Journal of Business Theory and Practice
ISSN 2372-9759 (Print) ISSN 2329-2644 (Online)
Vol. 7, No. 2, 2019
www.scholink.org/ojs/index.php/jbtp

Original Paper

Using Text Mining to Predicate Exchange Rates with Sentiment Indicators

Hassabelrasul Yusuuf ALtom Shihabeldeen[1*]

[1] College of Science & Humanities Studies at Al-Aflaj, Prince Sattam Bin Abdul-Aziz University, Saudi Arabia

[*] Hassabelrasul Yusuuf ALtom Shihabeldeen, College of Science & Humanities Studies at Al-Aflaj, Prince Sattam Bin Abdul-Aziz University, Saudi Arabia

Received: February 24, 2019 Accepted: March 6, 2019 Online Published: March 22, 2019
doi:10.22158/jbtp.v7n2p60 URL: http://dx.doi.org/10.22158/jbtp.v7n2p60

Abstract

Recent innovations in text mining facilitate the use of novel data for sentiment analysis related to financial markets, and promise new approaches to the field of behavioral finance. Traditionally, text mining has allowed a near-real time analysis of available news feeds. The recent dissemination of web 2.0 has seen a drastic increase of user participation, providing comments on websites, social networks and blogs, creating a novel source of rich and personal sentiment data potentially of value to behavioral finance. This study explores the efficacy of using novel sentiment indicators from Market Psych, which analyses social media in addition to newsfeeds to quantify various levels of individual's emotions, as a predictor for financial time series returns of the Australian Dollar (AUD)-US Dollar (USD) exchange rate. As one of the first studies evaluating both news and social media sentiment indicators as explanatory variables for linear and nonlinear regression algorithms, our study aims to make an original contribution to behavioral finance, combining technical and behavioral aspects of model building. An empirical out-of-sample evaluation with multiple input structures compares Multivariate Linear Regression models (MLR) with multilayer perceptron (MLP) neural networks for descriptive modelling. The results indicate that sentiment indicators are explanatory for market movements of exchange rate returns, with nonlinear MLPs showing superior accuracy over linear regression models with a directional out-of-sample accuracy of 60.26% using cross validation.

Keywords

text mining, exchange Rates, sentiment indicators, financial markets

1. Introduction

The foreign exchange market (FOREX) is the biggest financial market in the world, attracting a daily turnover of 4.0 trillion US dollars. Practitioners recognize the importance of this market for cross-border economic operations, speculative trading and risk hedging, which is open 24/7 and does not have any timely trading limitations (in contrast to the stock market), causing it to attract significant and continuous attention. While the various functions and market participants increase the complexity of this market, leading it to exhibit non-linear properties, the widespread participation also makes it subject to frequent analysis, varying in rigor from corporate analysis to individual's opinions shared as sentiment towards the market by millions of traders, through reports or in person, and more recently online through social media.

In the analysis of such sentiment, behavioral finance has made important contributions by demonstrating the impact of emotions and sentiments on economic and financial decisions, explaining certain market movements, e.g., in exchange rate prediction. The beginnings of this research field utilized weather conditions as a proxy for investor sentiment. Today, enabled by increasing computing power and textual analysis algorithms, major progress in this field has been achieved in the field of text (data) mining, which allows determining human sentiment directly from information shared in textual form across forums, blogs and social media networks across the internet. Recent inventions in the text mining industry led to sentiment indicators extracting the level of emotions contained in news items as well as comments made in social media networks, twitter and various topical blogs.

Although text mining has been used to extract information from new releases, ticker data, twitter or blogs before, we have found no other study assessing a combined index of market sentiment derived from both individual social media contributions and news feeds, as available in the form of the novel commercially available index provided by Market Psych through Thomson-Reuters. We conduct an empirical evaluation assessing the effect of different input variable selections of sentiment indicators, including joy, fear, buzz, etc. as behavioral indicators in forecasting the Australian Dollar versus US Dollar foreign exchange rate time series with respect to directional movements of continuous returns. To allow for both linear and/or nonlinear properties, we evaluate different input variables with Multivariate Linear Regressions (MLR) and nonlinear multilayer perceptron's (MLP) neural networks. As such, the paper aims to make an original contribution to research in behavioral finance and exchange rate prediction with sentiment indicators derived from social media.

2. Human Sentiment in Behavioral Finance

In classical financial theories, the predominant concept of the homo economics states that investors are determined to maximizing their individual utility function in a logically comprehensible and predictable manner. Their asset price perception equals the discounted value of expected cash flows in the future and is not influenced by any emotional irrationality. This hypothesis has led to two influential theories in their pioneering roles: the Efficient Market Hypothesis (EMH) and the Random Walk Hypothesis (RWH). Both theories postulate that asset prices are determined by available market information which is used by investors to form their future expectations. News is instantaneously incorporated and prices, therefore, follow a random walk pattern which is unpredictable.

In contrast, in behavioral finance the investors are no longer rational but their decision making is influenced by irrational behavior. Theoretical evidence shows that an individual's sentiment influences human decision-making. Various studies have used this hypothesis and attempted to discover meaningful relationships to financial markets. As one prominent example, people and their mood are greatly affected by the daily weather. As a consequence, behavioral finance found their investment decisions to be biased accordingly, so that weather phenomena like sunshine can serve as sentiment proxy with surprising indication of predictability of returns. However, careful research established that weather has different effects on different people and varies between locations. As weather proved less adequate for today's globalized FOREX markets due to the local nature of weather, proxies for sentiment were developed including explicit measurements, often in the form of surveys and polls which may contain some biases due to psychological effects of human self-reflection, and implicit indicators, developed to overcome these shortcomings. Implicit measurements of sentiments contain inter alia discounts on closed-end funds or liquidity. These measurements indicate market imperfections in asset pricing, often caused by the irrational behavior of noise traders. The price is not the discounted value of expected cash flows but formed by an unintuitive combination of supply and demand. Explicit sentiment indicators ask for the state of emotions, feelings or expectations of investors through questionnaires and surveys. Well-known sentiment indicators like the Consumer Confidence Index from Gallup are examined as well as individually built indices. Also a combination of implicit and explicit measurements was part of various studies. Reversed causation with market movements influencing consumer confidence is also found. However, research remains inconclusive as to the appropriate approach and more important, how fundamentals are extracted from these proxies to achieve a real sentiment measurement.

