

# A Sentiment Analysis of Twitter Content as a Predictor of Exchange Rate Movements

**SERDA SELIN OZTURK**

*Istanbul Bilgi University \**

**KURSAD CIFTCI**

*Istanbul Bilgi University*

Recently, social media, particularly microblogs, have become highly valuable information resources for many investors. Previous studies examined general stock market movements, whereas in this paper, USD/TRY currency movements based on the change in the number of positive, negative and neutral tweets are analyzed. We investigate the relationship between Twitter content categorized as sentiments, such as Buy, Sell and Neutral, with USD/TRY currency movements. The results suggest that there exists a relationship between the number of tweets and the change in USD/TRY exchange rate.

*Keywords:* Sentiment Analysis, Social Media, Foreign Exchange Markets, Data Mining

*JEL Classifications:* C82, G15, F31.

## 1 Introduction

According to Efficient Market Hypothesis (Fama, 1965), all available new and even hidden information is reflected in market prices because investors, being rational, want to maximize profits. The researchers widely accepted this hypothesis as a primary idea explaining the markets in general. However, the undeniable role of human sentiment, behaviors and emotions, which include the social mood in financial decision making, (Nofsinger2010), and behavioral finance as pioneered by Kahneman and Tversky (1979), have become a rising area of research challenging the existence of efficient markets (see, for example, Baker & Wurgler, 2007).

---

\* We thank Thanasis Stengos and an anonymous referee for helpful comments on an earlier draft of this paper. We are solely responsible for the remaining mistakes.

Serda Selin Ozturk: [serda.ozturk@bilgi.edu.tr](mailto:serda.ozturk@bilgi.edu.tr), Kursad Ciftci: [kursad.ciftci@gmail.com](mailto:kursad.ciftci@gmail.com)

© 2014 Serda Selin Ozturk and Kursad Ciftci. Licensed under the Creative Commons Attribution - Noncommercial 3.0 Licence (<http://creativecommons.org/licenses/by-nc/3.0/>). Available at <http://rofea.org>.

Sentiment analysis, also known as opinion mining, is a computer-based study that addresses opinion-oriented natural language processing. Such opinion-oriented studies include, among others, genre distinctions, emotion and mood recognition, ranking, relevance computations, perspectives in text, text source identification and opinion oriented summarization (Kumar and Sebastian, 2012). Through the growing application of internet technologies and the possibility of the implementation of practical applications in many areas, sentiment analysis is gaining importance. In this context, researchers have explored a variety of methods to compute indicators of the public sentiment and mood from large-scale online data.

There are three prominent types of online data sources that have been used for financial analysis. First, news sites have been perceived to be an important source for the sentiment of the investor (Tetlock, 2007). Second, search engine data have been informative before market fluctuations by the use of search volume of the stock (Da et al., 2010). Third and last, the source of online data is social media feeds. Social media feeds are becoming highly important for determining and measuring the social mood and investor behavior. Twitter is a popular microblogging service whereby users write messages called “tweets”. Millions of messages are written daily, and there is no limitation for the content.

The aim of this paper is to identify, investigate and evaluate valuable features and patterns in the price movement performance of the US Dollar/Turkish Lira exchange rate (USD/TRY) parity by analyzing the relevant data gained from Twitter with sentiment analysis. We will try to determine if an analysis of the messages found on Twitter can be used as an input in the decision-making process for investing decisions. The main contribution of the paper is the application of Twitter sentiment analysis on the foreign exchange market from an emerging economy. This study represents the first time that such an investigation has been done in that context, although there are many works examining the effects of Twitter-generated sentiments on stock market movements. One benefit of studying the foreign exchange market is that, due to the huge forex volume, the possibility of manipulation due to Twitter misinformation is expected to be lower than in stock market analysis.

In the next section, we present the relevant literature. In section 3 we describe data used and methodology. Our results are in section 4. The last section concludes.

## **2 Review of the Literature**

There are many studies that aim to identify a method to predict or even understand financial market movements properly. After the popularization of social media, such as blogs, micro-blogs and such sites as Facebook, there is now a new data source. This data source contains huge amounts of data; therefore, with the help of developed computer technologies, the use of sentiment analysis techniques has grown in recent years. Early papers analyzing the effects of sentiments on stock performance such as Wysocki (1998), Tumarkin and Whitelaw (2001),

Dewally (2003), Antweiler and Frank (2004), Gu et al. (2006) have determined, with various degrees of success, the role that sentiments in the form of micro blogs and/or micro stock message board activity have in the prediction of excessive stock returns. These findings have been supported by Tayal and Komaragiri (2009), who found that analyzing micro-blogs is more reliable for predicting the future performance of a company than larger regular blogs.

