Understand Bitcoin: Mining Software

December 8, 2015

Charlie Hume

Dustin Gay

Roma Koulikov

Background

- Bitcoin Advantages
  - No control by central monetary authority
  - Can be mined by anybody with computing resources
  - Completely digital - perfect for e-commerce, remittances, micro-payments

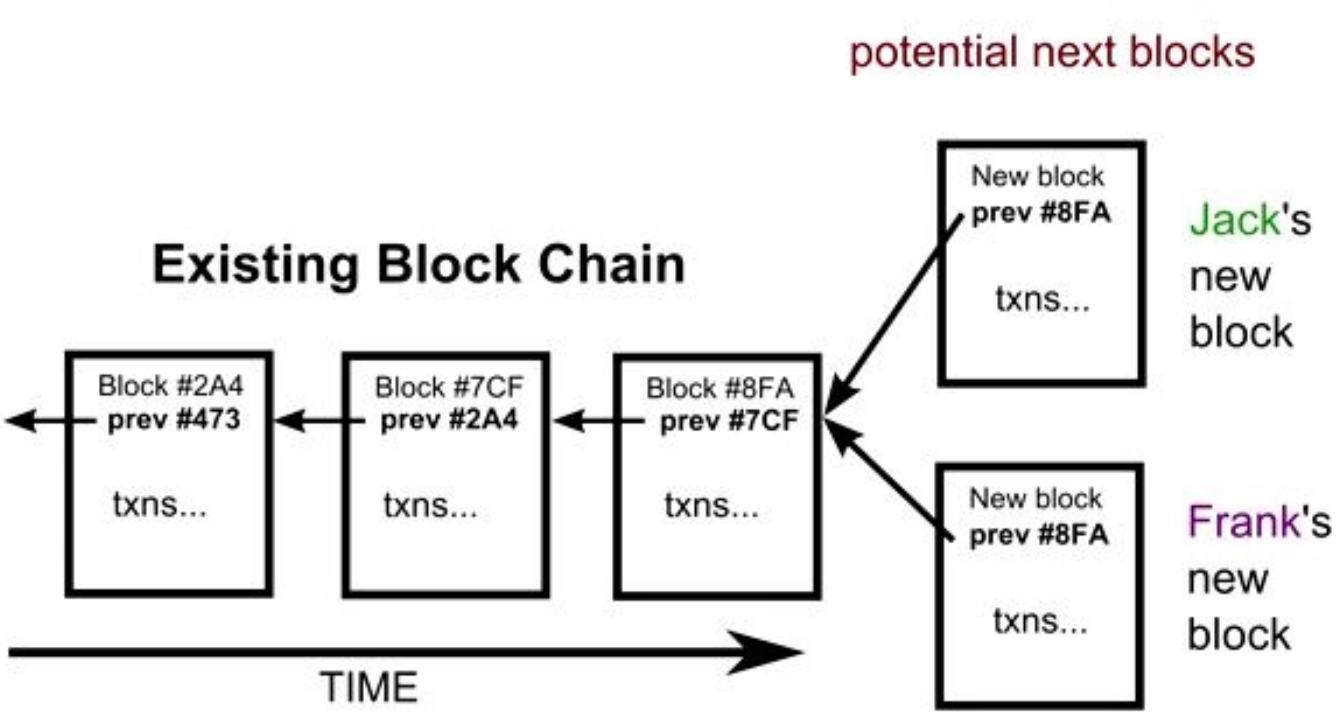
Research Question

Understand effect of BTC price changes on the types of commits for the most popular miners

Most Popular Mining Software

Miner	GitHub Commits
CGminer	7536
BFGminer	12632
BTCminer	NA
Bitminter	NA
Diablo miner	262
Poclbm	232

Bitcoin Mechanics

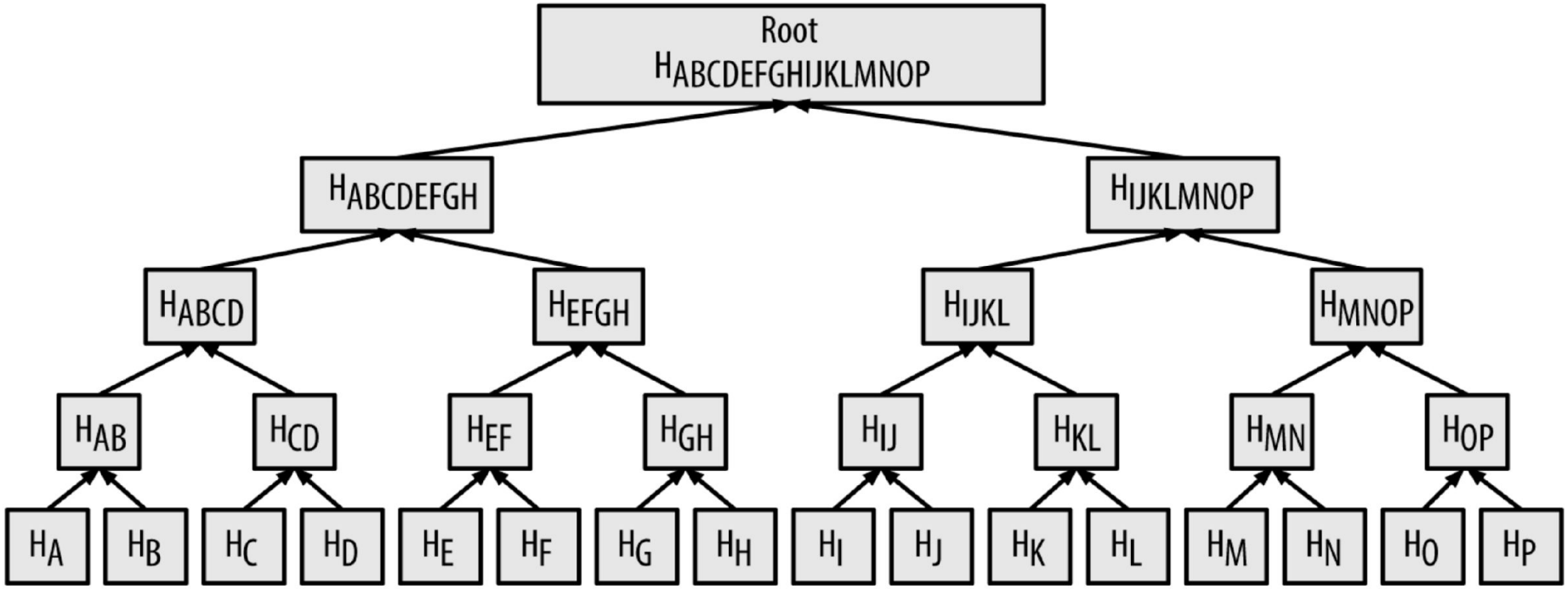


Mining Software Creates Blocks

Field	Purpose	Size (Bytes)
Version	Block version number	4
hashPrevBlock	256-bit hash of the previous block header	32
hashMerkleRoot	256-bit hash based on all of the transactions in the block	32
Time	Current timestamp as seconds since 1970-01-01T00:00 UTC	4
Bits	Current target in compact format	4
Nonce	32-bit number (starts at 0)	4

- SHA-256
- All fields are constant widths
- The Merkle Root is a single value that represents all transactions in the block

Merkle Root



^ Coinbase
- Your unique bitcoin address
- Block subsidy
- Extra nonce

Proof of Work

Field	Size (Bytes)
Version	4
hashPrevBlock	32
hashMerkleRoot	32
Time	4
Bits	4
Nonce	4

- Concatenate all fields as hex values into one string
- The hash of the string in hex must have at least ‘Bits’ 0’s on the end (little-endian)

- The nonce is incremented until this is achieved
- The extra nonce recalculates the Merkle Root

Inside Mining Software

- The main task is to prepare a block and send this block's unique proof-of-work problem to a device that solves and returns the answer
  - CPU
  - GPU
  - FPGA
  - ASIC
- The second most important aspect of the code is extensive driver and device compatibility

Inside Mining Software

- Four main categories of features
  - Diagnostic output
  - Documentation/updates
  - Drivers / Compatibility
  - Remote Access / Control

