Bitcoin Trading Strategies

Jing Guo

Strats Associate, Goldman Sachs

PhD in Financial Engineering, Columbia University

March 25, 2018

Outline

1. Bitcoin Introduction
2. Trading Bitcoin and Online Time Series Prediction
3. Bayesian regression and Bitcoin

Bitcoin

- Bitcoin is a peer-to-peer cryptographic digital currency that was created in 2009 by an unknown person using the alias Satoshi Nakamoto
- Bitcoin is unregulated and hence comes with benefits (and potentially a lot of issues) such as transactions can be done in a frictionless manner - no fees - and anonymously.
- It can be purchased through exchanges or can be 'mined' by computing/solving complex mathematical/cryptographic puzzles.
- As of February 2015, over 100,000 merchants and vendors accepted bitcoin as payment. Research produced by the University of Cambridge estimates that in 2017, there were 2.9 to 5.8 million unique users using a cryptocurrency wallet, most of them using bitcoin.

Outline

Overview

- Muhammad J Amjad (n.d.) introduce a theoretical framework for predicting and trading ternary-state Bitcoin price changes, i.e. increase, decrease or no-change.
- They present simple algorithms that achieve a high return on average Bitcoin investment, while consistently maintaining a high prediction accuracy $(60 - 70\%)$ and respectable Sharpe Ratio (> 2.0).
- They train their model on a different time frame, and achieve robust algorithm performance. They also provide a justification for why it makes sense to use classification algorithms.

Bitcoin Data

- They Bitcoin data, accessed via the OKCoin exchange using their APIs okc. All prices are reported in Chinese Yuans.
- The APIs return lists of **bid**$[t]$ and **ask**$[t]$ prices at the exchange.
- They cached several months of data from the exchange in year 2014, 2015 and 2016.
- An estimate of price as $p[t] = (\max(\textbf{bid}[t]) + \min(\textbf{ask}[t]))/2$, mid-price of the best bid and ask.
- Pricing data are sampled fairly frequently (5-10s intervals) at a low latency.

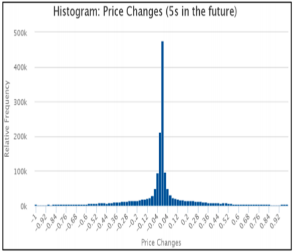
Trading Strategy

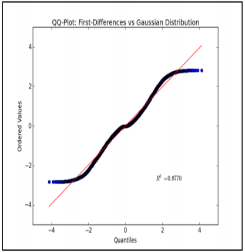
- Muhammad J Amjad (n.d.) limit the number of Bitcoins that can simultaneously be possessed to *one*.
- Formally, at time t, let $h[t] = 1$, if we are in possession of a Bitcoin and $h[t] = 0$, otherwise.
- Given $h[t]$ and a prediction of whether the price will go up, down, or stay about the same, the decision $d[t]$ at time t is given by:

$$d[t] = \begin{cases} \text{buy, if } h[t] = 0 \text{ \& price predicted up, with high confidence} \\ \text{sell, if } h[t] = 1 \text{ \& price predicted down, with high confidence} \\ \text{hold, otherwise.} \end{cases}$$

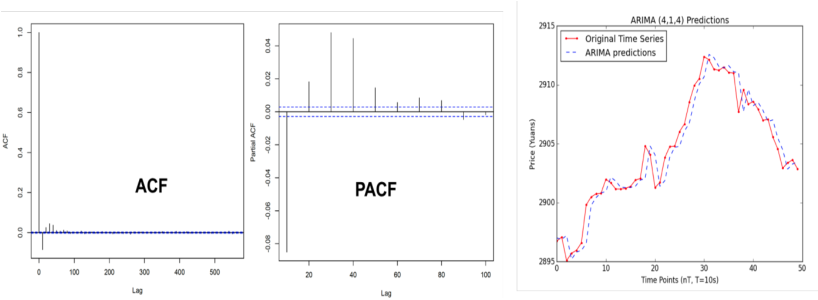
Stationary Analysis

- Let $p[t]$ be the price of Bitcoin at time t.
- Both ADF test and KPSS test reveal that the Bitcoin price time series is not stationary (p-value > 0.3 for ADF; p-value < 0.01 for KPSS).
- However, the time series produced by the first-differences, i.e., $y[t] = p[t] - p[t-1]$, is stationary with high confidence level (p-value < 0.01 for ADF; p-value > 0.1 for KPSS).





