The Economics of Cryptocurrencies

Zichao Yang

Dissertation submitted to the Faculty of the

Virginia Polytechnic Institute and State University

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Economics

Kwok Ping Tsang, Chair

Shaowen Luo

Bradley Paye

Alec Smith

Gerelt Tserenjigmid

April 9, 2021

Blacksburg, Virginia

Keywords: Bitcoin, Blockchain, Networks, Transaction Fees, Asset Returns

The Economics of Cryptocurrencies

Zichao Yang

(ABSTRACT)

This paper has four chapters. The first chapter serves as an introduction. The second chapter studies the transaction fees in the bitcoin system. The transaction fees and transaction volume in the bitcoin system increase whenever the network is congested and results from a simple VAR show that it is indeed the case. To account for the empirical findings, we build a model where users and miners together determine the transaction fee and transaction volume endogenously. Even though the fluctuating transaction fee mechanism in bitcoin introduces the extra cost of uncertainty to users, a back-of-envelope calculation shows that the cost of using the bitcoin network for transactions is still smaller than the cost of using the current conventional payment system with a fix transaction fee rate. The second chapter studies the time-varying price dispersion among different bitcoin exchanges. We identify the sources of price dispersion using a standard time-varying vector autoregression model with stochastic volatility. The results show that shocks to transaction fees and bitcoin price growth explain on average 20%, and sometimes more than 60%, of the variation of price dispersion. The third chapter studies the relationship between connections and returns in the bitcoin investor network. Using transaction data from the bitcoin blockchain, we reach three conclusions. First, on average, the annualized returns of connected

addresses in the network are 20.75% above those of their unconnected peers. Second, returns also differ among those connected addresses. By dividing the connected ad- dresses into ten deciles based on their centrality, we find that addresses in the two most-connected deciles earn higher returns than the other connected addresses. Third, eigenvector centrality is more related than degree centrality to higher returns, implying that quality of connections matters.

The Economics of Cryptocurrencies

Zichao Yang

(GENERAL AUDIENCE ABSTRACT)

This paper has four chapters. The first chapter serves as an introduction. The second chapter studies the transaction fees in the bitcoin system. The transaction fees in the bitcoin system can fluctuate given the amount of unconfirmed transactions in the bitcoin network. Our results show that the transaction fees and transaction volume in the bitcoin system increase whenever the network is congested. To account for this findings, we build a model and show that users and miners together can determine the transaction fee and transaction volume. Even though the fluctuating transaction fee mechanism in bitcoin introduces the extra cost of uncertainty to users, a back-of-envelope calculation shows that the cost of using the bitcoin network for transactions is still smaller than the cost of using the current conventional payment system with a fix transaction fee rate. The second chapter studies the price dispersion among different bitcoin exchanges. Our results show that transaction fees and bitcoin price growth can be important explanatory factors for the price dispersion among different bitcoin exchanges. The third chapter studies the relationship between connections and returns in the bitcoin investor network. Using transaction data from the bitcoin blockchain, we reach three conclusions. First, on average, those connected addresses in the network earn higher returns than their unconnected peers. Second, returns also

differ among those connected addresses. By dividing the connected addresses into ten groups based on their centrality, we find that addresses in the two most-connected groups earn higher returns than the other connected addresses. Third, eigenvector centrality, which measures the quality of connections, is more related than degree centrality, which measures the quantity of connections, to higher returns, implying that quality of connections matters.

Dedication

Dedicated to my parents

9684d08a8d320b34e92f5ba7b80a2100d56449c643ec6b0e6cc5459f1ea9fd9a

(This message is hashed using SHA-256, the same hash function used in bitcoin.)

Acknowledgments

I would like to express my deepest appreciation to my advisors, Dr. Kwok Ping Tsang, and Dr. Alec Smith. I still remember my first meeting with Dr. Tsang and Dr. Smith at Dr. Smith's office three years ago. And we decided to choose cryptocurrency as my research topic on that very day. Without my advisors' guidance and support in the past three years, this dissertation will not exist. Dr. Smith's rigorous attitude to research and optimistic attitude to life always inspire me and set an excellent example of how to be an outstanding scholar. I am also deeply indebted to Dr. Tsang. I remember the lunch meetings we had before the pandemic, the 11:30 pm Zoom call for my job market paper, and the two weeks we spent together to edit my paper sentence by sentence. Most importantly, Dr. Tsang shows me how to be a good advisor when I mentor my students in the future.

Besides my advisors, I would also like to extend my gratitude to my other committee members: Dr. Shaowen Luo, Dr. Bradley Paye, and Dr. Gerelt Tserenjigmid for their help in my dissertation writing process. This dissertation could not have been accomplished without their insightful comments and encouragement.

I would like to extend my sincere thanks to Dr. Suqin Ge, Dr. Sudipta Sarangi, Dr. Niloy Bose, Dr. Richard Cothren and Dr. Mark Liu for their kind supports, especially during this unprecedented period when the whole world is weathering COIVD-19. I would also like to thank Dr. Michael Taaffe at the Department of Industrial and

Systems Engineering for helping me find the right direction of my research.

I am also grateful to Dr. Steve Trost and Dr. Melanie Fox for showing me how to be a good lecturer and making the switch from in-person teaching to online teaching as smoothly as possible. I also would like to thank Ms. Amy Stanford, Mr. Will Bebout, and all the other staff at the Department of Economics. Thanks to their dedicated work, I can focus on my own research.

Special thanks to my friends. Abdelaziz (Zizo), I am lucky to meet you the first day I landed in the US, and your company in the last five years is a blessing. We went to different places and tried amazing or "interesting" food together. Thank you for standing by me all the time. Mayra and Alexis, I will never forget all the Movie Friday Nights or Board Game Saturday Nights we had together. Hopefully, we will eat the hotpot together soon in China. David and Seifeldeen, thank you for all those intellectual and entertaining debates about technology. Our "unthoughtful" trip to New York will always be part of the best memories I had. I will never forget the days I discuss VAR models with Yu. Despite the time difference, I appreciate all the conversations I had with Ming, who is like a big brother to me. I would also like to thank my other friends: Sunjin, Brian, Shan, Asal, Sakshi, Yao, He, Trevin, Chang Geun, Huang Phuong etc., for their moral support in the past five years.

Last but not least, I cannot end this acknowledgment without mentioning Masha, the cutest beagle in the world. She always looks at me with her puppy eyes while I am typing this dissertation and patiently wait for me to walk her in the next minute.

Contents

List of Figures

List of Tables

Attribution

Two Published Journal Articles (PJAs) are embedded in this dissertation.

Chapter 2: The Market For Bitcoin Transactions is published in the *Journal of International Financial Markets, Institutions & Money*

(https://doi.org/10.1016/j.intfin.2021.101282).

Chapter 3: Price Dispersion in Bitcoin Exchanges is published in *Economics Letters* (https://doi.org/10.1016/j.econlet.2020.109379).

My advisor, Kwok Ping Tsang, is the co-author of these two PJAs. Both authors have contributed equally.

Chapter 1

Introduction

In the past few years, hundreds of cryptocurrencies have been invented to serve different purposes. Among all these cryptocurrencies, bitcoin is, unarguably, the oldest and most important one. This dissertation seeks to study bitcoin from different aspects: The second chapter studies the bitcoin payment system as a replacement for our current fiat system. The third chapter studies the price dispersion among bitcoin exchanges, which is a unique phenomenon in the cryptocurrency market. In the last chapter, we examine the bitcoin investors' behavior in the bitcoin market, where the token itself is traded as a digital asset.

In the second chapter, we show that when the demand in the bitcoin network is high, the average transaction fees and transaction volume will also go up. The effect of a demand shock on transaction fees is most observable during the volatile period. Transaction volume first goes up then goes down during the early period, which means users postpone their small volume transactions when the network is congested. Then we propose a theoretical model account for the phenomenon observed in the VAR results. We argue that two plausible explanations for the empirical results are: (1) users' impatience; (2) the ratio of professional miners to amateur miners in the network across different periods. Furthermore, we calculated the cost of using bitcoin as a payment system with fluctuating transaction fee rates. Our conclusion shows that bitcoin is still a cheaper option compared with conventional payment systems.

When we treat bitcoin this token as a digital asset, the bitcoin market does not operate precisely as those conventional financial markets. For example, Bitcoin is traded in a number of exchanges, and there is a large and time-varying price dispersion among them. In the third chapter, we identify the sources of price dispersion among different bitcoin exchanges using a standard time-varying vector autoregression model with stochastic volatility. Using a TV-VAR-SV model, we find that shocks to transaction fees and bitcoin price growth explain, on average, 20%, and sometimes more than 60%, of the variation of price dispersion. One explanation could be that high transaction fees make investors reluctant to arbitrage and drive up price dispersion. Meanwhile, when the bitcoin price is volatile, investors may find it too risky to move between different exchanges to take advantage of the price dispersion.

The last chapter studies investors' trading behavior in the bitcoin market. We use the bitcoin address-level transaction data to identify the bitcoin investor network and find that the connected addresses in the network earn 20.75% higher returns than those unconnected addresses. We also find that the top 20% most connected addresses are earning significantly higher returns than their other connected peers. Furthermore, our results show that the quality of connections is more important than the quantity of connections.

Chapter 2

The Market For Bitcoin Transactions

2.1 Introduction

As a new currency initially created in the computer science community, bitcoin was born with some unique characteristics. Bitcoin is a currency that exists only on the internet[1]. It is accessible publicly, and anyone who is connected to the internet can create a bitcoin wallet and begin to use bitcoin. In this paper, we take some of these technical aspects of bitcoin into consideration and study how its transaction fees are determined in the bitcoin network.

In the early days, most bitcoin studies focus on technical or legal issues. However, recently bitcoin is getting more attention from economists, especially after the bullish period in 2017. Numerous studies are done focusing on the price discovery mechanism.[2] However, not much research is done on bitcoin as a payment system, which

[1]Some researchers view the bitcoin as the internet of money [2], but in this paper, we focus on the original intention of bitcoin: a digital currency.

[2]Brandvold et al. [8] is the first to study the bitcoin price discovery mechanism from the exchange level. By using the method of De Jong et al. [15] on seven exchanges, they find that MtGox and

is the original impetus of the creation of bitcoin.

In this paper, we first use the VAR method to examine the reactions to a congested network in terms of transaction volume and transaction fees. Then we try to build a simple structural model to capture what we observe in our VAR results. Our model focuses on the short-run dynamics inside the bitcoin system in response to a demand shock from the users' side. We allow the service rate from miners to be responsive to financial incentives, and, together with users' decisions, we will have a simple demand-supply model in which transaction fees and transaction volume are endogenously determined. Furthermore, we summarize three different costs in the bitcoin system and do a back-of-envelope calculation on the cost of using a payment system with a fluctuating transaction fee rate.

There are a number of related studies on the bitcoin market. Dimitri [18] models "the transaction fee as a Nash equilibrium outcome of an auction game with complete information," and the conclusion is that the optimal block size for miners depends on the distribution of users' willingness to pay. Basu et al. [6] argue that the transaction fee mechanism in the bitcoin system acts as a generalized first-price auction on multiple and identical items. Noda et al. [41] point out a potential vulnerability in the bitcoin payment system: the difficulty adjustment algorithm in the bitcoin network may fail to adjust the block arrival rate timely when a large group of miners suddenly leave the network. Auer [4] argues that the bitcoin payment system cannot be functional by solely relying on transaction fees. Once the payoff from block

---

Btce were price leaders in the early days. Kraaijeveld and De Smedt [31] find out Twitter sentiment can be used to predict the price return of bitcoin. Corbet et al. [14] also provide a comprehensive review of literature treating bitcoin as a financial asset.

rewards becomes zero, it may take months for transactions to be confirmed by miners. A counter argument is proposed by Garratt and van Oordt [21], the authors argue that because the mining flexibility is low in the bitcoin market, the fixed cost can keep miners staying in the bitcoin market even when the block reward declines. Most of these studies focus on the normative questions about the bitcoin system, like what the optimal block size should be, or how the vulnerability of this system should be dealt with. In this study, we instead take the current bitcoin transaction system as a given and study why the reactions to a congested network in terms of transaction volume and transaction fees may vary in different periods.

There are also studies using queueing theory to look into the mechanism of the Bitcoin system. Easley et al. [19] use queuing theory to study the performance of bitcoin as a transaction medium, and find out how transaction fees arise from the bitcoin network. Huberman et al. [24] embed queuing theory into their analysis of users' optimal strategies and establish a closed-form formula of users' transaction fees and waiting time given their different delaying costs. Again using queuing theory, Li et al. [33] examine the conditions under which different Nash equilibria exist in the bitcoin payment system, assuming that transaction fees preassigned to each user. Kasahara and Kawahara [27] and Kawase and Kasahara [28, 29] use a signal-server priority queue with batch service rate to model the transaction process in the bitcoin payment system and study the effect of block size on transaction-confirmation time. Kasahara and Kawahara [27] find that increased block size cannot effectively reduce transaction-confirmation time. By excluding the newly arrival transactions from the current mining block, Kawase and Kasahara [28, 29] show that increased block size can reduce the transaction-confirmation time of the high-priority group but not the low-priority group.

While queuing theory is clearly a suitable tool to understand the nature of the bitcoin system, to the best of our knowledge, the arrival rate of users and the service rate of miners in these bitcoin-related studies are assumed to be exogenous and constant for technical reasons.[3] The stable service rate from the miners' side can properly capture the supply side of the bitcoin system in the long run, given the existence of the difficulty adjustment algorithm embedded in the bitcoin blockchain protocol. However, the difficulty adjustment takes roughly two weeks, during which the supply rate is not constant. Another restriction of queuing theory is that the results only apply to the steady-state of the system, and not much is said on the adjustments of the bitcoin system. Our study focuses on the short-run dynamics inside the bitcoin system. We adopt a demand-supply model to allow the arrival rate from the users' side and the service rate from the miners' side to be both flexible, which can better capture the characteristics of the bitcoin system in the short-run.

We are also interested in the costs of employing the bitcoin transaction system in our society. Chiu and Koeppl [13] adopt the Lagos and Wright [32] model to study the welfare loss incurred by double-spending in the bitcoin system. They argue that the current bitcoin system is "generating a welfare loss of 1.4% relative to an efficient cash system," and the mining cost encouraged by the double-spending incentive is at least one of the main sources. In this study, we calculate the other costs due to

[3] Arrival rate or service rate can be time-dependent in the queuing theory, but the results are often intractable.

occasional congestion of the bitcoin system, costs that are present even without the double-spending issue.

2.2 Background

In 2008, a paper entitled "Bitcoin: A Peer-to-Peer Electronic Cash System" began to circulate on the internet under the pseudonym of Satoshi Nakamoto [40]. In this paper, Nakamoto proposes a decentralized electronic currency system, and its most innovative part is a mechanism called "Proof-of-Work" that solves the double-spend problem without a central clearinghouse. Under this system, every miner keeps an individual ledger of all the transactions, and miners reach a consensus on the state of transactions roughly every 10 minutes and then update their ledgers. Such a decentralized ledger system, compared with a central ledger system, is supposed to be more resilient to data loss and manipulation.

Based on this paper, Nakamoto created the bitcoin network in 2009. The network becomes a new way to distribute information and eliminates the need for a trusted third-party to facilitate the distribution process, and the trading parties do not need to reveal their identities to others. The anonymity helped bitcoin gain popularity beyond the computer science community, and people began to use it for transactions without revealing their identities.

In this section, we explain the technical details about bitcoin that are necessary for understanding the rest of the paper. We discuss three unique aspects of the bitcoin network that motivate this paper.

2.2.1 Mining

An important feature of mining is that, based on the computing power currently in the network, the bitcoin system automatically adjusts the difficulty of the Proof-of-Work algorithm to make sure that blocks are generated, on average, every 10 minutes. However, the adjustment is not immediate and takes 2016 blocks. After every 2016 blocks, all nodes will re-calibrate the difficulty of the Proof-of-Work algorithm based on the average mining time in the past 2016 blocks.

Due to the difficulty of the adjustment algorithm, we can reasonably assume that, when looking at the weekly or monthly horizon, the mining speed or service rate from the miners' side is relatively stable. However, when looking at a shorter period of time, we do see the mining speed changes based on how often the difficulty adjusts in a certain period (see Figure 2.1).

Another important feature of mining relates to the miners in the system. In the beginning, mining bitcoin is more or less a "hobby" to people in the computer science community. When bitcoin price took off in 2017, the competition of mining bitcoin became more intensive, as seen in the exponentially growing hash rate in Figure 2.1. It became no longer profitable to mine bitcoin using the CPU or GPU on a personal computer, and the miner community shifted to dedicated ASIC bitcoin mining equipment. Unlike a CPU or GPU, which can be used for other purposes, the mining ASIC equipment is specially designed to be used on mining bitcoins. The technology change implies that more professional miners join the bitcoin network recently, and mining bitcoin may become unprofitable for amateur miners. While we do not have direct evidence to prove the demographic change of bitcoin miners, Google trend data in Figure 2.2 on “antminer” and “bitmain” (the two biggest ASIC mining equipment producers) is indirect evidence on this mining technology change on the supply side.

2.2.2 Transaction volume

In the bitcoin system, there are two kinds of transactions: on-chain and off-chain transactions. On-chain transactions refer to the confirmed transactions stored in the blockchain, and anyone in the bitcoin network can verify them. Off-chain transactions are not stored in the blockchain, and their most common applications are Lighting Network transactions and exchange transactions. Off-chain transactions usually incur no transaction fees (e.g. transactions in the Lighting Network), or the transaction fee rate is preset by a third party (e.g. exchange transactions). However, the transaction fees in on-chain transactions fluctuate and are determined by demand (users) and supply (miners). Since the unique transaction fee mechanism is the main focus in our paper, it is natural that we only look at on-chain transactions.

Unfortunately, as the bitcoin system operates more like cash than bank accounts, it is challenging to measure precisely the volume of on-chain transactions. Suppose wallet A wants to send wallet B one bitcoin, and wallet A holds two bitcoins. The bitcoin ledger records the transaction as wallet A sending one bitcoin to wallet B and one bitcoin to wallet A itself. However, if the person owns both wallet A and C, then the ledger record may look like wallet A sends one bitcoin to wallet B and one bitcoin to wallet C, misleadingly looking like a transaction of two bitcoins.[4]

Hence, there is no foolproof way to filter out all of the "spurious" transactions and calculate the actual on-chain transaction volume. Some companies use different heuristics to parse on-chain transactions (see CoinMetrics for an example), and try to recover the accurate on-chain transaction volume. In this paper, we use the parsed on-chain transaction trade volume data compiled by CoinMetrics, but we want to remind our readers that the transaction volume used in this paper may still contain some measurement errors.

2.2.3 Transaction fees

In the bitcoin network, transaction fees have two essential roles in the transaction confirmation process:

(1) Security: transaction fees make it economically unattractive to attack the network with fake transactions.

(2) Motivation: transaction fees motivate miners to stay in the network, preventing the network from being controlled by a small group of people and becoming centralized.[5]

[4] A more detailed discussion on "spurious" transactions can be found at Tsang and Yang [48].

[5] Nowadays, miners' rewards still mostly come from Coinbase reward, which means the miner who successfully adds a block into the blockchain will be rewarded 12.5 BTC. However, Coinbase reward halves every 210,000 blocks. At this speed, approximately, the reward will become 0 in 2140, and then the only source of reward comes from the transaction fee.

Before 2015, transaction fees were never a concern to bitcoin users. In Figure 2.3, we can see that before June 2015, the average block size was always below 0.5 megabyte, which is half of the block size limit. During that period, people could even submit a transaction without offering any transaction fees, and some miners would still put the transaction into a half-empty block and get it confirmed. However, starting around 2016, a growing number of transactions needed to be confirmed, and the block size limit began to create competition among those unconfirmed transactions. From then on, people need to attach transaction fees in their transactions to motivate miners to promptly confirm their transactions.

Figure 2.4 shows that when the number of transactions waiting to be confirmed in the mempool is high, we also observe a higher average transaction fee recorded in the bitcoin network. Figure 2.5 shows that when average transaction fees are high, so is the average transaction volume. One plausible explanation for these phenomena is that users do not make a transaction decision based on the absolute level of transaction fees but the transaction fees as a percentage of the transaction amount (TFR henceforth). Users find smaller transactions more expensive, and their number decreases in the bitcoin payment system when the network is busy or when transaction fees are high. The positive correlation between average transaction fee and average transaction volume becomes not so obvious after 2018, and it becomes almost unobservable since the end of 2018. We think one plausible explanation is the gradual adoption of SegWit in the miner community. SegWit is a proposal to lift the block size limit. It enables miners to process more transactions in one block.

Due to the mining mechanism and transaction fee system, the relationship between transaction sizes and transaction fees is quite different from what we usually see in conventional payment systems. In the credit card system, the credit card company acts as a centralized clearinghouse, and sets the transaction fee unilaterally. In the bitcoin network, however, transaction sizes and transaction fees depend on how congested the network is. In the next section, we use VAR model to exam the reactions to a congested network in terms of transaction sizes and transaction fees.

2.3 Empirical Results

2.3.1 VAR with a certain ordering

In this section, we use a VAR model to study the effect of a demand shock on transaction fees and transaction volume in different periods.

