Anomaly-based Arbitrage of Cryptocurrencies

Nikhil Joglekar

Statistical Learning and Data Mining

Final Project

Cryptocurrencies

- A Cryptocurrency is a “medium of exchange designed around securely exchanging information which is a process made possible by certain principles of cryptography”
- Various different Cryptocurrencies exist, the largest being ‘Bitcoin’ (BTC)
- Cryptocurrencies are traded on ‘exchanges’ for other cryptocurrencies or even real currencies

Data Collection and Cleanup

- Acquired order-book from BTC-E, a large cryptocurrency exchange
- Tick-level data for price in USD, # bids, # asks, volume
- Data for BTC, LTC, NMC, NVC, and various other currencies
- ~2 gig CSV file, compressed into SQLite DB for use with Python
- Calculated hourly and daily returns of the different cryptocurrencies

$$\ln\left(\frac{BTC_{t+1}}{BTC_t}\right) \approx \frac{BTC_{t+1} - BTC_t}{BTC_t}$$

- Normality of returns – calculated returns, as the log method ensures normal distribution

Preliminary Investigation

- Cryptocurrencies and BTC have highly correlated returns
  - $r_{ltc,btc} = 0.55$
  - $r_{nmc,btc} = 0.89$
  - $r_{nvc,btc} = 0.79$
  - $r_{ppc,btc} = 0.73$
- Other crypto-pairs have similarly high correlations
- Can we detect anomalies in returns and profit off of them?

[1] 7 Day Price graphs for the top 4 coins by market cap. Can we quantify the similarity?

Regression → Coin Prediction

- Assumption: BTC is a *driver* of smaller Cryptocurrency price trends
  - *Corollary*: Smaller Cryptocurrency returns are mean reverting about BTC
- Naïve Regression (using 2/3 of dataset for training)

$$LTC_{Returns} = \alpha + \beta_1(BTC_{Returns}) + \beta_2(\sigma^2) + \beta_3(LTC_{Market\ Cap})$$

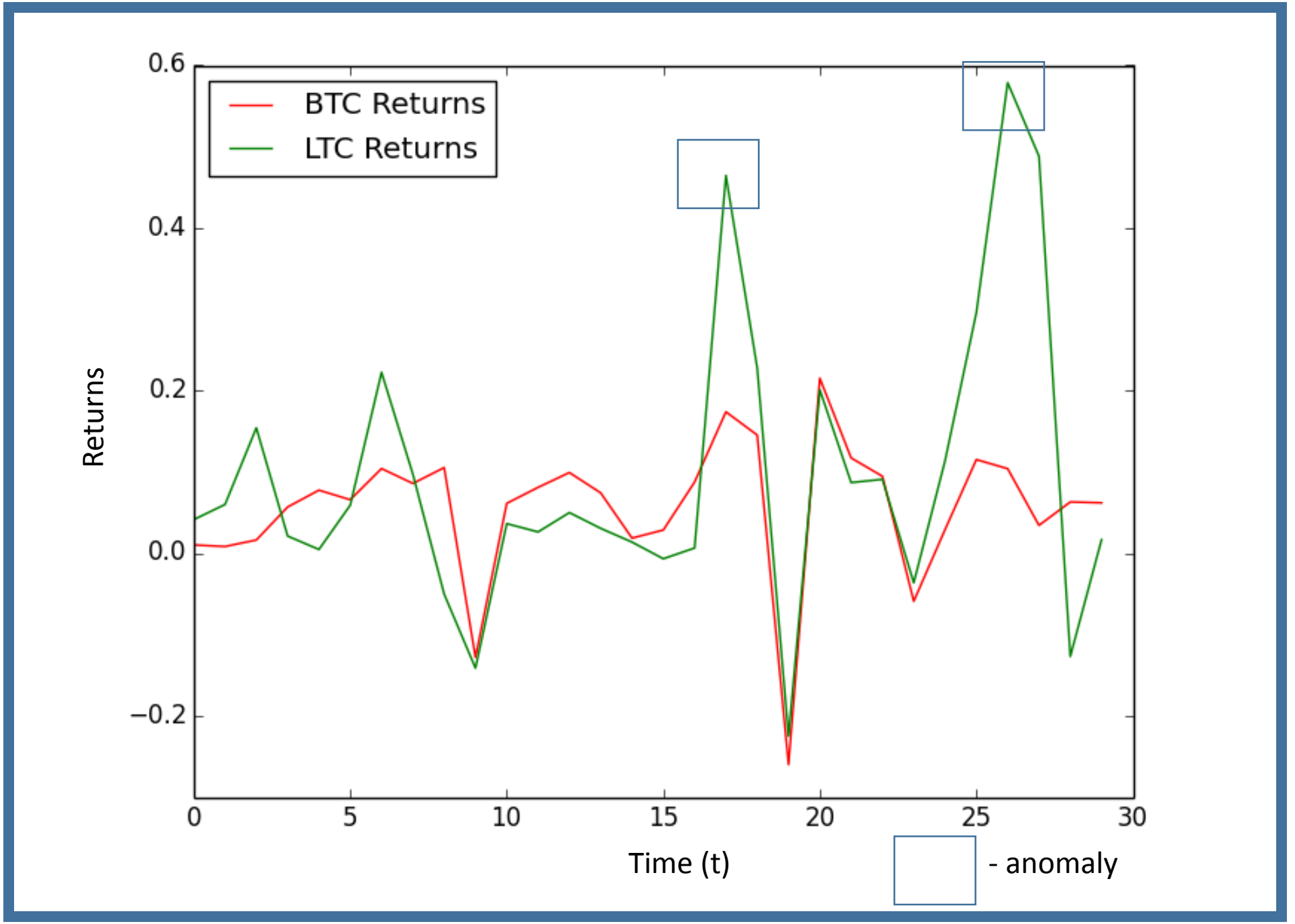
  - (exchange LTC with NMC, NVC, other small coins)
  - Coefficients chosen similar to (Fama French 92)[1]
    - $\beta_2(\sigma^2)$ is the risk premium for investing in a very risky (high variance) coin
      - Chosen over Beta from (FF92) as there is no "market index" of cryptocurrencies
    - $\beta_3(LTC_{Market\ Cap})$ factors for the size of the coin

