[[1]](#footnote-1)

Twitter Informed Trading on FOREX

Remy Lagrois, Daniel Lowery, Mike Sylvester

*Abstract*—Currency traders, like many others, have a presence on social media. They tweet about the trades they have made, are going to make, and what their position on different currency pairs are. We collected their tweets about the Euro/US Dollar pairing along with real time price updates from FOREX and loaded them into MongoDB. Code written in python was used to execute trades based on rules which took into account the number of tweets or the movement of the average rate. The pair was also traded using a buy and hold strategy and at random intervals. The results of these trading strategies were analyzed using SAS to determine if Twitter based trades outperformed the other methods and if not which was the optimal strategy.

*Index Terms*— Currency Trading, Dollar, Euro, FOREX, MongoDB, Python, SAS, Social Media, Twitter, USD

# INTRODUCTION

T

HIS century has seen a massive explosion of created and stored data. How to leverage this data is a new challenge with possibilities really only limited by imagination and the willingness to experiment. To that end we set out to try to use the social media platform Twitter to try to predict movements in the Foreign Exchange Market. Furthermore we wanted to use those predictions to actually turn a profit by trading between the US dollar and the Euro. This paper represents the first of many steps in accomplishing that goal. It details our methods and lessons learned which we can build off and experiment further on.

# FOREX and Twitter Data

The Foreign Exchange Market, more widely known as FOREX or FX, is the most traded currency market in the world. Trillions of dollars are traded daily globally by various organizations such as banks and individual investors. Unlike more traditional trading platforms the FOREX market is open 24 hours a day five days a week. This allows the various time zones to openly execute trades against all available currencies. The FOREX market is based on currency exchange rates and traders buy or sell selected currency pairs at a given rate known as the “Ask” price. Conversely, the currency pair can be sold at the “Bid” price. The goal is simple; buy at a lower price and sell at a higher price, or, sell at one price and buy at a lower price known as a short sell. One of the most common currency pairs is the euro against the dollar indicated by the symbol EUR/USD. Trading this pair consists of buying or selling one currency and selling the other, hence the idea of an exchange. While FOREX trading is rather simple there are many details of the process not described. This level of detail is not a fundamental aspect of this project, therefore, a general understanding of the concept is adequate.

Twitter is an avenue for millions of people to express their thoughts, feelings, or simply what is happening around them.  As such it represents a mine of information for the things happening in society and how those events affect it.  Researchers have leveraged this new resource to make quickly reacting predictions of greater social trends.  A well-publicized example is the ability to track the spread of the flu in real time and predict where an outbreak will occur.[1] Twitter data has also been used on less concrete events as well.  The volume of Tweets about a given Presidential candidate can correlate very strongly with their polling numbers.  One group of researchers found that analyzing the actual sentiment of the Tweets was no more effective than simply looking at the volume.[2] Given that we will only be taking volume into account.

## FOREX Data Collection

The project relies on two sources of data; FOREX and Twitter. Combined they form the basis for the project, however, they are gathered from different sources using slightly different technologies. Freely distributed application programming interfaces (API’s) are provided in both cases but the software and techniques used to capture and store the data are quite different between the two data sources.

The FOREX data API used integrates well with Microsoft.NET technologies. The particular API utilized is broken into two distinct categories; rate data interface and trading functions. The rate data interface is of primary interest in this project since it provides a direct TCP/IP socket interface to the price publication system. In other words, it provides the real-time FOREX data. The code required to connect and retrieve the data is quite simple but highly dependent on an interface key.



Fig. 1. FOREX API interface key

The example in Figure 1 is written in Visual Basic and illustrates the API interface key process. The variable “key” is session specific using a combination of user credentials. This key value is passed to the rate service as the primary method of authentication. Once connected to the rate service the live data feeds become available.

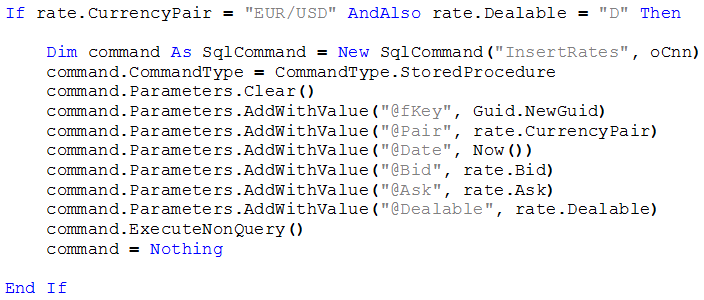


Fig. 2. Capturing FOREX data

The example in Figure 2, again written in Visual Basic, illustrates how the live data is captured and stored in a database. All available currency pairs are available, however, in the above example the EUR/USD is of interest and filtered accordingly. The “if” statement looks at only this pair and checks to ensure the pair is “Dealable”, i.e. currently tradeable. The transactional data is transmitted in simple tuple format with a single primary key value and various attributes such as the current bid and ask. The additional code presented in Figure 2 stores the tuples in a local SQL Server database using a stored procedure.