Goal: Determine Effect of Price on Commit Count

- Obtain all commits from November 2013 - November 2015 for CG Miner and BFG Miner
- Categorize commits
- Build linear regression models

Data Acquisition

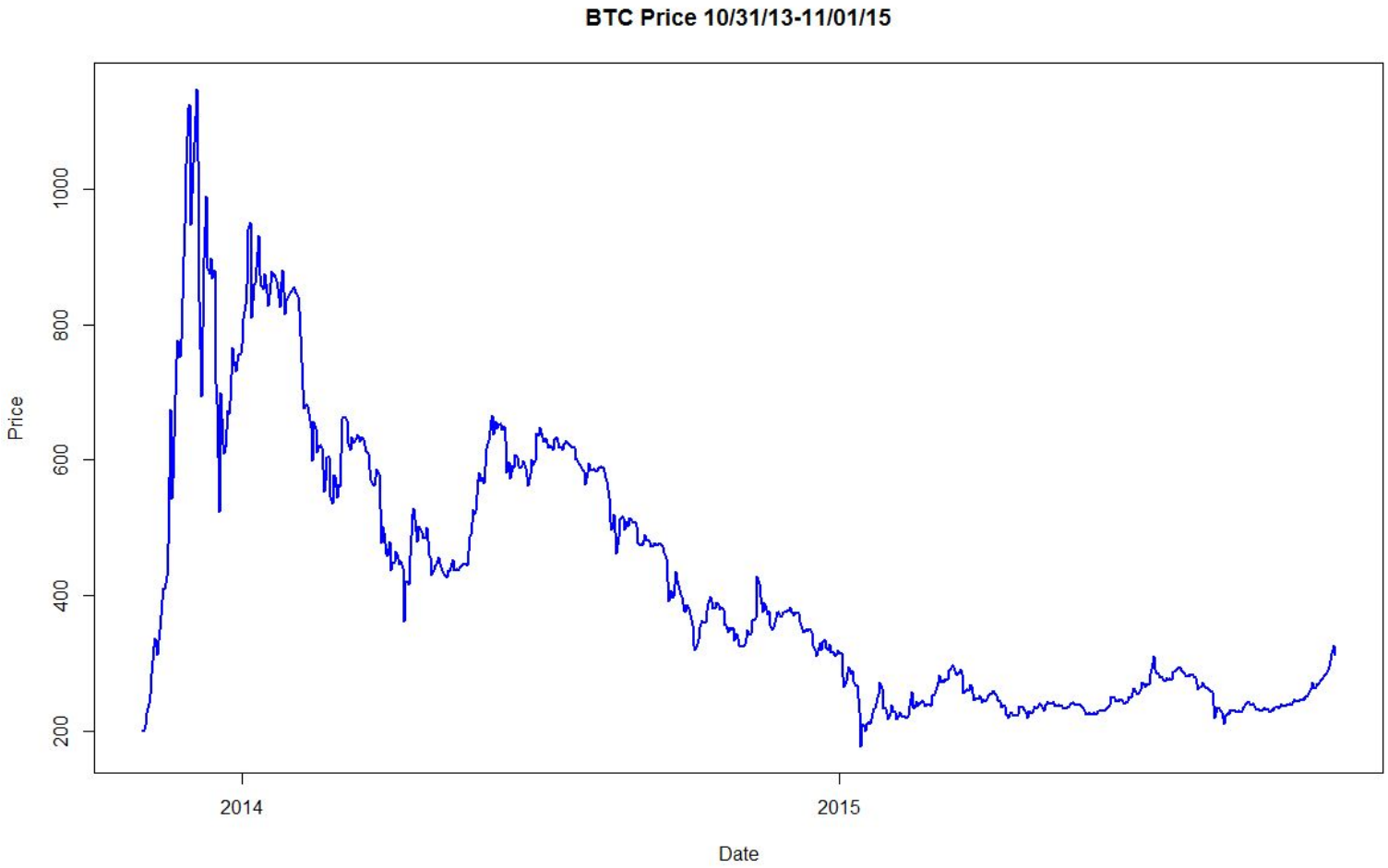
Github API

Branches

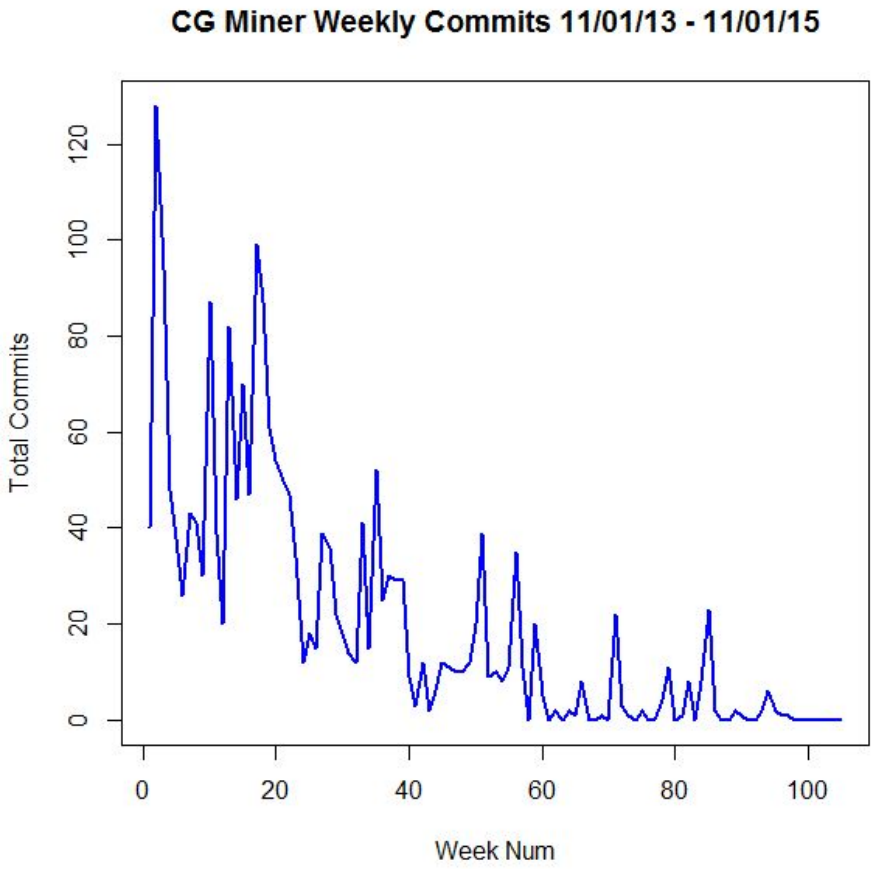
API Request Limit

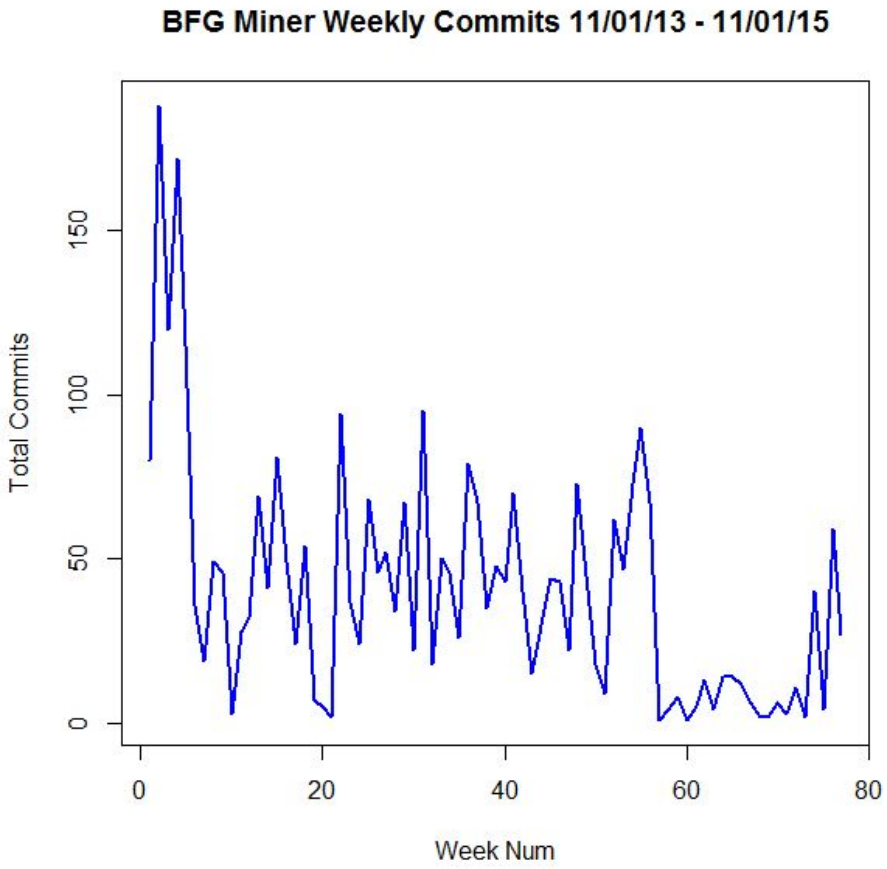
Stored in a mongodb

BTC Price