ARIMA Analysis



- ADF abd KPSS tests have already revealed that $d > 0$ fir stationary.
- Using the ACF and PACF plots in the previous plot, it can be safely assumed that $p, q \leq 8$.
- The lowest AIC/BIC values for the best fit were given by $(p, d, q) = \{4, 1, 4\}$ and $(p, d, q) = \{2, 1, 1\}$.

Further Model

- Given $\theta \in \mathbb{R}$, define

$$x[t] = \begin{cases} -1, & \text{if } y[t] < -\theta \\ 1, & \text{if } y[t] > \theta \\ 0, & \text{otherwise.} \end{cases}$$

- The task is to use price information $[t-d:t-1]$ to predict $x[t]$.
- The following features are extract to predict $x[t]$:
  - $x[t-1]$: the latest change in price.
  - $T_{\sigma,d}[t-1] = \sum_{s=t-d}^{t-1} x[s] = \sigma$: the tally-count of each quantized value in $x[t-d:t-1]$, for every d and $\sigma \in \{-1,0,1\}$.
  - $C_\sigma[t-1] = \max_k(k : \sigma = x[t-1] = x[t-1] = ... = x[t-k])$: consecutive run-length of each quantized value $\sigma \in \{-1,0,1\}$.

Empirical Trading Strategy

- Classification Algorithms Used
  - Random Forest (RF)
  - Logistic Regression (LR)
  - Linear Discriminant Analysis (LDA)
- Denote $h_d[t-1] := x[t-d : t-1]$. Trading decisions are determined:

$$\hat{x}[t] = \begin{cases} \sigma^*, & \text{if } \mathbb{P}^*(x[t] = \sigma^* | h_d[t-1]) \geq \gamma, \\ 0, & \text{otherwise.} \end{cases}$$

  where $\sigma^* = \text{argmax}_{\sigma \in \{-1,0,1\}} \mathbb{P}(x[t] = \sigma | h_d[t-1])$, and γ is the confidence threshold determined by validation.
- Combining multiple predictions

$$\hat{x}_w[t] = \sum_{d \in \mathbb{S}} \hat{x}_d[t] \times w_d, \quad \text{where } \sum_{d \in \mathbb{S}} w_d = 1.$$

Cumulative P & L

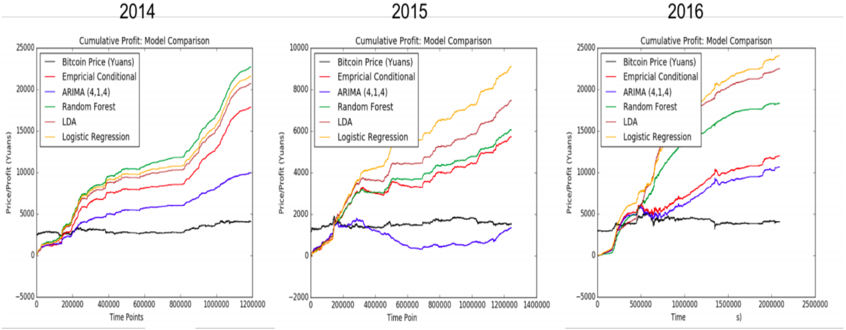


Figure: Cumulative Profit and Bitcoin Price in 2014, 2015, 2016 with $d \in \mathbb{S} = \{3, 4, 5\}$. γ selected via validation. Each time step represents 5s.

- Left: Training:2/16/14-3/14/14, Validation:3/15/14-3/31/14, Test:4/1/14-6/11/14.
  Center: Training:12/1/14-12/31/14, Validation:1/1/15-1/15/15, Test:1/16/15-3/31/15.
  Right: Training:2/26/16-4/15/16, Validation:4/16/16-5/15/16, Test:5/16/16-9/15/16.

Comparison of Algorithms

	Profit	Return	Sharpe	Accuracy
ARIMA	1366.7	0.9	0.43	0.49
EC	5721.1	3.7	2.17	0.64
RF	6049.9	3.9	2.56	0.78
LDA	7469.0	4.8	2.82	0.73
LR	9090.9	5.9	3.32	0.70

Table: Using a recent training period. Training: 12/1/14 - 12/31/14, Validation: 1/1/15 - 1/15/15, Test: 1/16/15 - 3/31/15. $d \in \mathbb{S} = \{3, 4, 5\}$. γ selected via validation.