To study the effect of demand shock in the bitcoin network, this VAR model consists of four variables: D_t, F_t, V_t and P_t.[6] D_t represents the transaction demand in the bitcoin network. Unfortunately, we cannot directly observe the exact demand for bitcoin transactions. Google trend and tweet data have been used as the proxy variables for bitcoin transaction demand in some studies [34, 52]. In this study, we choose to use the number of unconfirmed transactions in mempool (mempool size) to represent bitcoin transaction demand based on two reasons. The first reason is mempool size measures the materialized demand. Google trend and Twitter data

[6] Thanks to our referee's suggestion, we add bitcoin price into our model and it helps us better capture the effect of D_t on F_t and V_t.

may add more noise into the measure of demand because mass media sentiment may not eventually become real demand in the network. The second reason is that the Google trend adjusts their based line from time to time, which may bring more noise into our analysis. F_t represents average transaction fees per confirmed transaction in BTC. V_t represents the average transaction volume per confirmed transaction in BTC. P_t represents the bitcoin price in USD. Hence, the model can be written as:

$$\begin{bmatrix} a_{11}^0 & 0 & 0 & 0 \\ a_{21}^0 & a_{22}^0 & 0 & 0 \\ a_{31}^0 & a_{32}^0 & a_{33}^0 & 0 \\ a_{41}^0 & a_{42}^0 & a_{43}^0 & a_{44}^0 \end{bmatrix} \begin{bmatrix} D_t \\ F_t \\ V_t \\ P_t \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \end{bmatrix} + \begin{bmatrix} a_{11}^1 & a_{12}^1 & a_{13}^1 & a_{14}^1 \\ a_{21}^1 & a_{22}^1 & a_{23}^1 & a_{24}^1 \\ a_{31}^1 & a_{32}^1 & a_{33}^1 & a_{34}^1 \\ a_{41}^1 & a_{42}^1 & a_{43}^1 & a_{44}^1 \end{bmatrix} \begin{bmatrix} D_{t-1} \\ F_{t-1} \\ V_{t-1} \\ P_{t-1} \end{bmatrix} + ... + \begin{bmatrix} \epsilon_{D,t} \\ \epsilon_{F,t} \\ \epsilon_{V,t} \\ \epsilon_{P,t} \end{bmatrix}$$

The data comes from Coinmetrics.io and Blockchain.com, and the D_t, F_t, V_t, and P_t are all in log. Seasonal effects have been removed by adding the days of a week as exogenous variables, and the lags are chosen in different periods based on AIC.[7] We choose three lags for the early period and two lags for the rest two periods.

We split the data into three periods: the early period (from 2016-05-01 to 2017-01-01), the volatile period (from 2017-06-01 to 2018-02-01), and the recent period (from 2018-03-01 to 2018-10-01). The statistical description of data is shown in the following Table 2.1. These three periods are chosen based on two characteristics embodied in the bitcoin network: (1) From the miner side: the mining technology is evolving from general-purpose equipment (i.e. CPU, GPU) to dedicated mining equipment (i.e. ASIC). The latest expensive mining equipment makes mining bitcoin no longer

[7] Other information criterion rules give similar results, and do not change the final conclusions.

a cheap hobby for most people. We think the evolution of mining technology induces the demographic change of bitcoin miners: more professional miners joined the network recently. (2) From the user side: the uncertainty faced by users in the bitcoin network is changing, and the uncertainty is measured by bitcoin price volatility.

Based on the discussion in section 2.1, the whole industry begins to shift to relay on dedicated ASIC bitcoin mining equipment after 2017. By ending the early period before 2017, we can safely assume that the proportion of miners who mine bitcoins as a hobby is higher in the early period. Meanwhile, The volatile period is chosen by the standard deviation of the bitcoin price in USD. Based on the bitcoin historical price data on Blockchain.com, the standard deviation of the bitcoin price is \$4752.05 in the volatile period, which is much larger than that in the early period and the recent (\$108.48 and \$1209.4 respectively). Hence the early period represents a network with more amateur miners and low uncertainty; the volatile periods represents a network with more professional miners and high uncertainty; the recent period represents a network with more professional miners and low uncertainty, and is served as a benchmark to the other two periods. We also intentionally leave a gap on the time horizon to isolate these characteristics into different periods.

Due to the mempool data limitation, we do not have data earlier than 2016-05. We did not include the most recent data into our analysis due to the implementation of SegWit. SegWit is a proposal to lift the block size limit (1 MB) on the blockchain by removing the signature part out of the input section. Miners gradually begin to adopt SegWit since the 2017-07-21. However, the effect of SegWit is not so obvious before the end of 2018 because, most of the time, the average block size is still smaller than 1 MB (see Figure 2.3), which could be an indicator of a low adoption rate of SigWit among miners. However, from the end of 2018, we can observe that the average block size has become consistently larger than 1 MB. To minimize the effect of SigWit, we do not consider the data after 2018-10.

In the bitcoin network, miners are profit-oriented and tech-savvy. They monitor the bitcoin network 24/7 and respond immediately to the the network changes, which means miners will respond to the change of mempool size immediately by choosing to process more profitable transactions first. The mempool size, on the other hand, represents users' demand for transactions. When the transaction fee changes, users still need time to search for trade opportunities first. Meanwhile, if the transactions are already in the mempool, users cannot withdraw the transactions from the mempool, so the reaction from the mempool side to other shocks will be sluggish. For these reasons, we think it is reasonable to assume that transaction fees and transaction volume can be affected by a current shock to the mempool size, but the mempool size may not immediately respond to current shocks from transaction fees or transaction volume. Also, we think bitcoin price should be most sensitive to all the other shocks, so P_t can immediately react to all current shocks from other variables. Hence, we order the mempool size (D_t) first and the bitcoin price (P_t) last in our VAR model. The positions for F_t and V_t are arbitrary, and the final ordering in this main result is chosen as D_t, F_t, V_t, and P_t. In the following part, we will also show the ordering-free result. The level values of variables D_t and V_t are stationary in all three periods, F_t is stationary in the early and recent periods, but not in the volatile period, P_t is not stationary. To avoid the issue of spurious regressions, we use the first difference value of F_t in the volatile period and the first difference value of P_t across all three periods.[8]

To make the results comparable, all the shocks in the VAR results have been normalized to one-unit shocks from the number of unconfirmed transactions in mempool (D_t). First of all, we compare the responses across different periods. The impulse responses from the three periods are shown in Figure 2.6 and Figure 2.7. In Figure 2.6, we can see that, compared with the response in the recent period, the average transaction volume (V_t) experiences a period of dropping below the steady-state value 0 after the shock in the early period. Also the magnitude of the response from the average transaction fee (F_t) in the recent period is slightly larger. In Figure 2.7, the responses from V_t is slightly larger in the recent period. Also, the magnitude of the response from F_t is larger in the volatile period. In both graphs, we can see that demand shocks do not necessarily increase bitcoin price. This is understandable because the demand (D_t) we measured here is for bitcoin transactions, not for bitcoin itself. Some transaction demand shocks are caused by negative news about bitcoin, which may drive down bitcoin prices. In the following event study section, we will have further discussion on this topic. From what we observe in these two graphs, we can reach two conclusions: (1) a demand shock can induce a positive jump on the average transaction volume, and the reaction can turn into negative in the earlier period; (2) a demand shock can trigger a larger positive response from the average transaction fee in the volatile period.

[8]To make sure our results are not driven by how we treat F_t differently in the volatile period, we also use the first difference values and the level values of F_t across all three periods, the results are similar and do not change the conclusions we draw in this paper.

Then we compare the magnitudes of the responses from F_t and V_t within the same period in Figure 2.8). We can see only in the volatile period the magnitude of the response from F_t is observably larger than that from V_t. Meanwhile, as mentioned above, the negative response from V_t is much more observable in the early period.

2.3.2 VAR without a certain ordering

One concern about VAR results is that our impulse response results may be driven by the particular ordering chosen in our VAR. As mentioned in Primiceri [44], and following Diebold and Yilmaz [17] and Klößner and Wagner [30], now we try to be agnostic on the ordering and present average results obtained from all 24 possible orderings. The impulse response results are shown in Figure 2.9, Figure 2.10, and Figure 2.11. Comparing the early period and the recent period, we can see that the results are almost identical in Figure 2.9 and Figure 2.6. Figure 2.10 and Figure 2.7 show that the magnitude difference of the demand effect on average transaction fee (F_t) in the volatile period, and the recent period becomes slightly smaller in the average VAR results, even though the effect is still larger in the volatile period. Also, similar to the early period, the average transaction volume (V_t) experiences a period of dropping below the steady-state value in the volatile period. We also do not observe any noticeable demand effect on bitcoin price in the average VAR results. By comparing the magnitudes of the responses from F_t and V_t in the same period (see Figure 2.11), we still observe that, only in the volatile period, the magnitude of the responses from F_t is observably larger than that from V_t. All these average VAR results service as the robustness check showing that our main VAR results are not merely driven by a particular ordering.

2.3.3 Event Study

Beside the VAR results, we also examine events happened in the real world to see if the bitcoin market behaves as we observed in our VAR results. From the three periods mentioned in subsection 2.3.1, we pick out events that satisfied the following two criteria:

(1) The bitcoin transaction demand change should be larger than 50% on the event day.

(2) We can connect the event to a specific incident reported by media in ± 1 days.

Eventually, we find four events that roughly happened around the early period, six events for the volatile period, and six events for the recent period.[9] In these event studies, we define the inter-day movement as the log difference of a variable from its previous day. We find that the event study results are consistent with our VAR results.

Figure 2.12, Figure 2.13, and Figure 2.14 show the average event study results in the early period, volatile period and recent period respectively. The red line indicates

[9] We also observe two events that show a negative demand shock on 2017-10-25, Bitcoin Gold hard fork launch day (volatile period), and on 2018-02-15, Lunar New Year's Eve, a holiday celebrated in many Asian countries (recent period). The negative shock is associated with a drop on average transaction fees and average transaction volume. However, the data points are too sparse to draw a reliable conclusion.

the date when the event happened. The demand shocks in these three periods have been normalized to be a unit shock for easier comparison. Across these three periods, we do see a consistent pattern that a demand shock on bitcoin transactions (D_t) is associated with an increase in average transaction volume (V_t) and average transaction fees (F_t). Also, one unit demand shock is associated with a higher average transaction fee change in the volatile period than the other two periods. However, there is no clear pattern of how demand shock for bitcoin transactions affects bitcoin prices. If we dive into the individual event study figures, we can see that demand shock for bitcoin transactions can be associated with bitcoin price goes up (e.g. 2016-06-14, 2017-10-13 or 2018-07-18) or goes down (e.g. 2016-09-16, 2017-09-15 or 2018-06-27). This observation also explains why there is no noticeable effect of demand shock to bitcoin price in our VAR results. The inter-day movements in F_t, V_t and P_t for each event date and the specific incident that trigged the demand change can be found in Appendix.

Based on all the empirical results listed above, we have observed that, in response to a demand shock of bitcoin transactions (D_t), the responses from V_t and F_t behave differently in different periods. In the next section, we try to account for these differences by using a supply-demand model.

2.4 A Simple Model

As mentioned above, several studies use queuing theory to model the bitcoin transaction network [19, 24]. In such models, users usually arrive in the bitcoin network as a Poisson process with a rate λ, and the service time provided by miners is exponentially distributed with mean $1/\mu$. The models can tell us how the bitcoin network looks like when $T \to \infty$, which is the steady-state. In this paper, we leave the steady-state aside and propose a model to capture the short-run dynamics of the bitcoin network, in particular when the network is hit by a shock from the demand side. We hope this model can help us further understand the VAR results we observed in section 3.

2.4.1 Miners

Due to the bitcoin Proof-of-Work algorithm, most models assume that the service rate of miners is a Poisson process with an exogenous rate λ, which indeed is an accurate description when looking at a horizon longer than two weeks. Hence, if the time horizon we are interested in is shorter than two weeks, it is more appropriate to assume that the service rate as not determined by a fixed parameter μ. In this model, we allow the service rate to respond to transaction fees in any period.

Without loss of generality, we introduce a new concept called "effective service rate" into this model. We consider a case in which all miners join a mining pool, where the total reward is decided by the service rate provided by the mining pool. The reward is then divided based on how much computation power each miner puts into the mining pool, and by joining a mining pool, a miner can receive rewards based on how much effective service rate this miner provides, not on whether this miner successfully mined a block. The effective service rate ($q_{i,t}$) of miner i in any period

t is defined as:

$$q_{i,t} \equiv \frac{\text{Computation power from miner } i \text{ in period } t}{\text{Total computation power in the network in period } t} \times \text{Number of confirmed transactions in period } t$$

We assume no miner has market power, which means each miner's effective service supply is negligible compared with total demand and miners are price-takers. We can write down the individual supply function of miner i as:

$$q_{i,t} = f_t^{\alpha_i}$$

where $q_{i,t}$ is the effective service rate of miner i in period t, f_t is the transaction fees in period t, and α_i describes supply elasticity.

In Section 2.1, we point out that miners in the bitcoin network have changed from treating mining as a "hobby" to participating as professionals. Likewise, there are two kinds of miners in the model: professional miners and amateurs. For professional miners, their individual supply function is $q_{1,t} = f_t^{\alpha_1}$, and for amateurs, their individual supply function is $q_{2,t} = f_t^{\alpha_2}$. We assume $\alpha_1 > \alpha_2$, meaning that supply from professional miners is more flexible or price elastic.

If there are A_1 professional miners and A_2 amateurs, then market supply can be written as:

$$S'_t = A_1 f_t^{\alpha_1} + A_2 f_t^{\alpha_2}$$

However, S'_t only captures the financial incentive through transaction fees. The bitcoin system also rewards miners for providing their service, and this block reward is fixed in a short period.[10] Even without transaction fees from users, miners are still willing to provide a certain level of service for the rewards from the system. A more realistic way to write down market supply is:

$$S_t = S + A_1 f_t^{\alpha_1} + A_2 f_t^{\alpha_2} \tag{2.1}$$

where S is the fixed service rate provided by miners when there is no financial incentive from users, S_t is the total service rate provided by miners.

The elasticity of market supply is:

$$E_t = \frac{\partial S_t}{\partial f_t} \cdot \frac{f_t}{S_t} = \frac{\alpha_1 A_1 f_t^{\alpha_1} + \alpha_2 A_2 f_t^{\alpha_2}}{S + A_1 f_t^{\alpha_1} + A_2 f_t^{\alpha_2}}$$

The elasticity of the market supply will be in between α_2 and α_1, and its value is determined by A_1 and A_2, meaning that the ratio of the two kinds of miners in the network indirectly affects the elasticity of the market supply curve.

[10] The block reward in the bitcoin network halves every 210,000 blocks. A countdown for bitcoin halving can be found at https://www.bitcoinclock.com/.

2.4.2 Users

Another feature that is missing in queuing theory concerns the users (the demand side). In those models, a user may choose to stay in a queue indefinitely until she is served. However, when a user submits a transaction into the bitcoin network, it will first stay in a mempool. Every miner maintains his own mempool, which is costly to him, and as a result, a transaction may be cleared out of the mempool if it has stayed for too long. Typically, an unconfirmed transaction is allowed to stay in a mempool for less than a week.[11]

In our model, the bitcoin network exists for a total of T periods, and everybody knows that. To capture the mempool feature, we impose an "overlapping generations" structure that each transaction has to be confirmed either in the current or the next period. After that, it will be rejected from the bitcoin network. Users in the first $T-1$ periods can choose to finish a transaction in the current period, next period, or just leave the network. Users in the last T^{th} period can only process a transaction in the current period or leave the network.

To simplify the model, in each period, the transaction sizes of all the transactions satisfy a uniform distribution with the range $\bar{R} = (0, 1]$, and there is a continuum of users of mass Q_t, which is exogenous. We assume users can choose to do their transactions in their current period and gain utility $R_i - f_t$, postpone their transactions into next period and gain utility $\beta(R_i - f_{t+1})$, or leave the network and gain zero

[11] The bitcoin protocol does not specify how long an unconfirmed transaction can stay in a mempool, and the decision is made by miners themselves. The time limit is approximately 2 days [7].

utility. Here R_i is user i's transaction volume, the f_t is the equilibrium transaction fee at time t, and β is the discount rate. Suppose in each period t, the user who has a transaction volume R_t is indifferent between doing the transaction in the current period and postponing the transaction into the next period, and the transaction fee f_t paid by this user becomes the equilibrium transaction fee in the bitcoin network at time t.

In the T^{th} period, users behave as follow:

$$R_T = f_T \tag{2.2}$$

$$S + A_1 f_T^{\alpha_1} + A_2 f_T^{\alpha_2} = \begin{cases} (R_{T-1} - f_T)Q_{T-1} + (1 - R_T)Q_T & if \quad R_{T-1} > f_T \\ (1 - R_T)Q_T & if \quad R_{T-1} \leq f_T \end{cases} \tag{2.3}$$

Equation (2) says that, in the last period, no one can postpone transactions into the next period. The fees f_T directly determine that only the users whose transaction volume is larger than f_T will be willing and able to do transactions in the period T. Equation (3) says that if $R_{T-1} > f_T$, the service provided by miners will satisfy the demand from users who postponed their transactions from period $T-1$ and also the demand from users in period T who are willing and able to afford f_T. Moreover, only $(R_{T-1} - f_N)Q_{T-1}$ users from period $T-1$ will get their transactions confirmed in period T, the rest will be cleared out from the system. If $R_{T-1} \leq f_T$, all postponed transactions from period $T-1$ will be removed from the system.

For the second to the $T-1^{th}$ periods, users make decisions based on the following two rules:

$$R_t - f_t = \begin{cases} \beta(R_t - f_{t+1}) & if \quad R_t > f_{t+1} \\ 0 & if \quad R_t \leq f_{t+1} \end{cases} \tag{2.4}$$

$$S + A_1 f_t^{\alpha_1} + A_2 f_t^{\alpha_2} = \begin{cases} (R_{t-1} - f_t)Q_{t-1} + (1 - R_t)Q_t & if \quad R_{t-1} > f_t \\ (1 - R_t)Q_t & if \quad R_{t-1} \leq f_t \end{cases} \tag{2.5}$$

β in equation (4) represents the uncertainty of postponing transactions into the next period. In the bitcoin network, the value of β is primarily decided by how unpredictable the purchasing power of bitcoin is. This unpredictability can be measured by the volatility of bitcoin price in USD. When the volatility is low, then the purchasing power of bitcoin is more predictable, which means β is large, and people feel more comfortable to postpone their transaction into the next period.

Equation (4) says that if $R_t > f_{t+1}$, some users will choose to postpone their current transaction to the next period. The number of users who will postpone depends on how large the β is. When β is larger, users are more patient, and more users will postpone their transactions. If $R_t \leq f_{t+1}$, users anticipate that transaction fees in the next period will be higher, and all transactions smaller than f_t will eventually be removed from the bitcoin network.

In the first period, users follow:

$$R_1 - f_1 = \begin{cases} \beta(R_1 - f_2) & if \quad R_1 > f_2 \\ 0 & if \quad R_1 \leq f_2 \end{cases} \tag{2.6}$$

$$S + A_1 f_1^{\alpha_1} + A_2 f_1^{\alpha_2} = (1 - R_1)Q_1 \tag{2.7}$$

Equation (7) shows that, in the first period, miners only serve users who intend to do transactions in the current period. Eventually, the average transaction volume per transaction (V_t) in each period can be written as in equation (8).

$$V_t = \begin{cases} \frac{\frac{R_{t-1}+f}{2}(R_{t-1}-f_n)Q_{t-1}+\frac{1+R_t}{2}(1-R_t)Q_t}{(R_{t-1}-f_t)Q_{t-1}+(1-R_t)Q_t} & if \quad R_{t-1} > f_t \\ (1+R_t)/2 & if \quad R_{t-1} \leq f_t \end{cases} \quad t = 1, 2, ...T \tag{2.8}$$

Given equations (2) to (8), we can calculate the equilibrium average transaction volume V_t and the transaction fee f_t for any period t. However, to solve how the two variables change when the system is hit by demand shock to Q_t, closed-form solutions are hard to come by. To understand how the model works, we instead do some simple simulations in the following section.

2.5 Simulation Results

2.5.1 The evolution of the bitcoin network

We choose $T = 10$, corresponding to a period shorter than 2 weeks. In addition, we choose $S = 25$ and $Q_t = 25$ for $t \neq 2$, so that in the steady-state, transaction fees are zero and all transactions will not be postponed to the next period. The average transaction volume per transaction (V_t) is calculated as in equation (8), and since the steady-state V_t is 0.5, to bring it down to 0, we use $V_t - 0.5$ instead. The V_t curve shows the responses from the average transaction volume to the shock from demand Q_t, and the f_t curve shows the responses from the transaction fees to the shock from demand Q_t. Lastly, we choose $\alpha_1 = 2$ and $\alpha_2 = 0.5$ to differentiate the professional miners and amateurs.

