Brendan Saw (bsaw), Ramandeep Thind (rst6)

CMPT 353: Market Data Analysis

Do markets follow a random walk?

1. Introduction

In this project, we wish to utilize foreign exchange market (forex) information to come to conclusions about market volatility and determine whether there are patterns in market upturns or downturns. This information can be used alongside other data sources to analyze how global events can shape the market responses.

Note that this project will not analyze any individual stocks within the stock market since company-specific actions could dictate prices and affect stock volatility. Rather, we will be analyzing less-volatile commodities such as the entirety of the S&P 500 against the USD, or currency valuations against the USD. As this project focuses on market analyses, there will be some jargon that may be ambiguous to the reader - see Appendix for clarification on terms that may be unclear.

In this project, we wish to answer the question: **Do markets follow a random walk?**

2. Data Acquisition/Cleaning

We have 3 primary data sources: OANDA, Poloniex, and government websites. OANDA provides us data for common commodities (currencies, grain, gold, bond values). Poloniex provides information about cryptocurrency pairs. Government websites provide information about bond yields.

Data collected from OANDA was stored in Pickle formats in our Data Collection folder and contains data spanning from the start of 2010 to the end of 2020. Pickle files are named in a format similar to: EUR_USD_H4.pkl. This naming scheme represents a pair between the Euro and the US Dollar, collected every 4 hours (H4). Instead of H4 as the last symbol in the filename, there might be an H1 (collected every 1 hour), or D (collected once per day). These pairs are stored in the data/ folder.

Data from Poloniex was only collected once per day due to API restrictions and stored in the data/crypto/ folder, where they are named similar to USDT_BTC_D.pkl. Assume 1 USDT = 1 USD, which has been the case for the majority of USDT's existence. Data collected spans the 2016-2021 date range (5 years of data) as crypto is a relatively new asset as some cryptocurrencies did not exist before the year 2015. For cryptocurrency ticker disambiguation, see the Appendix.

We also collected bond yields mostly from government websites, but in the event where those weren't available (US 10Y bond yields), we collected them from MarketWatch. The UK government provided yields monthly, and the Canadian provided bond yields daily, so we had to group this data by month, taking the means of the bond yields.

3. Data Analysis

After collecting the data from OANDA and Poloniex, we need to come to some idea about whether markets follow a random walk. We will perform multiple tests to determine whether there is statistical significance that markets do/don't follow a random walk.

3.1. Statistical Analysis

There are various different statistical tests we performed to determine the randomness in the markets, and our expectations going into this was that if the market were **not random** then we would see some sort of significant patterns that occurred in the market.

All subsequent figures in this section use OANDA EUR/USD data from 2010 to 2020. The reason this is done is because EUR/USD is the most liquid and traded pair in the forex market. Furthermore, the US Dollar Index, which we were unable to get data for, is 57% composed of EUR/USD prices. The easiest and most straightforward way to detect patterns is looking at the effect of time in the marketplace. We started by determining which days of the week that were the first and second weekly extremes, finding that there was consensus across EUR/USD, GBP/USD, and SPX500/USD.

