Efficiency in the Bitcoin Market

Zichao Yang

Date: May 28, 2019

Motivation

The differences between bitcoin market and stock market:

- Bitcoin market is thinner than stock market.
- Bitcoin market does not have a real close or open time.
- There are hundreds of bitcoin exchanges operating in the world.
- Algorithm plays a role in deciding the pattern of bitcoin price.

In this paper, we focus on the last two points.

What does this paper do?

We take the technical aspect of bitcoin into consideration and study how the transaction fee functions differently in the bitcoin market.

We argue that there are two layers of efficiency in the bitcoin market:

- Efficiency on exchange level.
- Efficiency on the whole market level.

These two types of efficiency do not always move in the same direction.

Literature review

(1) Studies about the predictability of bitcoin return:

- People's attention and bitcoin return: results are inconclusive. (Kristoufek, 2015; Ciaian et al., 2016; Urquhart, 2018; Aalborg et al., 2018; Dastgir et al., 2019)
- Political or economic uncertainty and bitcoin return: results are inconclusive. (Dyhrberg, 2016; Ciaian et al., 2016; Bouri et al., 2017; Demir et al., 2018; Al-Khazali et al., 2018)
- Trade volume and bitcoin return: results are inconclusive. (Balcilar et al., 2017; Bouri et al., 2018; El Alaoui et al., 2018; Kjærland et al., 2018)

Literature review

(2) Studies testing weak form efficiency:

The results are inconclusive. Bitcoin is the most efficient one when compared with alter-coins, and efficiency is positively related with liquidity. (Urquhart, 2016; Nadarajah and Chu, 2017; Zhang et al., 2018; Bouri et al., 2018)

(3) Price across exchanges

- Brandvold et al. (2015) find that, in the period from April 2013 to February 2014, big bitcoin exchanges provide more information on price.
- Giudici and Abu-Hashish (2018) find that bitcoin price is highly interrelated among different exchanges.

Blocks and the blockchain

Blocks are defined as storage units which contain confirmed transaction data.

Blockchain is a sequence of blocks, where each block contains a piece of information about the location of its previous block.

Transaction View information about a bitcoin transaction

0d2946bfee28bfcd8d0da57b483dde1abd70865e780d71d2b8efc972af52ab02

3L4b2HAeukqEiEwDyJwzQs3p25hiF5Adnh (0.26226692 BTC - Output) → bc1qhfv57d4c3dz255q8hlnkc7ehnwd9wcq5pcg0ux - (Unspent) 0.00410706 BTC
bc1q836d3kcmhc25w6p7p6qdf2thxksj4rvxzhs0k9 - (Spent) 0.19419986 BTC

1 Confirmations 0.19830692 BTC

Summary	
Size	246 (bytes)
Weight	654
Received Time	2019-04-03 19:38:21
Lock Time	Block: 570071
Included In Blocks	570072 (2019-04-03 19:49:25 + 11 minutes)
Confirmations	1
Visualize	View Tree Chart

Inputs and Outputs	
Total Input	0.26226692 BTC
Total Output	0.19830692 BTC
Fees	0.06396 BTC
Fee per byte	26,000 sat/B
Fee per weight unit	9,779.817 sat/WU
Estimated BTC Transacted	0.00410706 BTC
Scripts	Hide scripts & coinbase

Mining and Proof-of-Work

The structure of block header:

Field	Description
Version	A version number to track software/protocol upgrades
Previous Block Hash	A reference to the hash of the previous block in the chain
Merkle Root	A hash of the root of the merkle tree of this block's transactions
Timestamp	The approximate creation time of this block
Target	The Proof-of-Work algorithm target for this block
Nonce	A counter used for the Proof-of-Work algorithm

This table is adopted from Table 10-3 in Antonopoulos (2017).

Block header hash: Hash Function

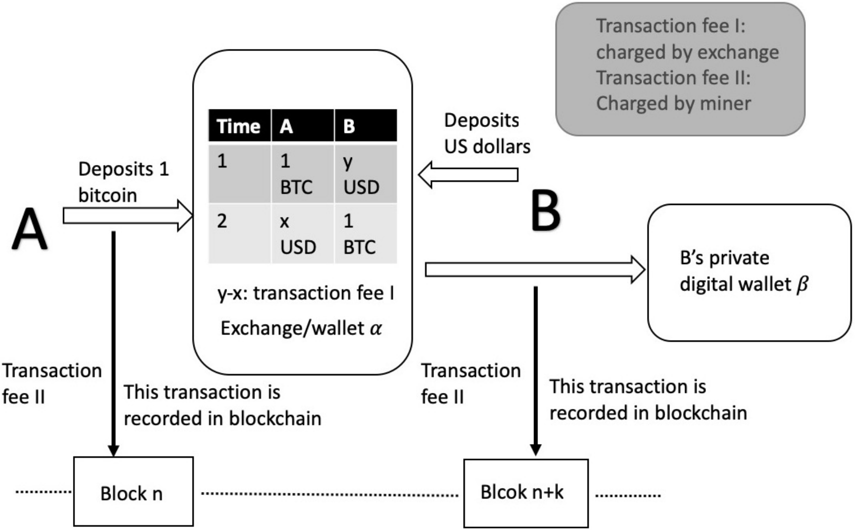
0000000000000000000ae165831e934ff763ae46a2a6c172b3f1b60a8ce26f

Mining: a process of hashing the block header repeatedly. In each time, the miner change the value of nonce a little bit until the block header hash satisfies the target set by the **Proof-of-Work** algorithm.