Now suppose there is a demand shock in period 2 of $Q_2 = 50$. By choosing different values of A_1, A_2, and β, we try to describe the bitcoin network under different situations. As mentioned in section 3, we intentionally choose three discrete periods to ensure each period can preserve different characteristics. Compared with the rest two periods, a larger portion of miners mine bitcoin as a hobby in the early period, so the ratio of professional miners to amateur miners will be low in this period. Hence, we choose $A_1 = 3, A_2 = 15$ to simulate the early period, and $A_1 = 15, A_2 = 3$ for the rest two periods. In the volatile period, people face higher uncertainty for postponing transactions into the next period, so we choose $\beta = 0.2$ to simulate the volatile period, and $\beta = 0.7$ for the other two periods. The numeric values of these parameters are arbitrary. We are not trying to match the absolute magnitudes of the VAR results. What we aim at are the changing directions and the relative magnitude changes among different variables. In particular, we focus on the ratio of A_1 to A_2 and β, and see how they contribute to the different responses we observed in the VAR results in different periods.

Comparing the responses across different periods: The simulation results corresponding to the VAR results can be found in Figure 2.15 and Figure 2.16. The simulation results for the early period and the recent period match the corresponding VAR results (Figure 2.6): the response from V_t switches to negative in the early period. This means many users choose to postponed transactions into their next period in the model, which drives down the average transaction volume. Also, in the simulation results, the respond from f_t is slightly larger in the recent period, which matches what we observe in the VAR results. The simulation results for the volatile period and the recent period also match the corresponding VAR results (Figure 2.7) : the magnitude of the responses from V_t is slightly larger in the recent periods, but the response from f_t is much stronger in the volatile period.

Comparing the responses within each period: The simulation results corresponding to Figure 2.8 can be found in Figure 2.17. We can see the magnitudes of the responses from V_t and f_t are almost the same in the early period and the recent period. However, the response from f_t in the volatile period is much larger than the response from V_t. These phenomena are also observed in the VAR results.

After matching the simulation results with the VAR results, the following question is what causes the different VAR results in different periods. In the following part, we will distinguish the roles of the composition of miners (ratio of A_1 to A_2) and users' patience (β) in our simulation model.

2.5.2 Comparative statics

The results above can be traced back to the changes in two parts of our model, one relates to the change in users' patience, and one relates to the change in the composition of miners.

Effect of β: The most obvious effect we can observe is that low β significantly increases the magnitude of the response from transaction fees. When people are more impatient, or when the future is more uncertain, the utility of postponing a transaction into the next period is almost 0. So users who have large transactions to make are willing to pay more to finish their transactions now.

Effect of the ratio of A_1 **to** A_2: When the ratio of A_1 to A_2 decreases, the most obvious effect is that V_t can drop significantly below 0 in the early period. This means more users will choose to postpone their transactions when there are more amateur miners and less professional miners in the bitcoin network. We think this is due to it is less costly to scale up service rates when there are more professional miners because they are more responsive to financial incentives. So it becomes reasonable to attract some users who initially will postpone their transactions into the next period to process their transactions in the current period by increasing the transaction fee a little bit. An increase in transaction fee will further deter users from postponing their transactions into the next period because they know the users in the next period are also willing to pay more transaction fee to boost up the service rate, so we do not observe a big drop on V_t when there are more professional miners (A_1) than armature miners (A_2). The unwillingness of postponing transactions drives up the fee and makes it only profitable for fewer transactions, that is why we can see the average transaction volume and transaction fee are higher in the recent period compared with that in the early period.

Based on the above analysis, we can see that: (1) the willingness of postponing transactions into the next period, which can be observed as a drop of V_t in the early period, is mainly caused by the low ratio of A_1 to A_2; (2) even though more professional miners in the network will deter users from postponing their transactions and drive up the transaction fee to some extent, the overreaction from f_t in the volatile period is mainly driven by a low β.

Till now, we have discussed the demand-supply mechanism behind the fluctuating transaction fee in the bitcoin system and explored the factors that can affect the transaction fee. In the next section, we try to measure the effect of this transaction fee system on bitcoin users.

2.6 Three Costs of Using the Bitcoin System

In the bitcoin network, every transaction has to be confirmed by miners. The price of confirming a transaction (transaction fee) is decided by both the demand side (users) and the supply side (miners). When the bitcoin network is congested, a demand shock can drive up the transaction fee. This unique transaction-fee-decision mechanism incurs three different kinds of costs when bitcoin is used as a payment system.

2.6.1 The transaction fee rate

At the end of 2017, the high transaction fee becomes a major concern inside the bitcoin community [9]. However, we argue that the transaction fee rate rather than transaction fee itself is a better index to measure how expensive a transaction network is. In Figure 2.18, we plot the daily transaction fee rate from 2016-01-01 to 2019-03-01 in the bitcoin network. We can see the transaction fee rates are actually relatively low. Even the highest rate at the end of 2017 is still lower than 0.1%. As a comparison, in 2019, the average credit card transaction fee rates in the U.S. range from 1.7% to 3.5% [43]. Hence, a congested bitcoin network may drive up the transaction fee rate. However, given the current historical data, the bitcoin network is still much cheaper than those conventional payment networks, like Visa and MasterCard.

2.6.2 The uncertainty of the transaction fee rate

Unlike a centralized transaction network, where transaction fee rate is pre-decided, the the transaction fee rate in the bitcoin network is decided by the supply-demand mechanism, which makes the transaction fee rate fluctuating. The uncertainty of the transaction fee rate imposes another cost on the bitcoin payment system.

By borrowing the approach from Lucas [36], we try to evaluate the potential cost of having a fluctuating transaction fee rate to a payment system. Consider a representative consumer is living in a world using bitcoin. Suppose the consumer's real consumption (C_t) satisfies lognormal distribution, with $log(C_t) \sim N(\mu, \sigma^2)$ and $E(C_t) = e^{\mu + \frac{\sigma^2}{2}}$. As in Lucas, the consumer's utility function of consumption is of CRRA form:

$$U(C_t) = \frac{C_t^{1-\gamma}}{1-\gamma} \qquad \text{with } \gamma \neq 1$$

where γ is the coefficient of risk aversion.

The expected utility of the consumer's real consumption is

$$E(U(C_t)) = \frac{e^{\mu(1-\gamma) + \frac{\sigma^2}{2}(1-\gamma)^2}}{1-\gamma} = U(e^{\mu + \frac{\sigma^2}{2}(1-\gamma)}).$$

Hence, the utility of the real consumption C_t in the bitcoin network equals to a certain consumption of $C_{eq} = e^{\mu + \frac{\sigma^2}{2}(1-\gamma)}$. The latter can be interpreted as the consumption level in an ideal payment system without transaction fees. Thus, the ratio of the consumption (λ) that the consumer is willing to give up to switch from this bitcoin system to this ideal transaction system is:

$$\lambda = \frac{E(C_t) - C_{eq}}{E(C_t)} = 1 - e^{-\frac{\sigma^2 \gamma}{2}}$$

when $\frac{\sigma^2\gamma}{2}$ is very small, then $\lambda \approx \frac{1}{2}\gamma\sigma^2$.

To estimate how large the λ can be, we need to find out the variance of C_t. Suppose the transaction fee rate is a random variable $r_{f,t}$. The representative consumer's income is W, and the consumer spends all the income, then: $C_t = (1-r_{f,t})W$. Then the variance of consumption is

$$\sigma^2 = Var(log(C_t)) = Var(log(1-r_{f,t}) + log(W)) = Var(log(1-r_{f,t})). \tag{2.9}$$

Equation (9) shows that the variance of C_t is only related to the variance of transaction fee rate, $r_{f,t}$. This conclusion also reconciles with the TFR rule discussed in subsection 2.2.3: people do not care about the absolute value of the transaction fee itself. What people care about is the transaction fee rate.

Using the whole bitcoin historical data, we can calculate that $\sigma^2 = 1.096e^{-8}$ in the bitcoin payment system. Even when we shrink the time window to only include the most fluctuating period, like from 2017-08-01 to 2018-04-01, $\sigma^2 = 2.815e^{-8}$.

If the consumer is extreme risk-averse, like $\gamma = 4$, in the utility function, and given $\sigma^2 = 2.815e^{-8}$, we can calculate that $\lambda = 5.630e^{-8}$. This means the cost of enduring the uncertain transaction fee rate in the bitcoin system only counts as $5.63e^{-6}\%$ of the consumer's consumption, which is negligible.

Now compare the bitcoin transaction system with the credit card system. If the consumer chooses to use credit card system, then the processing fee rate is certain, so the effective consumption in credit card system can be written as $C_{effective} = (1-r_{cc})e^{\mu}$, where r_{cc} is the flat processing fee rate charged by credit card companies and $\mu = log(W)$. In the bitcoin system, if the expected transaction fee rate is r_{btc}, then we can get $\mu' = log((1-r_{btc})W) = log(1-r_{btc}) + \mu$ and $C'_{eq} = e^{\mu' + \frac{\sigma^2}{2}(1-\gamma)}$.

$$C'_{eq} = C_{effective} \Rightarrow \sigma^2_{eq} = \frac{2ln(\frac{1-r_{cc}}{1-r_{btc}})}{1-\gamma} \tag{2.10}$$

Equation (10) shows how fluctuating the transaction fee rate in the bitcoin payment system could be that the representative consumer would still be indifferent between the bitcoin payment system and the credit card system. Based on previous discuss about the average credit card processing fees in the US, we assume the fee rate is 1% in the credit card system ($r_{cc} = 0.01$). And the representative consumer is extreme risk-averse ($\gamma = 4$). To make the argument strong, we consider the worst scenario in bitcoin system by choosing $r_{btc} = 0.001$, then we can get the condition for this consumer becomes indifferent between these two payment systems is $\sigma^2_{eq} = 6.03e^{-3}$. However, from above, we can see that, given the current transaction fee in the bitcoin system, the variance of consumption is $\sigma^2 = 2.815e^{-8}$, which is much smaller than $6.03e^{-3}$.

Hence, consumers may need to deal with a fluctuating transaction fee rate if they choose to use the bitcoin payment system. However, given the current historical data, the cost of the uncertainty from fluctuating transaction fee rates is still much smaller than the cost of using the conventional transaction networks, like Visa and MasterCard.

2.6.3 The cost of the crowding-out effect

In subsection 2.2.3, we talked about the TFR rule, which says people make transaction decisions based on how large the percentage of the current transaction fee will be given the sizes of their potential transactions. This rule implies that, when the transaction fee is high, only those big-size transactions can still maintain a reasonable transaction fee rate. The preference for big-size transactions during high-transaction-fee periods can be called the crowding-out effect. This crowding-out effect is caused by both users and miners. When the bitcoin network is congested, miners prefer to process the transactions which pay the highest fee. Meanwhile, the users realize that only when the sizes of their transactions are big enough can a high transaction fee be justified, which leads to the crowding out of those small-size transactions. The crowding-out effect may be a plausible explanation for what we observed in our VAR results that transaction fee and transaction volume become higher during the high demand period, and the bitcoin network is congested.[12] If crowding-out effect exists, it may cause two parts of costs that we failed to capture: (1) the crowding-out effect can block out a group of users who are willing to pay low transaction fees. This is the direct cost of the crowding-out effect. (2) We may underestimate the standard deviation of the transaction fees users are willing to pay. The crowding-out effect may stabilize the transaction fee rate observed in the historical data and cause us to underestimate the cost of having a fluctuating transaction fee system.

[12] Thanks to the referee for pointing out that there are other plausible explanations for the higher transaction fee and transaction volume during a high demand period. For example, there may be more trading and less payments activity during the high demand periods and all the trading involves higher transaction fee and transaction volume.

Unfortunately, due to the lack of data, we cannot directly measure the cost of the crowding-out effect.

In this section, we talk about three different kinds of costs associated with the bitcoin system. Using historical data, we find these costs are not unbearable compared with the conventional payment networks, like Visa and MasterCard. However, our back-of-envelope calculation may underestimate the costs because of the crowding-out effect.

2.7 Conclusion

In this paper, we study bitcoin as a payment system with fluctuating transaction fees. We find that a demand shock to bitcoin transactions is associated with higher average transaction fees and average transaction volume. This phenomenon is most obvious in the volatile period. We then propose a simple model to capture the behavior of users and miners in the bitcoin network, and find that the bigger effect of demand shock on average transaction fee in the volatile period may be caused by users' impatience in that period. More professional miners in the network can also deter users from postponing their transactions and drive up the average transaction volume. Furthermore, using historical data, we estimate the disadvantage of using a payment system with fluctuating transaction fee rates. We find that such cost of using the bitcoin system is negligible. However, our calculation does not consider the crowding-out effect on small volume transactions, which may impose an extra cost to certain group of users during the congested period.

This paper has several limitations. First, the trade volume data used in the VAR model are probably measured with error, as discussed in the paper. Second, if the bitcoin network can solve the scalability problem and significantly increase the number of transactions that can be processed in a given period, our current calculation based on the historical data may underestimate the variance of the transaction fee rate and also the cost of using the future bitcoin system. Third, in our model, the number of miners and the ratio of professional miners to amateur miners are exogenous, ignoring entry and exit of miners.

2.8 Tables

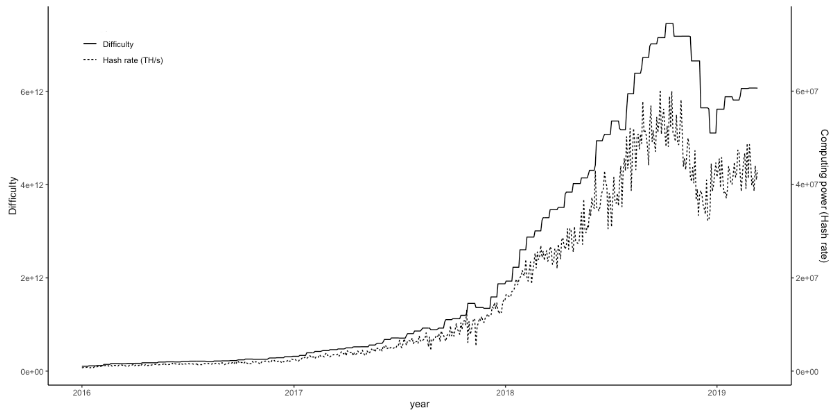
Table 2.1: Summary statistics

	Early period				Volatile period				Recent period			
	D_t	F_t	V_t	P_t	D_t	F_t	V_t	P_t	D_t	F_t	V_t	P_t
Mean	8.38	-8.15	1.89	6.45	9.89	-6.88	1.85	8.61	7.53	-8.96	1.13	8.92
Med	8.31	-8.15	1.88	6.46	10.08	-6.93	1.87	8.42	7.58	-8.97	1.12	8.89
Max	11.00	-7.50	3.16	6.88	12.11	-5.54	3.79	9.88	9.97	-6.89	1.72	9.35
Min	5.27	-8.63	1.13	6.08	5.08	-7.77	1.17	7.57	4.16	-9.64	0.55	8.68
S.D.	0.79	0.17	0.36	0.17	1.28	0.51	0.36	0.17	1.04	0.38	0.23	0.66
Obs.	246	246	246	246	246	246	246	246	215	215	215	215

This table reports the descriptive statistics of the data used in the three periods: the early period (from 2016-05-01 to 2017-01-01), the volatile period (from 2017-06-01 to 2018-02-01), and the recent period (from 2018-03-01 to 2018-10-01). D_t represents the unconfirmed transactions in the mempool, f_t represents the average transaction fee, V_t is the average transaction volume, P_t is the bitcoin price in USD. The data comes from Coinmetrics.io and Blockchain.com, and the variables are all in log.

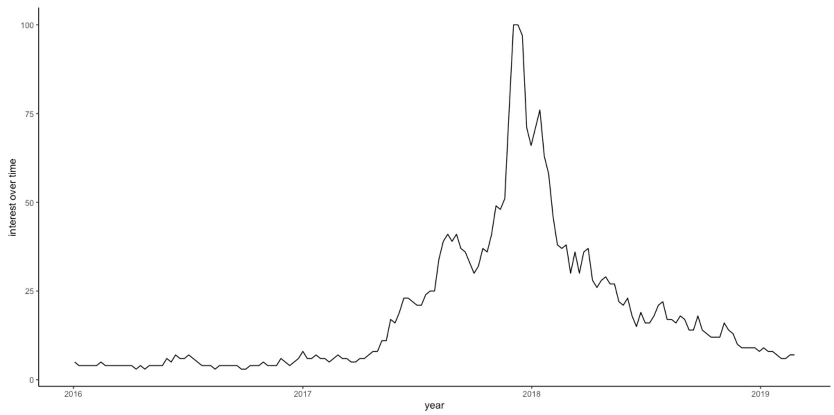
2.9 Figures

Figure 2.1: Difficulty and computing power (hash rate) since 2016



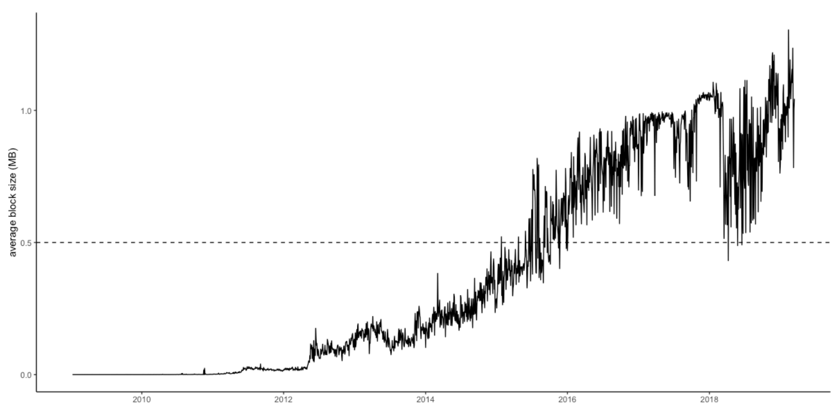
The data comes from blockchain.com. This figure shows the change of difficulty and computing power in the bitcoin network since 2016. The difficulty will be re-calibrated by the Proof-of-Work algorithm after mining every 2016 blocks to match with the computing power in the network. The computing power measures how much computing resource has been devoted into the network to mine blocks.

Figure 2.2: Searching popularity of "antminer" and "bitmain"



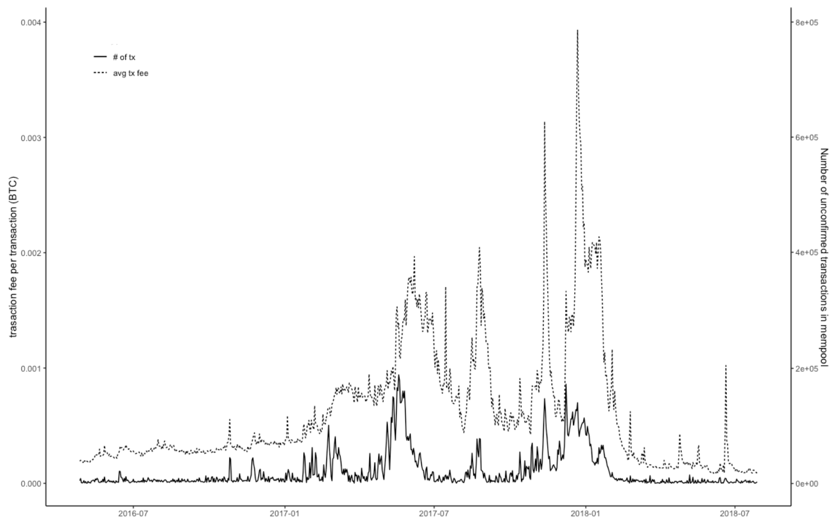
The data comes from the Google Trend. The Google trend data shows how many inquires to Google are related to "antminer" and "bitmain", which are the brands of the two biggest ASIC mining equipment producers. This figure shows the dynamic change of the popularity of "antminer" and "bitmain" since 2016.

Figure 2.3: Average daily block size



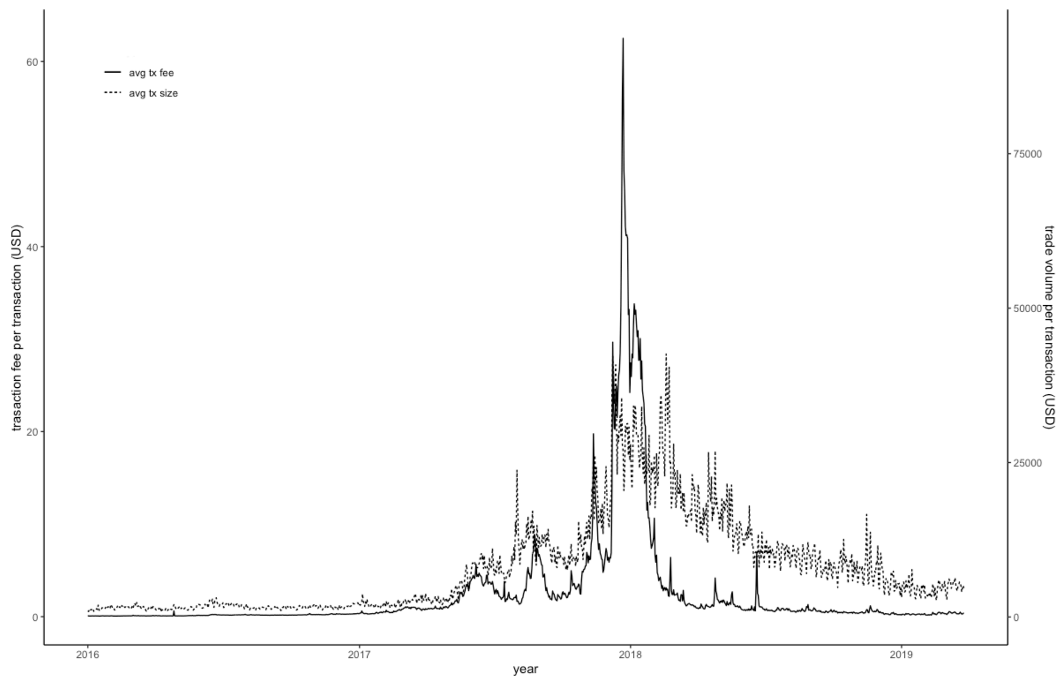
The data comes from blockchain.com. This graph shows the average block size on daily basis since the beginning the bitcoin network. Originally, the block size limit is set to be 1 MB by the bitcoin algorithm. We can see that, before June 2015, the blocks are always less than half full in the bitcoin network.