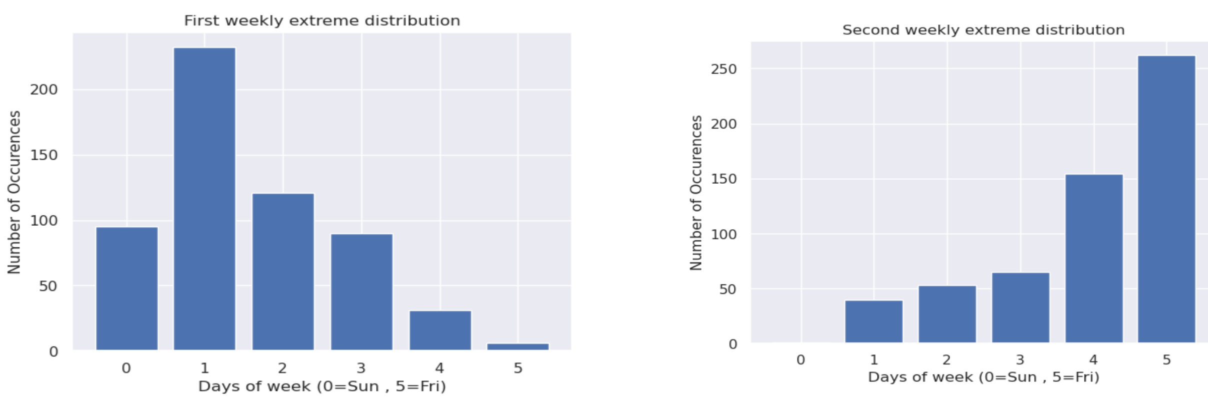
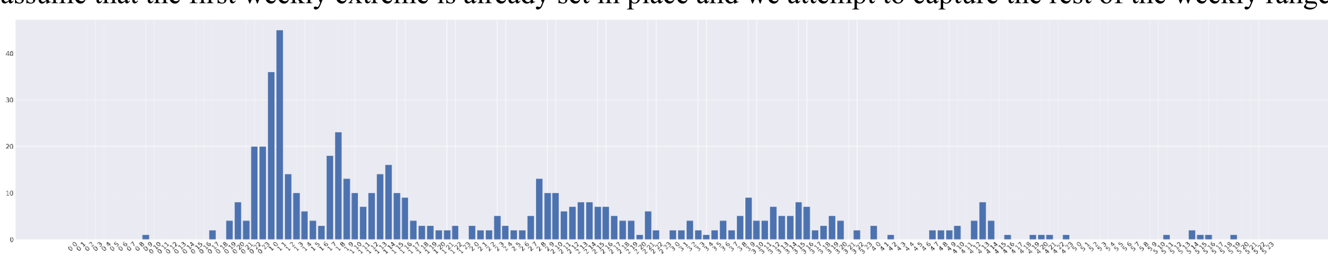


Figure 1: EUR/USD, first weekly extreme distribution (left) and second weekly extreme distribution (right)

Trading times of the global financial markets are open from Sunday 5pm - Friday 5pm in each market's respective time zones. Sunday and Monday very often created the first weekly extreme and Friday typically created the second weekly extreme for the weekly range. This information can prove useful in certain situations where we assume that the first weekly extreme is already set in place and we attempt to capture the rest of the weekly range.



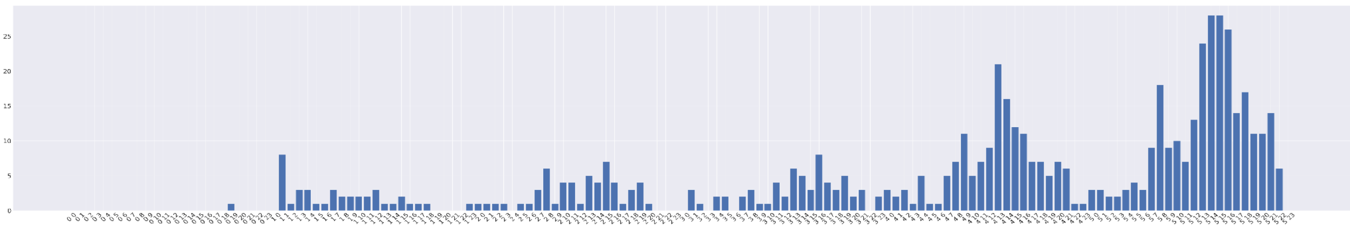
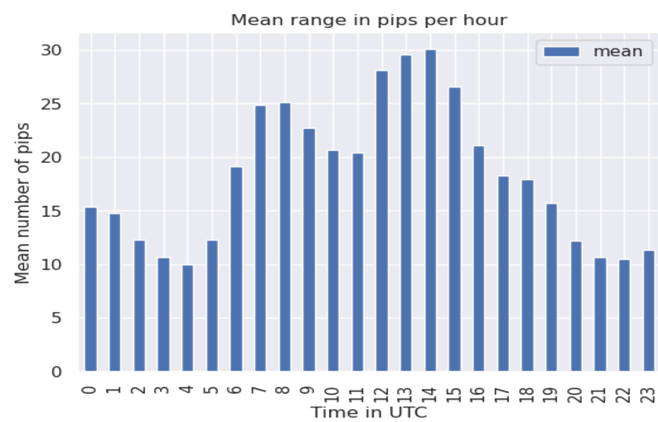


Figure 2: First hour extreme in the week (top, earlier page), second hour extreme in the week (bottom), UTC

The weekly range extremes can be examined in more detail by looking at the individual hours that create the week. The first weekly extreme occurs at Monday 0 UTC (top bar in top graph, Figure 2), with Monday 7 to 13 UTC trailing a close second place. For the second extreme of the week, Friday 12 to 16 UTC and Thursday 12 to 16 UTC tend to have the most occurrences to create the second extreme.