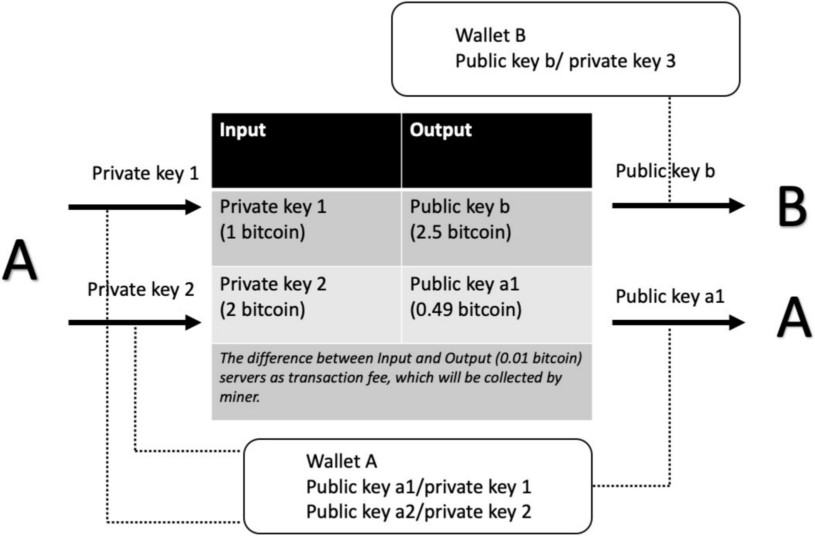
More on Proof-of-Work

- Difficult to find the nonce, but easy to verify the finding
- Makes the blockchain almost immutable (with mining competition)
- Auto-adjustable, generates blocks, on average, every 10 minutes
- Adjustment not symmetric See graph

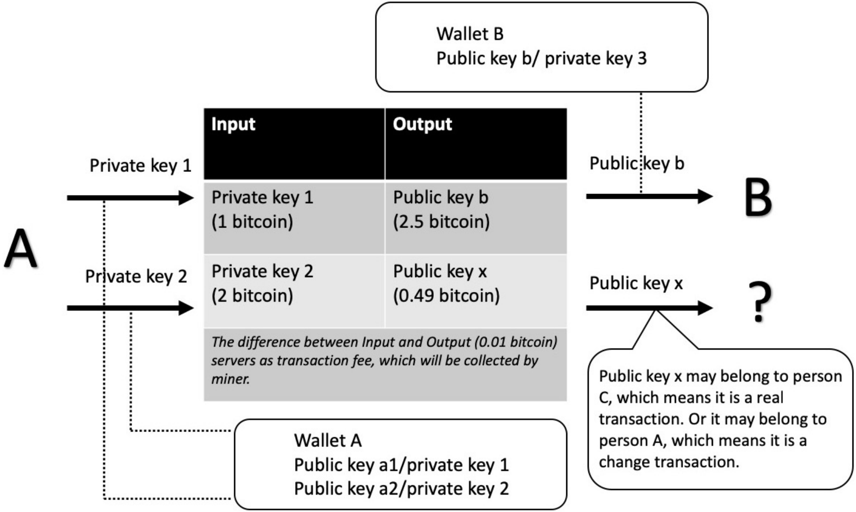
Exchange trade volume



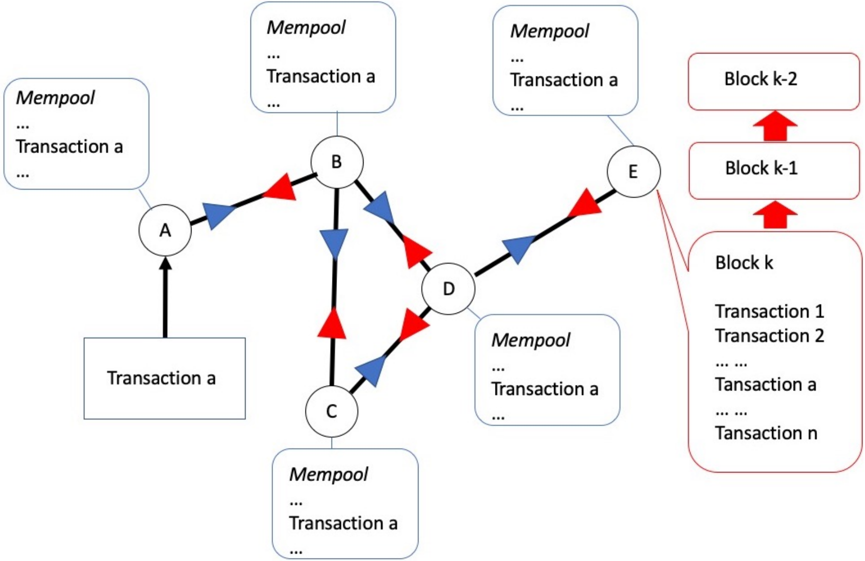
On-chain transaction volume



On-chain transaction volume



The procedure of transaction confirmation



Transaction fee

Transaction fees are left by users, then collected by miners.

When miners build their blocks, they select the transactions which have the highest reward-size ratio (RSR):

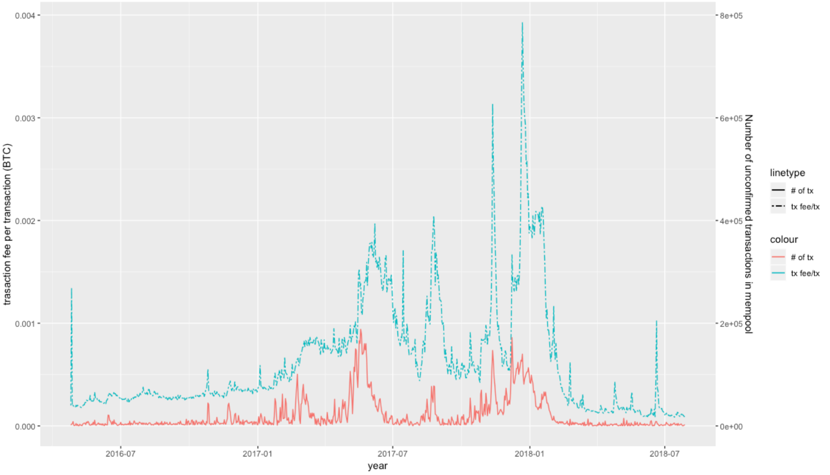
$$\text{RSR} = \frac{\text{transaction fee paid in this transaction}}{\text{the size of this transaction data}}$$

Users make transaction decisions based on the ratio of transaction value to transaction fee (TFR):

$$\text{TFR} = \frac{\text{transaction fee paid in this transaction}}{\text{the value of this transaction}}$$