BTC Price 10/31/13-11/01/15
Price
200
400
600
800
1000
2014
2015
Date

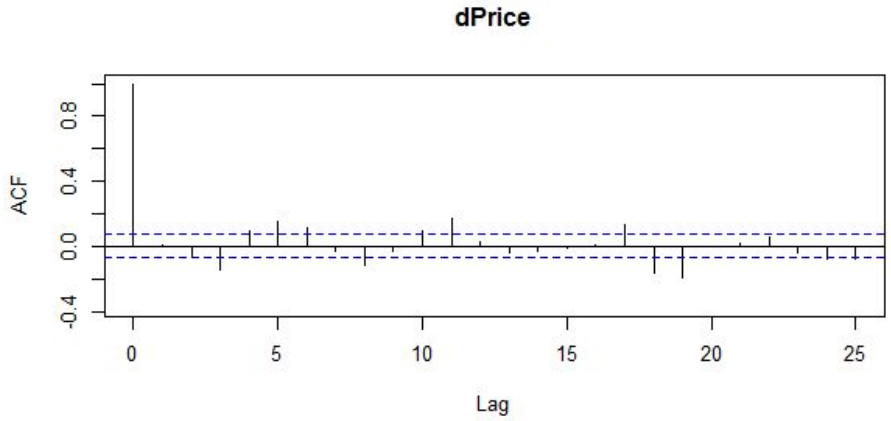

bitcoin

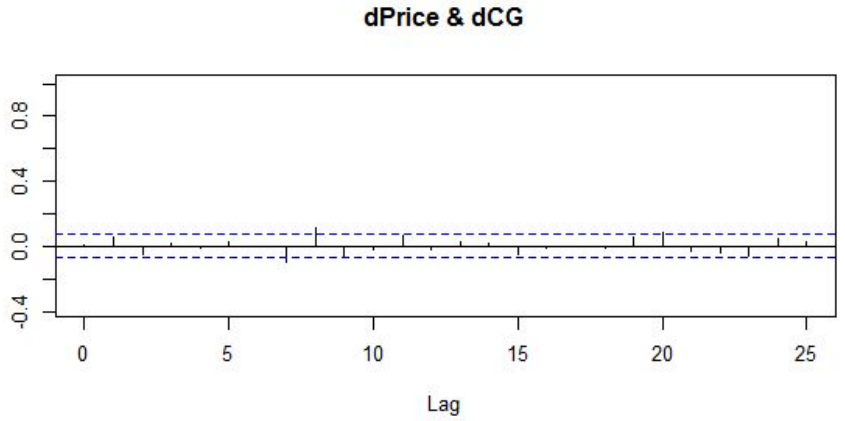

CG Miner Weekly Commits 11/01/13 - 11/01/15
Total Commits
0
20
40
60
80
100
120
0
20
40
60
80
100
Week Num

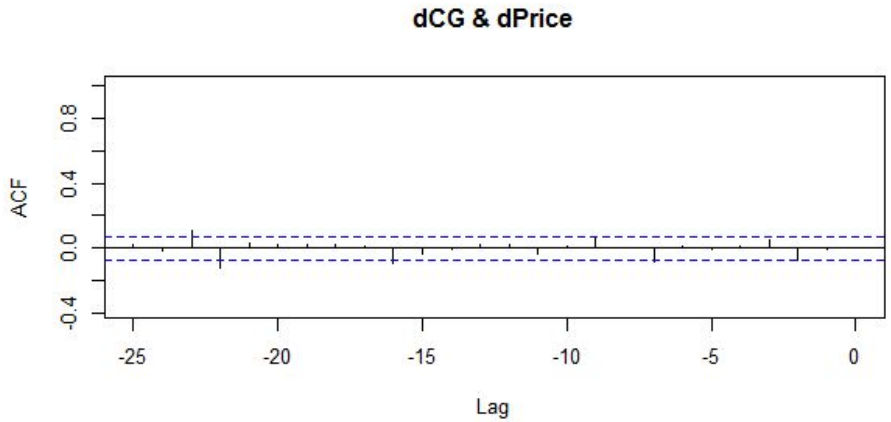

BFG Miner Weekly Commits 11/01/13 - 11/01/15
Total Commits
0
50
100
150
0
20
40
60
80
Week Num