Figure 3 (right): Mean number of pips per hour, UTC



There are specific times of day that induce large amounts of liquidity into the marketplace causing more volatility. New York's trading session starts at 12 UTC (7am New York) until 20 UTC. London's trading session starts at 7 UTC until 15 UTC. In Figure 3, the lower bars at 10 and 11 UTC represent the slight drop in volatility during a time called 'London Lunch'. There is an overlap between these sessions from 12 to 15 UTC, which has a higher number of pips, representing larger price ranges to work within.

To prove that greater volatility and trading occurs only during the sessions noted above, we determined the 'active' trading time range to be from 5 - 16 UTC, and 'dead' time range to be all other times of the day. The mean for the 'active' range was 77 pips, and the mean for the 'dead' time range was 60 pips. A T-Test utilized outputted a p-value of 9.4×10^{-38}, showing that means were statistically different.

Revisiting the weekly extremes, the first weekly extreme hour is at 7 UTC (London), equivalent to 2am for New York. The second weekly extreme generally forms on Thursday and Friday 12 - 16 UTC, which corresponds to the first four hours of the NY session (7 am to 11am NY time). In addition to this, we can also look at the first and second daily extremes. There is a ~30% chance that the first daily extreme occurs and an ~8% chance that the second daily extreme occurs in the London session. There is a ~13% and ~35% chance that the first and second daily extremes respectively occur in the NY session. Combining both these sessions, we have a 43% chance that the first extreme occurs in one of the NY/London sessions and a chance the 2nd extreme occurs in these periods.

In conclusion, despite the NY session making up only 4 hours or ~16.7% of the daily 24 hour trading window, it has a **35% chance** of creating the 2nd daily extreme! And this ties directly into the NY session also being a strong candidate to create the second weekly extreme on Friday or Thursday. There seems to be strong evidence that foreign exchange market is **very sensitive to time**.

3.2. Correlation Analysis

We can get a localized conclusion of the relationship between two pairs by taking the correlation of a pair (specifically concluding about 2010-2020 for OANDA pairs and 2016-2021 for Poloniex crypto pairs). This is important to determine, since it allows us to understand whether pairs are related to one another - enabling us to come to better conclusions about why an asset might be increasing/decreasing in value.

3.2.1. Common Commodity Pairs

We decided to run the analysis on many of the assets we had, and created heatmaps of pair correlations. In these graphs, lighter values = most positively correlated in the graph and darker values = less positively correlated pairs. It is important to see the colour legend on the right to determine exactly what colour hues represent in correlation values. Examples of heatmaps that we examine are below:

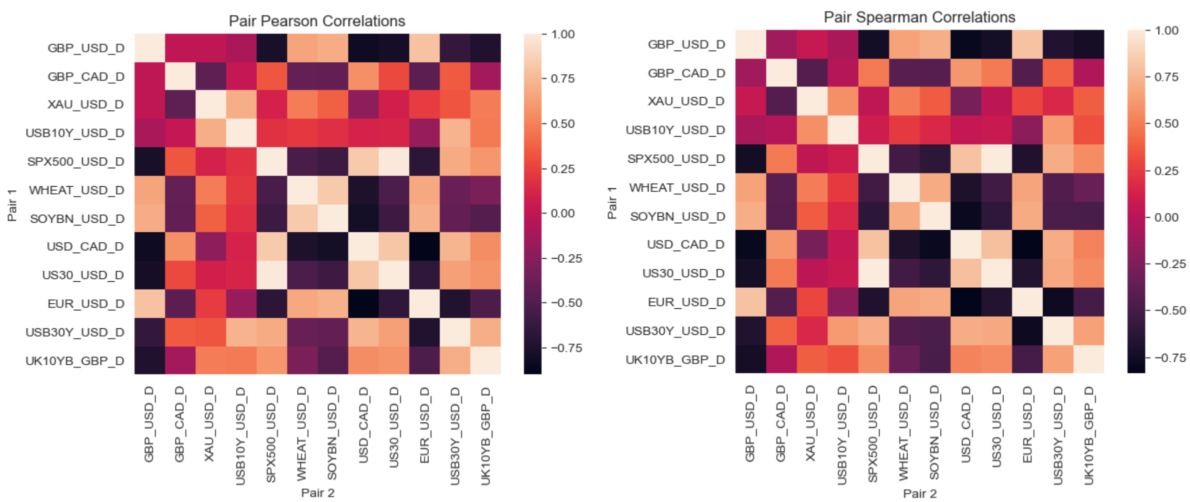


Figure 4: Heatmaps of Pearson (left) and Spearman (right) correlations of currency/commodity pairs