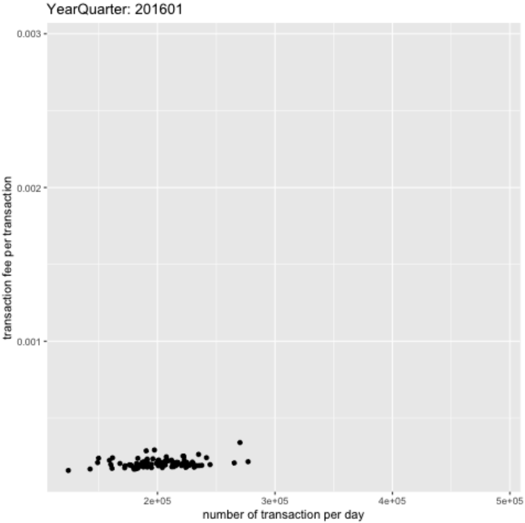
The gap between the goals of miners and users is bridged by the bitcoin wallet.

Mempool size and transaction fee per transaction

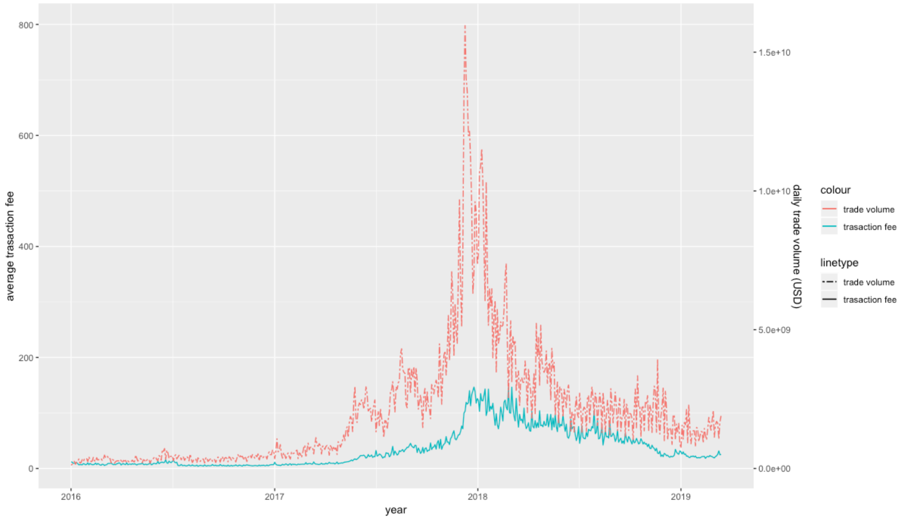


Go to conclusion

Number of transactions and transaction fee per transaction



Tx fee per tx and total tx volume



Exchange-level data

Exchange-level data comes from CryptoCompare. It includes the bitcoin close price in the seven largest exchanges (Bitfinex, Bitstamp, Coinbase, Gemini, Kraken, Lakebtc and Quoine) and market trade volume data from 2015-10-08 to 2019-03-03.

Table 2: Summary statistics of price

	Bitfinex	Bitstamp	Coinbase	Gemini	Kraken	Lakebtc	Quoine
	Unit: USD (dollar)						
Mean	3760.6	3749.2	3758.9	3755.4	3749.0	3973.3	3799.3
Median	2461.0	2502.0	2505.6	2497.6	2497.8	2598.3	2681.7
Maximum	19210.0	19187.8	19650.0	19500.0	19356.9	19315.7	19653.5
Minimum	243.6	243.4	243.4	243.9	242.7	240.6	244.5
Std. dev.	3781.3	3786.5	3808.0	3800.49	3783.6	3991.3	3871.8
Observations	1243	1243	1243	1243	1243	1243	1243

Price efficiency in a single exchange

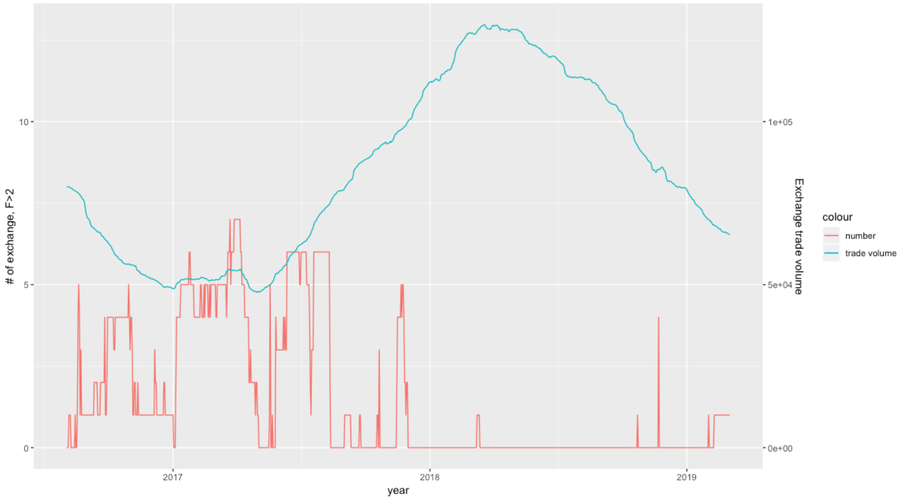
We apply the autoregressive analysis to study the dynamics of bitcoin return. The order of lag terms is chosen based on AIC. The autoregressive model can be written as:

$$R_{i,t} = Const + A(L)R_{i,t} + \mu_t$$

where $R_{i,t}$ denotes bitcoin return from exchange i at time t. $A(L)$ is lag operator and the order of lag terms is chosen based on AIC.

We combine the AR model with the rolling windows analysis (300-days window). Then we use an F-test to test the joint significance of all these lag terms in each regression. We count, in each period, how many exchanges have joint significant lag terms. There are seven exchanges, so the result in each period can range from 0 to 7.