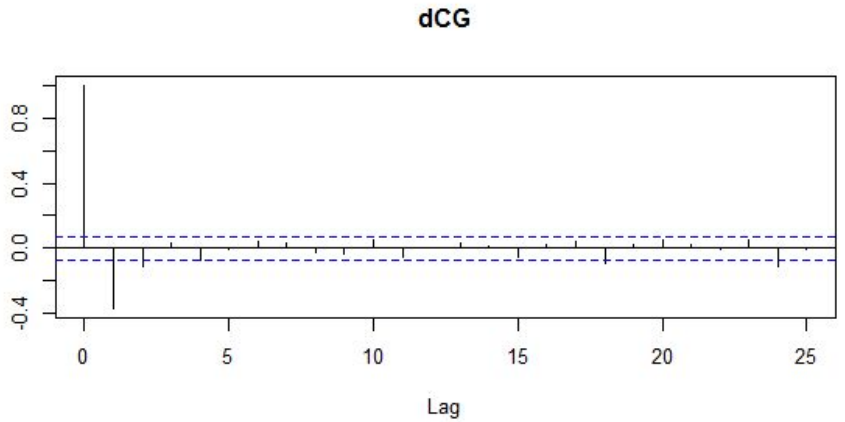

bitcoin

Time Series Analysis
Differences

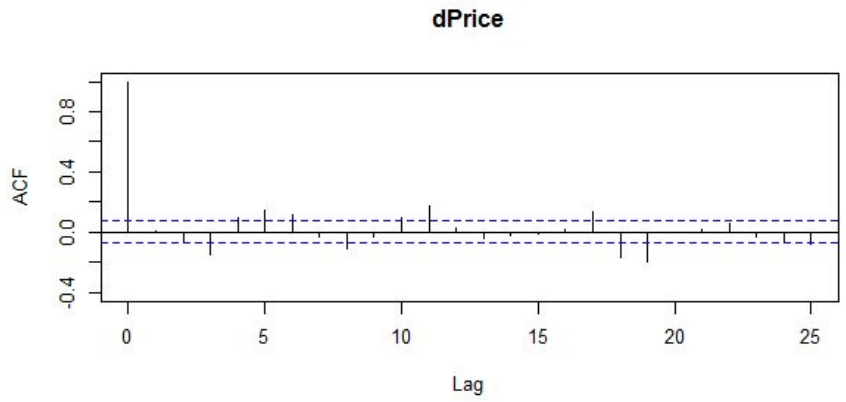


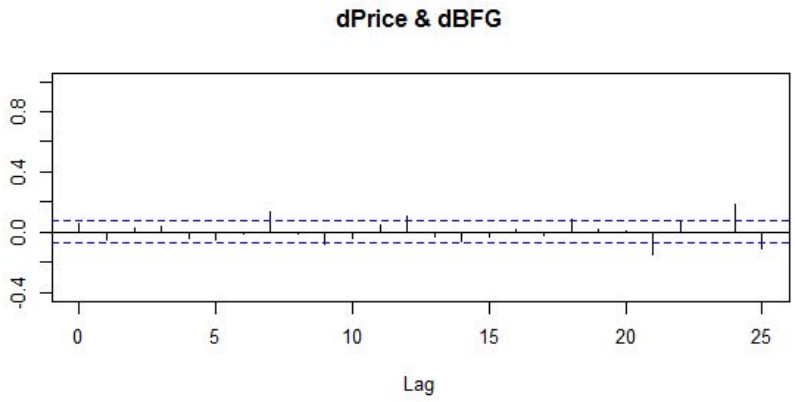


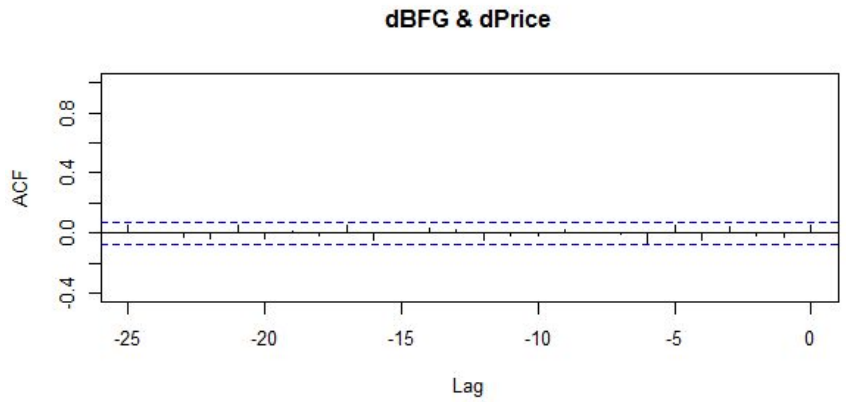


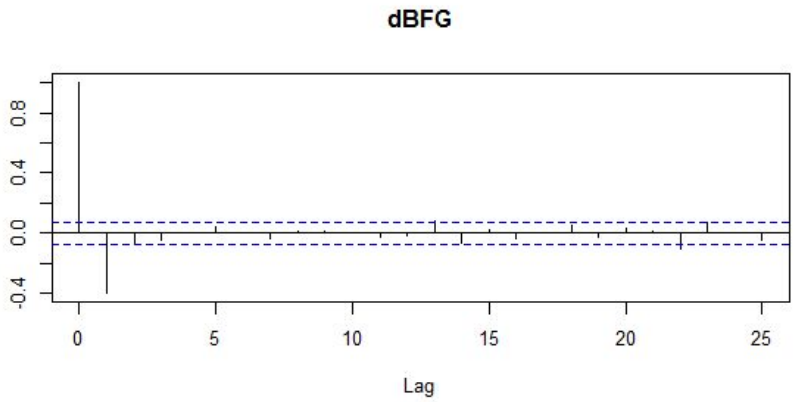


Differences Analysis - BFG Miner





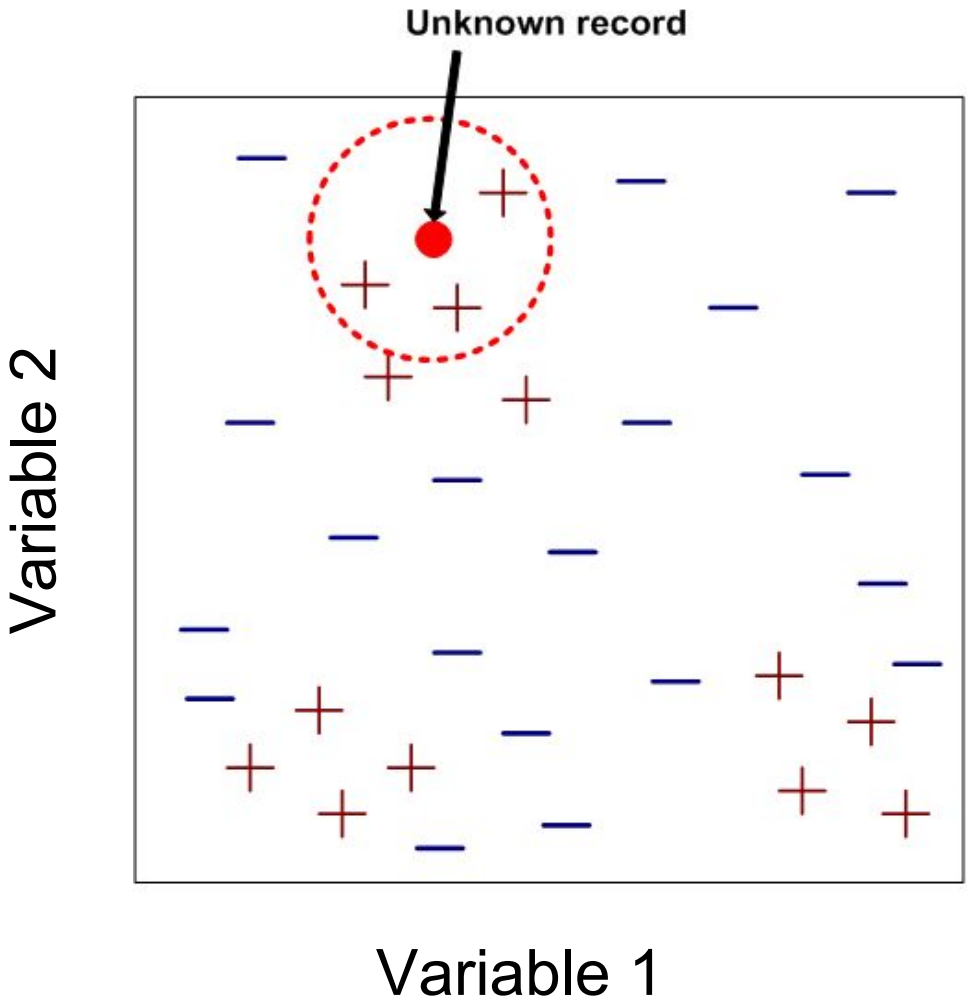




Clustering / Classification

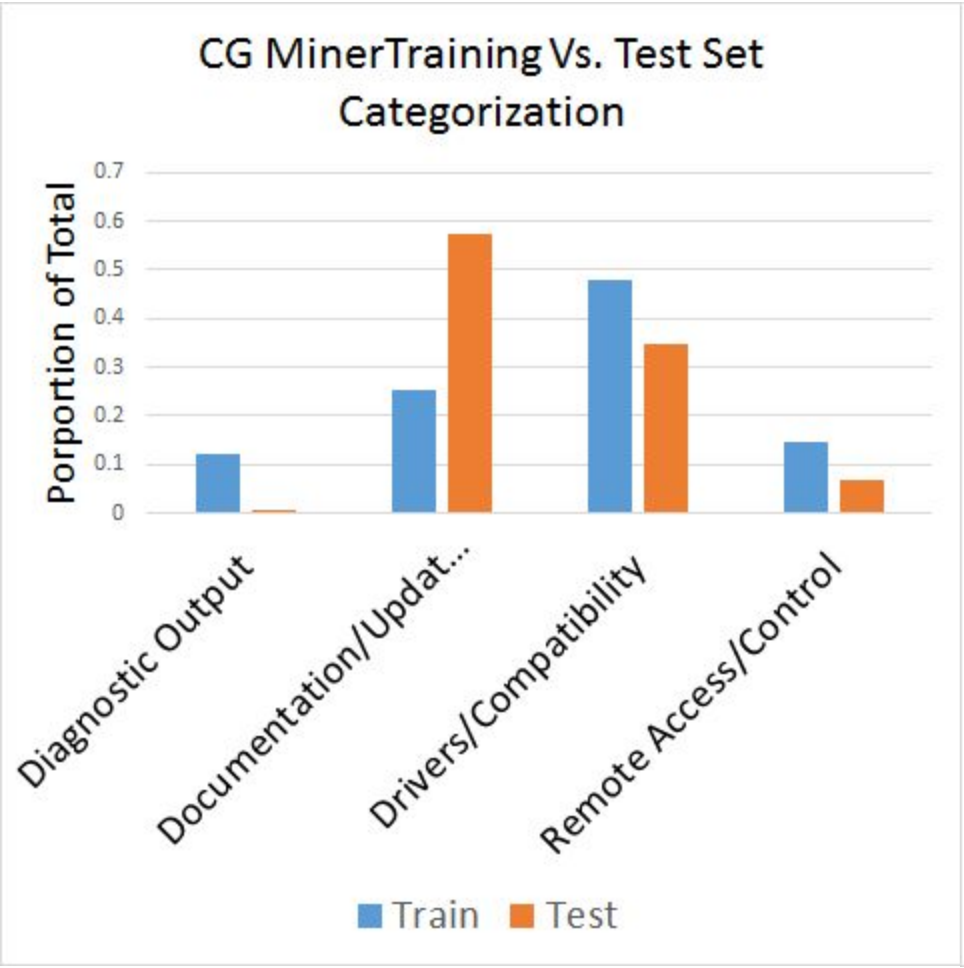
- First, create term-document matrix from all commit comments
- Convert to weighted matrix
- Deploy Algorithms
  - k Means Clustering
  - Naive Bayes