Comparison of Algorithms (Cont'd)

	Profit	Return	Sharpe	Accuracy
ARIMA	-4758.5	-3.1	-2.0	0.43
EC	5721.4	3.7	2.2	0.64
RF	7903.1	5.1	3.4	0.69
LR	8019.4	5.2	3.5	0.71
LDA	7934.1	5.1	3.3	0.69

Table: Using an eight month old Training period, same validation and test periods as those in previous Table. Training: 2/16/14 - 3/14/14. $d \in \mathbb{S} = \{3, 4, 5\}$. γ selected via validation.

Outline

Overview

- Devavrat Shah (n.d.) discuss the method of Bayesian regression and its efficacy for predicting price variation of Bitcoin.
- They devise a simple strategy for trading Bitcoin, which is able to nearly double the investment in less than 60 day period when run against real data trace.
- They also study short-term price variation pattern using clustering method.

Data

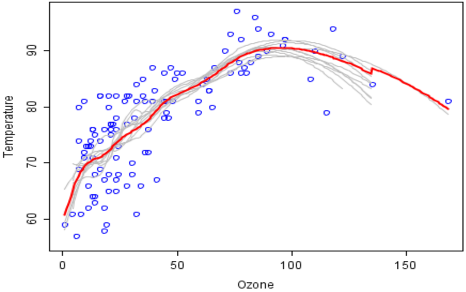
- Again, the authors used data related to price and order book obtained from Okcoin.com, one of the largest exchanges operating in China.
- The data concerns time period between February 2014 to July 2014. The total raw data points were over 200 million.
- The order book data consists of 60 best prices at which one is willing to buy or sell at a given point of time. The data points were acquired at the interval of every two seconds.
- They constructed a new time series with time interval of length 10 seconds; each of the raw data point was mapped to the closest (future) 10 second point.

Trading Strategy

- The trading strategy is very simple: at each time, they either maintain position of $+1$ Bitcoin, 0 Bitcoin or -1 Bitcoin.
- At each time instance, they predict the average price movement over the 10 seconds interval, denoted as Δp_t.
- Similar as in the previous section, a threshold θ is determined, such that
  - If $\Delta p_t > \theta$, buy a bitcoin if current bitcoin position is ≤ 0.
  - If $\Delta p_t < -\theta$, sell a bitcoin if current position is ≥ 0.
  - else, do nothing.

Non-parametric Regression

- Nonparametric regression is usually applied when non-linear relationship is observed.



- It in general requires larger sample sizes.
- Common nonparametric regression methods include Gaussian process regression and Kernel regression.

Kernel Regression

- The idea of kernel regression is straightforward: given a new x, predict $\hat{y}$ as the weighted sum of historical y_i, i.e.

$$\hat{y} = f(x) = \sum_i w_i y_i / \sum_i w_i$$

  where w_i depends on the distance between x and x_i. The closer x and x_i, the larger w_i.

- Devavrat Shah (n.d.) uses the following prediction equation:

$$\hat{y} = \sum_{i=1}^{n} \frac{\exp(-\frac{1}{4}||x - x_i||_2^2)}{\sum_{i=1}^{n} \exp(-\frac{1}{4}||x - x_i||_2^2)} y_i. \tag{1}$$

Predicting Price Change

- Recall that the time series of price variation of Bitcoin is over the interval of few months, measured in every 10 second interval.
- In order to apply equation (1), the authors generate three subsets of time-series data of three different lengths: previous 30 minutes, previous 60 minutes, and previous 120 minutes. and estimate ΔP^j as in equation (1), $j = 1, 2, 3$.
- Also construct feature $r_t = (\textbf{bid}[t] - \textbf{ask}[t])/(\textbf{bid}[t] + \textbf{ask}[t])$, where $\textbf{bid}[t]$ and $\textbf{ask}[t]$ are the *best* bid and ask at time t.
- The final estimation Δp is produced as

$$\Delta p = \beta_0 + \sum_{j=1}^{3} \beta_j \Delta P^j + \beta_4 r.$$

Predicting Price Change (Cont'd)

- The entire time duration is divided into three periods. The first time period is utilized to find patterns. The second period is used to learn parameters $\beta = \{\beta_0, ..., \beta_4\}$ and the last third period is used to evaluate the performance of the algorithm.