Anomaly Detection

- Gaussian Anomaly Detection Algorithm
  - Calculate μ and σ^2 for *k* time steps prior to current time

$$\left|\frac{\bar{x} - \mu}{\sigma}\right| > \varepsilon \Rightarrow Anomaly$$

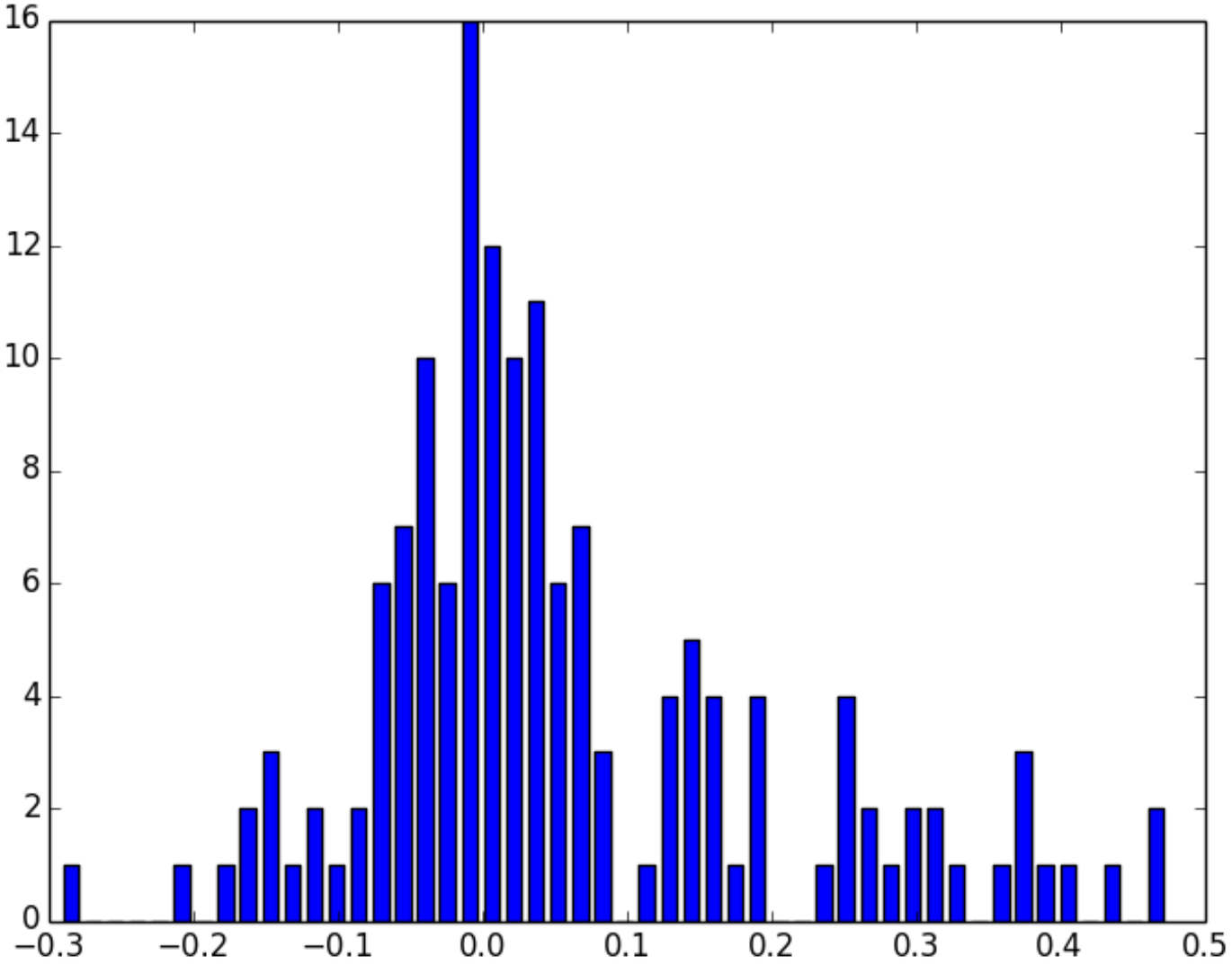
  - $\bar{x}$ = current mean returns, μ = model prediction of returns
  - If a certain point is ε standard deviations above the mean, it is an anomaly
  - Pros:
    - Easy to compute, fast
- K Nearest Neighbors (WIP)


BTC Returns
LTC Returns
Returns
Time (t)
0.6
0.4
0.2
0.0
−0.2
0
5
10
15
20
25
30
- anomaly

Anomaly Based Trading Model (WIP)

- When an anomaly is detected, classify whether the anomaly indicates an overvalued or undervalued price.
- If $\frac{\bar{x}-\mu}{\sigma} > \varepsilon$, investment is overvalued, long Cryptocoin
- At point $t + k$, where $t = current\ time, k > 1$, and $\left|\frac{\bar{x}-\mu}{\sigma}\right| < \varepsilon$, close investment

Anomaly Reversion (WIP)



Distribution of reversion returns of 7 time stpes for each anomaly