## Twitter Data Collection

There are several ways to access Twitter data and the streaming method was selected for the project. The Streaming API provides very low latency read only access to the entire global network. The API also relies on interface keys freely accessible to anyone with a Twitter account.

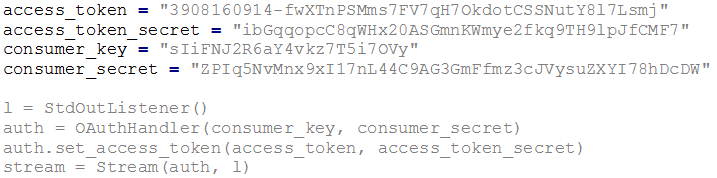


Fig. 3. Twitter API interface key

The example in Figure 3 is written in Python and illustrates the various keys required to connect to the streaming Twitter data. Once connected the data becomes available via various commands. Targeting data in an important aspect of unstructured data and the API allows for stream filtering. This significantly reduces the amount of unrelated data and allows the project to target Tweets related to “EUR/USD”. The filtered data is presented in Java Script Object Notation (JSON) format and while it is in readable format additional parsing of the data is required prior to locally storing in a database.

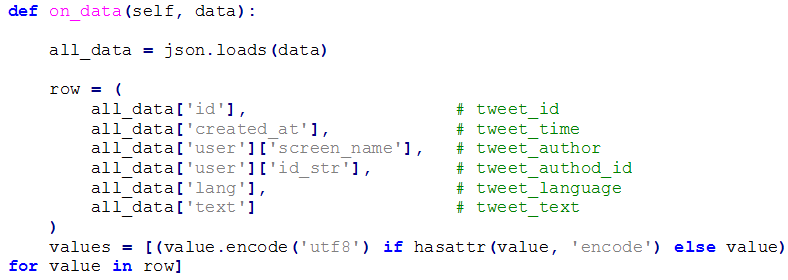


Fig. 4. Capturing Twitter Data

The example in Figure 4, again written in Python, illustrates the streaming and parsing of data. The JSON data is stored in the variable “all\_data” and various attributes/value pairs of interest are extracted such as screen name and Tweet text. The values are stored in a tuple and formatted accordingly for later storage in a local database.

# MongoDB

The data was loaded into MongoDB using the *mongoimport* tool. Data was stored in a database with two collections, a twitter collection for tweet data and a forex collection for forex data. From here the data was pulled into Python to apply the trade rules and determine a profit or loss (P/L).

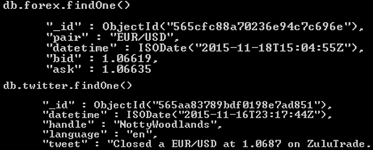


Fig. 5. FOREX and Twitter Schema

## Trading Period

The FOREX data was collected at the tick level. This means every time the price of the currency changed, the new price was stored in the database. These price changes can occur as much as multiple times a second. FOREX traders will normally select a particular time period to trade depending on their strategy. For instance, a short-term trader may decide to trade on the fifteen minute period. This kind of trader would most likely focus on the fifteen minute close of the pair they are trading. The close is the last recorded price (tick) that the currency pair traded for. In this project, it was decided to use an hourly closing price. In the FOREX markets, this trading period could be considered an intermediate-term where typically no decision to trade would be made until the end of the hour.

# Trading Rules

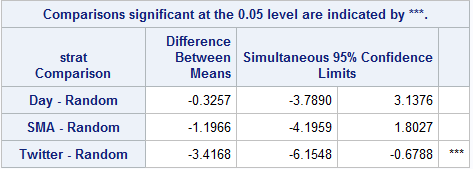
## Buy and Hold

The buy and hold strategy is one a lot of investors use in the stock market. It is commonly thought that if one holds on to an investment the price will eventually go up. While this may be true in many scenarios, the investment will eventually need to be closed out. For instance, if the investment was entered during a market peak and then sold twenty years later during a multi-year recession, money could actually be lost.

Given the time constraints of this project, it was impossible to hold on to a currency pair for an extremely long period. In FOREX, many traders would consider a day as a long time to hold a trade, therefore, for the buy and hold strategy, the EUR/USD was bought at the first hourly close of the day (00:59) and sold at the last hourly close (23:59). A total of four trades were made with a mean P/L of -$0.0012725.