Above are two correlation heatmaps, using Pearson and Spearman correlations. The type of correlation did not affect the heatmap greatly. From this, it is safe to assume that the pairs' monotonic relationship (Spearman) is quite similar to their linear relationships (Pearson). These results are expected as it makes sense for a linear relationship to exist between variables since time is linear. As a result, we will be using Pearson linear correlations to evaluate similarities between pairs going forward.

In the Pearson correlation heat map (Figure 4, left), immediately apparent are negative correlations between USD/CAD and many pairs. This is most likely because the order of the USD-CAD pair has been reversed (normally USD is the second pair but in this case it is the first). We can also see significant positive correlations between US30/USD (Dow Jones), and SPX500/USD (S&P 500) (linear correlation of 0.99) which is a good indicator that these tests are providing relevant data for us to examine (both these pairs represent groups of the largest stocks in the United States). Pairs that are relatively highly positively correlated include: EUR/USD and GBP/USD (correlation of 0.79), SOYBN/USD and WHEAT/USD (corr: 0.83), USB30Y/USD and US10YB/USD (corr: 0.72). Let's take a look at some of these pairs in more detail:

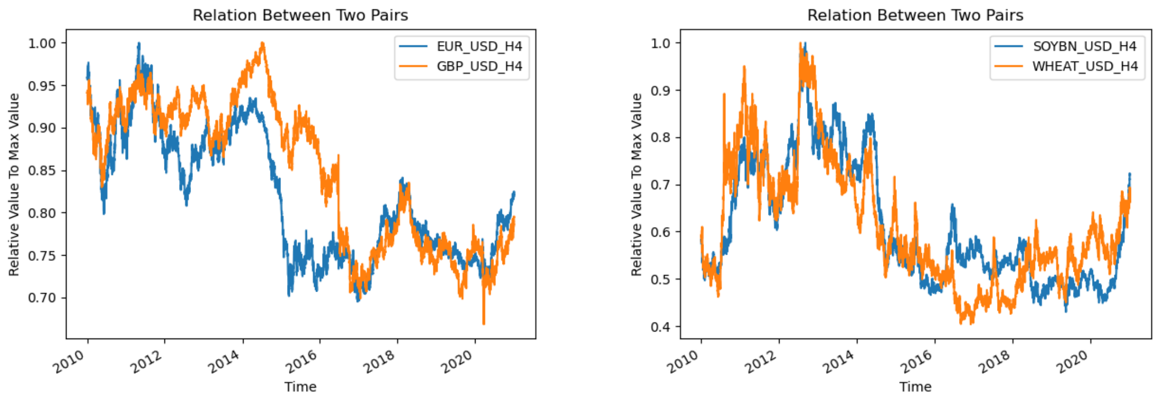


Figure 5: Relative value to Series maximums for EUR and GBP (left), soybeans and wheat (right).

Items like soybeans and wheat follow each other's prices since they both are food products, while EUR and GBP are very similar after 2016. Before 2016, GBP was trading higher at higher highs than EUR, but we can come to an interesting observation: the original Brexit vote occurred in mid-2016, which perhaps could explain why the GBP dropped so significantly during that year, and then stayed in-line with the Euro currency valuation changes afterwards.

From these results, we can see **established (long-term) assets similar to one-another will behave similarly**. Though what about newer assets like cryptocurrencies? We took a look at cryptocurrency pairs from Poloniex and found that there are high positive correlations between cryptocurrencies of similar market caps (ex: BTC and ETH), though as always, correlation != causation since there could be many factors affecting both currencies' gain/decline. Nevertheless, **many commodities regardless of volatility and age, appear to have correlations with other pairs**.

3.2.2. Bond Spreads

Bond spreads are the difference in the yield on two different bonds, and central banks can change interest rates which affect bond yields. We'll be examining how a pair from OANDA representing exchange rates of currencies (ex: CAD/USD) is related to the bond spread between the two countries. Below are the results that were obtained by taking the difference between the bond yields of two countries, and plotting it against the currency of those two countries.