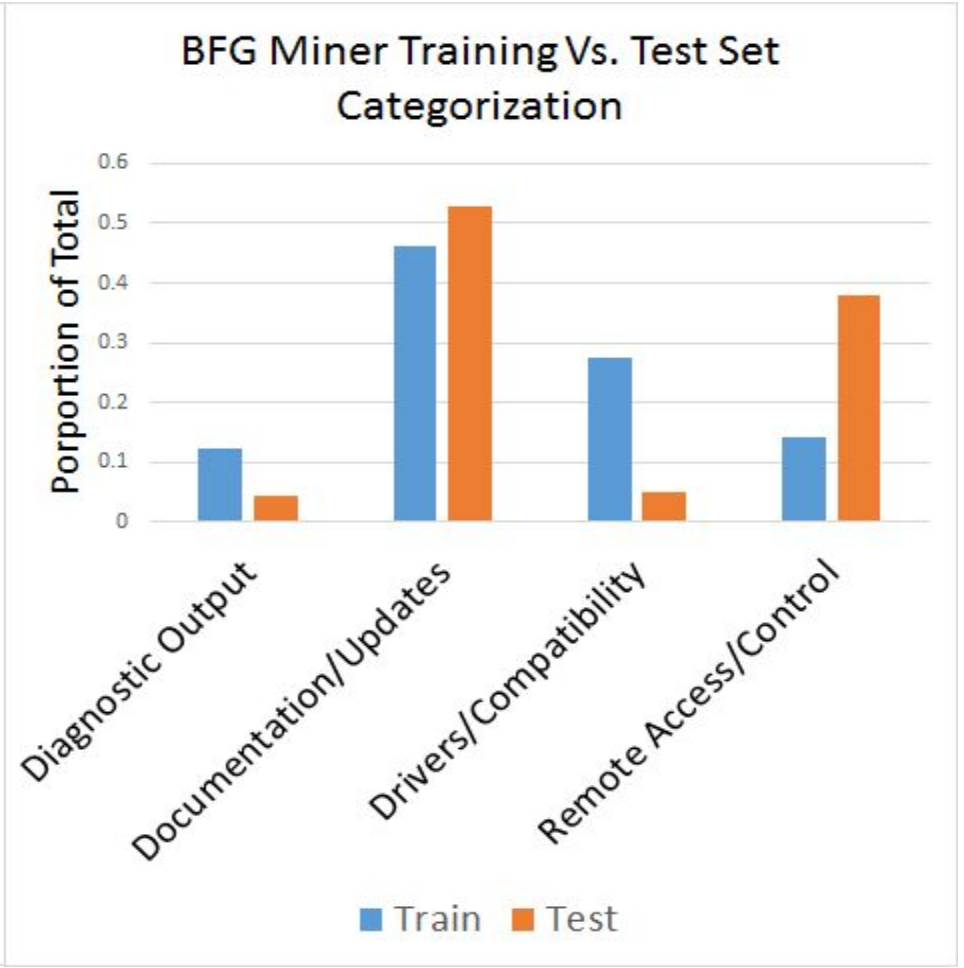
k Nearest Neighbors



1. Compute distance to other training records.
2. Identify k nearest neighbors.
3. Assign test record label from majority label of k-nearest neighbors.

All models are wrong but...