Plot results with the mean of exchange trade volume



Price correlation among different exchanges

We still use bitcoin return data from the seven largest exchanges and conduct the test for Granger causality on the bitcoin return data for every two exchanges.

$$R_{i,t} = Const + A(L)R_{i,t} + B(L)R_{j,t} + \mu_t$$

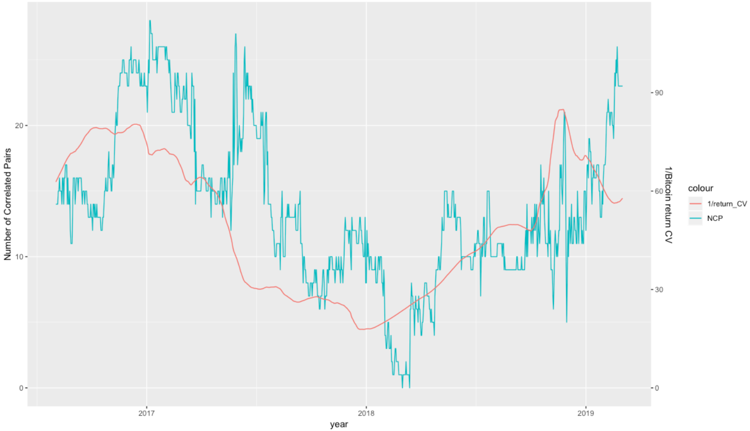
where $R_{i,t}$ and $R_{j,t}$ denote bitcoin return from exchange i and j at time t, $A(L)$ and $B(L)$ are lag operators, which are chosen based on AIC.

We combine the Granger causality test with the rolling windows analysis. We calculate the number of correlated pairs (NCP) of exchanges by counting the number of exchanges, from which the return data can Granger cause the return of other exchanges in each period. In our dataset, there are seven different exchanges, so the maximum value of NCP is $7 \times 6 = 42$.

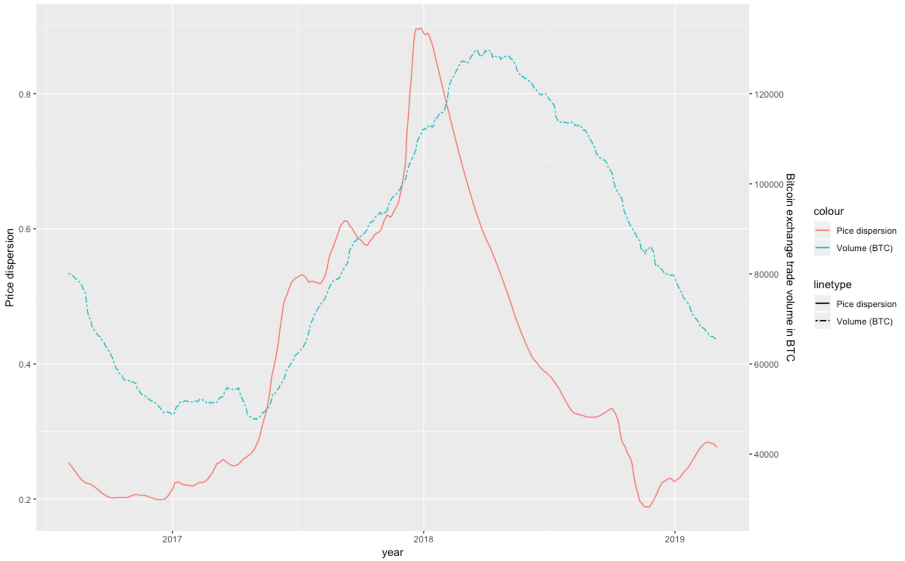
Plot results with the inverse of CV of bitcoin price

The coefficient of variation (CV) of price can be defined as:

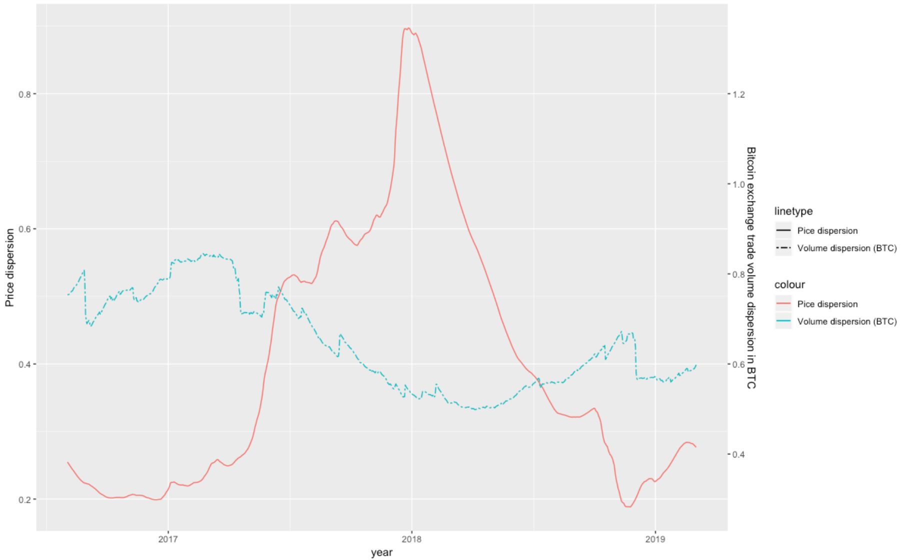
$$CV_t = \frac{\text{Standard Deviation of the price in } t^{th} \text{ period}}{\text{Mean of the price in } t^{th} \text{ period}}$$



Price dispersion and trade volume

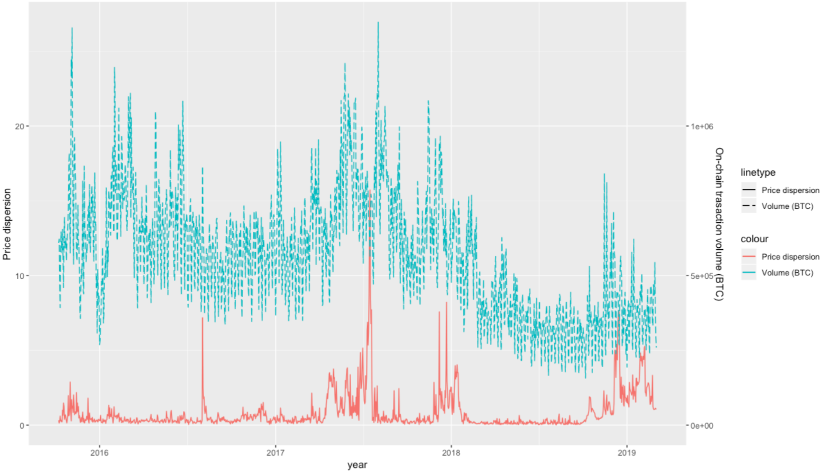


Price dispersion and trade volume dispersion



Price dispersion and on-chain transaction volume

When on-chain trade volume is high, price dispersion becomes high and exchanges become more isolated with each other.



