Bitcoin: Statistical Analysis and Price Forecasting

• What are bitcoins?
• Introduction to the bitcoin exchange.
• Available Bitcoin Data Sets.
• Possible Methods for Analysis.
• Literature Review
What are Bitcoins?

• Bitcoins are the most popular form of cryptocurrency.
• Cryptocurrencies are a form of virtual currency, which allows two parties to directly exchange monetary units without a central payment system.
• In addition to bitcoins there are other forms of cryptocurrencies, such as litecoin.
What are Bitcoins?

- Bitcoins (litecoins, etc.) are created through the process called “mining”.
- Mining is a virtual process that allows a user with specific computing hardware and software to find solutions to certain complex mathematical problems.
- The user is then rewarded for these computations with bitcoins.

- Bitcoins can also be bought through online exchanges or received as payment for products or services.
Bitcoin/Litecoin Exchange

- Many different exchanges for trading: Bitstamp (USD), OKCoin (CYN), Cavirtex (CND).
- I currently use the Canadian exchange Cavirtex (https://www.cavirtex.com).

- It is a much smaller exchange, but can be used with any Canadian bank account.
- Can exchange CND for bitcoin or litecoin, trade bitcoin for litecoin, etc.
The exchange rate for bitcoin can be extremely volatile.

BitcoinWisdom (https://bitcoinwisdom.com) is a great resource for up to date information on all the major bitcoin exchanges.
Bitcoin Data Sets

- There is a large amount of bitcoin data available on Quandl ([https://www.quandl.com/collections/markets/bitcoin-data](https://www.quandl.com/collections/markets/bitcoin-data)).

![BTC Price in USD (24hr Average)](chart.png)
Bitcoin Data Sets

- The data includes bitcoin prices on various exchanges.
- There are also many other variables that could be related to bitcoin price.

- Some price forecasting models consider only the past prices of bitcoin.
- Including these variables in the model could increase forecast skill.
- It can be difficult to determine what variables to use and how much time lag to use for each variable.
- This leads to a complex predictor selection problem.
Possible Methods for Analysis

- A random forest model in R could be used to help with predictor selection problem.
- FDA regression could be used for time series data.
Literature Review


Statistical Analysis of the Exchange Rate of Bitcoin

- Provides a statistical analysis of the log-returns of the exchange rates of bitcoin vs. USD.

**Log-Return**: if $S(t-1)$ and $S(t)$ are two consecutive observations in a time series, the log-return is defined as:

$$r(t) = \ln \left( \frac{s(t)}{s(t-1)} \right) = \ln[s(t)] - \ln[s(t-1)]$$

It shows the relative changes in the variable and can be used to compare directly with other variables with different base values. (Also called continuously compounded return)
Statistical Analysis of the Exchange Rate of Bitcoin

• The data set used for this study is BTC Exchange Rate Bitstamp (BTC/USD) Sept. 13, 2011 – May 8, 2014. (Quandl).

• They fit many different distributions that are commonly used in finance to this data.

• The generalized hyperbolic distribution is shown to give the best fit.

\[
f(x) = \frac{(\gamma/\delta)^{1/2} \alpha^{1/2-\lambda}}{\sqrt{2\pi K_{\lambda}(\delta\gamma)}} \left[\delta^2 + (x - \mu)^2\right]^{1/2} K_{\lambda-1/2}\left(\alpha \sqrt{\delta^2 + (x - \mu)^2}\right)
\]

for \(-\infty < x < \infty, -\infty < \mu < \infty, -\infty < \lambda < \infty, \delta > 0, \alpha > 0\) and \(\beta > 0\), where

\[
\gamma = \sqrt{\alpha^2 - \beta^2}.
\]
Statistical Analysis of the Exchange Rate of Bitcoin

Fig 7. Empirical histogram and fitted pdf of the generalized hyperbolic distribution (left), Empirical pdf and fitted pdf of the generalized hyperbolic distribution plotted on log scale (right).
Statistical Analysis of the Exchange Rate of Bitcoin

Finally, we give predictions for the exchange rate of Bitcoin. Let $Y_i$ denote the exchange rate on the $i$th day counting from the 13th of September 2011. Then $X_i = \ln Y_i - \ln Y_{i-1}$ is the log-return on the $i$th day. We can write the exchange rate on the $n$th day (counting from the 13th of September 2011) as

$$Y_n - Y_0 = \exp \left( \sum_{i=1}^{n} X_i \right).$$

We suppose $Y_0$ is a deterministic variable taking the value 5.97, the value suggested by the data set. So,

$$Y_n = 5.97 + \exp \left( \sum_{i=1}^{n} X_i \right) = 5.97 + \exp (T)$$