Increasing computing power and the recent progress in text mining research in general, and automatic text interpretation in the form of text mining in particular, have given further opportunities to reveal different levels of sentiment from written text. This resulted in the development of two different approaches for textual usage. The first technique directly links the statements of news articles with stock movements, while the second method uses text mining techniques to extract human sentiment from language and quantify individual levels of emotions. Even though many researchers doubt that news contains any novel information which may be helpful for prediction models, other studies show, occasionally, that news can, in fact, augment prediction models and, hereby, disprove EMH and RWH. Drawbacks arise from algorithmic issues of interpreting seldom used or ambiguous expressions and scalability problems. To overcome limitations in understanding complete texts, simpler approaches such as the unordered bag-of-words approach, or improved uses of manipulating textual information using a combination of verb and noun or semantic coherence are employed.

In many cases news articles also contain sentiment information which can be extracted by text mining. They can serve as sentiment indicators. But not only news articles, but also comments in social media and blogs contain sentiment which could provide additional predictive indicators. Another approach test the prediction power of sentiment expressed in financially and non-financially related social media sites. In particular, Twitter seems to be a popular choice of investigation and led to the opening of a Twitter specific hedge fund. Considering text mining approaches in foreign exchange markets, the FOREX market has also been subject of various studies. External factors like macroeconomic variables or interest rates seem to have prediction power on returns. Asymmetric effects due to sentiment are also evident in this complex market which seems to be better forecasted with non-linear models like neural networks, which regularly demonstrate superior in- and out-of-sample errors over linear autoregressive or random walk models. In this context, this paper presents novel research by filling the gap between sentiment analysis based on text mining and foreign exchange forecasting. It both enhances linear and nonlinear methods with newly invented sentiment indicators searching both articles and social media and representing various levels of emotions.

3. Experimental Design

3.1 Data Exploration

The data set for the exchange rate AUD-USD contains 783 daily closing prices over 3 years from 4th September 2009 to 4th September 2012, recorded at 4pm New York time (data provided by Thomson Reuters). The data contains daily closing prices of trading days, excluding Saturday and Sunday, but inclusive any bank holidays. Using the formula below we derive continuous returns *rt*.

$$r_t = \ln\left(\frac{p_t}{p_{t-1}}\right) \qquad (1)$$

with *pt* the exchange rate price in time *t*, *pt*-1 the price lagged one period in time *t*-1 generates a time series with 782 explorations in Figure 1. These returns align with stylized facts of financial returns in terms of non-normal distribution, no correlation between consecutive returns and serious correlation in squared returns. For the independent variables we consider 14 different sentiment indicators computed as indices, in particular Carry Trade, Conflict, Fear, Joy, Optimism, Peg Instability, Price Forecast, Price Momentum, Price Up, Trust, Uncertainty, Urgency, Violence, and one buzz index. The indicators are created by scanning English-written social media and newspaper for expressions of emotion, available from Market Psych and published on Thomson-Reuters.

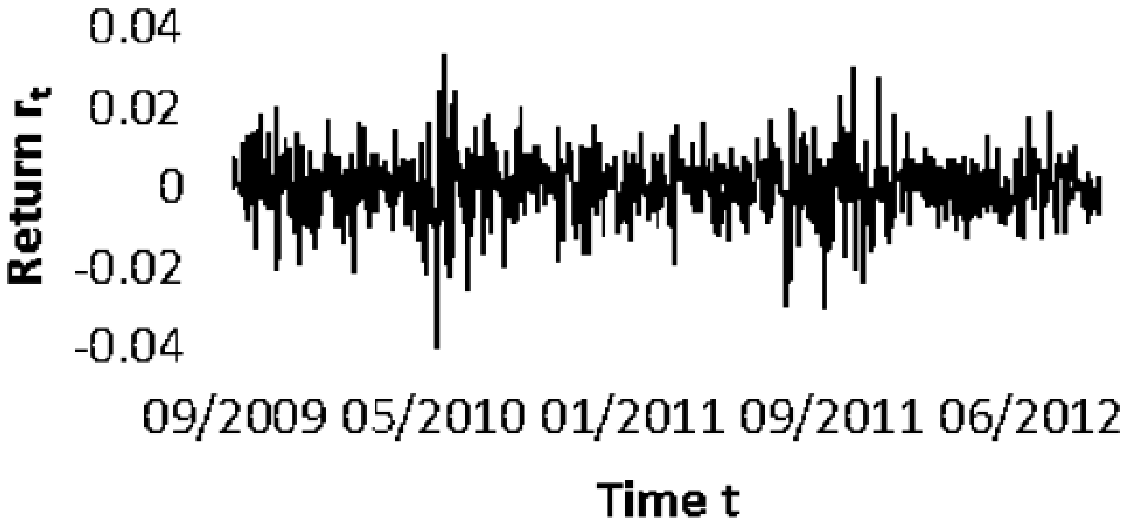

Figure 1. Time Series Plot Continuous Returns

News data is extracted from Reuter's and mainstream news from top international business sources, while the social media information contains the top 30% blogs, microblogs and other social media and is gathered exclusively by Market Psych. Due to company non-disclosure policy no further insights on the sources of data, how it is quantified, or what processes underlie the data collection and quantification are available.

3.2 Time Lags of Indicators

A preliminary analysis of the indices shows various indices being significantly (positively or negatively) correlated with each other, but at least on a simultaneous basis there is no evidence for multicollinearity. The bivariate analysis for each indicator with continuous returns tries to reveal any linear relationships. It commences with scatterplots with the squared Pearson Correlation Coefficient R2 for linear relationships to visualize linear or non-linear correlation between sentiment index and asset. Sentiment (R2=0.078) and Price Up (R2=0.087) for instance, show contemporaneous significant positive correlation as visualized in the scatterplots of Figure 2.