The use of Twitter-based micro-blog data has been applied in areas other than the prediction of stock behavior (see, for example, Asur and Huberman, 2010, for the prediction of movie box-office revenue by the analysis of relevant tweets). Twitter-based stock market analysis is found in papers by Sprenger and Welpel (2010), Bar-Haim et al. (2011), Bollen et al. (2011), Hsu et al. (2011), Mao et al. (2011), Zhang et al. (2011), Ruiz et al. (2012) and Rao and Srivastava (2013). These studies have found that Twitter sentiments have a significant predictive ability for stock behavior, although not always a strong one. In a related study of the forex market, Vincent and Armstrong (2010) found that Twitter data are more useful for predicting break points than classical methods. However, their analysis did not attempt to evaluate the predictive impact of Twitter-based sentiments on the forex market, nor did they analyze the forex market of an emerging economy, which we undertake in our paper.

### 3 Data and Methodology

#### 3.1 Data

In this paper, data mining and text mining techniques are used to predict changes in the exchange rate based on the relevant Twitter content. Data mining is a process to identify new knowledge from existing large data sets (Chakrabarti et al., 2004). Text mining refers to the process of discovering interesting patterns from text documents (Tan, n.d.). We use data mining techniques, such as regression and dimension reduction, and text mining techniques, such as sentiment analysis, are used to analyses parity movement performance.

The analysis is based on the data for the year 2013. The daily exchange rate of USD/TRY, i.e. the number of Turkish lira per one dollar were retrieved from Central Bank of the Republic of Turkey ([www.tcmb.gov.tr](http://www.tcmb.gov.tr)), for the period 01.01.2013 – 31.12.2013. Return series are calculated by the formula below:

$$Ret = \ln \frac{p_t}{p_{t-1}}$$

where:  $p_t$  denotes the exchange rate at the end of day  $t$

The reason for using return data instead of price data is to determine the direction of currency movement. If the return value is positive, the value of the dependent variable is defined as 1,

and if return value is negative or equal to 0, it is defined as 0. Table 1 below summarizes the descriptive statistics for return of USD/TRY.

Table 1. Descriptive Statics for Return of the USD/TRY Exchange Rate

	Return
Mean	0.0005
Standard Deviation	0.0111
Median	0.0000
Maximum	0.1294
Minimum	-0.1292
Skewness	-0.0685
Kurtosis	111.1495
Jarque Berra	163748.8
P-value	0.0000

Twitter sentiments have been collected through Twitter with the keywords USD/TRY, #USD/TRY, Dollar, #Dollar from 01.01.2013 to 31.12.2013. All tweets are sorted into 3 different categories depending on their sentiment, which are Buy, Sell and Neutral. The terms defined as Buy, Sell and Neutral represent the number of tweets by the given date. Tweets are calculated for each day and distributed for the three categories depending on their sentiment, as mentioned previously. Table 2 below displays used Twitter content as an example. When the first comment was made the exchange rate was around 2, therefore it indicates a BUY sentiment.

Table 2. Example Classification of Twitter Content

User Name	Date	Twit	Sentiment
@Tembel_Trader	31/10/2013	In March #USDTRY will be 2.2 #BIST100 69000	BUY
@selmanozalp	20/07/2013	After a recent upward movement, USDTRY will sharply decrease	SELL
@DirencNokta	16/08/2013	The margin of USDTRY is get too narrow	NEUTR

Finally, to combine return series and Twitter data, the number of positive, negative and neutral sentiments are calculated separately for each day.<sup>1</sup>

### 3.2 Methodology

We use a binary dependent variable, logit model, to conduct the analysis of the Twitter data. The logit regression model is displayed below:

---

<sup>1</sup> The data are available from the authors upon request.

$$Z_t = \alpha + \beta_1 Buy_{t-1} + \beta_2 Sell_{t-1} + \beta_3 Neut_{t-1} + u_t$$

$$P_t = f(Z_t)$$

$$Z_t = \begin{cases} 1 & \text{if return at time } t > 0 \\ 0 & \text{otherwise} \end{cases}$$

$Buy_{t-1}$  = Number of positive tweets on the preceding day

$Sell_{t-1}$  = Number of negative tweets on the preceding day

$Neut_{t-1}$  = Number of neutral tweets on the preceding day

$u_t$  = iid disturbance term

The sentiments in Twitter, which are the Buy, Sell and Neutral data, are used from the 24 hours in the preceding day since we expect a delay in the effects of Twitter data on the exchange rate.

#### 4 Results

The results of the logit regression are shown in Table 3.

Table 3. Regression Results

Variable	Coefficient	Prob.
C	-0.686	0.002
Buy <sub>t-1</sub>	0.09	0.000*
Sell <sub>t-1</sub>	-0.17	0.000*
Neut <sub>t-1</sub>	0.042	0.089
McFadden R-squared		0.168

\* Significant at the 5% level

The results indicate that Twitter sentiment affects the exchange rate. The coefficient on the Buy<sub>t-1</sub> variable is positive, and the coefficient on the Sell<sub>t-1</sub> variable is negative. Both are significant at the 1% level. The coefficient on the Neut<sub>t-1</sub> variable is insignificant.