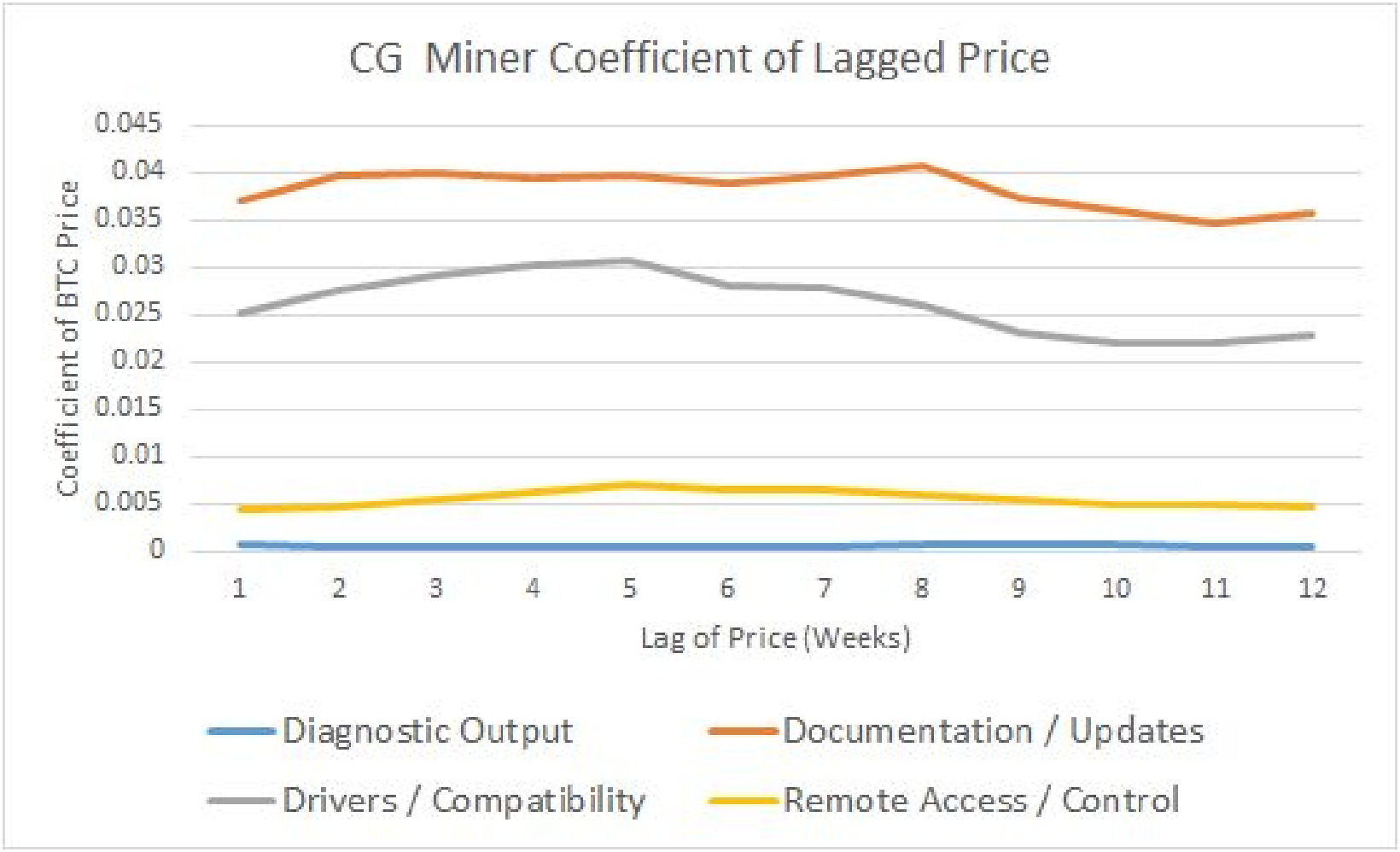


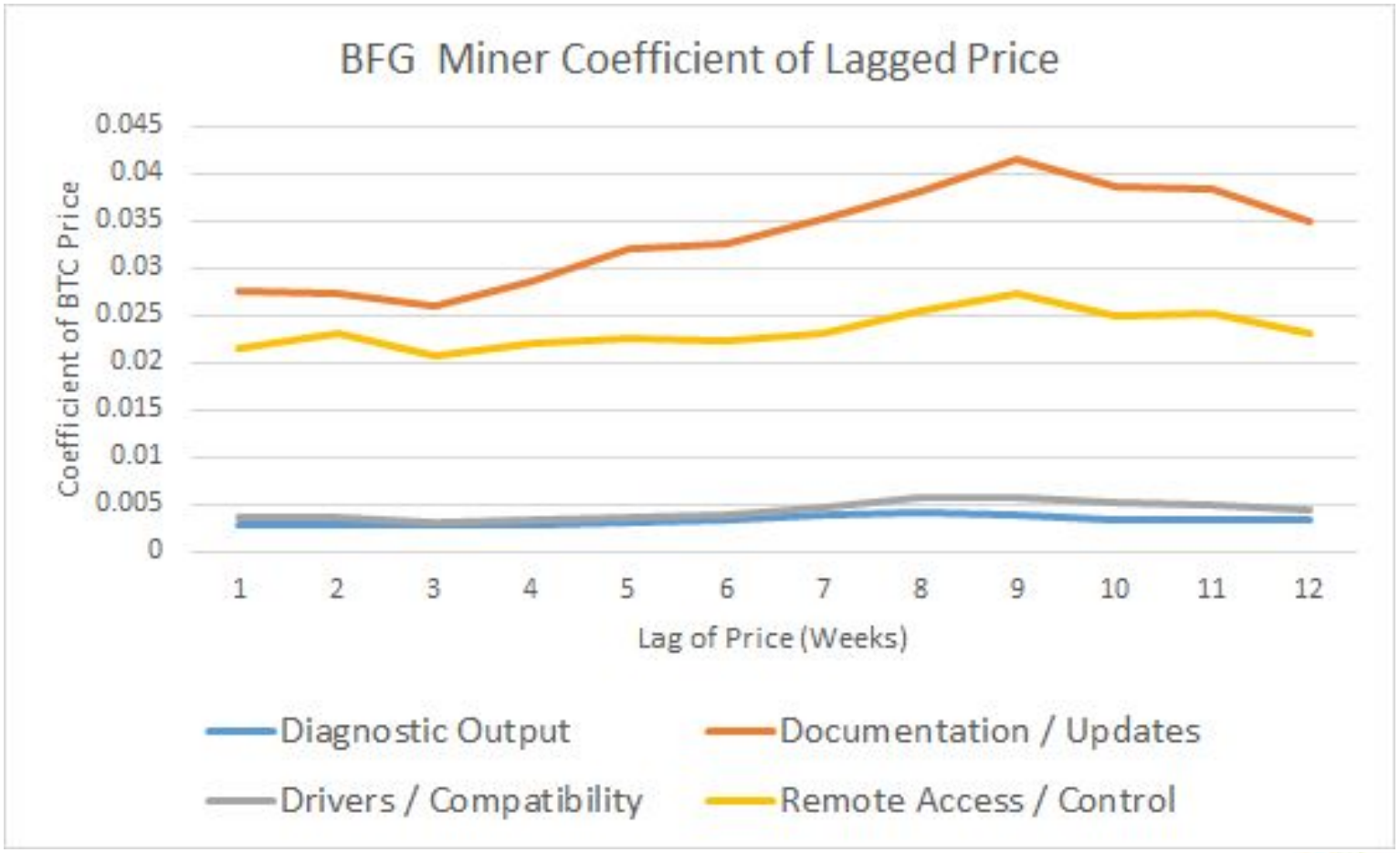
Method

Linear models for lags j = 1-12 weeks:

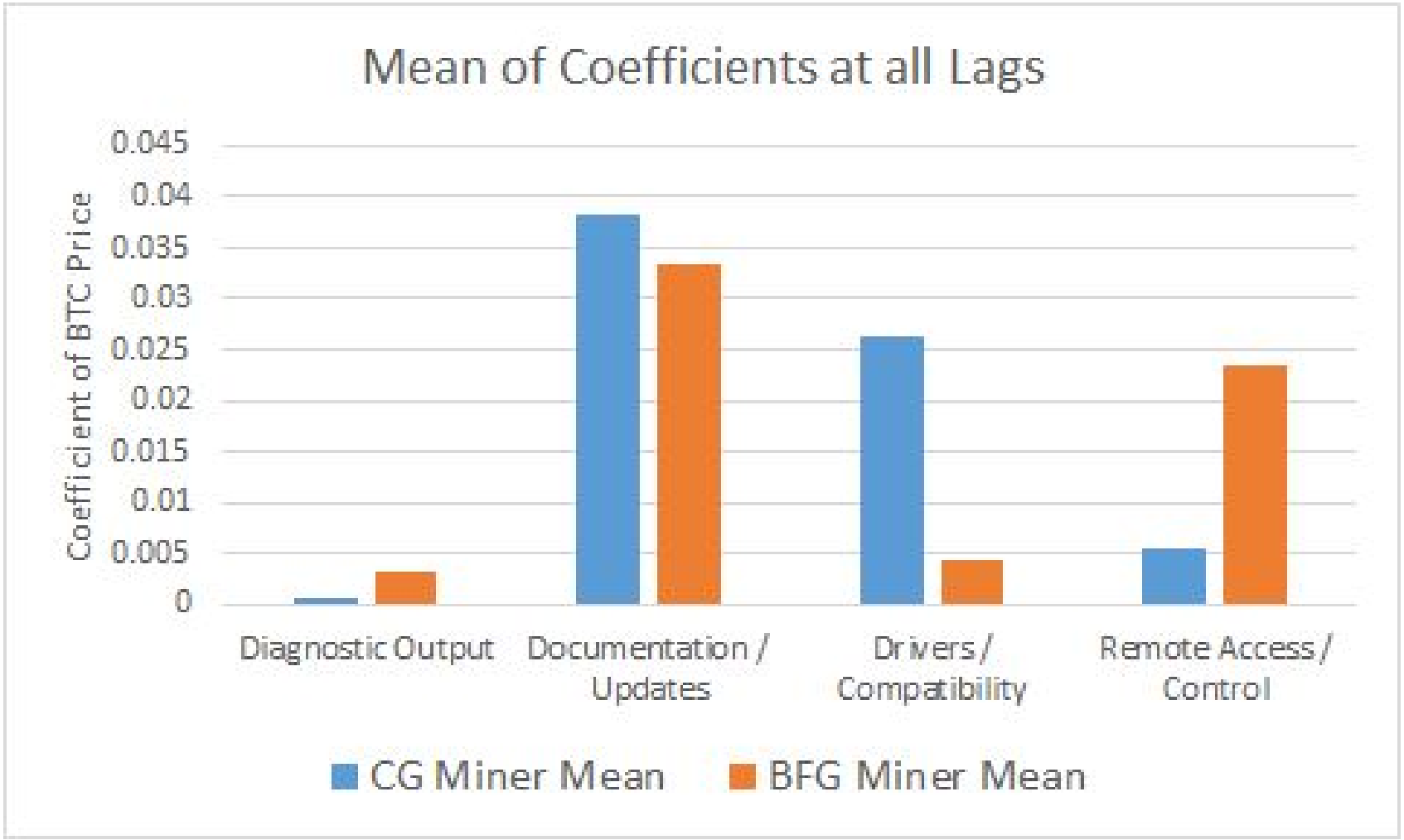
Post Count of Category $i_t = \beta$ Price $_{t-j}$