As most of the time series are not normally distributed, the better known Pearson's correlation coefficient is deemed. Consequently, the analysis is extended by providing Kendall's tau-b non-parametric correlation coefficients, favored over Spearman's rho as confidence intervals are less reliable and interpretable. The coefficient is computed for the concurrent values of sentiment and asset and, hereby, represents a descriptive correlation. The cross correlation with a time lag up to four days demonstrates the suitability of sentiment indicators for forecasting purposes, but also indicate a reversed causation. Table 1 shows that returns seem to be followed by high values of Sentiment (0.336), Optimism (0.139), Joy (0.153), Price Up (0.369) or Market Momentum (0.197). Fear is negatively correlated with an exceptional value on the following day (-0.252).

3.3 Binning of Indicators

An additional working hypothesis assumes that extreme values in the tails of the sentiment indicators may have more predictive power, i.e., negative sentiment on negative returns, while observations around the average may not be indicative for market movements, i.e., not neutral or positive sentiment on positive returns. In consequence, we propose rescaling and transforming the original Market Psych indicators for better interpretability. In order to simplify the analysis of such values, different binning concepts were used to categories the values separated by one or two positive and negative standard deviations and model them as categorical variables.

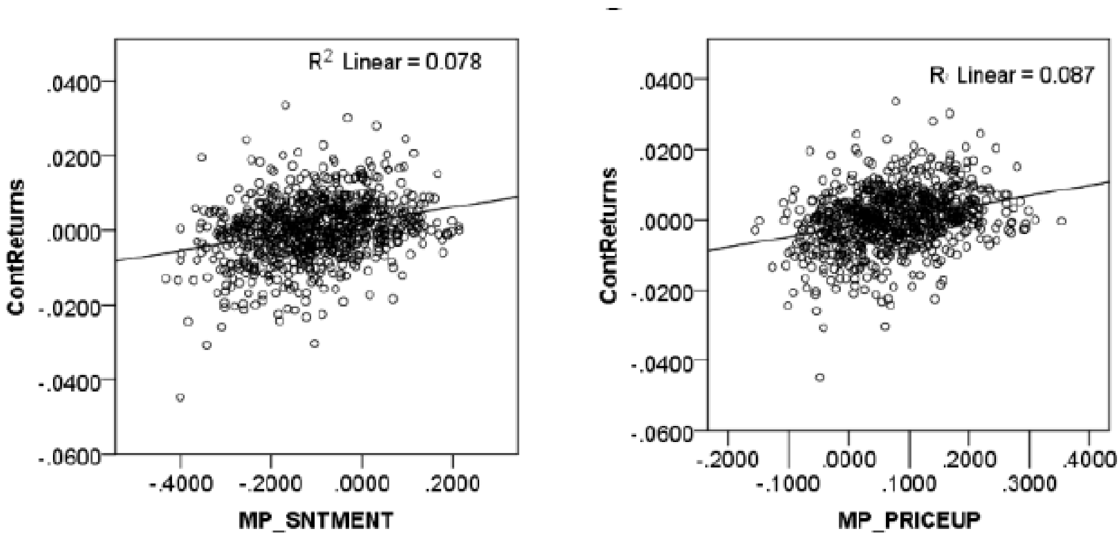

Figure 2. Scatter Plot Sentiment-Continuous Returns

Boxplots were used in a visual analysis to determine differences in the median and percentiles. Occasionally, exceptional values of sentiment showed significant correlation with returns, which were consequently modelled as categorical input variables instead of the original metric indices.

3.4 Modelling Set-Up

In order to facilitate comparisons between the linear and the non-linear forecasting models, various general settings are kept identical for both. The data partition follows the recommendation by who finds the best performance for MLPs with a large training set. 60% of the data or approx. 469 data points are used to train the parameters of the MLP, 20% or approx. 156 observations in the validation set then serve to pick the best-performing model compared to others. This choice is then evaluated on test data it has not seen before in the generalization set with the remaining 20%. More than 100 observations in the out-of-sample test ascertain statistical validity and reliability.

Table 1. Cross Correlation Kendall's Tau-B Continuous Returns

Lags	-4	-3	-2	-1	0
Buzz	-0.028	-0.001	-0.034	-0.04	-0.023
Sentiment	0.01	0.003	-0.017	-0.029	0.177**
Optimism	-0.01	-0.004	0.024	-0.024	0.082**
Fear	-0.007	-0.002	0.007	0.046	-0.065**
Joy	-0.008	-0.005	-0.024	-0.006*	0.075**
Trust	0.011	0	-0.003	0.017	-0.022
Violence	-0.003	0.012	-0.016	-0.004	0.038
Conflict	-0.066**	0.013	0.015	-0.005*	-0.057*
Urgency	-0.033	0.022	-0.027	0.013	0.033
Uncertainty	0.041	0.001	0	0.012	0.027
Price Up	0.021	0.01	-0.014	-0.026	0.196**
MarketForecast	-0.025	-0.003	-0.013	-0.002	0.062*
Carry Trade	0.002	0.048	-0.02	0.011	-0.009
Peg Instability	-0.005	0.007	-0.006	-0.02	0.011
Market Mommentum	-0.012	-0.008	0.009	-0.039	0.101**

The data is not pre-processed, so outliers, trends and non stationarities as well as potential seasonality remain in the time series, with the exception of scaling data for network training.

All MLR and MLP models are estimated in the forecasting standard software Intelligent Forecaster (IF) on an Intel Core i7 CPU 1.73 GHz 4GB RAM.

3.5 Manual Selection of Input Parameters

In case of a manual selection of input parameters to build MLRs and MLPs particular combination of manually identified information from the data exploration is employed:

1) History of lagged realizations of returns

2) Choice of continuous sentiment indicators based on bivariate analysis

3) Concurrent continuous sentiment indicators with highly significant Kendall´s tau-b correlation at the 0.01 level

4) Only Sentiment as parameter

5) Only Price Up as parameter

6) Full model with all continuous sentiment indicators

7) Binned sentiment with pattern from analysis

8) Choice of continuous and binned sentiment indicators based on bivariate analysis

9) Full model with all continuous and binned sentiment indicators

It should be noted that these represent descriptive models, which use a contemporaneous sentiment indicator *it* for the same day of the predicted return *rt* in addition to time lags up to a history of one week. This means it is assumed that the value of the sentiment in *t* is presumed to be known, as in explorative or descriptive modelling and unlike forecasting exercises. The objective of the analysis is therefore a descriptive approach to assess whether the return on currencies and its corresponding sentiment indices show common behavior.