Daily data

Daily blockchain-related data come from CoinMetrics. The date range is chosen from 2015-01-01 to 2018-01-01.

Table 4: Summary statistics of blockchain-related data

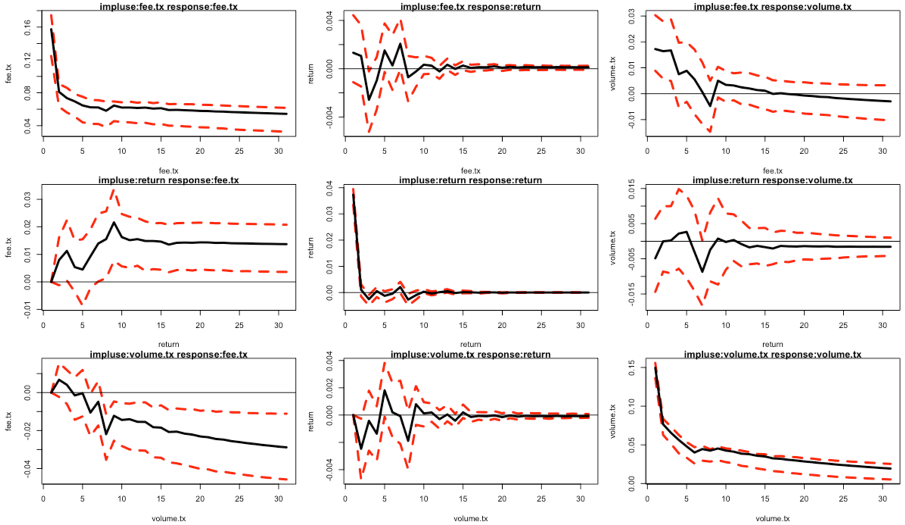
	on-chain tx volume/tx log-form	tx fee/tx log-form	bitcoin return log-form
Mean	1.032	-7.998	0.003
Median	1.022	-8.211	0.003
Maximum	2.124	-5.539	0.224
Minimum	0.250	-9.165	-0.236
Std. dev.	0.310	0.752	0.038
Observations	1096	1096	1096

Structural VAR

The structural VAR model is comprised of daily data on the log-form of bitcoin daily return (R_t), the daily average on-chain transaction volume per transaction in BTC(V_t), the daily average transaction fee per transaction in BTC(F_t).

$$\begin{bmatrix} a_{11}^0 & 0 & 0 \\ a_{21}^0 & a_{22}^0 & 0 \\ a_{31}^0 & a_{32}^0 & a_{33}^0 \end{bmatrix} \begin{bmatrix} F_t \\ R_t \\ V_t \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} + \begin{bmatrix} a_{11}^1 & a_{12}^1 & a_{13}^1 \\ a_{21}^1 & a_{22}^1 & a_{23}^1 \\ a_{31}^1 & a_{32}^1 & a_{33}^1 \end{bmatrix} \begin{bmatrix} F_{t-1} \\ R_{t-1} \\ V_{t-1} \end{bmatrix} + \ldots + \begin{bmatrix} \epsilon_F \\ \epsilon_R \\ \epsilon_V \end{bmatrix}$$

Results


impluse:fee.tx response:fee.tx
impluse:fee.tx response:return
impluse:fee.tx response:volume.tx
impluse:return response:fee.tx
impluse:return response:return
impluse:return response:volume.tx
impluse:volume.tx response:fee.tx
impluse:volume.tx response:return
impluse:volume.tx response:volume.tx
fee.tx
return
volume.tx

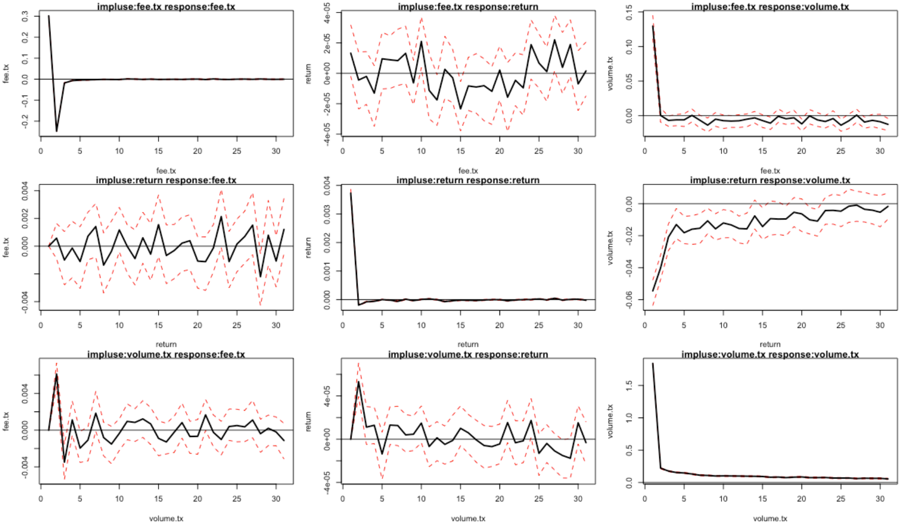
Block-level data

We match minute-level trade volume and price data from Bitstamp with block-level total transaction fees and number of transactions data from BlockSci. Proxy Variable Explain