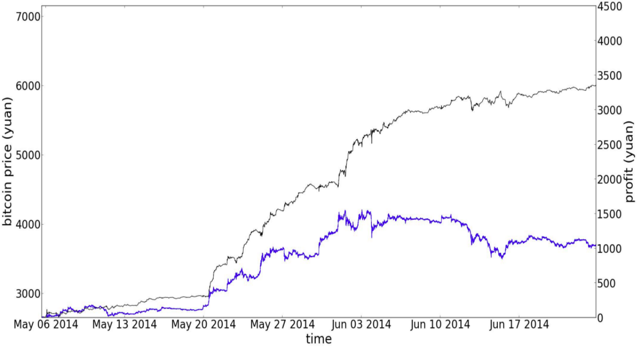
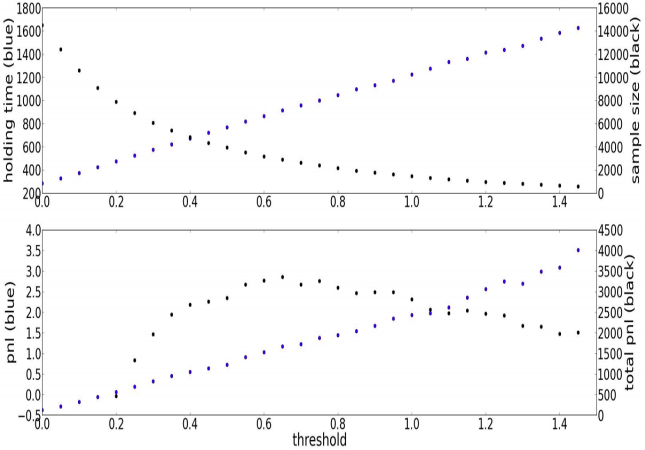


Figure: The cumulative profit of the strategy starting May 6, 2014 versus the price of Bitcoin.

Predicting Price Change (Cont'd)



- Figure: The effect of different threshold on the number of trades, average holding time and profit.

Predicting Price Change (Cont'd)

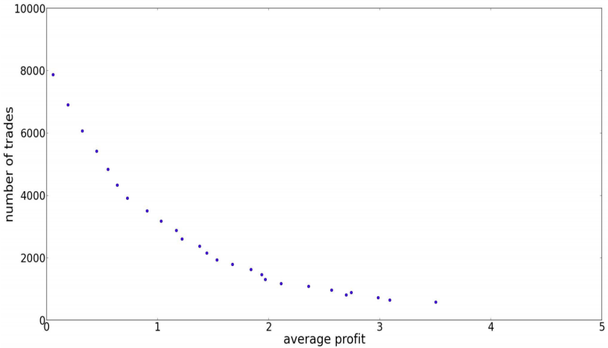
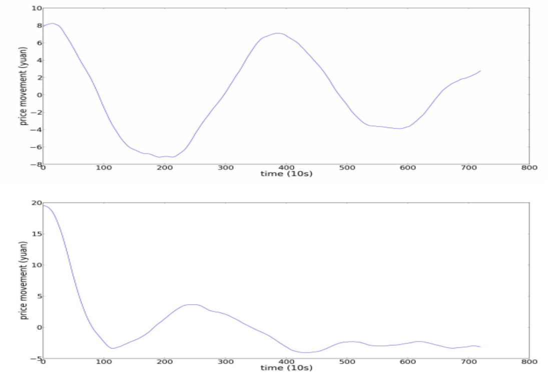


Figure: The inverse relationship between the average profit per trade and the number of trades.

Bitcoin Price Pattern

- The patterns utilized in prediction were clustered using the standard $k-$means algorithm. The cluster centers (means as defined by $k-$means algorithm) found are reported in the following:



References I

Devavrat Shah, K. Z. (n.d.). Bayesian regression and bitcoin, *52* .

Muhammad J Amjad, D. S. (n.d.). Trading bitcoins and online time series prediction.