3.6 Evaluation of Accuracy

Three different error measurements are chosen to compare models based on the validation set. In general, the error *et+h* for time *t* and forecasting horizon *h* is defined as the difference between the actual *yt+h* and the predicted value *ŷt+h*. In the case of the Mean Absolute Error (*MAE*) the error is estimated as:

$$MAE = \frac{1}{n} \sum_1^n |e_{t+h}| \quad . \tag{2}$$

The MAE treats positive and negative errors equally, but is not suitable to detect the forecasting bias. In contrast, the commonly used Root Mean Square Error (*RMSE*).

$$RMSE = \sqrt{\frac{1}{n} \sum_1^n e_{t+h}^2} \tag{3}$$

Weights large errors more than small errors, and overweighs big errors compared with the MAE, yet remains on the same scale as the time series' values.

In addition to interval scaled error metrics we consider the Directional Inaccuracy of Returns (*DIAR*), which indicates how often the algebraic sign (+ or -) is forecasted incorrectly as a percentage of all predictions, in the interpretation of correct classification of direction, and, therefore, appears to be particular suitable for interpreting returns. DIAR on the validation set serves as the main error metric for comparison, as the directional forecast seems most crucial for an investment decision. Afterwards, the types of models are compared with each other with respect to RMSE, which also appears to be somewhat suitable for financial decisions as investors want to avoid exceptional losses. The remaining errors serve for direct comparison with the naïve method, i.e., the random walk.

3.7 Multivariate Linear Regression

The MLR uses the found linear correlation between the sentiment indicators and the returns to estimate linear relationships using Ordinary Least Square (OLS) estimation. Both the explanatory and the returns are stationary time series. The variables are entered with two different selection techniques, (a) being the automated stepwise forward method with a threshold to enter α=0.05, and the other (b) based on the manual data exploration. The stepwise technique only includes significant parameters according to their *p*-values, but bears significant risks of misjudgments. A constant is either included or neglected. The variables are transformed to fit in specific ranges in order to achieve consistency in the input data so that variables with larger range (Buzz indicator) do not outweigh others. Simultaneously by using linear scaling instead of standardization the variables remain as authentically as possible with respect to data variability:

- No Scaling
- Linear Scaling with range -1; 1
- Linear Scaling with range 0.2; 0.8
- Linear Scaling with range -0.6; 0.6

Overall, 15 different lag structures are defined with stepwise selection. With respect to the experimental design settings this constructs distinct regression 120 models (15 Lag structures * 4 Transformations * 2 Constants). Manual selection results in 39 different lag structures and 312 models (39 Lag structures * 4 Transformations * 2 Constants).

3.8 Multilayer Perception

The MLP models exploit potential non-linear relationships in the sentiment indicators and lagged returns; a general introduction to neural networks is given by. The MLP remains the most widely used neural network architecture, a feed-forward network with input, hidden and output layers with various processing nodes in each. Feed-forward, hereby, refers to the direction of information flow from the input via the hidden to the output nodes, essentially modelling a nonlinear autoregressive (AR) intervention model of order p, NARX(p).

This MLP aims to predict the continuous return in t using past returns t-1, t-2, …, t-n and chosen sentiment indicators up to time t. The network employs one output node, as determined by the experiment to forecast one single return rt. Theoretical experiments have shown that one hidden layer is sufficient to approximate any complex non-linear function with high accuracy. The number of hidden nodes is determined by extensive enumeration. A trial-and-error procedure is chosen despite some well-known rules of thumb which seem problem specific. Too few nodes leave out important information, while too many lead to over fitting. A step size of 2 increases the number of nodes from n= 2, ..., 12.

In order to receive valuable and correct results the input and output factors are linearly scaled in the range 0.2; 0.8. This represents an external scaling where the training data is fit into a specified range. The headroom (difference to the range 0; 1) aims to avoid saturation effects in the asymptotic limits of the activation function. This setting is crucial for the choice of activation function and allows using the logistic function for the hidden layer. Only training and validation set is used for parameterization of the MLPs, as the observations of the test set should solely serve for the out of sample performance evaluation, and not for training (and vice versa).

All input vectors are presented randomly to the MLP, without replacement. A conventional back propagation learning algorithm is employed, minimizing a squared error loss function for a suitable local minimum. The learning rate, as the magnitude of the weight changes, is chosen as 0.5 to balance slow learning processes from small learning rates versus large rates which cause oscillations, with a high momentum of set to 0.7. We employ a decreasing factor to the learning rate to facilitate convergence by 0.99 after each epoch. The MLP employs a MAE error function to stop the learning algorithm when the error converges to a low value. If the relative improvement is below 0.01% the learning process is stopped. Squared error measures are avoided due to their theoretical limitations mentioned in the literature. The learning algorithm is given starting weights between -1 and 1. In order to allow robust network selection, the network is to be initialized more than 10 times with random starting weights. 30 starting points appear as a good choice. The target function is linear to penalize all errors equally. Overall, 7020 MLP variants and architectures are computed (39 Lag structures * 30 Starting Points * 6 Topology Variants).

4. Experimental Results

First, each type of model is evaluated with respect to its different modifications of lag-structures, used input clusters, variable transformations and included constants. Second, the best model is described with details about coefficients. Afterwards, the best models of a category are compared with each other. Different input parameters for the automatic MLR are compared with respect to continuous (Lag-G1: 0), binned (Lag-G1: 1) and a combination of continuous and binned (Lag-G1: 2) variables. The validation error shows that continuous variables outperform binned and their combined approach.