The McFadden R-squared value is 0.168 in this model. Normally, the value of McFadden ranges from 0 to 1, and if the results are between 0.2-0.4, it is accepted as the perfect fit. The value of McFadden in this study is very close to 0.2, so it can be accepted as a strong model. Because normally there are many factors that can affect the exchange rate, and they are left out in this study to concentrate on the effects of tweets, a perfect fit should not be expected. The mean values of the Twitter sentiment in the analyzed period are summarized in table 4 below.

Table 4. Mean Values for the Number of Each Tweet Classifications

	Mean
Buy <sub>t-1</sub>	7.68
Sell <sub>t-1</sub>	4.53
Neut <sub>t-1</sub>	10.49
Z	3.576

The probability of an increase in the exchange rate based on the Z value calculated when all the variables are at their mean values is approximately 42 percent.

To calculate the effect of a one-unit increase in the variables, the formulas used are the following:

$$\text{Marginal Effect of Buy}_{t-1} = f(z)\beta_1$$

$$\text{Marginal Effect of Sell}_{t-1} = f(z)\beta_2$$

$$\text{Marginal Effect of Neut}_{t-1} = f(z)\beta_3$$

The  $f(z)$  value is equal to 0.244. The calculated marginal effects can be found in Table 5.

Table 5. Marginal Effects of Each Tweet Classifications

	Marginal Effect
Buy <sub>t-1</sub>	0.022
Sell <sub>t-1</sub>	-0.041
Neut <sub>t-1</sub>	0.010

These results imply that, when the number of tweets are at their mean values, a 1% increase in the number of tweets expressing positive sentiment in the preceding day raises the probability of an increase in the exchange rate by 0.17%, and a 1% increase in the number of in the

number of tweets expressing negative sentiment raises the probability of a decrease in the exchange rate by 0.19%.

Table 6 summarizes the estimated equation analysis for logit estimation.  $P(Z=1) \leq 0.5$  shows the number of times that the predicted probability of  $Z$  being equal to 1 is less than or equal to 0.5;  $P(Z=1) > 0.5$  shows the number of times it is less than 0.5. We treat prediction as correct when either  $P(Z=1) \leq 0.5$  and  $Z=0$  or  $P(Z=1) > 0.5$  and  $Z=1$ . Using this criterion the model correctly predicts 58% of increases in the exchange rate and 82% of decreases. These results suggest that the model fit is reasonable, even though the only variables used to explain the movement of the exchange are the number of tweets in the previous day.

Table 6. Estimated Equation Analysis

	Estimated Equation		
	Z=0	Z=1	Total
$P(Z=1) \leq 0.5$	156	61	217
$P(Z=1) > 0.5$	34	84	118
Total	190	145	335
Correct	156	84	240
% Correct	82.11	57.93	71.64
% Incorrect	17.89	42.07	28.36

## 5 Conclusions

The Efficient Market Hypothesis was challenged by behavioral finance by observing the importance of human emotion, sentiment and mood in financial decision-making. Thus, a question arose: What is the best model to predict the behavior of the financial markets? Previous research tried to analyze many different types of content such as surveys and news. After the growth of internet technologies and as a result of social media, researchers started paying attention on the predictive ability of social media. In many areas, social media data are analyzed for purposes of prediction, such as stock markets, box-office returns and brand value; however, there is no research on the foreign exchange markets to date. This paper's contribution to the literature is the analysis of USD/TRY data with the relevant Twitter content.

In this paper, we study Twitter data for the period 01.01.2013 to 31.12.2013 to analyze exchange rate sentiment. The returns of USD/TRY are calculated on a daily basis, and the logit model is used to analyze Twitter content and currency data.

We find that there is significant relationship between Twitter sentiment and USD/TRY movements. Our study indicates that it is possible to improve the prediction of exchange range movements using Twitter data.

