Prices? *The Quarterly Journal of Economics*, *125*(3), 1145–1194. https://doi.org/10.1162/qjec.2010.125.3.1145

Cheong, C., Kim, Y.-J., & Yoon, S.-M. (n.d.). Can We Predict Exchange Rate Movements at Short Horizons? *Journal of Forecasting*, *31*(7), 565–579. https://doi.org/10.1002/for.1236

De Bodt, E., Rynkiewicz, J., & Cottrell, M. (2004). Some known facts about financial data. *European Journal of Economic and Social Systems*, *17*, 167–182.

Dickey, D. A., & Fuller, W. A. (1986). Likelihood ratio statistics for autoregressive time series with a unit. *Ecometrica, Vol. 49, No. 4.*

Donahue, J., Anne Hendricks, L., Guadarrama, S., Rohrbach, M., Venugopalan, S., Saenko, K., & Darrell, T. (2015). Long-Term Recurrent Convolutional Networks for Visual Recognition and Description (pp. 2625–2634). Presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.

Dunis, C. L., & Williams, M. (2002). Modelling and Trading the EUR/USD Exchange Rate: Do Neural Network Models Perform Better? *Derivatives use, trading and regulation* 8(3), 211-239.

Dupernex, S. (2007). Why might share prices follow a random walk? *Student Economic Review, Vol. 21.*

Eddelbuettel, D. (2017). anytime: Anything to "POSIXct" or "Date" Converter. *R package version 0.3.0.* Retrieved from https://CRAN.R-project.org/package=anytime

El Shazly, M. R., & El Shazly, H. E. (1997). Comparing the forecasting performance of neural networks and forward exchange rates. *Journal of Multinational Financial Management*, *7*(4), 345–356. https://doi.org/10.1016/S1042-444X(97)00018-2

Everitt, B. (2002). *The Cambridge dictionary of statistics* (2nd ed), Cambridge University Press.

Ferraro, D., Rogoff, K., & Rossi, B. (2015). Can oil prices forecast exchange rates? An empirical analysis of the relationship between commodity prices and exchange rates. *Journal of International Money and Finance*, *54*, 116–141. https://doi.org/10.1016/j.jimonfin.2015.03.001

Fischer, T., & Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. *European Journal of Operational Research*, *270*(2), 654–669. https://doi.org/10.1016/j.ejor.2017.11.054

Ghosal, D., Bhatnagar, S., Akhtar, M. S., Ekbal, A., & Bhattacharyya, P. (2017). IITP at SemEval-2017 Task 5: An Ensemble of Deep Learning and Feature Based Models for Financial Sentiment Analysis. In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)* (pp. 899–903). Vancouver, Canada: Association for Computational Linguistics. Retrieved from http://www.aclweb.org/anthology/S17-2154

Gradojevic, N., & Yang, J. (2006). Non-linear, non-parametric, non-fundamental exchange rate forecasting. *Journal of Forecasting*, *25*(4), 227–245. https://doi.org/10.1002/for.986

Graves, A., Mohamed, A. r, & Hinton, G. (2013). Speech recognition with deep recurrent neural networks. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing* (pp. 6645–6649). https://doi.org/10.1109/ICASSP.2013.6638947

Greff, K., Srivastava, R. K., Koutník, J., Steunebrink, B. R., & Schmidhuber, J. (2017). LSTM: A Search Space Odyssey. *IEEE Transactions on Neural Networks and Learning Systems*, *28*(10), 2222–2232. https://doi.org/10.1109/TNNLS.2016.2582924

Hafezi, R., Shahrabi, J., & Hadavandi, E. (2015). A bat-neural network multi-agent system (BNNMAS) for stock price prediction: Case study of DAX stock price. *Applied Soft Computing*, *29*, 196–210. https://doi.org/10.1016/j.asoc.2014.12.028

Hagenau, M., Liebmann, M., & Neumann, D. (2013). Automated news reading: Stock price prediction based on financial news using context-capturing features. *Decision Support Systems*, *55*(3), 685–697. https://doi.org/10.1016/j.dss.2013.02.006

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation, 9(8),* 1735-1780. https://doi.org/10.1162/neco.1997.9.8.1735

Hua, G. B., & Pin, T. H. (2000). Forecasting construction industry demand, price and productivity in Singapore: the Box-Jenkins approach. *Construction Management and Economics*, *18*(5), 607–618. https://doi.org/10.1080/014461900407419

Huang, W., Lai, K. K., Nakamori, Y., & Wang, S. (2004). Forecasting foreign exchange rates with artificial neural networks: A review. *International Journal of Information Technology & Decision Making*, *03*(01), 145–165. https://doi.org/10.1142/S0219622004000969

Khadjeh Nassirtoussi, A., Aghabozorgi, S., Ying Wah, T., & Ngo, D. C. L. (2015). 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. https://doi.org/10.1016/j.eswa.2014.08.004

Kilian, L., & Taylor, M. P. (2003). Why is it so difficult to beat the random walk forecast of exchange rates? *Journal of International Economics*, *60*(1), 85–107. https://doi.org/10.1016/S0022-1996(02)00060-0

Kim, K. (2003). Financial time series forecasting using support vector machines. *Neurocomputing*, *55*(1), 307–319. https://doi.org/10.1016/S0925-2312(03)00372-2

Lachtermacher, G., & Fuller, J. D. (1995). Back propagation in time-series forecasting. *Journal of forecasting, 14*(4), 381-393.