Figure 2.4: Number of transactions in *mempool* and average transaction fees



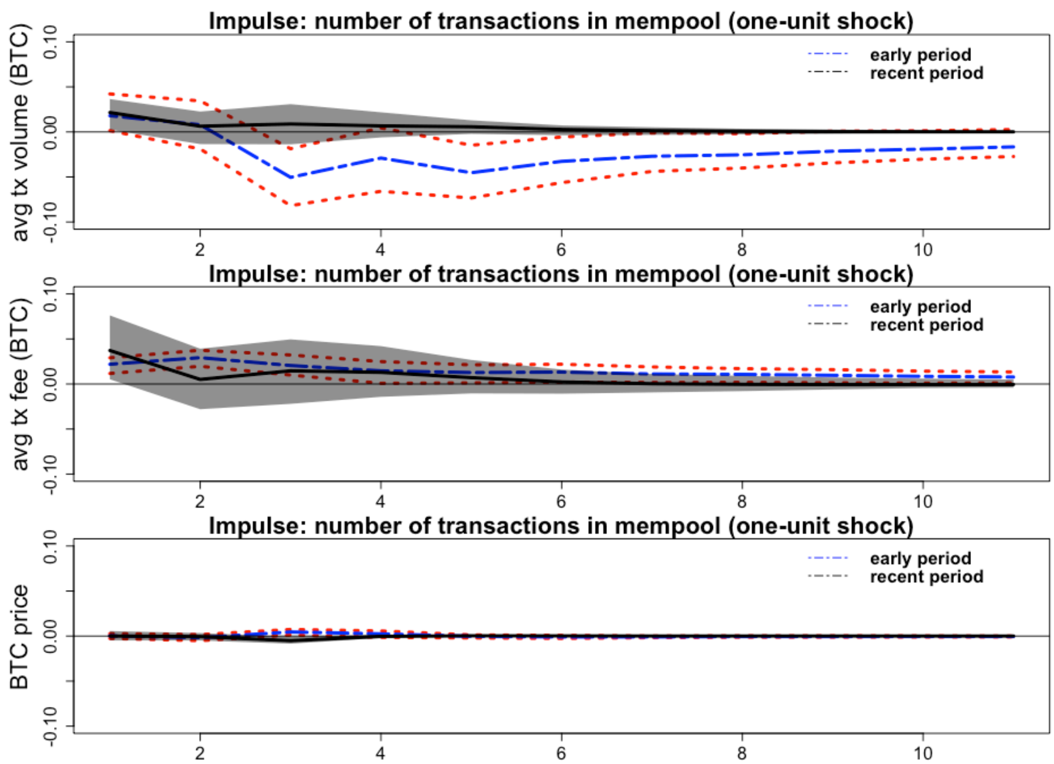
The data comes from blockchain.com. This graph shows the relationship between the number of transactions in mempool (unconfirmed transactions) and the average transaction fees paid in those confirmed transactions in the bitcoin network. When the number of unconfirmed transactions is high, we can observe an increase in transaction fees in those confirmed transactions.

Figure 2.5: Average transaction volume and average transaction fees



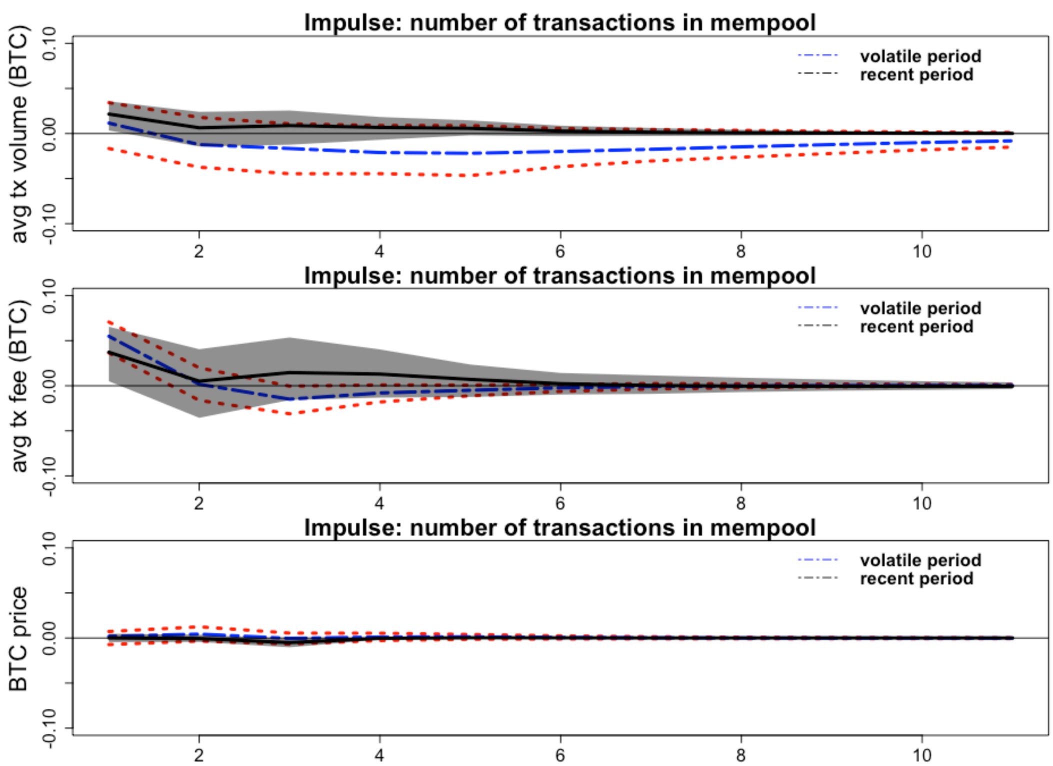
The data comes from blockchain.com. This graph shows the relationship between the average transaction volume and average transaction fees. We can see that when transaction fees are high, so is the average transaction volume. This relationship becomes not so obvious since 2018, and it becomes almost unobservable since the end of 2018. One plausible explanation is the gradually adoption of SegWit inside the miner community.

Figure 2.6: Impulse responses in the early period and the recent period



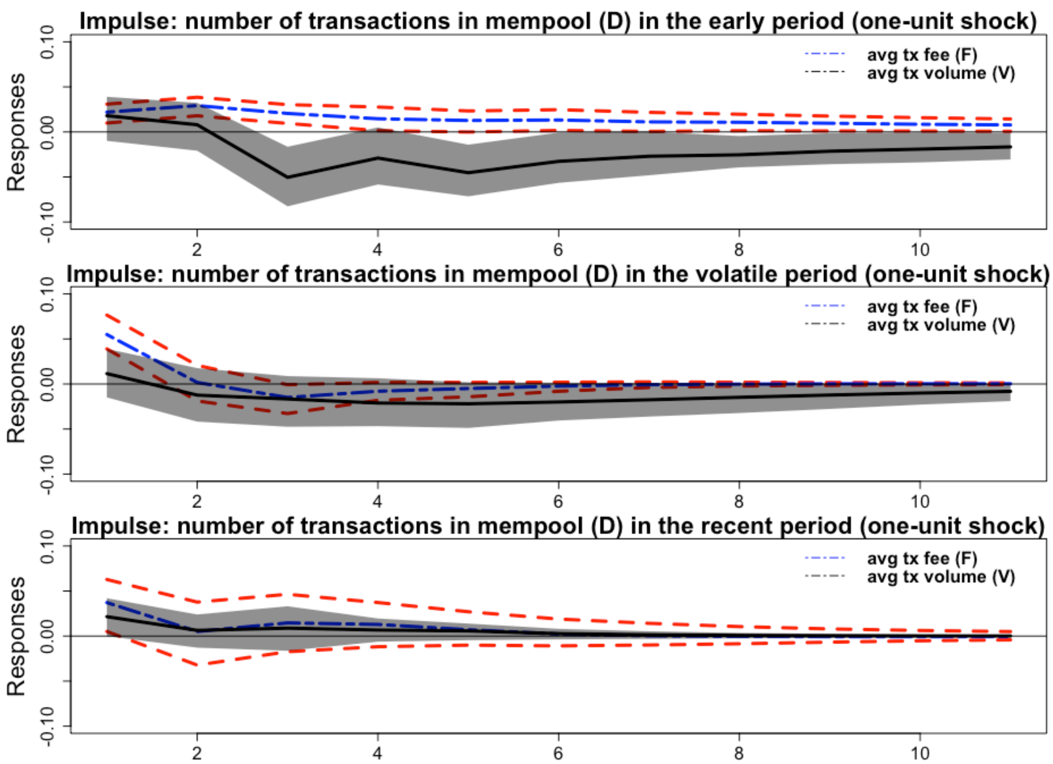
This graph compares the responses to a one-unit shock from demand in the early period and the recent period. Compared with the responses in the recent period, we cans see that V_t experiences a period of dropping below the steady-state value 0 after the shock in the early period. The magnitude of the response from F_t is slightly larger in the recent period.

Figure 2.7: Impulse responses in the volatile period and the recent period



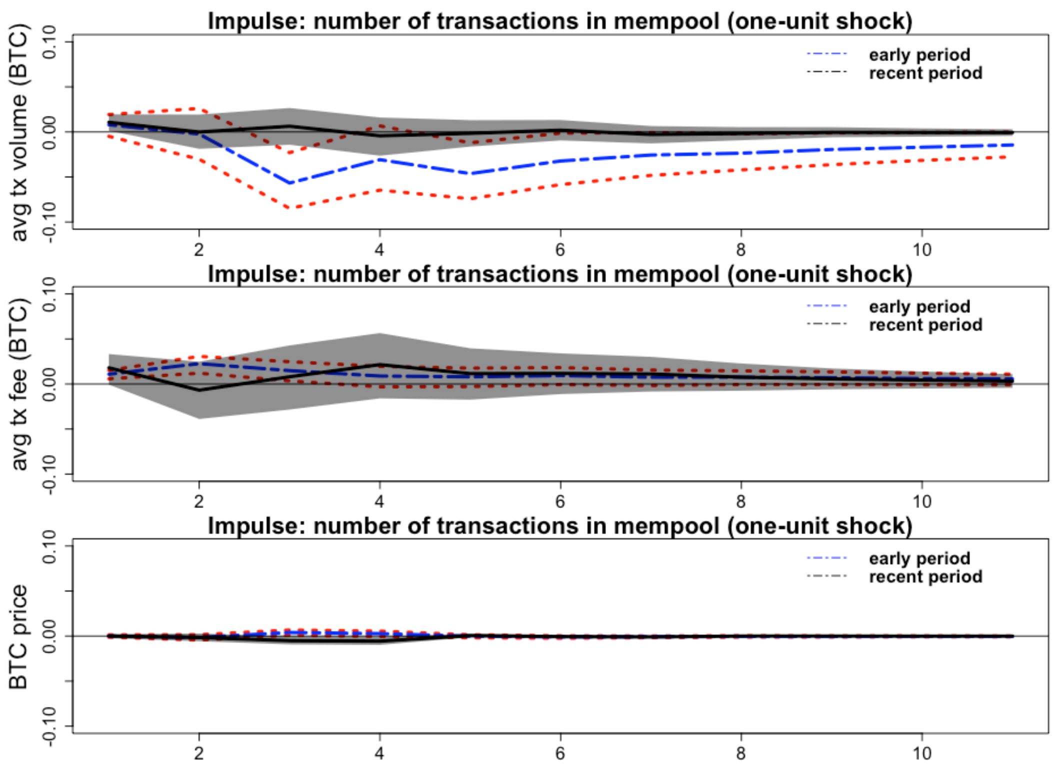
This graph compares the responses to a one-unit shock from demand in the volatile period and the recent period. The responses from the V_t is slightly larger in the recent period. However, the magnitude of the response from F_t is larger in the volatile period.

Figure 2.8: Impulse responses in the three periods



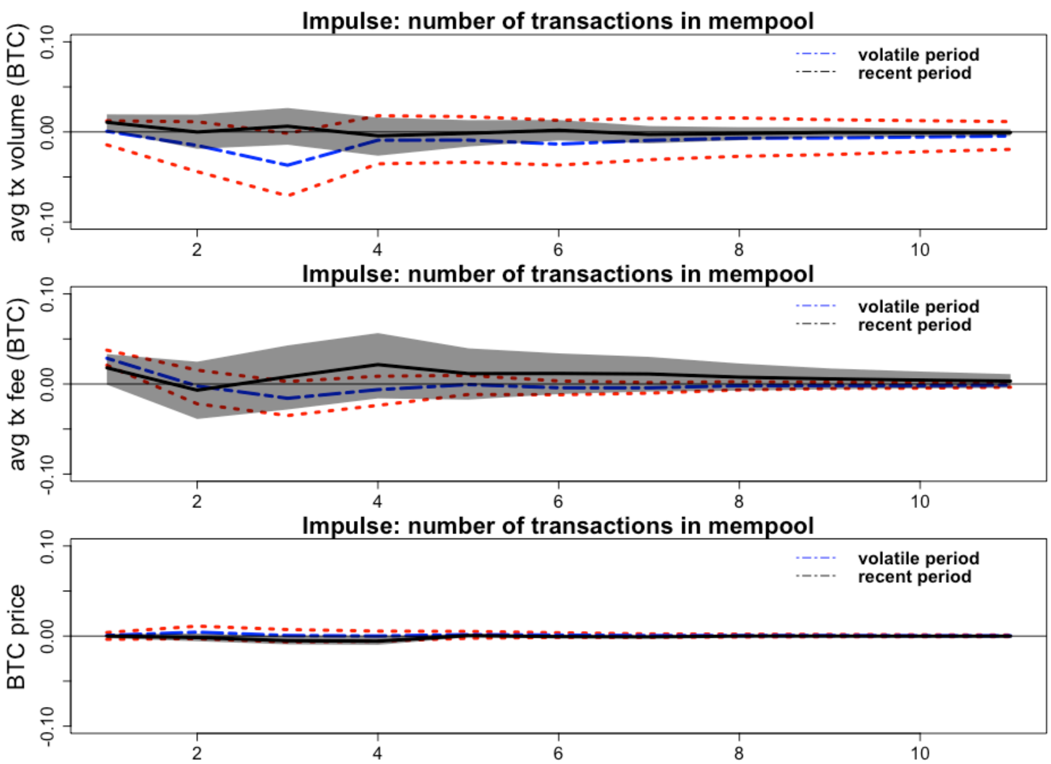
This graph compares the magnitudes of the responses from F_t and V_t within the same period. We can see only in the volatile period that the magnitude of the response from F_t is obviously larger than that from V_t. Meanwhile, the response from V_t turns into negative after the initial positive response is much more observable in the early period.

Figure 2.9: Average impulse responses in the early period and the recent period



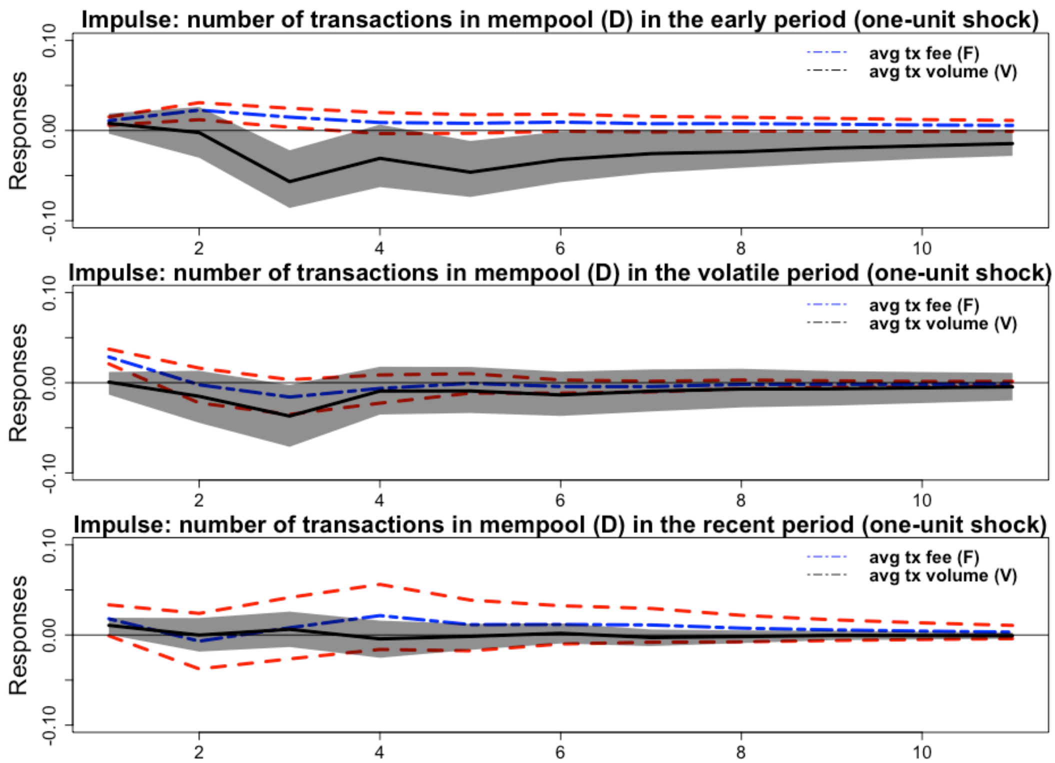
This graph compares the responses to a one-unit shock from demand in the early period and the recent period. Compared with the responses in the recent period, we can see that V_t experiences a period of dropping below the steady-state value 0 after the shock in the early period. The magnitude of the response from F_t is slightly larger in the recent period.

Figure 2.10: Average impulse responses in the volatile period and the recent period



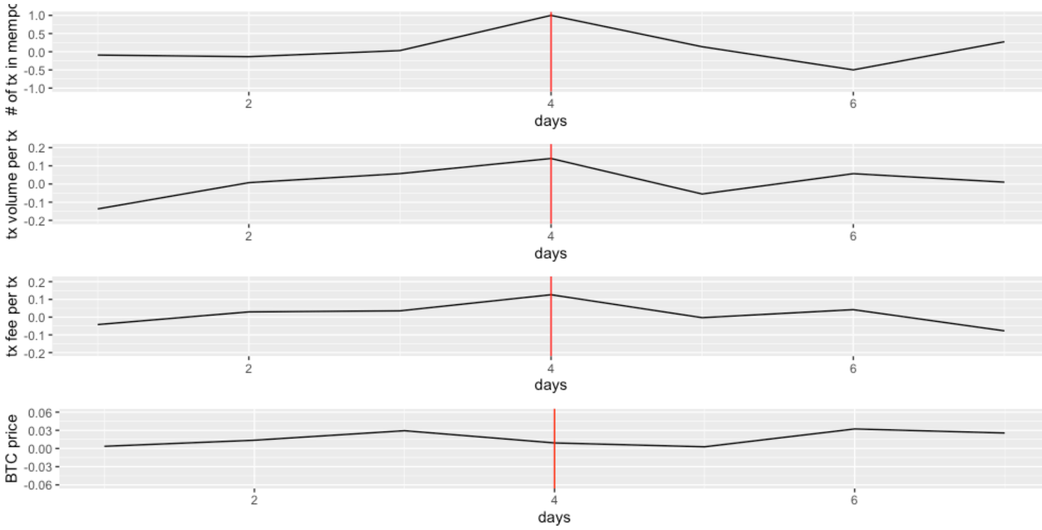
This graph compares the responses to a one-unit shock from demand in the volatile period and the recent period. Compared with the responses in the recent period, we can see that V_t experiences a period of dropping below the steady-state value 0 after the shock in the volatile period. However, the magnitude of the response from F_t is larger in the volatile period.