Table 5: Summary statistics of blockchain-related data on block level

	Bitstamp tx volume/tx log-form	tx fee/tx log-form	bitcoin return log-form
Mean	-4.600	0.000	0.000
Median	-4.352	-0.003	0.000
Maximum	5.992	6.329	0.070
Minimum	-18.787	-6.290	-0.152
Std. dev.	2.037	0.411	0.004
Observations	150,785	150,785	150,785

Results


impluse:fee.tx response:fee.tx
impluse:fee.tx response:return
impluse:fee.tx response:volume.tx
impluse:return response:fee.tx
impluse:return response:return
impluse:return response:volume.tx
impluse:volume.tx response:fee.tx
impluse:volume.tx response:return
impluse:volume.tx response:volume.tx
fee.tx
return
volume.tx

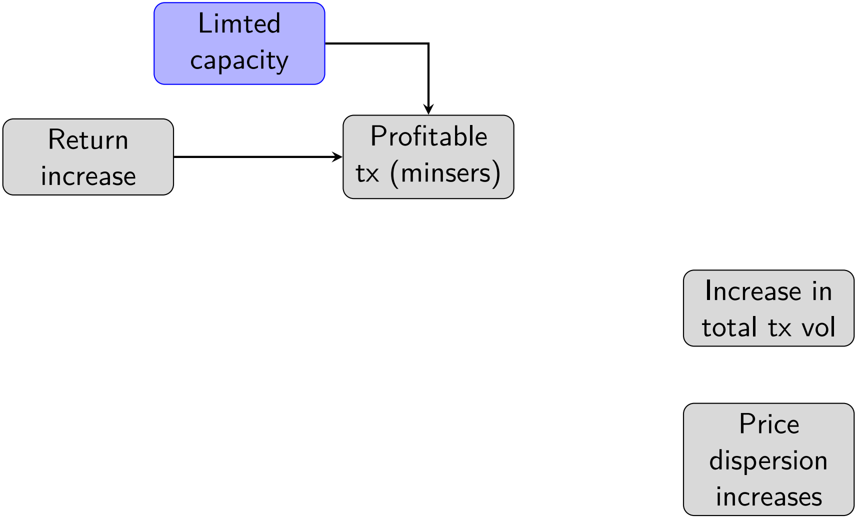
Mechanism behind results

Increase in total tx vol