Leetaru, K. (2014). GDELT Project. Retrieved June 24, 2018, from https://www.gdeltproject.org/

Liu, L., & Wang, W. (2008). Exchange Rates Forecasting with Least Squares Support Vector Machine. In *2008 International Conference on Computer Science and Software Engineering* (Vol. 5, pp. 1017–1019). https://doi.org/10.1109/CSSE.2008.140

Malkiel, B. G., & Fama, E. F. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. *The Journal of Finance*, *25(2)*, 383–417. https://doi.org/10.1111/j.1540-6261.1970.tb00518.x

Meese, R. A., & Rogoff, K. (1983). Empirical exchange rate models of the seventies: Do they fit out of sample? *Journal of International Economics*, *14*(1), 3–24. https://doi.org/10.1016/0022-1996(83)90017-X

Montañez, S., & Antonio, J. (2011). *A contribution to exchange rate forecasting based on machine learning techniques* (Ph.D. Thesis). Universitat Ramon Llull. Retrieved from https://www.tdx.cat/handle/10803/48492

Moritz, S. (2018). imputeTS: Time Series Missing Value Imputation in R (Version 2.7). *The R Journal, 9(1).* Retrieved from https://CRAN.R-project.org/package=imputeTS

Mozetič, I., Gabrovšek, P., & Novak, P. K. (2018). Forex trading and Twitter: Spam, bots, and reputation manipulation. *ArXiv:1804.02233 [Cs]*. Retrieved from http://arxiv.org/abs/1804.02233

Nag, A. K., & Mitra, A. (2002). Forecasting daily foreign exchange rates using genetically optimized neural networks. *Journal of Forecasting*, *21*(7), 501–511. https://doi.org/10.1002/for.838

Rudy, J., Dunis, C., Giorgioni, G., & Laws, J. (2010). Statistical Arbitrage and High-Frequency Data with an Application to Eurostoxx 50 Equities. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2272605

Schumaker, R. P., Zhang, Y., Huang, C.-N., & Chen, H. (2012). Evaluating sentiment in financial news articles. *Decision Support Systems*, *53*(3), 458–464. https://doi.org/10.1016/j.dss.2012.03.001

Sun, A., & Chang, J.-F. (2017). Application of radial basis function neural network to predict exchange rate with financial time series. *International Journal on Smart Sensing and Intelligent Systems*, *10*(2), 308–326. https://doi.org/10.21307/ijssis-2017-213

Tay, F. E. H., & Cao, L. (2001). Application of support vector machines in financial time series forecasting. *Omega*, *29*(4), 309–317. https://doi.org/10.1016/S0305-0483(01)00026-3

Walczak, S. (2001). An Empirical Analysis of Data Requirements for Financial Forecasting with Neural Networks. *Journal of Management Information Systems*, *17*(4), 203–222. https://doi.org/10.1080/07421222.2001.11045659

Wang, J. (2008). Why are exchange rates so difficult to predict? *Economic Letter*, *3*. Retrieved from https://ideas.repec.org/a/fip/feddel/y2008ijunnv.3no.6.html

Wang, Z., & Bovik, A. C. (2009). Mean squared error: Love it or leave it? A new look at Signal Fidelity Measures. *IEEE Signal Processing Magazine*, *26*(1), 98–117. https://doi.org/10.1109/MSP.2008.930649

Wickham, H., Chang, W., & RStudio. (2016). ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics *R package version 2.2.1*. Retrieved from https://CRAN.R-project.org/package=ggplot2

Wickham, H., François, R., Henry, L., Müller, K., & RStudio. (2018). dplyr: A Grammar of Data Manipulation. *R package version 0.7.5*. Retrieved from https://CRAN.R-project.org/package=dplyr

Wickham, H., Henry, L., & RStudio. (2018). tidyr: Easily Tidy Data with “spread()” and “gather()” Functions. *R package version 0.8.1*. Retrieved from https://CRAN.R-project.org/package=tidyr

Willmott, C. J., & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research*, *30*(1), 79–82. https://doi.org/10.3354/cr030079

Wong, W. K., Xia, M., & Chu, W. C. (2010). Adaptive neural network model for time-series forecasting. *European Journal of Operational Research*, *207*(2), 807–816. https://doi.org/10.1016/j.ejor.2010.05.022

Wuthrich, B., Cho, V., Leung, S., Permunetilleke, D., Sankaran, K., & Zhang, J. (1998). Daily stock market forecast from textual web data. In *1998 IEEE International Conference on Systems, Man, and Cybernetics, 1998* (Vol. 3, pp. 2720–2725 vol.3). https://doi.org/10.1109/ICSMC.1998.725072

Yao, J., & Tan, C. L. (2000). A case study on using neural networks to perform technical forecasting of forex. *Neurocomputing*, *34*(1), 79–98. https://doi.org/10.1016/S0925-2312(00)00300-3

Zhang, G., & Hu, M. Y. (1998). Neural network forecasting of the British Pound/US Dollar exchange rate. *Omega*, *26*(4), 495–506. https://doi.org/10.1016/S0305-0483(98)00003-6

APPENDIX A

The preprocessing of the data was implemented with the R language and the following packages were used: dplyr (Wickham, François, Henry, Müller, & RStudio, 2018), ggplot2(Wickham, Chang, & RStudio, 2016), tidyr(Wickham, Henry, & RStudio, 2018), imputeTS (Moritz, 2018) and anytime(Eddelbuettel, 2017).