Figure 2.11: Average impulse responses in the three periods



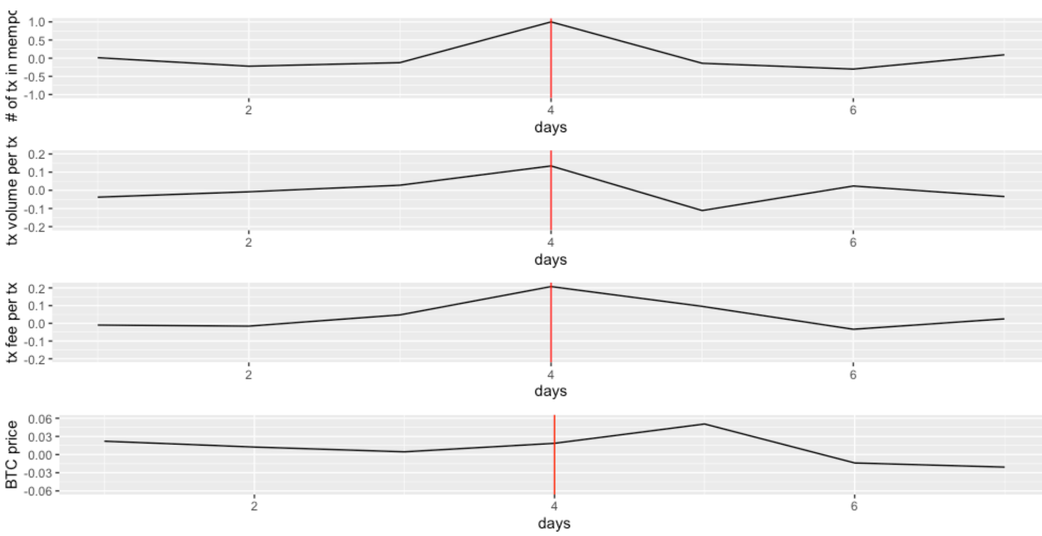
This graph compares the magnitudes of the responses from F_t and V_t within the same period. We can see only in the volatile period that the magnitude of the response from F_t is observably larger than that from V_t. Meanwhile, the initial positive response from V_t turns negative in the following up periods becomes observable in the volatile period, even though it is still more obvious in the early period.

Figure 2.12: Event study in the early period



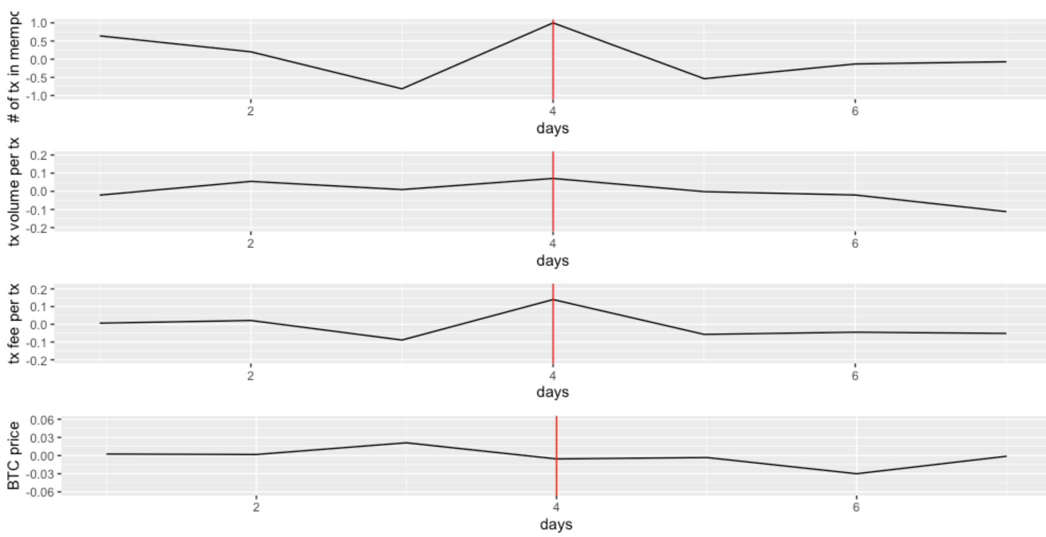
This graph shows the average event study result in the early period, the event dates includes in the early period are 2016-05-27, 2016-06-14, 2016-09-16, and 2016-10-09. The specific incidents happened in these dates can be found in Appendix.

Figure 2.13: Event study in the volatile period



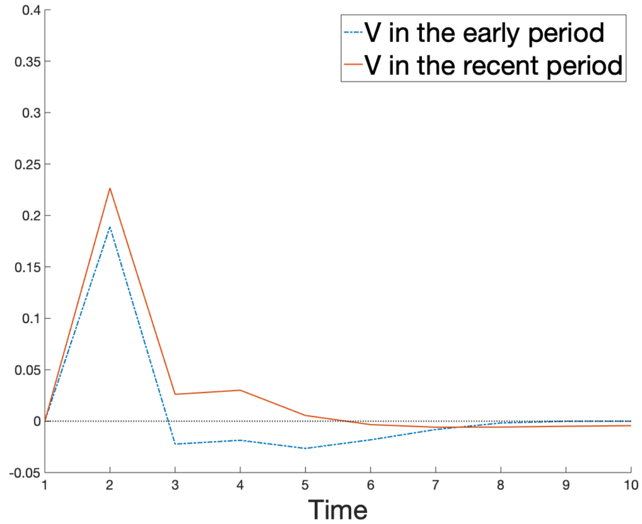
This graph shows the average event study result in the early period, the event dates includes in the early period are 2017-05-09, 2017-08-14, 2017-09-15, 2017-10-12, 2017-12-07, and 2017-12-20. The specific incidents happened in these dates can be found in Appendix.

Figure 2.14: Event study in the recent period

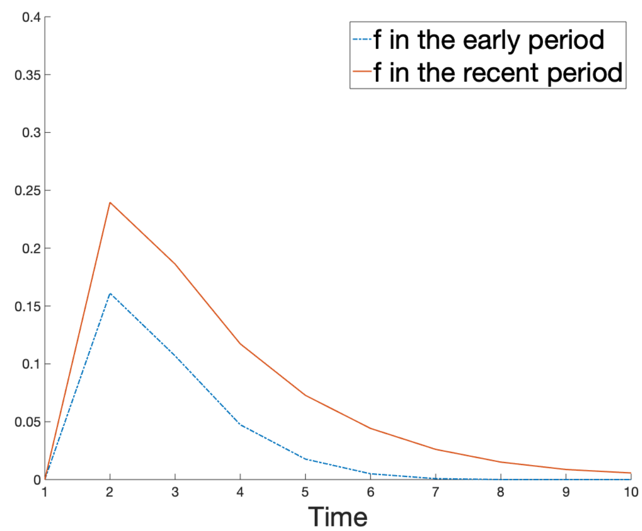


This graph shows the average event study result in the early period, the event dates includes in the early period are 2018-03-13, 2018-04-24, 2018-05-23, 2018-06-13, 2018-06-27, and 2018-09-04. The specific incidents happened in these dates can be found in Appendix.

Figure 2.15: Responses in the early period and the recent period



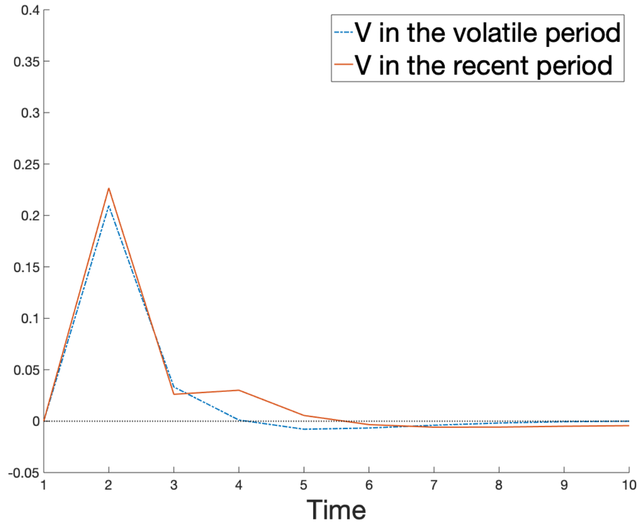
(a) Average transaction volume



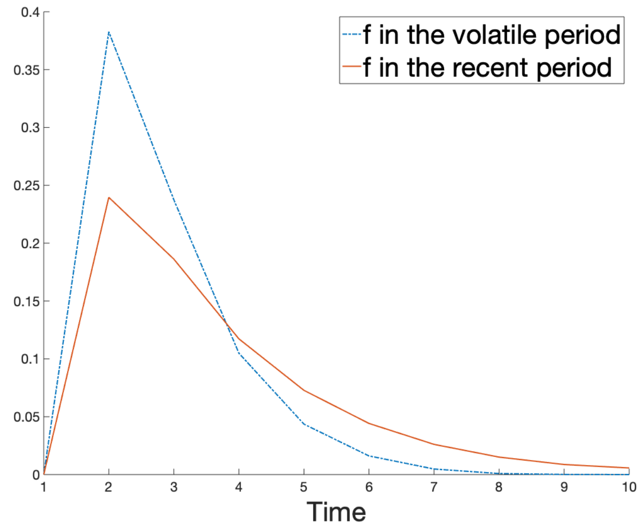
(b) Average transaction fee

The graph shows the simulation results from the early period and the recent period. The response from V_t switches to negative in the early period. Meanwhile, in the simulation results, the respond from f_t is slightly larger in the recent period.

Figure 2.16: Responses in the volatile period and the recent period



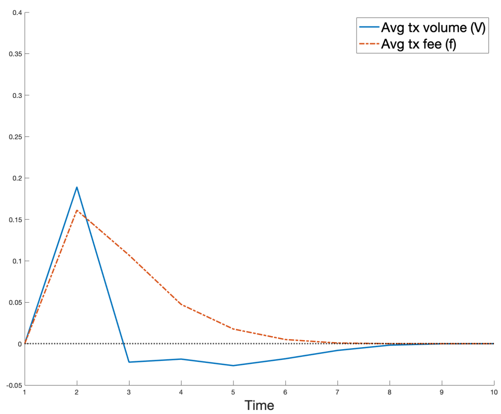
(a) Average transaction volume



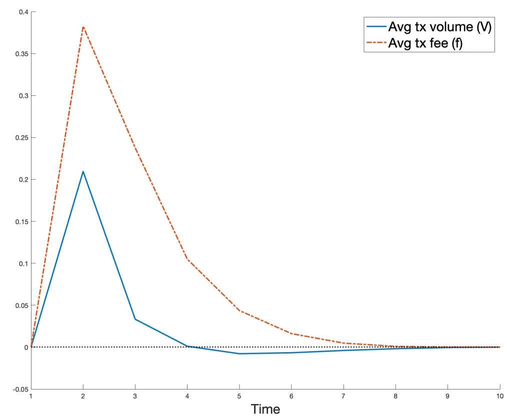
(b) Average transaction fee

The graph shows the simulation results from the volatile period and the recent period. The magnitude of the responses from V_t are very similar in these two periods, but the response from f_t is much stronger in the volatile period.

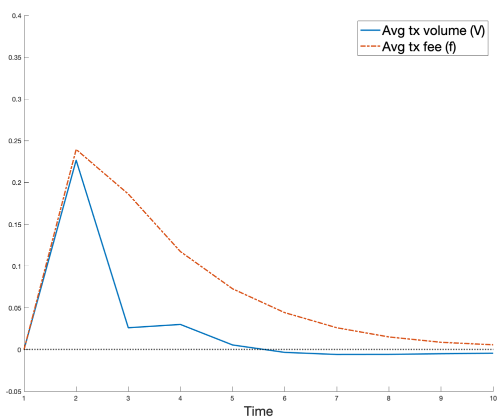
Figure 2.17: Simulation responses in each period



(a) Early period



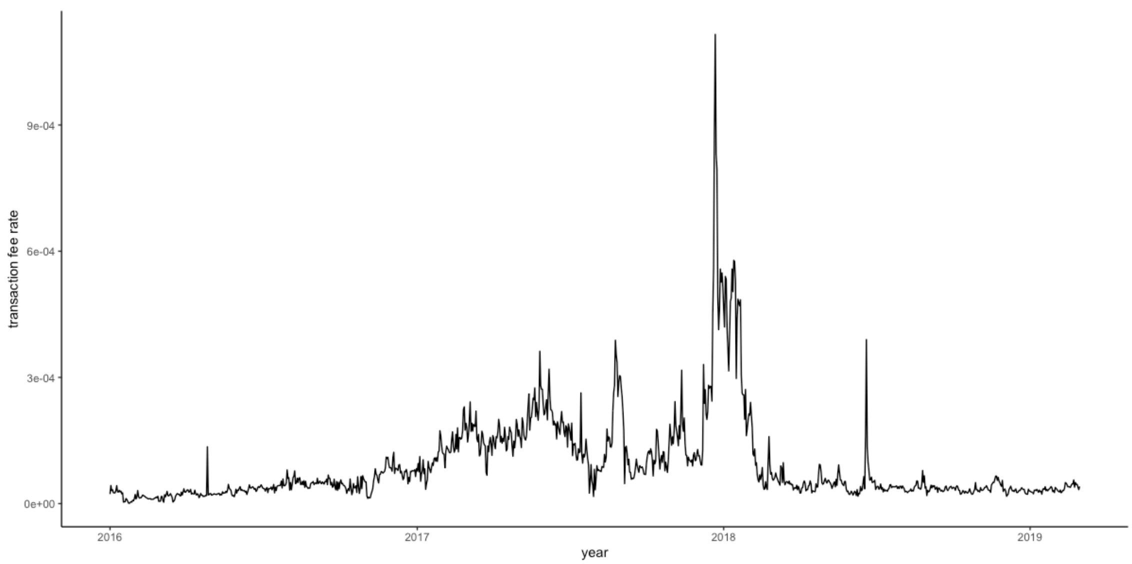
(b) Volatile period



(c) Recent period

This figure shows the simulation results inside each period. The magnitudes of the responses from V_t and f_t are almost the same in the early period and the recent period. However, the response from f_t in the volatile period is much larger than the response from V_t. These phenomena are also can be observed in the VAR results.

Figure 2.18: Transaction fee rate



The data comes from blockchain.com. This figure shows the daily transaction fee rate in the bitcoin network during the period 2016-01-01 $\sim$ 2019-03-01. We can see the transaction fee rate is actually relatively low. Even the highest rate is still lower than 0.1%.

Chapter 3

Price Dispersion in Bitcoin Exchanges

3.1 Introduction

Bitcoin has attracted a lot of attention from economists, especially after the volatile period in 2017. Many studies have been focused on the price discovery process and market efficiency in the bitcoin market. For example, Brandvold et al. [8], Giudici and Abu-Hashish [22] find that bitcoin price is highly interrelated among different exchanges, and large bitcoin exchanges provide more information in the price discovery process. Cheah et al. [11], Urquhart [49] find that the bitcoin market is not efficient. However, Nadarajah and Chu [39], Tiwari et al. [46] show that the bitcoin market is actually efficient.

One major difference between bitcoin and other financial assets is that there are multiple bitcoin exchanges. And an interesting phenomenon among these exchanges is that price dispersion among them is large and changing over time. In this paper, we use the closing prices in eight bitcoin exchanges that have the longest available data to calculate the price dispersion. We provide basic descriptions of bitcoin exchanges in Table 3.1 and summary statistic of price dispersion among these exchanges in Table 3.2. We find the price dispersion among these exchanges can range from 0.10% to 5.27% of the current indexed price. Figure 3.1 gives us a sense of how the price dispersion changes from 2017. The price dispersion reaches the highest value between the end of 2017 and the beginning of 2018 when bitcoin price reaches a historical high and the market is volatile. During that period, the price dispersion among different exchanges can be as high as 5.27% of the bitcoin price and the price difference was as large as $3,000. After that, price dispersion goes down, but it increases again around the first quarter of 2019 after a fall in bitcoin price.

This paper tries to find out what drives price dispersion. We estimate a time-varying vector autoregression model with stochastic volatility (TV-VAR-SV) model, and from the variance decomposition and impulse responses, we will argue that transaction fees and bitcoin price growth rate are the major factors that drive price dispersion across the eight exchanges.

3.2 Data and Methodology

In the TV-VAR-SV model, time-varying coefficients capture the dynamic change of the effects of different factors in our model, and stochastic volatility allows for time-varying behavior in the shocks (discussion of this model can be found at Del Negro and Primiceri [16], Primiceri [44], among many others). We find the flexibility of the model suitable for a drastically changing market like that of bitcoin.

The priors are set in a similar manner as in Primiceri [44]. As our model contains five variables, the degrees of freedom are set to 56 for Q, 6 for W, and [2 3 4 5] for the four blocks of S. The priors are written as:

$$
\begin{aligned}
&B_0 \sim N(\hat{B}_{OLS}, 4 \cdot V(\hat{B}_{OLS})), && S_1 \sim IW(k_S^2 \cdot 2 \cdot V(\hat{A}_{1,OLS}), 2),\\
&A_0 \sim N(\hat{A}_{OLS}, 4 \cdot V(\hat{A}_{OLS})), && S_2 \sim IW(k_S^2 \cdot 3 \cdot V(\hat{A}_{2,OLS}), 3),\\
&log\sigma_0 \sim N(log\hat{\sigma}_{OLS}, I_n), && S_3 \sim IW(k_S^2 \cdot 4 \cdot V(\hat{A}_{3,OLS}), 4),\\
&Q \sim IW(k_Q^2 \cdot 56 \cdot V(\hat{B}_{OLS}), 56), && S_4 \sim IW(k_S^2 \cdot 5 \cdot V(\hat{A}_{4,OLS}), 5),\\
&W \sim IW(k_W^2 \cdot 6 \cdot I_n, 6).
\end{aligned}
$$

Following Primiceri [44] and the standard time-invariant VAR results, we pick $lag = 2$, $k_Q = 0.1$, $k_S = 0.1$ and $k_W = 0.01$. To produce variance decomposition and impulse responses, we need to choose an ordering of the variables in our TV-VAR-SV model. As mentioned in Primiceri [44], and following Diebold and Yilmaz [17] and Klößner and Wagner [30], we are agnostic on the ordering and instead present average results obtained from all 120 possible orderings.

Due to missing data in some exchanges on certain days, and due to computational challenge, we choose to use weekly data instead of daily data in the model.[1] Besides average transaction fees per transaction (F_t), we also consider other variables that

[1] In daily data standard VAR, the information criterion usually suggests to pick 7 to 15 lags. Its means we need to calculate at least 180 posterior values in the B matrix of the TV-VAR-SV model, which we find to be infeasible.

may explain price dispersion ($P_{d,t}$), the inter-exchange price volatility ($P_{v,t}$), the average inter-exchange trade volume dispersion (V_t), and the bitcoin price growth rate ($P_{g,t}$).

We use Wednesday data from the eight exchanges mentioned above to build the weekly dataset of F_t, $P_{d,t}$, and V_t.[2] Transaction fees (in USD) and the number of daily transactions recorded in bitcoin blockchain are used to calculate F_t (in log). We define price dispersion as $P_{d,t} = \frac{SD(P_{j,t})}{mean(P_{j,t})}$, where $P_{j,t}$ is the closing price of exchange j at time t. V_t is calculated as the log of the largest trade volume minus the log of the second smallest trade volume (due to the missing data in some exchanges on certain days). $P_{v,t}$ is calculated as the log of the highest price minus the log of the lowest price from last Thursday to current Wednesday among these exchanges, and it shows how volatile bitcoin is over the past 7 days. $P_{g,t}$ is calculated by using the bitcoin price index data to measure the weekly price growth from last Wednesday to current Wednesday. All the exchange-level data is obtained from `https://www.cryptocompare.com`. The bitcoin blockchain-related data and the price index data are available on `https://www.blockchain.com`. The sample period we are examining is from 2016-06-01 to 2019-11-01, and the summary statistics are in Table 3.2.[3]

[2]We also used data from other weekdays for robustness check and the results are qualitatively similar. Those results are available upon requests.

[3]The first 56 observations in our sample are used to calculate the priors of the parameters in the TV-VAR-SV model, so the estimates reported in the following sections begin from the first week of July 2017.

3.3 Empirical Results

We present the variance decomposition results for price dispersion over time in Figure 3.2. We can see that shocks from the four variables ($P_{v,t}$, V_t, $P_{g,t}$, F_t) have substantial explanation power on the forecast error variance of price dispersion ($P_{d,t}$) during the volatile period (from the end of 2017 to the beginning of 2018), and most of that comes from transaction fees and price growth. Throughout the whole sample, these two variables on average explain 20% of the forecast error variance, but they account for more than 60% of the forecast error variance in the volatile period. However, the explanation power of these four variables goes down after March 2018, around the time when the bitcoin price has dropped more than 50% from the previous peak.[4]

To examine how the shocks of the four variables are related to price dispersion, we plot the impulse response functions (IRF) for price dispersion for a few specific dates. The weeks chosen are the first week of December 2017 (week I), the first week of January 2019 (week II), and the first week of Auguest 2019 (week III).[5] Week I comes from the most volatile period, during which transaction fees and price growth account for more than 60% of the forecast error variance of price dispersion. Week

[4]Thanks to the suggestion of a referee, we also examined network difficulty as another possible factor that may explain price dispersion. Replacing price volatility or volume dispersion, two variables that we find to be relatively unimportant, we find that network difficulty does not seem to be an important factor. Since network difficulty changes only every two weeks, we believe that its lack of variation is why it is not explaining much of price variation. Also, transaction fees and price change remain the major drivers of price dispersion. These additional results are available upon request.