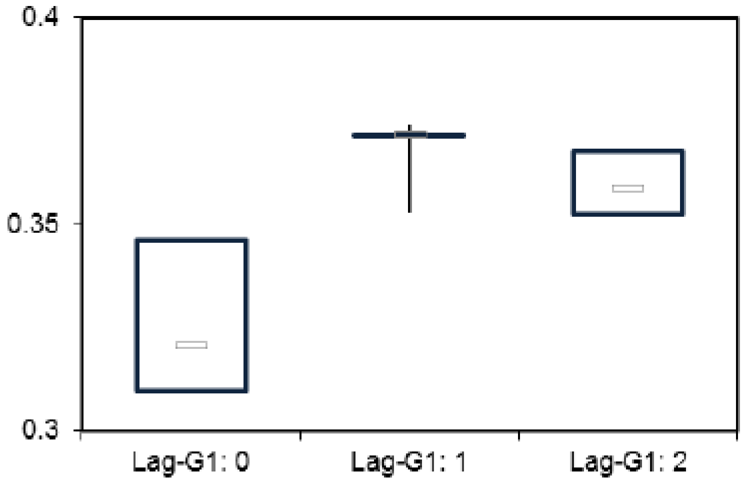

Figure 3. Stepwise MLR Input Clusters

Different variable transformations or a constant have no impact at all on the error measurement. According to DIAR on the validation set the best descriptive model is the following:

Table 2. Stepwise MLR Best Models

Lag structure	Tra E	Val E	Gen E
MLR_sw_Original_-4tol_d(0;1)	0.39400	0.30970	0.41940

The descriptive model uses the continuous variables of one week to achieve a validation error of 31%. The generalization error is significantly higher with 42%. Its generalization error is with 52% slightly above 50%. The automatic selection results in the following combination of input parameters.

Table 3. MLR_SW_Original_-4TO1_D

Time Series	Lag	p-Value	Coefficient
Constant Term		0.0356	-0.0028
MP_Priceup_3	1	0.0000	0.0302
Cont Returns	0	0.0000	-0.3003
MP_Trust_3	1	0.0159	-0.0489
MP_Sntment_3	-1	0.0109	0.0144
MP_Optimsm_3	-1	0.0185	0.0000
MP_Sntment_3	1	0.0347	0.0144
MP_Fear_2	0	0.0242	0.0000
MP_Mktfcst_3	1	0.0400	-0.0911

The previous return has the greatest impact with negative coefficient of -0.30 followed by the concurrent value of Market Forecast with -0.09. All parameters have a p-value below 0.05 and are statistically significant at the 95% confidence interval.

The manual selection of input parameters for MLR gives a similar interpretation. The different variable transformations give that no scaling is slightly superior. Including a constant results in the lowest error in the validation set this time (Reg-G2: 0). However, the majority of the models with constant term have overall a lower error in the validation set.

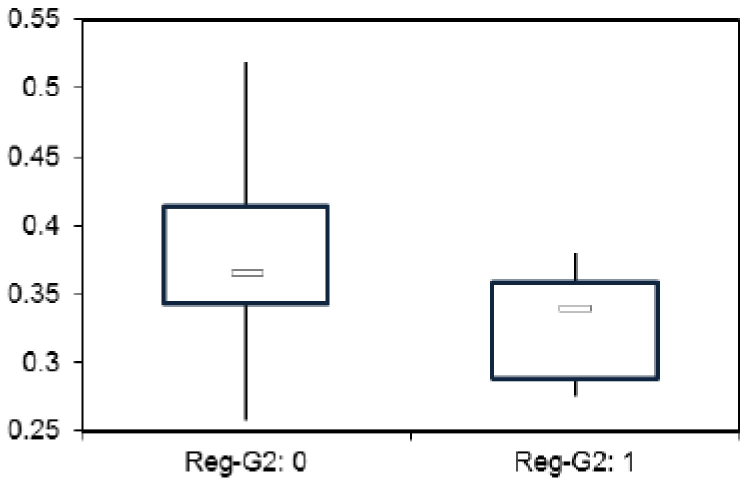

Figure 4. Manual MLR Including Constant

Various differences are visible comparing the input parameters. A few models have directional errors above 0.5 in all partitions. In general, the models with the best parameters with significant Kendall's tau-b correlation (Lag-G1: 2) outperform in the validation set with the lowest median. The best model uses only Price Up as input parameter (Lag-G1: 5).

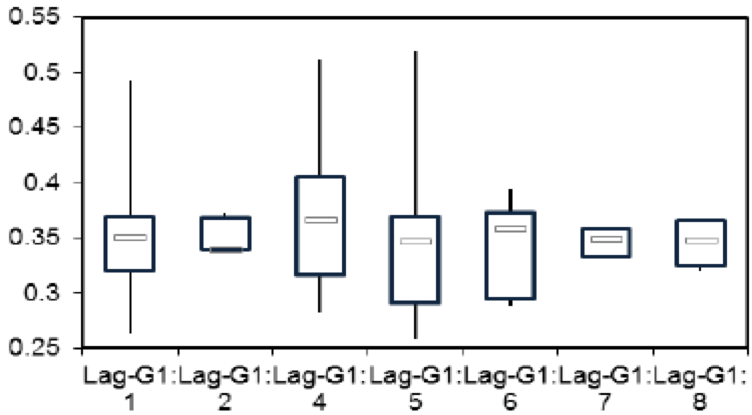

Figure 5. Manual MLR Continuous Returns Input Clusters

The best descriptive model of the manual MLR is summarized in table IV. The model is better than its correspondent with automatic selection. This means that a judge mental approach outperforms the automatic stepwise selection and indicates that the characteristics found during the data exploration are usable for the model set-up. The best descriptive model with manual selection uses the past history of Price Up of one week achieving a validation error of 26% and generalization error of 37% compared to 31% and 42% of the best model with automatic selection.

Table 4. Manual MLR Best Models

Lag structure	Tra E	Val E	Gen E
MLR_m_PriceUp_-4tol_d(5;1)	0.4111	0.2581	0.3742

The model abstains from including a constant. The previous returns enter the equation with different signs. Lag0 and lag-2 are negative while the remaining coefficients are positive. For Price Up the concurrent value is positive while previous coefficients are negative. Statistically significant, however, are only the concurrent Price Up (p-value 0.0001), the previous return (p-value 0) and Price Up (p-value .03334).