## Moving Average

Using a moving average is a very simple and popular method used to determine whether prices appear to be going up or going down. Trade rules are usually made based on whether the current price of the currency pair is above or below the moving average. For instance, a ten period moving average would look at the average of the last ten closing prices. If the current price of the currency pair is above the moving average, it would suggest that the price is going up because it is literally above average. On the other hand, if the price is below the moving average, price would seem to be

down. Once the price is considered up or down compared to the moving average, it is expected that price will continue in that direction for some period of time.

For the moving average strategy, the EUR/USD was bought if the price had risen above the ten period moving average and sold once the price dipped below the ten period moving average. A total of nine trades were made with a mean P/L of -$0.00031

## Random Buy and Sell

Fig. 5. Box plot of the log P/L results

Fig. 6. Results of Dunnett’s test

In many people’s minds, investing in any market, and especially the FOREX market, is the same as gambling. The thought is that no matter what strategy the investor uses, they could get the same results randomly buying and selling.

This project uses a random buy and sell strategy as a control. A random day and hour were selected for when to buy the currency pair and then a random day and hour after the buy were selected to sell the currency pair. This was done using the “randrange” function from the “random” Python module. Given a range of numbers, a random number will be generated within that range. To generate a random day, the dataset’s range of days was used. To generate a random hour, 0-23 was used. A total of ten trades were made with a mean P/L of $0.001917.

## Trading Based on Tweet Volume

The thought behind this trading strategy is that if a currency pair receives a large number of tweets then a large price change would soon follow. The EUR/USD was bought whenever it was mentioned in over 99 tweets for a given hour. If a buy position had been entered, the position would then be sold at the close of any hour where the EUR/USD was mentioned less than 100 times. A total of nine trades were made with a mean P/L of -$0.0008467.

# Statistical Analysis

In order to determine if these strategies produced significantly different results and, which was the best, several statistical tests were applied to the data. All the tests used a log transformation of the P/L resulting from the trading rules. A standard alpha of 0.05 was used to determine statistical significance. After the tests were run at α = 0.05, levels of 0.10, 0.25, and 0.40 were also tested to see if more nuance (while likely not as significant) could be pulled out of the data.

## Analysis of Variance

The first test done was an analysis of variance (ANOVA) to which shows whether or not at least one of the strategies produced a different result from the others. A p-value of 0.0330 was found which is below our alpha of 0.05; therefore we reject the null hypothesis that each of these strategies produced the same results and instead accept the hypothesis that at least one of these strategies differs from the rest. The ANOVA alone does not tell us which strategy or strategies did not equal the others.

## Dunnett’s Test

In our experiment the Random Strategy was used as a control. A Dunnett’s test was performed on the data to compare each strategy to our control. Dunnet’s test simply carries out a T-test between each strategy and the Random method. Only one had a statistically significant difference from Random; that strategy was the one based on Twitter. The 95% confidence interval for the difference between the log of the means was -6.154 to -0.6788 (Twitter minus Random). This indicates that the random strategy will outperform Twitter the vast majority of the time.

## Scheffe’s Method

The final statistical test performed was Scheffe’s Method which compares each pairwise combination of strategies using a T-test. While we know only Twitter differed from the Random strategy, it is still possible that other pairs may also show a difference. Under the alpha of 0.05 no other pairs had a significant difference other than the Random and Twitter strategies.

## Tests With Increased Alpha

The same tests were also performed using alpha levels of 0.10, 0.25, and 0.40. These tests were done to see if more differences could be found in more tolerant conditions (that is to say a lower bar for rejecting the null hypothesis of no difference). The results of all tests were exactly the same as detailed above at alphas of 0.10 and 0.25; that is only the Random and Twitter results showed any significant difference. At the alpha level of 0.40 Twitter underperformed compared to every other strategy.

Given the results of these tests we can confidently conclude that Random will outperform twitter. It also seems more likely than not that the other strategies will also outperform Twitter though we cannot conclude that with as much confidence. The results also imply that over the time period examined the random strategy would produce the most profit but would differ little from the Moving Average and Day strategies would be very small.

References

1. H. Achrekar, A. Gandhe, R. Lazarus, Ssu-Hsin Yu and B. Liu, 'Predicting Flu Trends using Twitter data', *2011 IEEE Conference on Computer* USA: Abbrev. of Publisher, year, ch. *x*, sec. *x*, pp. *xxx–xxx.*
2. Opera Solutions, 'Predicting US Primary Elections with Twitter', NIPS 2012 Workshop, 2012.

1. This paper was submitted for grading on December 10th, 2015 for the ‘File Organization and Database Management’ course from Southern Methodist University.

   The authors have received no financial support in the research, conduction, or writing of this paper.

   All authors are current students in the Masters of Data Science program of Southern Methodist University, Dallas Texas 75205. [↑](#footnote-ref-1)