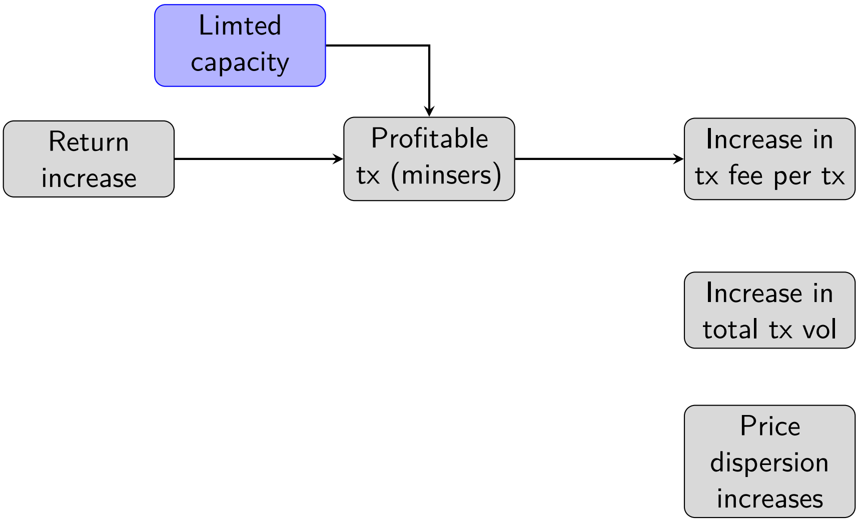
Price dispersion increases

Mechanism behind results

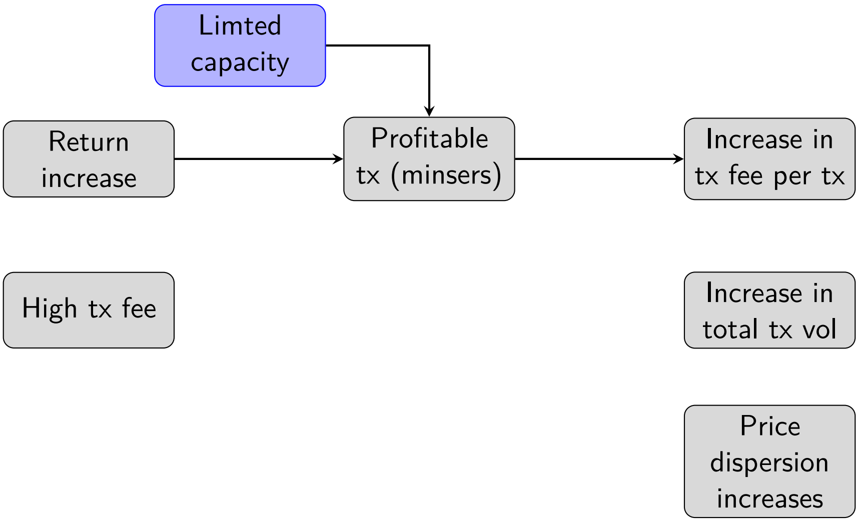
Mechanism behind results



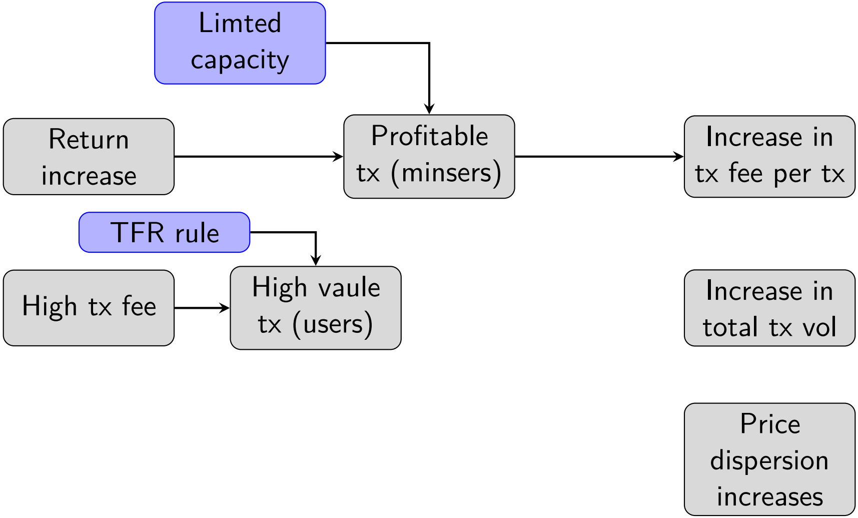
Mechanism behind results



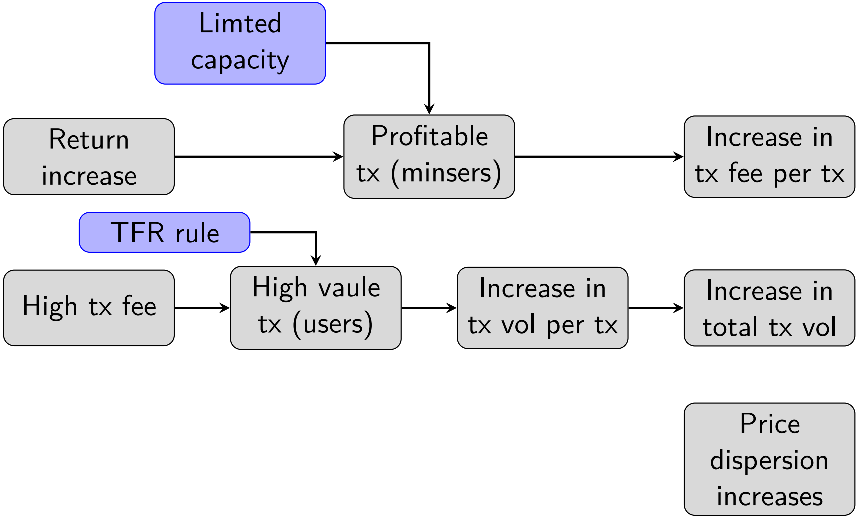
Mechanism behind results



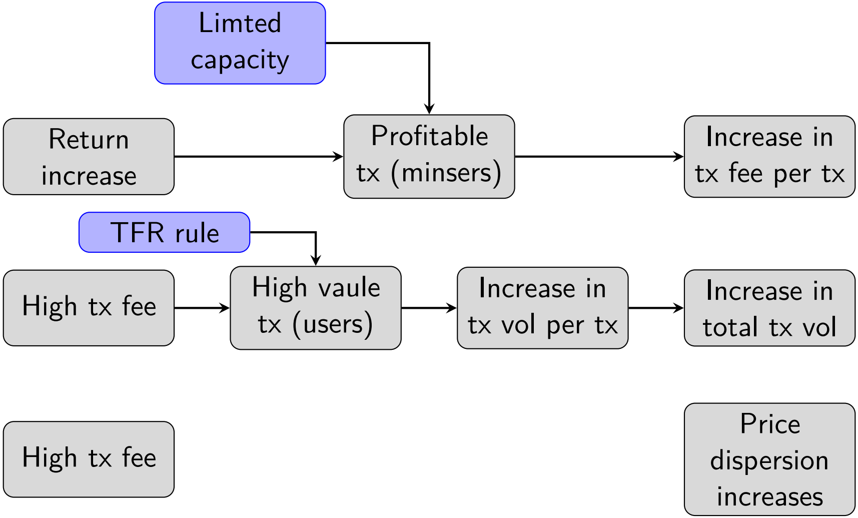
Mechanism behind results



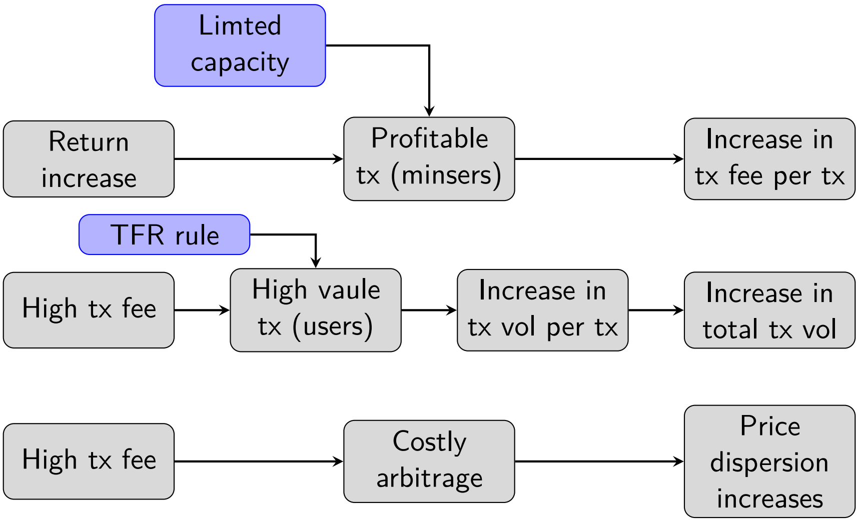
Mechanism behind results



Mechanism behind results



Mechanism behind results



Extensions

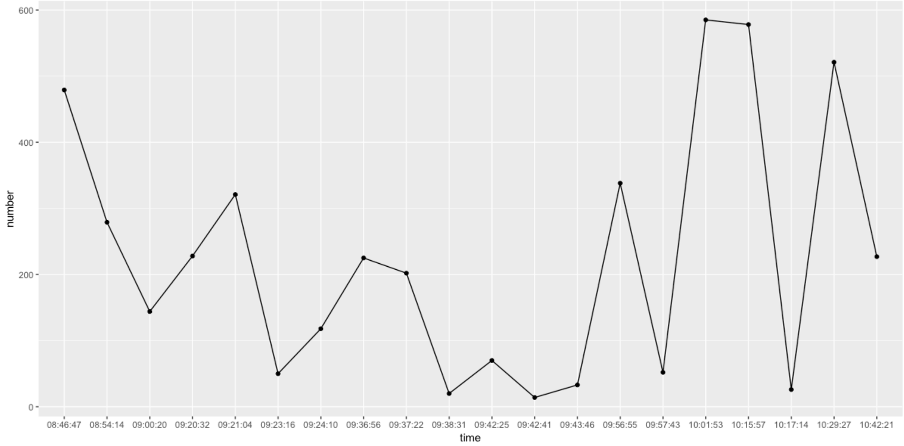
- Get more accurate block-level on-chain transaction volume data from BlockSci.
- Introduce structural breaks into our analysis.
- Use local projection method as a robustness check.