Table 5. MLR_M_PriceUp_-4TO1_D

Time series	Lag	Reg.Para	Std.errors	T-Statistics	P-values
Cont Returns	0	-0.2017	0.0524	-3.848	0.0001
	-1	0.0947	0.0528	1.7915	0.0739
	-2	-0.0538	0.0526	-1.0235	0.3066
	-3	0.0174	0.0516	0.3364	0.7367
	-4	0.0167	0.0513	0.3258	0.7447
MP_Priceup	1	0.0369	0.0052	7.1317	0.0000
	0	-0.0115	0.0054	-2.1341	0.0334
	-1	-0.0058	0.0054	-1.0817	0.2800
	-2	-0.0025	0.0054	-0.4568	0.6480
	-3	-0.0027	0.0054	-0.5012	0.6165
	-4	-0.0045	0.0047	-0.9624	0.3364

Total computation time for the MLPs was 1h 48 min. Unique for this type of model is the possibility to combine the ten best models of one error measurement into one model (Lag-G2: 0). This may lead to overall improvements. In fact, the error is lower.

The number of hidden neurons does not seem to be very sensitive and similar errors are achieved. MLPs with four neurons (Topology: N4) have the lowest deviation in the validation error while having the same median as the other. The only exception is with two neurons which has a higher error than the others as it seem to be too simple to capture all properties of the exchange rate. Nevertheless, the minimum error is identical for all neuron combinations.

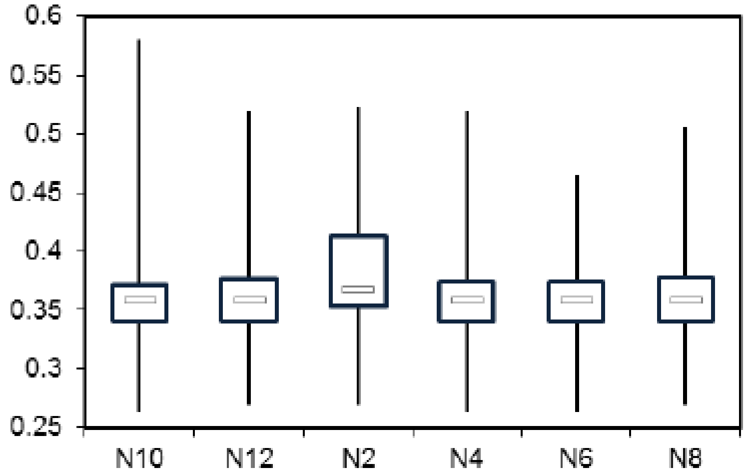

Figure 6. MLP Topology

With respect to input clusters the best errors with the lowest median are achieved by combination of the ten best models, however, their deviation is greater than for others. The models using Price Up and Sentiment as input (Lag-G1: 2) or only Sentiment (Lag-G1: 5) have the same minimum error. But the former has a lower range between minimum and maximum.

The best model uses Price Up exclusively and is, hereby, similar to the manual MLR, but outperforms the model in the validation set with error 24% compared with 26%. Surprisingly, it is slightly beaten in the generalization set with errors of 40% to 37%.

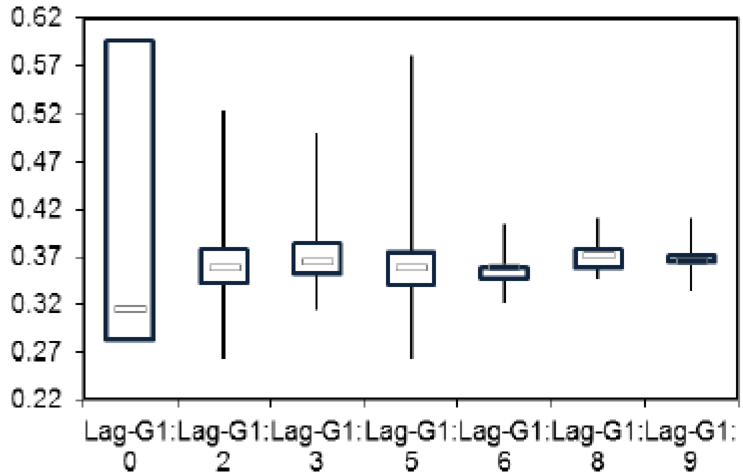

Figure 7. MLP Input Clusters

Table 6. MLP Best Models

Lag structure	Topology	Tra E	Val E	Gen E
MLP_m_PriceUp_01_d	N2	0.40510	0.24360	0.39740

Non-linear MLPs seem to outperform the linear regressions model in the relevant validation error as visible in Table 7. This aligns with previous findings in research. The automatic MLR achieves the best training error while the manual MLR is superior in the generalization error.

Table 7. Compare Best Models

Error	Automatic MLR	Manual MLR	MLP
Training	0.39400	0.41110	0.40510
Validation	0.30970	0.25810	0.24360
Generalisation	0.41940	0.37420	0.39740

In order to show that the more complex techniques of the MLP are beneficial for exchange rate prediction compared to simpler models three benchmarks are chosen. The "Only Positive" is a model which predicts only positive returns. In financial terms it represents a buy-and-hold portfolio in a long position in the currency pair. The contrasting strategy of a short position is "Only Negative" with only negative returns. Zero returns are always neglected and lead to an inaccurate prediction in both models. This is economically plausible as in case of a zero return an investor would have transactions costs which lead to an overall negative return. The aim of the comparison with these benchmark models is to show that the accuracy rate of the MLP is not due to regular patterns in the exchange rate. The naïve method uses the return of the previous day for prediction of the following.