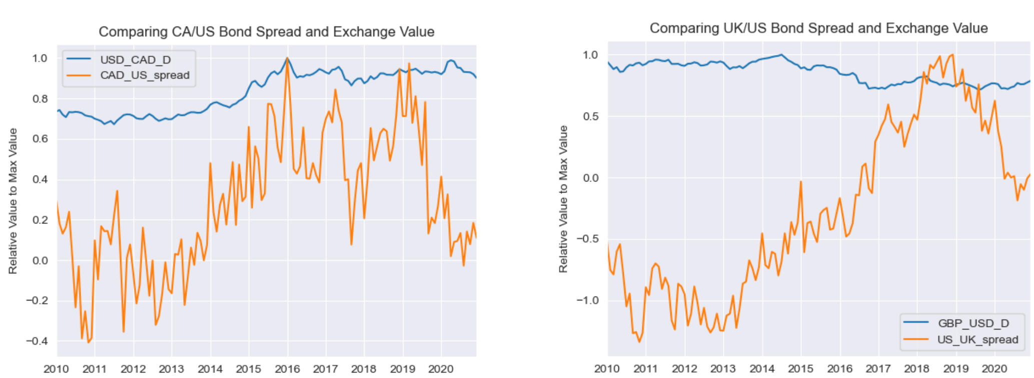
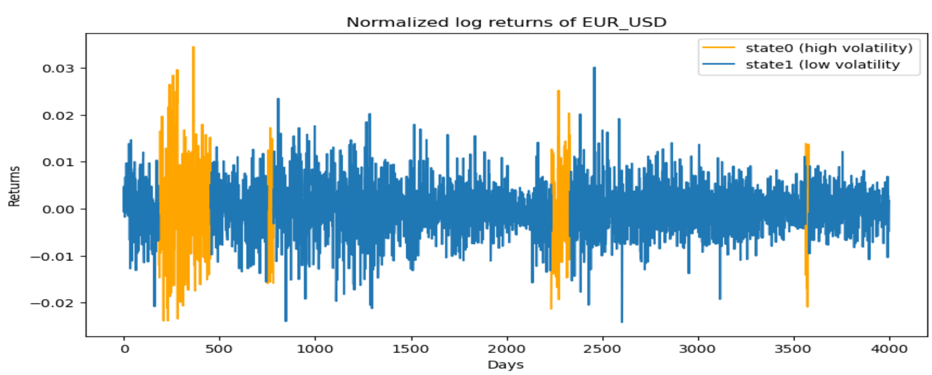


Figure 6: Bond Spread to Exchange Values, all values scaled down by max value in Series. CA/US bond spread vs exchange value (left), UK/US bond spread vs exchange value (right)

Correlations between Series in both instances resulted in statistically significant linear correlation results: CA/US (corr: 0.72, p-value: 3.5×10^{-34}), UK/US (corr: -0.82, p-value: 1.8×10^{-22}). The reason why the UK/US correlation was significantly negatively correlated was due to the backwards order of the pairs in the spread calculations. As a result of this, we can conclude that the dollars and bond yields are related to one another. **Since central banks dictate interest rates which affect bond yields, they could have a statistically significant impact on the value of its currency relative to another currency based on what the central bank sets the bond yield rates to be**.

3.3. Volatility Clustering

We decided to investigate the idea of volatility clustering after we determined time of day affects volatility above in our statistical analysis section. Using a Hidden Markov model (HMM), we decided to attempt to identify clusters of volatility in EUR/USD returns. The model was inspired and adapted from an Udemy course on Financial Engineering. The high volatility can be seen visually in periods of high variance. Since HMMs are stationary models and market data is non-stationary, the price data was normalized by taking the log of daily returns. The HMM initially switched a lot from the high volatility and low volatility state, but after optimizing the transition matrix to have a higher probability in staying in the current state vs switching to the other state we were able to see much more consistent states of high and low volatility. Since HMMs are a form of unsupervised learning it's our job to associate meaning behind the clusters it creates. It's clear in the figure below that the model is attempting to cluster the periods of high variance and low variance. Overall, it would be interesting to attempt to label periods of high and low variance, and then see how well the model could perform in correctly labelling these states.



4. Results

Overall, throughout the report, we have found multiple patterns in the data that we could use to answer our initial question.