Price Effect on Post Counts



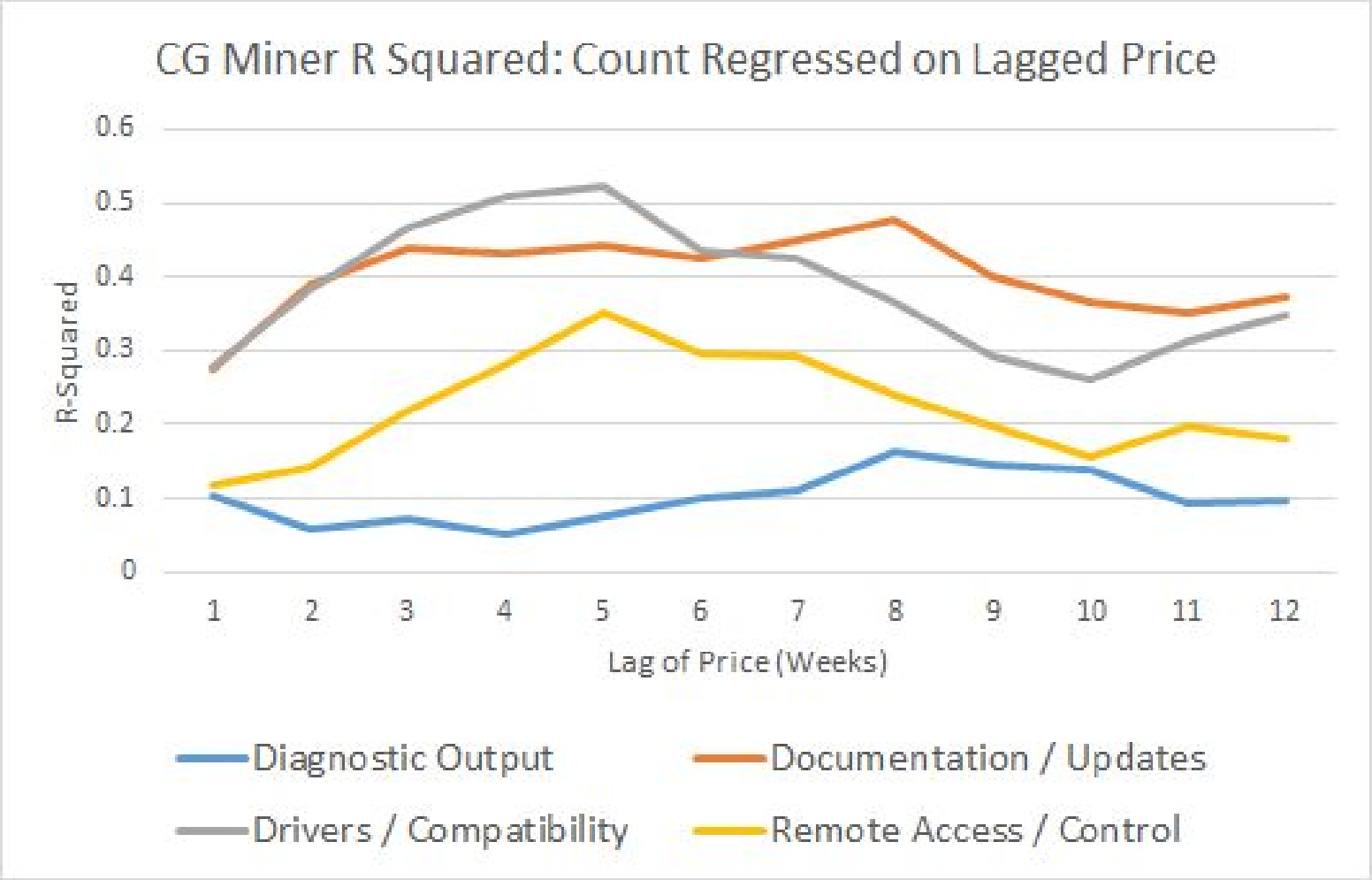

BFG Miner Coefficient of Lagged Price
Coefficient of BTC Price
0.045
0.04
0.035
0.03
0.025
0.02
0.015
0.01
0.005
0
1
2
3
4
5
6
7
8
9
10
11
12
Lag of Price (Weeks)
Diagnostic Output
Documentation / Updates
Drivers / Compatibility
Remote Access / Control

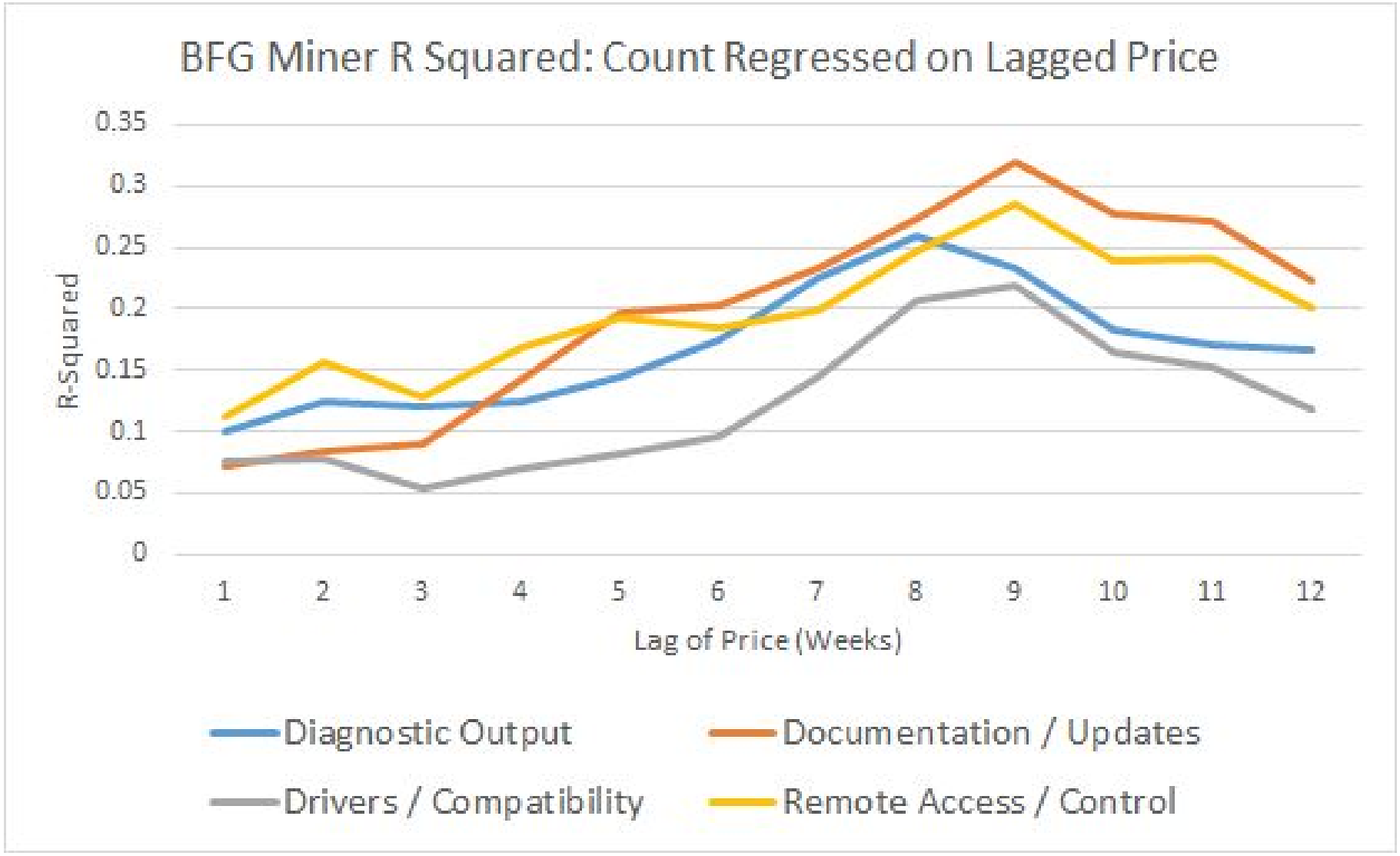

bitcoin


Mean of Coefficients at all Lags
Coefficient of BTC Price
0.045
0.04
0.035
0.03
0.025
0.02
0.015
0.01
0.005
0
Diagnostic Output
Documentation / Updates
Drivers / Compatibility
Remote Access / Control
CG Miner Mean
BFG Miner Mean


bitcoin

Explanatory Power




BFG Miner R Squared: Count Regressed on Lagged Price
R-Squared
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
1
2
3
4
5
6
7
8
9
10
11
12
Lag of Price (Weeks)
Diagnostic Output
Documentation / Updates
Drivers / Compatibility
Remote Access / Control