Future researches

- Cryptocurrency Confidence Survey
- A theoretical multi-exchanges model
- A network model using the data from BlockSci (block-level individual/clustered data)

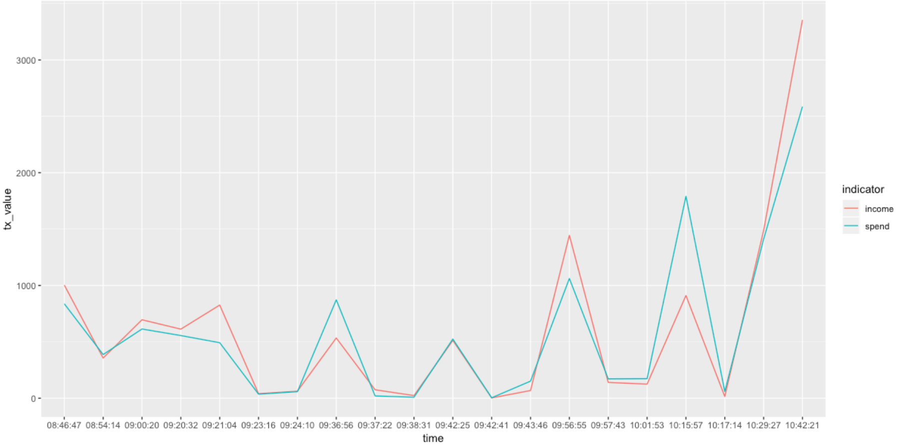
BlockSci: how many clusters are involved in every block

Mt. Gox filed for bankruptcy at 9:30 am UTC, 2014-02-28.

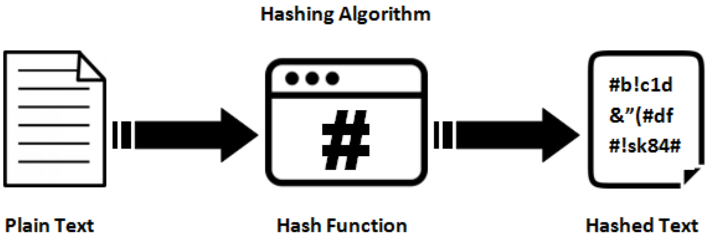


BlockSci: the value of transactions in every block

Mt. Gox filed for bankruptcy at 9:30 am UTC, 2014-02-28.



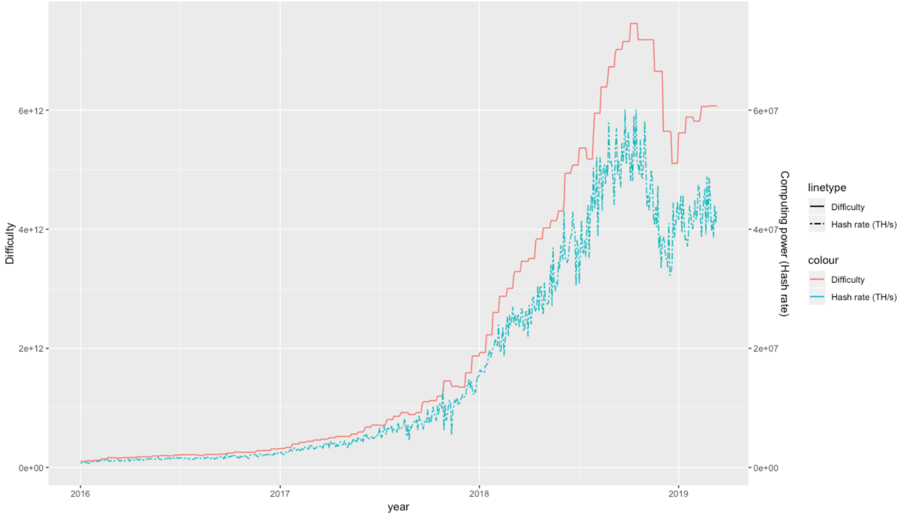
Hash function



Plain Text: arbitrary length
Hashed Text: certain length (like 32 bytes)

Mining and Proof-of-Work

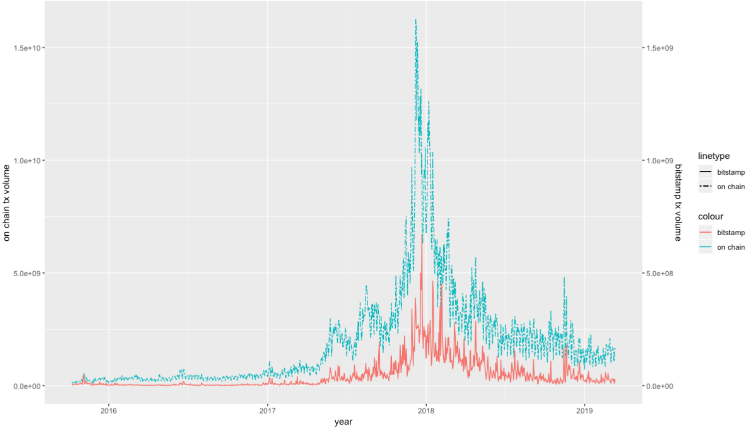
Difficulty and computing power (hash rate)



Proof-of-Work

On-chain transaction volume VS. Exchange trade volume

We match daily on-chain transaction volume data from CoinMetrics with minute-level trade volume data from Bitstamp. The correlation is 0.886.



Block-level Data