[5]The weeks are chosen arbitrarily, but the main conclusion does not change by moving them forward or backward a few weeks.

II comes from the period with the biggest rebound of transaction fees after the most volatile period, and bitcoin price also has reached a recent low in this period (see Figure 3.1). Week III comes from the recent period when bitcoin price goes back to above $10,000.

The effects of the transaction fee shock and the price growth shock on price dispersion are statistically different in week I compared with the other two weeks (see Figure 3.3 and Figure 3.4). The upper left subplot in each figure shows the impulse responses of price dispersion to one specific shock, and the other subplots show pairwise differences between the impulse responses on different weeks. In week I, transaction fees shock reduces price dispersion first, but eventually it increases price dispersion among different exchanges. However, transaction fees shock only reduces price dispersion in the other two weeks. On the other hand, price growth shock increases price dispersion in week I but reduces in the other two weeks.[6]

In the bitcoin system, transaction fees affect price dispersion through two different channels. High fees deter users from arbitraging when price dispersion is small. On the other hand, higher fees encourage miners to confirm more transactions and reduce price dispersion. In Figure 3.5 we plot transaction fees along with the daily number of transactions. Around the time of week I, fees go up to 50 USD per transaction, and the number of daily transactions is almost the same as that around the time of weeks II and III when fees are around 2 USD per transaction. The number of transactions in mempool (Figure 3.6) serves as another piece of evidence.[7] Consistent with the

[6]For the other two variables, price volatility and trade volume dispersion, the IRF results do not show any significant time variation.

[7]The mempool is recorded by snapshots, the absolute number should not be treated as the real

conclusion in Tsang and Yang [47], these figures show that high fees around time of week I are mainly driven by the demand side (users) and the supply side (miners) have difficulty catching up with the extremely high demand (or that their supply of service is inelastic). Transaction fees shock with inelastic supply makes users reluctant to arbitrage during this period and drives up price dispersion. As supply is becoming more elastic later in weeks II and III, and fees shock actually slightly reduces price dispersion.[8]

Since transaction fees are not related to the transaction volume in the bitcoin system, when bitcoin price is increasing substantially transaction is likely to be cheaper, and it encourages people to trade and reduce price dispersion. Hence, in weeks II and III we see from the impulse responses that price growth shock eventually having a negative impact on price dispersion. However, in week I when bitcoin price was volatile, users find it too risky to move between different exchanges to take advantage of the price dispersion. Such uncertainty explains why price growth shock increases price dispersion around that period.

3.4 Conclusion

Using a TV-VAR-SV model, we show that a substantial proportion of price dispersion in the bitcoin markets can be explained by transaction fees and price growth.

---

demand. However, the relative magnitude of these numbers can tell us how dramatic the demand has changed.

[8] There are two possible reasons for more elastic supply: (1) the demand from users is relatively low and does not hit the miners' processing limit and (2) the introduction of SegWit increases miners' processing capacity.

Furthermore, the relationship between price dispersion and the two shocks changes over time. Still, there is a large proportion of price dispersion that is unaccounted for, and we believe that more micro-level data are needed to pin down the other sources.

3.5 Tables

Table 3.1: Summary of bitcoin exchanges

	Bitfinex	Bitstamp	Cex.io	Coinbase
Headquarter	Hong Kong	London	London	San Francisco
Currency	USD, EUR CNH, JPY GBP	USD, EUR	USD, EUR RUB, GBP	USD, EUR GBP
Time zone	UTC+8	UTC+0	UTC+0	UTC-8
	Gemini	**Kraken**	**Poloniex**	**Exmo**
Headquarter	New York	San Francisco	Wilmington	Moscow
Currency	USD	USD, EUR, CAD GBP, CHF, JPY AUD	USD	USD, EUR RUB
Time zone	UTC-5	UTC-8	UTC-5	UTC+3

This table shows the some basic information about the bitcoin exchanges used in our study. Even though these exchanges are in different time zones, data used in our study are all collected at UTC+0 this time point.

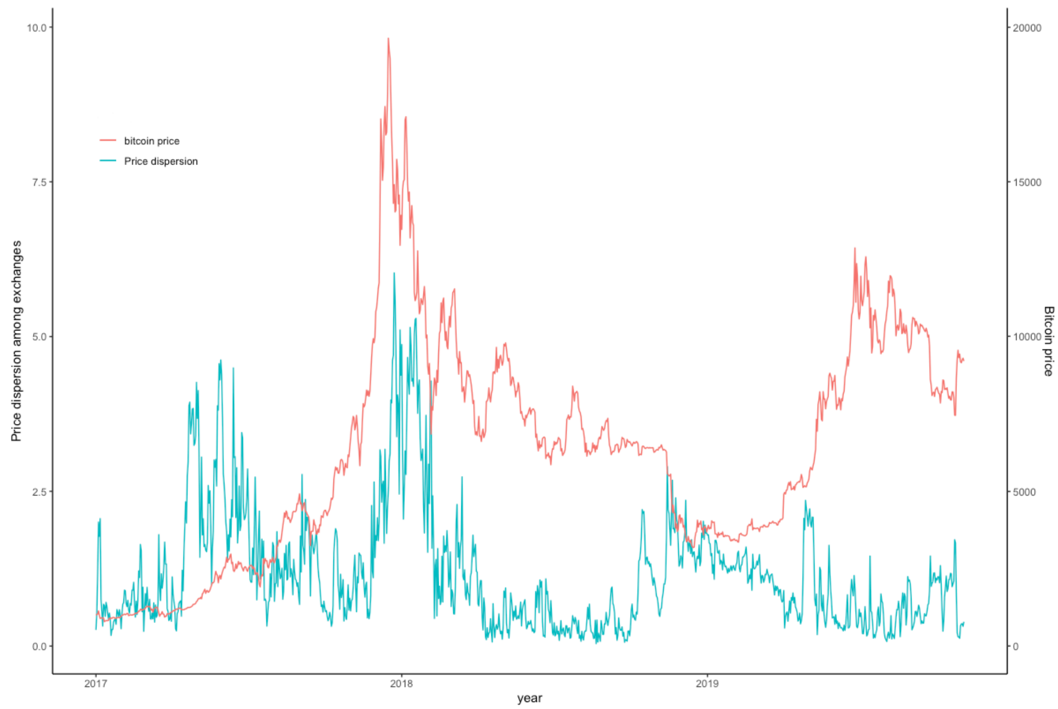
Table 3.2: Summary statistics

	$P_{d,t}$	F_t	$P_{v,t}$	$P_{g,t}$	V_t
Mean	1.093	0.028	0.171	0.016	3.431
Med	0.755	-0.095	0.146	0.012	3.410
Max	5.274	3.711	1.302	0.356	4.990
Min	0.099	-2.054	0.025	-0.364	2.055
S.D.	0.985	1.266	0.132	0.112	0.516
Obs.	178	178	178	178	178

This table reports the summary statistics of the data used in this study. The data period is from July 2017 to October 2019. $P_{d,t}$ represents the price dispersion among eight bitcoin exchanges: Bitfinex, Bitstamp, Cex.io, Coinbase, Exmo, Gemini, Kraken and Poloniex. F_t represents the average transaction fees per transaction (in log), $P_{v,t}$ represents the average price volatility among these exchanges, $P_{g,t}$ is the bitcoin price growth rate and V_t is the average trade volume dispersion among these exchanges. The data comes from cryptocompare.com and blockchain.com.

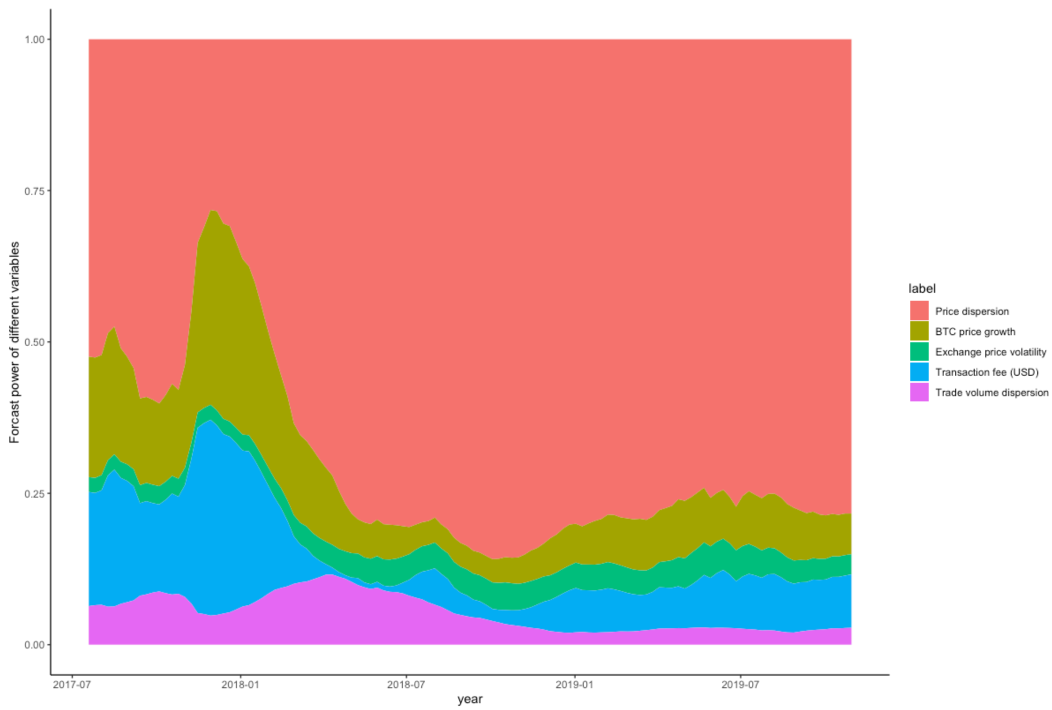
3.6 Figures

Figure 3.1: Price dispersion among exchanges and the average BTC price



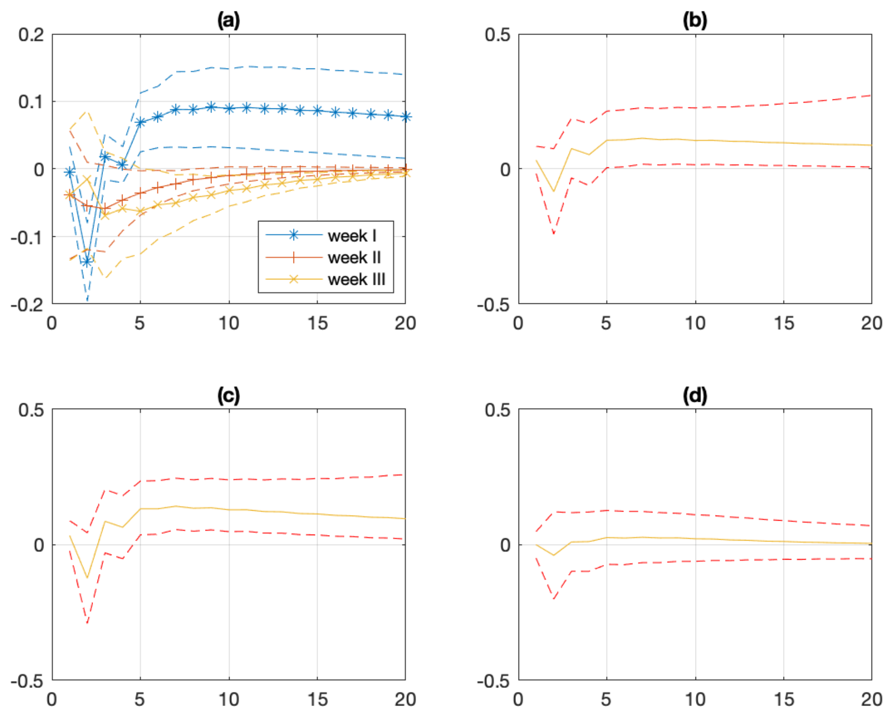
This figure shows how the price dispersion ($P_{d,t}$) among different bitcoin exchanges changes along with time.We can see the price dispersion reached the highest value when the bitcoin price reached historical high at the end of 2017.

Figure 3.2: Power figure of variance decomposition results



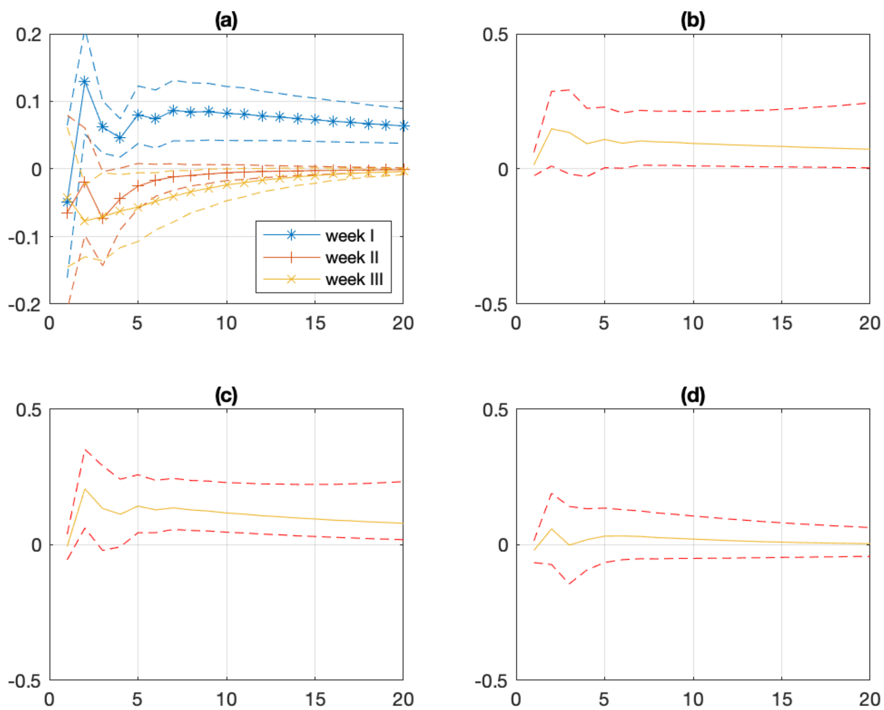
This figure shows the average results for the 20-step ahead variance decomposition with rolling windows analysis. The results show that transaction fees (F_t) and price growth ($P_{g,t}$) count for more than 60% of the forecast error variance of price dispersion in the volatile period.

Figure 3.3: Impulse response of price dispersion to transaction fees shock



(a) Impulse responses of price dispersion to transaction fee shock in week I, week II, and week III, (b) difference between the responses in week I and week II, (c) difference between the responses in week I and week III, (d) difference between the responses in week II and week III.

Figure 3.4: Impulse response of price dispersion to price growth shock

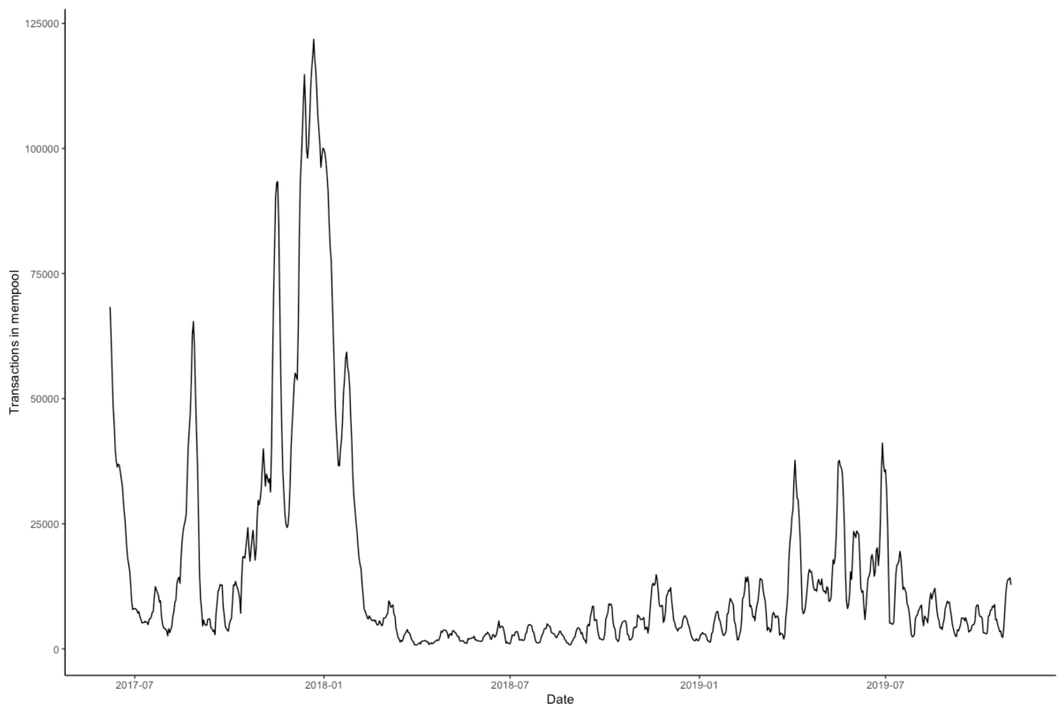


(a) Impulse responses of price dispersion to transaction fee shock in week I, week II, and week III, (b) difference between the responses in week I and week II, (c) difference between the responses in week I and week III, (d) difference between the responses in week II and week III.

Figure 3.5: Average transaction fees and confirmed transactions per day

This figure shows the seven-days moving average change of the average transaction fees and the daily total confirmed transactions.

Figure 3.6: Number of transactions in mempool



This figure shows the seven-days moving average of the daily number of transactions in mempool. The absolute daily number does not represent the real demand, but the relative magnitude of these numbers shows us how dramatic the demand has changed along with time.

Chapter 4

Do Connections Pay Off in the Bitcoin Market?

4.1 Introduction

This paper asks three questions: (1) Comparing with bitcoin addresses that are unconnected in the bitcoin investor network, do connected addresses earn a higher return? (2) Do more connected addresses earn a higher return than less connected ones? (3) When we use centrality to measure an address' connectedness, which centrality is more related to higher returns?

To answer the first question, we construct a bitcoin investor network using raw transaction data from the bitcoin blockchain. Based on the network, all the qualified transaction addresses are clustered into a connected group and an unconnected group. We find that, on average, over the period from June 2016 to May 2019, the return of the connected group is 20.75% higher than that of the unconnected group.

To answer the second question, we further divide the connected addresses into ten deciles based on their degree centrality or eigenvector centrality. We find that the addresses in the two most-connected deciles defined using either centrality measure consistently show higher returns than all the other connected addresses.

To answer the third question, we evaluate which centrality measure matters more in the two most-connected deciles, and we find that eigenvector centrality better explains the returns of those most-connected addresses. This result shows that the quality of connections is more important than the quantity.

We also conduct a series of robustness checks. We choose three different thresholds to identify the network connection between addresses, and half-monthly and weekly returns are used to replace the monthly returns used in the main results. All the results in these robustness checks are consistent with our main results. To see how the results change as the bitcoin market expanses and matures, we also break the sample into three periods, and the results are essentially the same.

A number of studies focus on the characteristics or behaviors of bitcoin investors. Xi et al. [51] conduct a web-based revealed preference survey to profile Initial Coin Offering (ICO) investors' characteristics in China and Australia, and they find the investing motivation factors are different in these two countries. Cahill et al. [10] suggest that investors may confuse bitcoin with its underlying technology, blockchain, and use the performance of bitcoin to predict the outlook of this new technology. After analyzing the bitcoin trading data from Mt. Gox, Glaser et al. [23] conclude that new bitcoin users from 2011 to 2013 tend to treat bitcoin as an alternative investment option rather than a new currency. This result aligns with the survey result conducted by Mahomed et al. [37] in South Africa. Using the Mt. Gox exchange trading data, Gandal et al. [20] and Chen et al. [12] demonstrate that there are market manipulations inside the Mt. Gox exchange which cause unprecedented spikes in bitcoin price. To study the behaviors of bitcoin investors, we think exchange data provides a more accurate and comprehensive picture. However, exchange data are not publicly available. The leaked Mt. Gox exchange trading data has been widely used in several studies, but the drawback is that it only covers the period from April 2011 to November 2013. In this paper, we are using the data obtained directly from the bitcoin blockchain. Bitcoin blockchain raw data is publicly accessible, and it allows us to look at the more recent periods where the market is more established and mature. However, bitcoin blockchain raw data also comes with its own problem: the transaction data are connected to bitcoin addresses, not individual investors. In this paper, we propose a simplified parsing procedure to mitigate this drawback.

Most studies on investor networks focus on stock markets. By exploiting a dataset that includes all account-level trading information on the Istanbul Stock Exchange in 2005, Ozsoylev et al. [42] find that investors' investment returns are positively correlated with their centrality in the network. Rossi et al. [45] show that investment managers who are better connected tend to have a better portfolio performance. Ahern [1] analyzes an illegal insider trading network and finds that people in the network earn a 35% return over 21 days, and more central traders in the network earn even higher returns. Walden [50] introduces a dynamic noisy rational expectations model and finds out that an agent's profitability is determined by Katz centrality. However, stock markets are generally more regulated and mature. In this paper, we look at the bitcoin market, which is relatively new and less regulated. Liu and Tsyvinski [35] find that coin returns are positively related to cryptocurrency network growth rates, and current cryptocurrency prices include information about the expected network growth. As a complement to their study, we focus on how returns differ among investors inside the bitcoin network using address-level transaction data.