Insight

- BTC price explains greater portion of variation in CG Miner vs. BFG Miner
- CG Miner price effect - lag of 5 weeeks
- BFG Miner price effect - lag of 9 weeks
- Actual effect of price change on post count is small

Model Improvement

- Improve classification by training on greater number of records
- Examine linear regression assumptions
- Tune ks for kNN algorithm
- Fix Naive Bayes model
- Incorporate control variables

Appendix

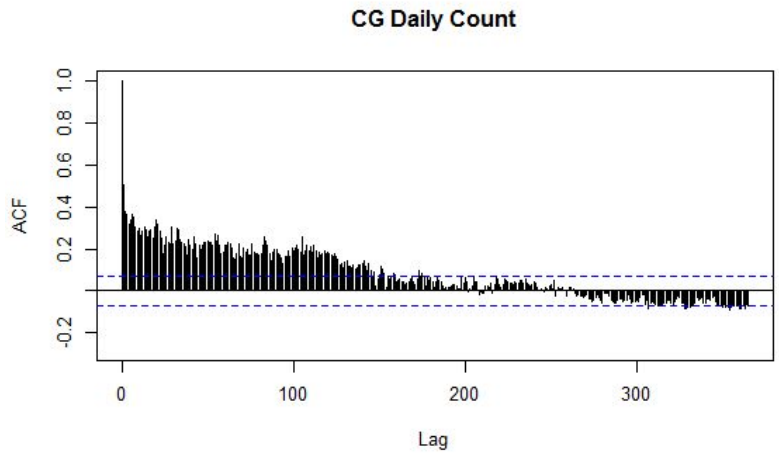
CG Miner P-Values

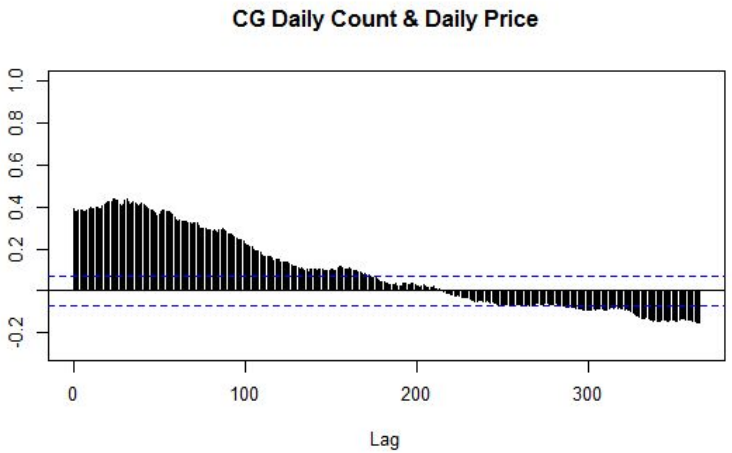
Lag	Diagnostic Output	Documentation / Updates	Drivers / Compatibility	Remote Access / Control
1	7.25E-04	6.44E-09	4.43E-09	2.53E-04
2	1.32E-02	7.09E-13	1.30E-12	6.64E-05
3	5.76E-03	1.02E-14	7.34E-16	4.29E-07
4	2.21E-02	2.38E-14	1.32E-17	5.46E-09
5	4.32E-03	1.37E-14	3.85E-18	3.34E-11
6	1.09E-03	9.42E-14	3.63E-14	2.71E-09
7	6.78E-04	1.30E-14	1.26E-13	4.16E-09
8	2.98E-05	1.10E-15	2.18E-11	1.84E-07
9	1.00E-04	1.65E-12	7.21E-09	3.43E-06
10	1.52E-04	3.31E-11	6.17E-08	5.36E-05
11	2.05E-03	1.09E-10	1.85E-09	4.52E-06
12	2.06E-03	3.56E-11	1.91E-10	1.45E-05

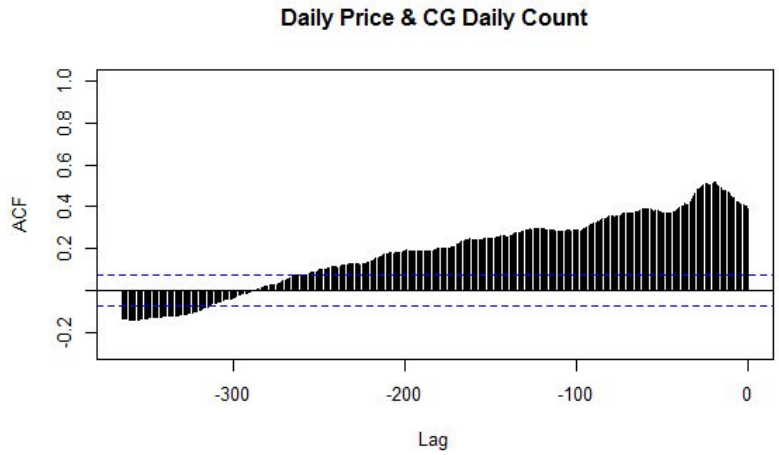
BFG Miner P Values

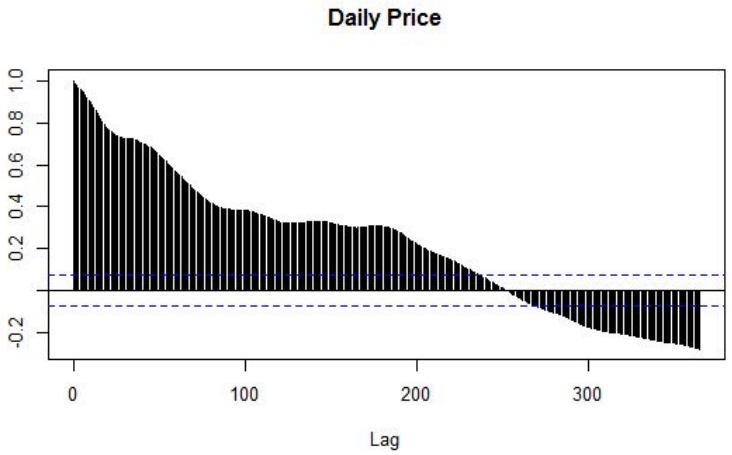
Lag	Diagnostic Output	Documentation / Updates	Drivers / Compatibility	Remote Access / Control
1	8.67E-04	5.38E-03	4.11E-03	4.22E-04
2	2.02E-04	2.59E-03	3.54E-03	2.46E-05
3	2.75E-04	1.87E-03	1.61E-02	1.73E-04
4	2.40E-04	7.45E-05	6.30E-03	1.30E-05
5	6.91E-05	2.27E-06	3.06E-03	3.27E-06
6	1.10E-05	1.89E-06	1.47E-03	5.69E-06
7	4.75E-07	2.64E-07	7.71E-05	2.53E-06
8	5.16E-08	1.92E-08	1.75E-06	1.19E-07
9	3.84E-07	8.66E-10	9.21E-07	1.07E-08
10	9.67E-06	2.21E-08	3.13E-05	2.87E-07
11	2.51E-05	4.04E-08	6.73E-05	2.91E-07
12	3.22E-05	1.09E-06	5.65E-04	4.36E-06

Autocorrelation at Higher Lags









bitcoin