Table 8. Benchmark Models

Error	Naïve	Only Positive	Only Negative	MLP
Training	0.5224	0.4573	0.5534	0.4051
Validation	0.4679	0.4872	0.5128	0.2436
Generalization	0.5577	0.5128	0.5000	0.3974

The errors are summarized in the previous table. The MLP is superior in all data sets and outperforms all simple benchmark models. Simplified, the directional errors of the benchmark models are around 50% each. This aligns with previous findings. The long-term average return of the time series is zero, so positive and negative returns are equally evident. MLP as well as manual and automatic MLR justify their complexity with significantly lower inaccuracy rates.

5. Conclusion & Critical Consideration

The previous modeling has shown that sentiment indicators are able to describe market behavior. With respect to continuous returns the non-linear descriptive model can achieve a directional accuracy of 75.64% in the validation set and a real-world rate of 60.26% in the generalization set. The empirical results led to the conclusion that non-linear models outperform linear regressions and also simple benchmark models like the naïve method. It is superior to both long and short buy-and-hold strategies. This aligns with previous findings in forecasting exchange rates due to the characteristics of the foreign exchange market. For subsequent research it would be interesting to investigate whether the indices contain information with a timely advantage prior to a market movement on which a trading strategy can be based on. We hypothesize that high Sentiment may lead to high returns on the following trading day.

The bivariate analysis has also shown that the relation between the exchange rate and various sentiment increases after market movements. This leads to the assumption that there is a relationship which might not be predictive for returns, but rather indicates a reversed causation. The question is whether the market movements, and in particular, extraordinary returns or changes in uncertainty actually influence the mood of people. Further research based on this data is required to answer this question.

The analysis has exposed the ability to describe financial time series using sentiment indicators, but further research is required whether it is also able to predict them. The trading window of daily prices might be too rough for a highly liquid market as the FOREX. Already today, the market is dominated by automatic algorithmic trading which use any kind of available information—including sentiment indicators. Even though these technologies bear doubtlessly its own risks, they contribute to market liquidity and adhoc corrections in order to achieve market equilibrium without long-term arbitrage opportunities. High-frequency data might contain novel information for shorter time periods which may enable profitable trading strategies.

Acknowledgments

We would like to thank the deanship of scientific research at Prince Sattam Bin Abdul-Aziz University for supporting us in this research.

References

Adya, M., & Collopy, F. (1998). How effective are neural networks at forecasting and prediction? A review and evaluation. *Journal of Forecasting*, *17*(5-6), 481-495. https://doi.org/10.1002/(SICI)1099-131X(1998090)17:5/6<481::AID-FOR709>3.0.CO;2-Q

Akhtar, S., Faff, R., & Oliver, B. (2011). The asymmetric impact of consumer sentiment announcements on Australian foreign exchange rates. *Australian Journal of Management*, *36*(3), 387-403. https://doi.org/10.1177/0312896211410723

Anderson, S. et al. (2013). Fear and Closed-End Fund discounts. *Applied Economics Letters*, *20*(10), 956-959. https://doi.org/10.1080/13504851.2013.767969

Antweiler, W., & Frank, M. Z. (2004). Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards. *The Journal of Finance*, *59*(3), 1259-1294. https://doi.org/10.1111/j.1540-6261.2004.00662.x

Azoff, E. M. (1994). *Neural Network Time Series Forecasting of Financial Markets*. Chichester: John Wiley and Sons.

Baker, M., & Stein, J. C. (2004). Market liquidity as a sentiment indicator. *Journal of Financial Markets*, *7*(3), 271-299. https://doi.org/10.1016/j.finmar.2003.11.005

Bank of International Settlement. (2010).

Bathia, D., & Bredin, D. (2012). An examination of investor sentiment effect on G7 stock market returns. *The European Journal of Finance*, 1-29.

Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, *2*(1), 1-8. https://doi.org/10.1016/j.jocs.2010.12.007

Brooks, C. (1996). Testing for non-linearity in daily sterling exchange rates. *Applied Financial Economics*, *6*(4), 307-317. https://doi.org/10.1080/096031096334105

Brown, G. W., & Cliff, M. T. (2004). Investor sentiment and the near-term stock market. *Journal of Empirical Finance*, *11*(1), 1-27. https://doi.org/10.1016/j.jempfin.2002.12.001

Chen, H. et al. (2012). Customers as Advisors: The Role of Social Media in Financial Markets. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2024086

Chen, S. S. (2011). Consumer confidence and stock returns over market fluctuations. *Quantitative Finance*, *12*(10), 1585-1597. https://doi.org/10.1080/14697688.2011.565363

Crone, S. F. (2005). Stepwise selection of artificial neural network models for time series prediction. *Journal of Intelligent Systems*, *14*(2-3), 99-122. https://doi.org/10.1515/JISYS.2005.14.2-3.99

Crone, S. F., & Dhawan, R. (Eds.). (2007). Forecasting Seasonal Time Series with Neural Networks: A Sensitivity Analysis of Architecture Parameters. In *Neural Networks, 2007*. IJCNN 2007. International Joint Conference on. https://doi.org/10.1109/IJCNN.2007.4371282

Crone, S. F., & Pressmar, D. B. (Eds.). (2006). *An extended evaluation framework for neural network publications in sales forecasting*.

Crone, S. F. (Ed.). (2002). Training artificial neural networks for time series prediction using asymmetric cost functions. In *Neural Information Processing, 2002*. ICONIP 02. Proceedings of the 9th International Conference on (5). https://doi.org/10.1109/ICONIP.2002.1201919

Cybenko, G. (1992). Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals and Systems*, *5*(4), 455. https://doi.org/10.1007/BF02134016

Fama, E. F. (1965). THE BEHAVIOUR OF STOCK-MARKET PRICES. *Journal of Business*, *38*(1), 34-105. https://doi.org/10.1086/294743

Fan, W., & Watanabe, T. (n.d.). Dynamic prediction of forthcoming trends in stock prices from news articles. In *Proceedings of the 2nd International Conference on Web Intelligence, Mining and Semantics* (pp. 16.1-16.9).

Flom, P. L., & Casell, D. L. (Eds.). (2007). *Stopping stepwise: Why stepwise and similar selection methods are bad, and what you should use*. North East SAS Users Group Inc 20th Annual Conference: 11-14th November 2007, Baltimore, Maryland.