The remainder of the paper is structured as follows: Section 2 introduces the basic technical background of bitcoin. Section 3 explains data sources and the blockchain data parsing procedure. Section 4 constructs the bitcoin investor network. Section 5 shows the empirical results about the effect of connectedness and centrality on bitcoin investment returns. Section 6 concludes the paper.

4.2 Background

The original idea of Nakamoto [40] is to build bitcoin as a decentralized payment system that does not need a centralized clearinghouse yet still keeps transaction records secure and immutable. Bitcoin and other cryptocurrencies have gained the attention of regulation institutions, researchers, and investors in the past few years. In the following sections, we will explain some technical details about bitcoin that are necessary to understand the rest of this paper.

4.2.1 Structure of a bitcoin block

Blocks are storage units that contain confirmed transaction data. Each block in the blockchain, which is a sequence of blocks, contains a piece of information about the location of its previous block. The size of each block is set to be less than 1 MB to ensures that each block can only contain a certain number of transaction records. Meanwhile, every block includes a block header that includes six fields: version, previous block hash, merkle root, timestamp, target, and nonce. The complete structure of a bitcoin block is summarized by [3] and reproduced in Table 4.1. In this paper, the relevant data comes from the transactions section in every bitcoin block.

4.2.2 Bitcoin transactions

Every transaction in the bitcoin system consists of two parts: transaction input and output. Except for coinbase transactions, a transaction input usually should come from an unspent transaction output (UTXO) generated from a previous transaction.[1] In Table 4.2, we list the major components in a bitcoin transaction.

When a user receives bitcoins from other users, what it means is that this user's bitcoin wallet has detected a new UTXO that can be spent by one of the private keys controlled by this wallet. One feature of UTXO is that it is an indivisible chunk of bitcoin, just like a dollar bill. Users cannot spend a part of a UTXO in one transaction in the same sense that they cannot use a part of a dollar bill in one transaction. However, users only need to tell their bitcoin wallet the desired transaction amount and transaction fee, and then their bitcoin wallet will automatically select from all

[1]The coinbase transaction in a block sends the mining reward to the miner who successfully mined this block. The transaction input in a coinbase transaction is not linked to a UTXO. We remove all coinbase transactions from our dataset.

the UTXOs controlled by the wallet to compose an amount equal or greater than the desired transaction amount.

To verify the ownership of a UTXO, bitcoin relies on the public-key cryptography. A digital wallet can generate almost unlimited pairs of public key and private key, and each pair includes one public key and one private key. A public key is included in every UTXO, and it tells the whole bitcoin network the receiver of that UTXO. On the other hand, a private key is a certification telling miners who is the real owner of that UTXOs as one public key can only be deciphered by its corresponding private key. To simplify the analysis, here we can think of each UTXO as a dollar bill with a face value equal to the amount of bitcoin embedded in this UTXO. Figure 4.1 illustrates a bitcoin transaction.[2] In user A's digital wallet, there are 2 pairs of public and private keys (public key 1/private key1, public key 2/private key2). Public key 1 is linked to UTXO1 (1.3 bitcoins) and UTXO2 (0.7 bitcoins) respectively, and public key 2 is linked to UTXO3 (1 bitcoin). Now user A wants to send 2.5 bitcoins to user B (public key 3). User A cannot send 1.5 bitcoins from public key 1 and 1 bitcoin from public key 2 to public key 3. What user A can do is send out all 3 bitcoins: 2.5 bitcoins to public key 3 and 0.49 bitcoin to her own public key. The 0.01 bitcoin difference serves as the transaction fee collected by the miner who successfully confirmed this transaction.

In the above transaction, user A sends back the change of 0.49 bitcoins to her own address (public key 1), and such transactions are usually referred as change trans-

[2]Here we use public keys instead of bitcoin addresses (public key hashes) to illustrate this simple transaction. In practice, most transactions actually use bitcoin addresses instead of public keys on the output side.

actions. Change transactions are easy to be detected if users keep sending back the changes to the addresses that showed up on the input side. However, directly sending back changes to one of the already used addresses may allow other people to trace back all the transactions related to user A. To protect her privacy, user A can use the bitcoin wallet to generate a new pair of public key 4 and private key 4 and send the rest 0.49 bitcoin to public key 4 (see Figure 4.2). By doing so, other people can never tell if the 0.49 bitcoin transaction is a change transaction or not. This practice produces two challenges: (1) It is difficult to cluster the transaction data into individual-investor level because investors can generate new addresses for every new transaction to cover their traces. (2) It is difficult to filter out all the change transactions from the transaction dataset as investors can always generate new addresses to receive changes. Some solutions are proposed in the following section to address these issues.

4.3 Price and transaction data

4.3.1 Data source

In the bitcoin transaction raw data, the input side does not include the real transaction amount. Instead, the input side includes the UTXOs that are supposed to be used to fulfill the transaction. It is the miners' responsibility to verify if these UTXOs contain enough bitcoins to fulfill this transaction. The raw transaction data stored in the blockchain is rather hard to interpret. Luckily, Google has pre-parsed the raw data and made it more user-friendly, so we use Google's bitcoin transaction dataset as our main data source. Considering the massive amount of data that needs to be processed, we choose to use Google's BigQuery service to pre-process the whole dataset and use R for econometric analysis.

We also use the minute-level bitcoin price data from Bitstamp and Coinbase to construct a block-level bitcoin price dataset. If the price data is available on both platforms, we use the average value; if the price data is only available on either one platform, we use the available data to represent the block-level bitcoin price data. If the price data is not available from both sources, the price will be imputed based on the average value of neighboring available prices.

4.3.2 Data parsing

As mentioned in Section 2, users can generate a new pair of public and private keys for each transaction to hide their identities. One challenge brought up by this privacy feature is that we can never be certain if two addresses are owned by the same person or owned by two different people. [38] propose two heuristic rules to parse bitcoin addresses data: (1) the public keys used as inputs in the same transaction should be considered as being owned by the same person; (2) the change address on the output side in a transaction should be recognized as being controlled by the person who initiates the transaction.[3] Although these two rules seem plausible, they are not fool-proof in some situations. [26] point out that rule (1) does not apply to CoinJoin

[3] Change address refers to the output address(es) in a change transaction.

transactions.[4] Meanwhile, it is very difficult to detect all the change transactions. One way proposed by [38] to detect the change address is if that address has only been used once. However, this method may mislabel some long-term investors and create false super-clusters. In particular, more than 50% of addresses during the sample period only appear once in our dataset. If we follow the above rule, then most of the transactions in our dataset will be falsely labeled as change transactions.

Due to the availability of the minute-level price data, we begin the sample from 2016-06-01, and the sample ends on 2019-05-31. In this 3-years period, bitcoin price changed dramatically, and bitcoin was popularized by mass media outlets in 2017, especially when its price reached almost 20,000 USD at the end of 2017. To address the concern that our dataset may be dominated by transactions from a certain time window, we also divide the whole sample into three periods: period 1 (2016-06-01 to 2017-05-31), period 2 (2017-06-01 to 2018-05-31), and period 3 (2018-06-01 to 2019-05-31). Period 1 represents the bitcoin market when it is not popularized by mass media outlets. Period 2 represents the period when bitcoin became a mainstream investment option and experienced an astonishing price jump. Period 3 represents the period when the public hype about bitcoin was dying out, and bitcoin had the largest price drop in its history.

Table 4.3 shows how many unique transaction addresses are recorded into the bitcoin blockchain. More than 300 million unique transaction addresses are recorded into the bitcoin blockchain in our sample period. Furthermore, more than 90% of addresses

[4] CoinJoin transactions combine transactions from different spenders into one transaction, in this way outsiders cannot tell which spender paid which recipient(s).

showed up less than three times. After dividing the full sample into three periods, we can see the number of addresses and the percentages of addresses that appear at certain times are stable across these three periods, which means our dataset is not dominated by transactions from a certain time window. To further address this concern, we will report whole sample regression results along with the results in each period in this paper.

In this paper, we choose to use the following three steps to parse the bitcoin transaction data:

Step 1: Subsample the original transaction dataset to only include addresses that appear at least ten times during a given period.[5]

Step 2: In each transaction, we remove all the output addresses that also show up on the input side as they are likely to be change addresses.

Step 3: In each transaction, we treat all the addresses on the input side of a transaction as being controlled by the same person.[6]

In the first step, we filter out all the users who only use addresses once for any transaction and may care more about transaction privacy than returns, along with the users who do not trade frequently. In this study, we use monthly returns, half-monthly

[5]In this subsample, each transaction record may no longer preserve all the involved addresses. However, our goal is calculate address level return, so transaction level incompleteness will not affect our final results.

[6]As we mentioned at the beginning of this subsection, CoinJoin transactions can be exceptions. However, comparing with joining in a CoinJoin transaction to protect privacy, a much easier way is to generate different addresses for different transactions. The addresses in this dataset have already been used more than ten times, we think their owners will be less likely to use CoinJion transactions to protect their identities.

returns, and weekly returns to measure investors' performance on a short-term horizon. If the address does not trade frequently, we can safely assume this investor does not care that much about the short-term horizon. Meanwhile, removing those less frequent addresses from the dataset can also significantly reduce the computing pressure and also reduce the number of false super-clusters. In the second step, we did not adopt more aggressive methods to parse out change transactions because we may end up mislabeling some legitimate transactions. Step three also helps us reduce the size of our dataset. These three steps will certainly miss some connections between addresses and mislabel some addresses owned by the same person as they have different owners. However, this simplified method can significantly reduce computing pressure by reducing our sample size from more than 35 million addresses to around 6 million addresses. Meanwhile, these steps will not falsely connect some addresses that are not owned by the same person, so we can avoid false super-clusters in our dataset.

In this next section, we will use the parsed dataset to construct the bitcoin investor network.

4.4 The Bitcoin Investor Network

4.4.1 Definition

We mostly follow Ozsoylev et al. [42] to define the bitcoin investor network. However, the bitcoin market is different from the stock market in the following aspects: (1)

In the bitcoin market, there is only bitcoin this one asset to trade. (2) Bitcoin transactions are confirmed in batches. When a new block is successfully mined in the bitcoin network, all the transactions inside this block are confirmed at the same time. As a result, we need to modify the approach from Ozsoylev et al. [42] and define the bitcoin investor network as:

Definition The bitcoin investor network , $\varepsilon^{\Delta b,M}$, is defined such that for each pair of bitcoin addresses, i, $j \neq i$, $\varepsilon^{\Delta b,M} = 1$ if and only if i and j traded in the same direction within Δb blocks at least M times over the period.

To calculate the bitcoin investor network, we choose among 10, 30, and 50 for M. Meanwhile, we set $\Delta b = 1$, which means we keep the time window as 10 minutes on average. The first reason for keeping the time window Δb short is that the information-driven trading is usually called “fast” trading and happens in a short time window, and we want to separate it from other types of trading. The second reason is to save computing time.

We construct the bitcoin investor network separately for three different periods (see subsection 4.3.2). The reason behind this decision is that we are concerned about the change in the composition of bitcoin investors given the landscape of the bitcoin market has changed a lot during these three periods. For example, some addresses may be active in period 1 but not in period 2. By estimating the network using the whole sample, we may under-sample the addresses that are only active in a certain period. Meanwhile, if there is no dramatic investor change, estimating the network by periods should give us similar results as using the whole data sample. If there is significant investor change, then estimating the network by periods can better capture the dynamic change in the bitcoin investor network. Table 4.4 provides the summary statistics of the bitcoin investor network for different periods and values of M. We can see that, firstly, the number of links decreases when the value of M increase. The increase of M means the criteria of being connected increase, so we expect fewer addresses will be counted as connected in the dataset. Secondly, the number of links is similar in the first two periods and slightly larger in the third period. Meanwhile, the fraction of links is small in all three periods, which means the network is sparse and a small group of addresses are connected in the network. Overall, we can say the bitcoin investor network is sparse, and the statistics are mostly similar across these three different periods.

4.4.2 Node centrality and returns

After identifying the bitcoin investor network, we can now measure the centrality of every node (address) in it. While many methods have been proposed to measure network centrality, we follow Ozsoylev et al. [42] and only consider degree centrality and eigenvector centrality in this paper as they directly measure how connected a node is in the network.[7] Degree centrality measures how connected a node is in the network by counting how many other nodes are connected with it. Eigenvector

[7] A thorough discussion about network centrality, please see Jackson [25].

centrality measures a node's connectedness by considering the importance of all the neighbors connected with it. As we can see, these two measures describe different characteristics of the nodes in this bitcoin investor network. The degree centrality measures how widely connected a node is in this network. The eigenvector centrality tells us how important a node's neighbors are in this network.

To measure bitcoin returns, we follow the same approach as in Barber et al. [5] and Ozsoylev et al. [42]. We choose $\Delta B = 4320$ as the window length, which is roughly a month given that bitcoin blocks are generated every 10 minutes on average. Defining the window in terms of blocks can help us precisely match bitcoin returns with the transaction data, which is also confirmed in terms of blocks.[8] For each trade, z, of individual i, we define the return as:

$$\mu_{i,z} = sign * \left(\frac{P^{b+\Delta B} - P^b}{P^b} - r_{tbill} \right) \tag{4.1}$$

where $P^{b+\Delta B}$ is the bitcoin price B blocks later after the trade happened in block b. When we set $B = 4320$, $\frac{P^{b+\Delta B} - P^b}{P^b}$ gives us roughly the monthly return rate of each trade. r_{tbill} is the daily U.S. Treasury 3-month yield rate, and the data for weekends is imputed based on the neighboring available values.[9] Both rates have been annualized to make sure they are comparable.[10] The symbol $sign$ indicates the trade direction,

[8] In the Appendix, for the robustness check, we also consider 2160 blocks and 1008 blocks, which correspond to 15 days and 7 days. And the results are essential the same.

[9] The daily U.S. Treasury 3-month yield rate can be found on the U.S. department of the Treasury website. Equation (1) represents excess returns. However, because the magnitude of r_{bill} is much small than the first term in equation (1) and r_{bill} also changes less frequently than the first term, adding r_{tbill} or not does not have an observable impact on our results.

[10] To annualize monthly return rates, we multiply the data with 365/30.5. Following the same logic, we use 365/15 for half-monthly return rates and 365/7 for weekly return rates. We did not

which is negative for input addresses and positive for output addresses. To further reduce the computing burden and lower the computing memory requirement, we calculate the average returns for an address if there are multiple trades from this address in a month.

Due to the difficulty of linking all the addresses that are owned by an individual, it is not possible to accurately calculate the trade amount of every investor. As a result, the weighted returns used in Ozsoylev et al. [42] are highly inaccurate in this study. For example, if there are three addresses, a, b and c, owned by an investor. Suppose c is a safe vault, and the investor usually transfers a large amount of bitcoin into c and moves a small amount of bitcoin out of c when the bitcoin price is increasing rapidly. Without knowing that these addresses are owned by the same person and all the transactions among them are change transactions, the weighted returns will be heavily biased by returns caused by those change addresses. To mitigate the effect of these undetected change transactions, we think the definition of returns in equation (1) is a better candidate to measure investors' performance. Equation (1) measures how often an address can show up at the right side of a trade at the right time and ignores the trading amount due to incomplete trading information on the individual investor level.

use the formula $(1+r)^t - 1$ because return rates can be very high in some months, $(1+r)^t - 1$ may make the performance of an address in a high bitcoin return month dominates its whole year performance.

4.4.3 Stability

We have already stated the benefits of estimating the bitcoin investor network individually for three different periods. We follow Ozsoylev et al. [42] and consider two different methods to test the stability of the bitcoin investor network.

In the first method, the null hypothesis assumes that the network links are generated randomly. Suppose a network contains N nodes and k_1 links, then the possibility of two different nodes are linked is k_1/K, where $K = N(N-1)/2$ is the total possible links in the network. Suppose the network in two different periods has k_1 and k_2 links, and $k_1 << K, k_2 << K$, then the expected number of overlap links ($E_{random}[y]$) in these two periods is:[11]

$$E_{random}[y] \approx \frac{k_1 k_2}{K}$$

where k_1 is the number of links in the first period, k_2 is the number of links in the second period, and K is all the possible links among all the addresses in the network.

In the second method, the null hypothesis assumes the bitcoin investor network is not truly randomly generated, and the degree distribution has heavy tails. Hence, we take degree into consideration, and the degree-adjusted expected number of overlap links ($E_{degree-adjusted}$) between period 1 and period 2 is written as:

$$E_{degree-adjusted} = \frac{k_2}{2k_1 N} \sum_{N}^{i=1} (D_i - 1)^2$$

[11]The condition that $k_1 << K, k_2 << K$ is satisfied in our dataset as the fraction of links reported in Table 4.4 is less than 0.005%, which means the bitcoin investor network is very sparse.

where N is the number of investors in the network, and D is the degree distribution of the network in period 1.

Table 4.5 shows the stability of the bitcoin investor network for $M = 50$.[12] We can see that the actual number of overlap links between any two periods is significantly higher than the expected number of overlap links when the network links are randomly generated. For the degree-adjusted method, the actual number of overlap links is still significantly higher than the second null hypothesis predicted. Both methods suggest that the connections in this network are relatively stable across these three periods. Hence, the bitcoin investor network does not change dramatically in our sample period.

We also try $M = 10$ and $M = 30$ for robustness checks, and the results are reported in the appendix (Table B.1, Table B.2). We still can see that, under the random generation hypothesis, the real overlap link data still reject the null hypothesis that the network is not stable over time. However, the picture is not so clear when we compare the results for $M = 10$ using the degree-adjusted method. These results may indicate that: 1. when we set $M = 10$, the threshold of being identified as connected is low, and many false connections are included in the dataset; 2. high-connected, more central nodes stay active in the bitcoin investor network longer than the rest nodes. Hence, when we set $M = 50$ and rule out more less-connected nodes, we can observe a more stable network. Given the above reasons, we use $M = 50$ in our main results.

[12] The number of links in Table 4.5 is slighted different from the number in Table 4.4. This is because Ozsoylev et al. [42] counted self-links in their method. To maintain consistency, we also include self-links in this table.

4.5 Results

In this section, we try to answer three questions: (1) Comparing with addresses that are unconnected in the bitcoin investor network, do connected addresses earn a higher return? (2) Do more connected addresses earn a higher return than less connected ones? (3) When we use centrality to measure an address' connectedness, which centrality measure is more correlated to higher returns. In the rest of this paper, we will report our main results using $M = 50$ and use $M = 10, 30$ for robustness check. Also, we will report the results for the whole data sample along with the results for each period to provide more detailed information.

4.5.1 Do connected addresses earn higher returns?

We divide the addresses into two groups based on their connectedness. The unconnected group includes all the addresses that are not connected with any other addresses in the sample. The connected group includes addresses that have at least one connection with other addresses. To reduce the influence of outliers, we truncate the addresses at the top two percent of connectedness in the dataset. We calculate the average monthly return of all the addresses in each group for every month in the sample period, and the results are shown in Figure 4.3, Figure 4.4, and Figure 4.5 correspond to $M = 10, 30$, and 50.

When $M = 10$, we can see that the average returns of the connected group are higher than that of the unconnected group when the bitcoin market return, in general, is increasing. The most dramatic example is in November 2017, where the annualized average monthly return of the connected group is around 791%, and that of the unconnected group is only 539%. In contrast, the connected group may have lower returns during the period when bitcoin returns, in general, are dropping. When we increase the threshold to $M = 30$ or 50, the pattern remains consistent, but the return difference between these two groups shrinks. This result is not surprising because more addresses are categorized as unconnected when we raise the threshold M.

To check whether the results are sensitive to the time horizon of returns, we also present results using half-monthly and weekly returns under $M = 10, 30, 50$ (see Figure B.1-Figure B.6 in appendix). All results consistently show that the connected group earns higher returns than the unconnected group when bitcoin returns, in general, are increasing. Occasionally, when the bitcoin market return is dropping, most of which happen in period 3 in the dataset, the connected group may do worse than the unconnected group.

Figures are straightforward and give us a first impression of the relationship between connectedness and bitcoin returns. However, figures can also be misleading if we miss other important factors that can also affect returns. In the next step, we run multivariate regressions at the address level to see how connectedness is related to returns. We control the monthly market trade volume, the monthly market number of trades, the monthly individual trade volume, the monthly individual number of trades, and the monthly bitcoin price volatility, which is calculated as the standard deviation of bitcoin prices in each month. All these variables are in log. The summary statistics of these variables are reported in Table 4.6.

The regression results in Table 4.7 show that connectedness always matters using the whole sample because the coefficient of the connectedness variable shows that, on average, the annualized returns of connected addresses in the network are 20.75% above those of their unconnected peers. Then we break the sample into the three periods mentioned before, and the conclusion still holds. Even though the return difference between the connected and unconnected groups is relatively small in period 1, during which the bitcoin was only popular in some small circles. The difference becomes obvious in period 2, when bitcoin finally entered the public eye and gained worldwide popularity. During period 2, the addresses in the connected group, on average, earn an additional 44.75% annualized returns compared with their peers in the unconnected group. After controlling other factors, we can see that connected addresses still earn higher returns than their unconnected peers in period 3, during which we observe some largest bitcoin price drops in history.