While looking at the foreign exchange market, it is very clear that there are patterns related to time of day and day of week that influence how the market behaves. It's very likely that traditional trading sessions bring in large amounts of liquidity and volatility through the large entities that may be executing orders into the market and it's clearly seen in the data we collected. We can hypothesize that many funds may be looking to close open positions before the weekend to reduce risk since the market can open on Sundays with large gaps due to unforeseen events on the weekends - this closing of positions likely causes the second extreme of the week to form on Fridays and Thursdays as seen in our data. We were able to prove that intraday price movement is dictated by time; however, a further study could be analyzing how the daily price range forms in combination of price and time (we only examined time in this report). Another topic of further study could be studying seasonality in foreign exchange markets in terms of price movement and volatility.

Looking through the lens of the correlation analysis, we found that commodities that are similar to one-another will behave similarly, which allows us to understand that pairs might influence one-another. By looking at correlated pairs such as UK/USD and EUR/USD, when they diverge we know that they must return to the mean since they appear to be related based on our analysis. Additionally, since bond spreads had a statistical linear correlation with the currency valuation, we also know that central banks can indirectly have a hand in influencing how much the currency exchange value is worth by manually changing the interest rates.

The Hidden Markov Model was a useful tool in identifying the periods of high and low variance that are stemming from volatility clustering. The possibility of volatility clustering also helps to further show that there is a pattern in the way volatility occurs in the forex market. We can hypothesize that large quick volatile moves inject excitement, greed, fear and liquidity into the marketplace while long periods of slow consolidation allow for large entities to slowly build up positions. This duality between periods of low and high variance likely is a reflection of investor and trading sentiment.

Due to this, we can conclude that markets do not follow a true random walk, as we can use correlated pairs and other trends to predict prices of the asset.

5. Accomplishment Statements

Brendan

- Created scripts to acquire and clean data from Poloniex for cryptocurrency pairs and government websites for bond yields
- Performed correlation analysis on volatile and non-volatile pairs to get an idea for correlation within asset classes
- Analyzed bond spreads to determine how they are related to currency valuation

Raman

- Created scripts to acquire and clean data from OANDA for standard commodity and currency pairs
- Performed statistical analysis on Forex pairs to determine statistical temporal patterns regarding the foreign exchange market
- Created an Hidden Markov model to help identify periods of high and low volatility in the foreign exchange market

6. Resources

https://developer.oanda.com/rest-live-v20/development-guide/
https://docs.poloniex.com/
https://www.dmo.gov.uk/data/ExportReport?reportCode=D4H
https://www.bankofcanada.ca/rates/interest-rates/canadian-bonds/
https://www.marketwatch.com/investing/bond/tmubmusd10y/download-data?startDate=1/1/2010&endDate=12/31/2020&countryCode=bx
https://www.udemy.com/course/ai-finance/

7. Appendix

As this project involves analysis on financial data, here are some terms and their definitions that will commonly be used - to clear up any ambiguity when talking about outcomes and what we can infer from them:

- Random walk hypothesis: The idea that stock market prices follow a random and unpredictable path
- OANDA: A financial services company that provides trading services for forex
- Poloniex: A cryptocurrency exchange
- Pip: Smallest whole unit measurement of the difference between bid and ask spread
- Weekly range: The highest high and lowest low created during the week's trading
- Session:
- First weekly extreme: The low of the weekly range for bullish weeks and the high of the weekly range for bearish weeks
- Second weekly extreme The high of the weekly range for bullish weeks and the low of the weekly range for bearish weeks

Charts Download links:

- First Weekly extreme: https://drive.google.com/file/d/1ojI6gP4m3O3Omzvgy7Fl5Eujq6hQ2lDC/view?usp=sharing
- Second Weekly extreme: https://drive.google.com/file/d/1lhXgQFxVV4FGL9T87CrpJr523XHF8ikt/view?usp=sharing

OANDA Currencies/commodities:

EUR: Euro	GBP: Pound sterling	USD: US dollar	CAD: Canadian dollar
XAU: Gold	US30: Dow Jones	USB30Y: US 30Y Bonds	SPX500: S&P 500
SOYBN: Soybeans	UK10YB: UK 10Y Gilt	USB10Y: US 10Y Bonds	

Poloniex Cryptocurrencies:

BTC: Bitcoin	ETH: Etherium	LTC: Litecoin	DASH: Dash
XMR: Monero	XRP: Ripple	STR: Stellar (XLM)	USDT: Terra USD