## References

- Agresti, A. (2007), *An Introduction to Categorical Data Analysis*, Wiley&Sons, 2nd Edition.
- Antweiler, W. & Frank, M. (2004), Does Talk Matter? Evidence From a Broad Cross Section of Stocks. Working Paper, University of Minnesota.
- Asur, S. & Huberman, B. (2010), *Predicting The Future With Social Media*. In Proceedings of 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT), vol 1, 492-499..
- Baker, M. & Wurgler, J. (2007), Investor Sentiment in the Stock Market. *Journal of Economic Perspectives*, 21(2), 129-152.
- Bar-Haim, R., Dinur, E., Feldman, R., Frsko, M. & Goldstein, G. (2011), *Identifying and Following Expert Investors in Stock Microblogs*. In Proceedings of the Conference on Empirical Methods in Natural Language Processing, 1310–1319.
- Bollen, J., Mao, H. & Zeng, X. (2010), Twitter Mood Predicts the Stock Market. *Journal of Computational Science*, 2(1), 1–8.
- Bommel, J. (2003), Rumors. *The Journal of Finance*, 58(4), 1499–1519
- Chakrabarti, S., Ester, M., Fayyad, U., Gehrke, J., Han, J., Morishita, S., Piatetsky-Shapir, S., & Wan, W. (2006), *Data Mining Curriculum : A Proposal (Version 1.0)*, ACM SIGKDD, Curriculum Committee.
- Da, Z., Engelberg, J. & Gao, P. (2014), The Sum of All Fears: Investor Sentiment and Asset Prices. *The Review of Financial Studies*, 66(5), 1461-1499.
- Dewally, M. (2003), Internet Investment Advice: Investing with a Rock of Salt. *Financial Analysts Journal*, 59(4), 65–77.
- Fama, E.F. (1965), The Behavior of Stock-Market Prices. *The Journal of Business*, 38(1), 34–105.
- Gu, B., Konana, P. & Liu, A. (2006), Identifying Information in Stock Message Boards and Its Implications for Stock Market Efficiency. Workshop on Information Systems and Economics, Los Angeles, CA.
- H. Stock, J. & W. Watson, M. (2010), *Introduction to Econometrics*. Addison-Wesley, 3rd Edition.
- Cohen, J., Cohen, P., West, S. G. & Aiken, L. (2013), *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*. Routledge.
- Hsu, E., Shiu, S. & Torczynski, D. (2011), Predicting Dow Jones Movement with Twitter. CS229 Final Project, Autumn.
- Kahneman, D. & Tversky, A. (1979), Prospect Theory: An Analysis of Decision Under Risk. *Econometrica*, 47(2), 263-291.
- Karabulut, Y. (2013), *Can Facebook Predict Stock Market Activity?* In Proceedings of American Finance Association 2013 Meetings.

- Kumar, A. & Sebastian, T.M. (2012), Sentiment Analysis: A Perspective on its Past, Present and Future. *International Journal of Intelligent Systems and Applications*, 4(10), 1–14.
- Mao, H., Counts, S. & Bollen, J. (2011), *Predicting Financial Markets: Comparing Survey, News, Twitter and Search Engine Data*. arXiv.org, Quantitative Finance Papers 1112.1051.
- Nofsinger, J.R. (2005), Social Mood and Financial Economics. *Journal of Behavioral Finance*, 6(3), 37–41.
- Rao, T. & Srivastava, S. (2013), *Modeling Movements in Oil, Gold, Forex and Market Indices Using Search Volume Index and Twitter Sentiments*. In Proceedings of the 5th Annual ACM Web Science Conference.
- Ruiz, E. J., Hristidis, V., Castillo, C., Gionis, A., & Jaimes, A. (2012), *Correlating Financial Time Series with Micro-blogging Activity*. In Proceedings of the fifth ACM international conference on Web search and data mining.
- Speed, T.P. & Yu, B. (1993), Model Selection and Prediction: Normal Regression. *Annals of the Institute of Statistical Mathematics*, 45(1), 35–54.
- Sprengrer, T.O. & Welppe, I.M. (2014), Tweets and Trades: The Information Content of Stock Microblogs. *European Financial Management*, 20(5), 926-957
- Tan, A. (1999), *Text Mining : The state of the art and the challenges Concept-based*. In Proceedings of the PAKDD 1999 Workshop on Knowledge Discovery from Advanced Databases.
- Tayal, D. & Komaragiri, S. (2009), Comparative Analysis of the Impact of Blogging and Micro-blogging on Market Performance. *International Journal*, 1(3), 176–182.
- Tetlock, P.C. (2007), Giving Content to Investor Sentiment : The Role of Media in the Stock Market. *The Journal of Finance*, 63(3), 1139-1168.
- Tumarkin, R. & Whitelaw, R.F. (2001), News or Noise? Internet Postings and Stock Prices. *Financial Analysts Journal*, 57(3), 41–51.
- Vincent, A. & Armstrong, M. (2010), Predicting Break-Points in Trading Strategies with Twitter. Available at SSRN 1685150.
- Wysocki, P. (1998), Cheap Talk on the Web: The Determinants of Postings on Stock Message Boards. University of Michigan Business School Working Paper 98025.
- Zhang, X., Fuehres, H. & Gloor, P. (2011), Predicting Stock Market Indicators Through Twitter “I Hope It is not as Bad as I Fear.” *Procedia-Social and Behavioral Sciences*, 26, 55–62.