Harrell, F. E. (2001). Regression modeling strategies. In *With applications to linear models, logistic regression, and survival analysis*. New York: Springer. https://doi.org/10.1007/978-1-4757-3462-1

Hens, T., & Rieger, M. O. (2010). Financial economics. In *A concise introduction to classical and behavioral finance*. Heidelberg, New York: Springer. https://doi.org/10.1007/978-3-540-36148-0

Hirshleifer, D., & Shumway, T. (2003). Good Day Sunshine: Stock Returns and the Weather. *Journal of Finance*, *58*(3), 1009-1032. https://doi.org/10.1111/1540-6261.00556

Kaya, M. I. Y., & Karsligil, M. E. (Eds.). (2010). Stock price prediction using financial news articles. In *Information and Financial Engineering (ICIFE), 2010 2nd IEEE International Conference on*. https://doi.org/10.1109/ICIFE.2010.5609404

Kendall, M. G., & Gibbons, J. D. (1990). *Rank Correlation Methods* (5th ed.). USA: Oxford University Press (Charles Griffin Book Series).

Lavrenko, V. et al. (Eds.). (2000). *Language Models for Financial News Recommendation*. 354756: ACM.

Lee, V. C. S., & Tshung Wong, H. (2007). A multivariate neuro-fuzzy system for foreign currency risk management decision making. *Advanced Neuro computing Theory and Methodology Selected papers from the International Conference on Intelligent Computing 2005 (ICIC2005) International Conference on Intelligent Computing 2005*, *70*(4-6), 942-951. https://doi.org/10.1016/j.neucom.2006.10.025

Malkiel, B. G. (1973). *A random walk down wall street*. New York: W.W. Norton and Company Ltd.

Malkiel, B. G. (1977). The Valuation of Closed-End Investment-Company Shares. *The Journal of Finance*, *32*(3), 847-859. https://doi.org/10.1111/j.1540-6261.1977.tb01993.x

Ord, J. K., & Fildes, R. (2013). *Principles of business forecasting*. Mason, OH: South-Western Cengage Learning.

Panda, C., & Narasimhan, V. (2007). Forecasting exchange rate better with artificial neural network. *Journal of Policy Modeling*, *29*(2), 227-236. https://doi.org/10.1016/j.jpolmod.2006.01.005

Persky, J. (1995). The Ethology of Homo Economics. *Journal of Economic Perspectives*, *9*(2), 221-231. https://doi.org/10.1257/jep.9.2.221

Peterson, R. L. (2007). Inside the Investor's Brain. In *The Power of Mind over money*. Hoboken, New Jersey: John Wiley & Sons, Inc.

Rodgers, J. L., & Nicewander, W. A. (1988). Thirteen Ways to Look at the Correlation Coefficient. *The American Statistician*, *42*(1), 59-66. https://doi.org/10.2307/2685263

Sarle, W. S. (Ed.). (1994). *Neural Networks and Statistical Models*.

Schmeling, M. (2009). Investor sentiment and stock returns. Some international evidence. *Journal of Empirical Finance*, *16*(2), 394-408. https://doi.org/10.1016/j.jempfin.2009.01.002

Schumaker, R. P. et al. (2012). Evaluating sentiment in financial news articles. *Decision Support Systems*, *53*(3), 458-464. https://doi.org/10.1016/j.dss.2012.03.001

Schumaker, R. P., & Chen, H. (2009). A quantitative stock prediction system based on financial news. *Information Processing & Management*, *45*(5), 571-583. https://doi.org/10.1016/j.ipm.2009.05.001

Schwartz, H. H. (2007). A Introduction to Behavioral Economics: The Complicating but Sometimes Critical Considerations. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.960222

Shu, H. C. (2010). Investor mood and financial markets. *Journal of Economic Behavior & Organization*, *76*(2), 267-282. https://doi.org/10.1016/j.jebo.2010.06.004

Tang, Z., Almeida, C. de., & Fishwick, P. A. (1991). Time series forecasting using neural networks vs. Box-Jenkins methodology. *Simulation*, *57*(5), 303-310. https://doi.org/10.1177/003754979105700508

Taylor, S. J. (2011). *Asset Price Dynamics, Volatility, and Prediction*. Princeton University Press. https://doi.org/10.1515/9781400839254

Tetlock, P. C. (2007). Giving Content to Investor Sentiment: The Role of Media in the Stock Market. *Journal of Finance*, *62*(3), 1139-1168. https://doi.org/10.1111/j.1540-6261.2007.01232.x

Thomson, R. (2012). *Thomson Reuters Market Psych Indices*. Output Image Format Guide.

Tolle, K. M., & Chen, H. (2000). Comparing noun phrasing techniques for use with medical digital library tools. *Journal of the American Society for Information Science*, *51*(4), 352-370. https://doi.org/10.1002/(SICI)1097-4571(2000)51:4<352::AID-ASI5>3.0.CO;2-8

Vincent, A., & Armstrong, M. (2010). Predicting Break-Points in Trading Strategies with Twitter. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.1685150

Wright, W. F., & Bower, G. H. (1992). Mood effects on subjective probability assessment. *Organizational Behavior and Human Decision Processes*, *52*(2), 276-291. https://doi.org/10.1016/0749-5978(92)90039-A

Yao, J., & Tan, C. L. (2000). A case study on using neural networks to perform technical forecasting of forex. *Neuro computing*, *34*(1-4), 79-98. https://doi.org/10.1016/S0925-2312(00)00300-3

Young, M. (1989). *The Technical Writer's Handbook*. Mill Valley, CA: University Science.

Zaiyong, T., & Fishwick, P. A. (1993). Feed forward Neural Nets as Models for Time Series Forecasting. *ORSA Journal on Computing*, *5*(4), 374. https://doi.org/10.1287/ijoc.5.4.374

Zhang, G., Patuwo, E. B., & Hu, Y. M. (1998). Forecasting with artificial neural networks: The state of the art. *International Journal of Forecasting*, *14*(1), 35-62. https://doi.org/10.1016/S0169-2070(97)00044-7