In Table B.3 to Table B.10, we also conduct the same regression using $M = 10, 30$ and replace monthly returns with half-monthly returns or weekly returns for robustness check, and the results consistently show that the connected addresses have significantly higher returns than the unconnected ones. The difference is most obvious in period 2, when the bitcoin price rapidly increased and reached its historically high.

4.5.2 Do more connected addresses earn higher returns?

We have shown that connected addresses in the network have higher returns than their unconnected peers, and the next question to ask is whether being more connected matters in this network.

We use centrality to measure the connectedness of addresses. As mentioned before, we use both degree centrality and eigenvector centrality in this paper. Degree centrality measures how connected a node is by counting the number of nodes connected to it. Presumably, a more connected node plays a more critical role in spreading information in the network as an information hub. Being a hub may give this node access to more exclusive information about the market and help the node earn higher returns. However, it is possible that some addresses may not have high degree centrality, but they may become essential information hubs by occupying important positions in the network and being connected to some influential nodes. That is why we also consider the eigenvector centrality, which measures how important a node is by looking at how important its neighbors are in this analysis. Eigenvector centrality is calculated using iGraph package in R, and the value has been normalized to be between 0 and 1. Due to limited computing power, we only report the results for $M = 50$.[13]

Figure 4.6 shows how degree centrality is related to bitcoin returns in the full sample. We divide the addresses in the connected group into ten deciles based on their degree

[13] In this study, Eigenvector centrality is calculated on Virgina Tech ARC servers using 20 cores (128 GB shared memory). Given the size of the network, more computing resources are needed to calculate the Eigenvector centrality when M = 30 or 10.

centrality values, and then we calculate the average monthly returns of each decile over the whole period. We can see from this figure that the two most-connected deciles earn even higher returns than the other connected ones. Changing the time horizon of returns by using half-monthly and weekly returns does not change the conclusion. Using the same approach, Figure 4.7 shows the performance of addresses with different eigenvector centrality values in the full sample, and the results are very similar.

To compare the return difference over time, we break the full sample into the three different periods mentioned in subsection 4.3.2. From Figure B.7 to Figure B.12 we can see that no matter which centrality measure we use, addresses in the two most-connected deciles are earning higher returns than their other connected peers in the first two periods, with the difference being more dramatic in the second period. However, in the third period, addresses in the two most-connected deciles are doing worse than the addresses in the unconnected group. However, directly drawing conclusions from the figures in different periods may be misleading because we may ignore other important factors that can also affect returns.

To carefully examine the relationship between returns and centrality among connected addresses, we run regressions for returns on deciles with different centrality while controlling for other factors. In this regression, the ten connected deciles are categorized using either degree centrality or eigenvector centrality and labeled by ten dummies. Hence, the base group in the regression result is the unconnected addresses, and the 1-10 groups in regression results represent the least connected decile to the most connected decile. In the regressions, we control bitcoin price volatility, market number of trades, market trade volume, individual trade volume and individual number of trades. From the regression results in Table 4.8, we can reach two conclusions: (1) Most connected groups have higher returns than the unconnected base group, with the exceptions in decile 6 and 7 in the first period. We do not see this kind of exception in more recent periods. (2) The two most-connected deciles earn higher returns than the others in all periods regardless of which centrality measure is used.

Table B.11 and Table B.12 replace the monthly returns with half-monthly returns and weekly returns for the robustness check. To visualize all the results, we plot out the coefficients of group dummies and show them in Figure 4.8 and Figure 4.9. We can see a jump on the coefficient values at the last two deciles in these two figures. This pattern also matches the upward trend we have seen in Figure 4.6 and Figure 4.7. Again, the pattern does not depend on the time horizon of returns.

We also plot out the coefficients in the three individual periods. Figure B.13 to Figure B.16 show that the two most-connected deciles consistently have larger coefficient values in periods 1 and 2, no matter whether the groups are identified using degree centrality or eigenvector centrality. In Figure B.17 and Figure B.18, we can see the two most-connected deciles are having higher monthly returns than the rest addresses in period 3 because the coefficients for these two groups are larger than the rest. Among all these regressions, the two most-connected deciles only seem to perform worse than their other connected peers when we use weekly returns. Even in this case, the addresses in the $80-90\%$ range still have the highest returns among all the groups. It is only the top 10% that does not show high weekly returns.[14]

The results so far suggest that the two most-connected deciles have even higher returns than the others. Next, we focus on the two most-connected deciles and see which centrality measure matters more to these most-connected addresses.

4.5.3 Which centrality measure matters more?

So far, we have found that the addresses that belong to the two most-connected groups earn higher returns than the other connected addresses, and the results do not depend on which centrality is used. One explanation is that these two centrality measures are correlated, but they may not be equally important. In this subsection, we are interested in testing which centrality measure matters more. We run regressions for returns on degree centrality, eigenvector centrality, and their interaction for addresses in the top two most-connected deciles in terms of both centrality measures. Following the previous regressions, we also include bitcoin price volatility, market number of trades, market trade volume, individual trade volume, and individual number of trades as control variables. The summary statistics of these variables are in Table 4.9.

[14]A plausible explanation for the abnormal performance of the top 10% in period 3 is that we accidentally include some exchanges or other institutions in this group. These institutions, like bitcoin exchanges, may have goals other than pursuing high returns in the bitcoin market. In this regression dataset, we truncate the top two percent most connected addresses in the original dataset to avoid influence by outliers, and we get the most connected address in the first period has 7707 connections, in the second period has 5896 connections. However, in the third period, the most connected address still has 16082 connections. If we only include the addresses with less than 7000 degrees into the regression, then the results will be similar to what we see in the first two periods. For the sake of consistency, in this study, we choose to only truncate the top 2% most connected addresses in all three periods.

We can see from the regression results in Table 4.10 that, compared with degree centrality, eigenvector centrality is a better explanatory variable for returns of the most-connected addresses in the bitcoin investor network. The coefficient of eigenvector centrality is positive and significant, which means higher eigenvector centrality is related to higher returns among those most-connected addresses. However, with the eigenvector centrality in the regression, the coefficient of degree centrality becomes negative even though it is not significant in the full sample, which means higher degree centrality is not related to higher returns in our sample. The results from the different period samples make the above conclusion even more obvious: the coefficient of degree centrality becomes negative when we break down the full sample into three periods. The conclusion also holds no matter what time horizon of returns we use in the regressions.

Furthermore, the coefficient of the interaction term is negative and significant, and there are two interpretations. If we can interpret high eigenvector centrality as having more high-quality connections, then higher degree centrality, when holding eigenvector centrality constant, means potentially having more noisy connections. First, as degree centrality increases, the marginal impact of eigenvector centrality goes down. When nodes connect to more and more other nodes, they will eventually reach more and more high-quality connections in the network and lack less and less crucial information to make the right investment decision, and this is why the marginal return related to the increase of eigenvector value decreases when nodes have more and more connections. Second, when an address has high eigenvector centrality, higher degree centrality is related to lower returns. A node that receives both high-quality and noisy connections will have lower returns than other nodes that only have high-quality connections because this node's investment decision will be affected by the noisy connections. Hence, adding more connections may not be beneficial to those most-connected addresses. What these addresses need is connecting with more well-connected addresses like themselves.

4.6 Conclusion

This paper reaches three conclusions. First, compared with their unconnected peers, connected investors earn higher returns. Second, the return difference also exists among the connected addresses. We divide the connected addresses into ten deciles based on their centrality, and addresses in the two most-connected deciles earn even higher returns than their rest connected peers. Third, we also find that high eigenvector centrality is more related to high returns than degree centrality, which means the quality of connections is more important than quantity.

In this paper, our data parsing rules almost certainly mislabel some addresses that are controlled by one person as owned by different people. As mentioned in the paper, to mitigate the effect of this mis-categorization issue on results, we use each address's returns instead of weighted returns in our analysis. With the improvement of data parsing technology, such as BlockSci [26], we may be able to cluster bitcoin addresses into the individual level with higher accuracy in the future. We can then more confidently calculate an individual investor's weighted returns, which may give us a more precise picture of how connectedness and centrality are related to return in the bitcoin market.

4.7 Tables

Table 4.1: The structure of a bitcoin block

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

This table is adopted from Table 9-1 and Table 9-2 in Antonopoulos [3]. Hash function can transfer data of arbitrary size into data of a fixed size. And the output can be called hash. Hash function is pre-image resistance, which means as long as the input is the same, you can always get the same output, but you are not able to deduce the input from the output. Merkle root is the root of merkle trees, which are data structures that helps nodes quickly verify transactions and reduce data transaction. BH: Block Header.

Table 4.2: The structure of a bitcoin transaction

Field	Description
Input: Transaction Hash	Pointer to the previous transaction that contains the UTXO
Input: Output Index	The location of the UTXO in the previous transaction
Input: Unlocking Script	The script to fullfill the condition set by the UTXO locking script
Output: Amount	The amount of bitcoin that should be sent to the output address
Output: Locking Script	The script that set the condition to spend the bitcoin

This table is adopted from Table 6-1 and Table 6-2 in Antonopoulos [3].

Table 4.3: Number of unique addresses in each period under different criteria

Times	Full sample	(%)	Period 1	(%)	Period 2	(%)	Period 3	(%)
≥1	370,118,349	100	116,796,653	100	139,381,552	100	121,402,453	100
≥2	203,841,132	55.07	63,967,311	54.77	74,344,299	53.34	65,959,905	54.33
≥3	24,805,525	6.70	7,511,554	6.43	10,481,970	7.52	7,979,337	6.57
≥5	12,947,120	3.50	4,135,734	3.54	5,241,400	3.76	4,143,966	3.41
≥10	6,209,197	1.68	2,144,685	1.84	2,206,217	1.58	2,025,334	1.67
≥30	1,957,869	0.53	733,001	0.63	574,290	0.41	637,315	0.52
≥50	1,106,451	0.30	418,306	0.36	308,356	0.22	354,599	0.29

This table shows how many unique transaction addresses are recorded in the bitcoin blockchain. The Full sample period covers Jun 2016-May 2021. Period 1 covers Jun 2016-May 2017, period 2 covers Jun 2017-May 2018, and period 3 covers Jun 2018-May 2019. Times $\geq n$ means the number of addresses that appear at least n time in a given period. (%) shows the percentage of addresses that appear at least n times in a given period.

Table 4.4: Summary statistics for the bitcoin investor network in different periods

	Statistics	M=10	M=30	M=50
Period 1: Jun 2016 - May 2017	Number of links	696,588,746	203,830,771	108,242,916
	Average number of links	325	95	50
	Fraction of links	0.0152%	0.0044%	0.0023%
	Maximum number of links	1,092,437	351,897	193,247
Period 2: Jun 2017 - May 2018	Number of links	736,537,035	209,150,592	92,857,510
	Average number of links	334	95	42
	Fraction of links	0.0151%	0.0043%	0.0019%
	Maximum number of links	1,234,352	311,727	162,539
Period 3: Jun 2018 - May 2019	Number of links	764,699,333	256,855,523	145,163,621
	Average number of links	378	127	72
	Fraction of links	0.0187%	0.0063%	0.0036%
	Maximum number of links	1,120,383	332,901	168,979

Fraction of links is equal to average number of links divided by the number of all the potential links a node can have, which is 2144684 in period 1, 2206216 in period 2, and 2025333 in period 3. We can see the bitcoin investor network is sparse, and network characteristics are mostly similar across different periods.

Table 4.5: Stability of the bitcoin investor network, M=50

Compared periods	period 1 - period 2	period 2 - period 3	period 1 - period 3
Number of links in period 1, k_1	114,014,182		114,014,182
Number of links in period 2, k_2	98,628,776	98,628,776	
Number of links in period 3, k_3		150,934,887	150,934,887
Overlap links, y	10,077,079	12,961,294	5,821,282
$E_{random}[y]$	675	894	1,033
$y/E_{random}[y]$	14923.97	14499.97	5633.55
$E_{degree-adjusted}[y]$	118,705	108,186	181,658
$y/E_{degree-adjusted}[y]$	84.89	119.81	32.05

This table shows the stability of the bitcoin investor network across the three periods for $M = 50$. Period 1 is from Jun 2016 to May 2017, period 2 is from Jun 2017 to May 2018, and period 3 is from Jun 2018 to May 2019. Two addresses are linked if they trade in the same direction for M times. Overlap links, y, shows the number of intersecting links between two different periods. $E_{random}[y]$ is the expected number of intersecting links between two periods if the network is random. $E_{degree-adjusted}[y]$ is another measure of the expected number of intersection links between two periods that takes the degree distribution of the original network into consideration.

Table 4.6: Summary statistics for regression variables

			Full sample	Period 1	Period 2	Period 3
Dependent Variable: Returns	Monthly Returns	Mean	0.8309	0.9910	1.1386	0.3568
		Median	0.1935	0.5444	0.0371	-0.0011
		S.D.	2.9367	2.0735	3.9475	2.2095
	Half-monthly Returns	Mean	0.8333	1.0399	1.2516	0.2002
		Median	0.3128	0.5936	0.4413	0.0192
		S.D.	3.6836	2.5052	4.8621	2.9995
	Weekly Returns	Mean	0.7619	0.9567	1.19498	0.1244
		Median	0.2882	0.4844	0.6016	0.0111
		S.D.	4.1812	2.9226	5.4118	3.5508
Independent Variables: Address Level	Trade Volume	Mean	-2.3845	-2.1297	-2.2253	-2.7991
		Median	-2.3618	-1.7852	-2.3075	-2.8483
		S.D.	3.0048	3.2893	2.7680	2.9309
	# of Trades	Mean	1.5207	1.6912	1.4057	1.4690
		Median	1.3863	1.6094	1.3863	1.3863
		S.D.	1.1093	1.1675	1.0615	1.0714
Independent Variables: Market Level	Trade Volume	Mean	18.0391	18.3343	18.2454	17.5509
		Median	18.1413	18.3185	18.3838	17.5233
		S.D.	0.4608	0.1941	0.4498	0.1403
	# of Trades	Mean	15.8895	15.8902	15.8465	15.9343
		Median	15.9061	15.9290	15.8652	15.9192
		S.D.	0.1967	0.1255	0.2287	0.2033
	Price Volatility	Mean	5.5174	3.8423	6.7238	5.7736
		Median	5.6333	4.2279	6.8567	5.6333
		S.D.	1.4528	0.9759	0.7612	0.7860
obs.			26,242,261	8,046,594	9,314,410	8,793,441

This table shows the summary statistics of the control variables in the regression analysis. Definition of these variables can be found in subsection 4.5.1. All the independent variables are in log.

Table 4.7: Regression results for connectedness and monthly returns, M=50

	Full Sample	First Period	Second Period	Third Period
Address Level Variables				
Connectedness	0.2075	0.0606	0.4475	0.1665
	(>20)	(>20)	(>20)	(>20)
Trade Volume	-0.0543	-0.0764	-0.0995	0.0000
	(<-20)	(<-20)	(<-20)	(-0.1342)
# of Trades	-0.0486	0.0812	-0.0654	-0.1074
	(<-20)	(>20)	(<-20)	(<-20)
Market Level Variables				
Price Volatility	0.0520	0.1621	0.8592	-0.4957
	(>20)	(>20)	(>20)	(<-20)
Trade Volume	1.7569	0.3722	7.1016	6.8990
	(>20)	(>20)	(>20)	(>20)
# of Trades	2.2585	4.8400	-8.2764	0.5978
	(>20)	(>20)	(<-20)	(>20)
$\bar{R}^2$	0.1203	0.1404	0.2057	0.2464
obs.	26,242,261	8,046,594	9,314,410	8,793,441

This table display the results from the regression of monthly returns on connectedness, log address-level trade volume, log address-level number of trades, log bitcoin price volatility, log market-level trade volume, and log market-level number of trades. All the results are reported with robust standard errors, and the numbers in the parentheses are the t-statistics.

Table 4.8: Regression results for groups with different centrality and monthly returns, M=50

	Degree Centrality				Eigenvector Centrality			
	Full Sample	First Period	Second Period	Third Period	Full Sample	First Period	Second Period	Third Period
Group 1	0.1876	0.1011	0.4158	0.1462	0.1386	-0.0450	0.4078	0.1160
(0-10%)	(>20)	(>20)	(>20)	(>20)	(>20)	(-8.34)	(>20)	(>20)
Group 2	0.2054	0.0549	0.4071	0.1479	0.1925	0.1200	0.3243	0.1173
(10-20%)	(>20)	(9.36)	(>20)	(>20)	(>20)	(>20)	(>20)	(>20)
Group 3	0.2195	0.0619	0.3565	0.1728	0.2190	0.0269	0.2836	0.1204
(20-30%)	(>20)	(12.33)	(>20)	(>20)	(>20)	(5.21)	(>20)	(>20)
Group 4	0.1570	0.0516	0.3520	0.1661	0.1999	0.1017	0.2870	0.1807
(30-40%)	(>20)	(10.87)	(>20)	(>20)	(>20)	(>20)	(>20)	(>20)
Group 5	0.1212	0.0347	0.3663	0.1518	0.2068	0.0625	0.3672	0.1559
(40-50%)	(>20)	(7.96)	(>20)	(>20)	(>20)	(12.69)	(>20)	(>20)
Group 6	0.1174	-0.1156	0.4063	0.1325	0.1516	-0.0080	0.3900	0.1673
(50-60%)	(>20)	(<-20)	(>20)	(>20)	(>20)	(-1.77)	(>20)	(>20)
Group 7	0.1832	-0.0496	0.3573	0.1109	0.1407	-0.0013	0.4386	0.1383
(60-70%)	(>20)	(-10.36)	(>20)	(>20)	(>20)	(-0.32)	(>20)	(>20)
Group 8	0.2049	0.0410	0.3660	0.1794	0.1479	-0.0106	0.4685	0.1935
(70-80%)	(>20)	(8.19)	(>20)	(>20)	(>20)	(-2.46)	(>20)	(>20)
Group 9	0.2262	0.1139	0.5564	0.2044	0.3357	0.0832	0.8512	0.2632
(80-90%)	(>20)	(>20)	(>20)	(>20)	(>20)	(>20)	(>20)	(>20)
Group 10	0.4414	0.2264	0.9113	0.2424	0.3241	0.2391	0.6802	0.2175
(90-100%)	(>20)	(>20)	(>20)	(>20)	(>20)	(>20)	(>20)	(>20)
$\bar{R}^2$	0.1203	0.1403	0.2057	0.2461	0.1203	0.1402	0.2058	0.2461
obs.	26,242,261	8,046,594	9,314,410	8,881,202	26,242,261	8,046,594	9,314,410	8,881,202

This table displays the results from the regression of monthly returns on groups with different centrality, controlling for log address-level trade volume, log address-level number of trades, log bitcoin price volatility, log market-level trade volume, log market-level number of trades (not reported here). Centrality is measured using either degree centrality or eigenvector centrality. All the results are reported with robust standard errors, and the numbers in the parentheses are the t-statistics.

Table 4.9: Summary statistics for regression variables in the two most-connected groups

			Full sample	Period 1	Period 2	Period 3
Dependent Variable: Returns	Monthly Returns	Mean	1.2412	1.2343	2.030	0.2739
		Median	0.5878	0.8255	1.0963	-0.0405
		S.D.	2.8193	1.8929	3.8116	2.0970
	Half-monthly Returns	Mean	1.2563	1.2820	2.1771	0.0621
		Median	0.9779	0.9219	2.1325	-0.0616
		S.D.	3.2378	2.1734	4.1974	2.7632
	Weekly Returns	Mean	1.1511	1.1498	2.1030	-0.0298
		Median	0.6767	0.6549	2.2595	-0.1317
		S.D.	3.3104	2.3794	4.2448	2.8255
Independent Variables: Address Level	Degree	Mean	8.2431	8.1647	8.1241	8.4167
		Median	8.2815	8.2522	8.1426	8.4798
		S.D.	0.4877	0.5215	0.2973	0.7885
	Eigenvector	Mean	-3.0913	-3.3058	-3.4362	-2.6858
		Median	-3.4289	-3.4325	-3.9889	-2.4117
		S.D.	1.4450	1.0562	1.3153	1.9972
	Trade Volume	Mean	-1.7646	-2.0215	-1.4546	-1.5154
		Median	-1.9147	-2.1718	-1.7810	-1.9039
		S.D.	2.6426	2.9809	2.2625	2.6075
	# of Trades	Mean	2.9812	3.1123	2.9139	2.9791
		Median	3.3322	3.4012	3.3322	3.1355
		S.D.	1.1137	1.1731	1.0801	1.1260
Independent Variables: Market Level	Trade Volume	Mean	18.1417	18.3336	18.3160	17.5365
		Median	18.2352	18.3185	18.5099	17.4930
		S.D.	0.4410	0.1946	0.4238	0.1350
	# of Trades	Mean	15.8788	15.8850	15.8681	15.9084
		Median	15.9061	15.9290	15.8652	15.9017
		S.D.	0.1812	0.1252	0.2159	0.2010
	Price Volatility	Mean	5.2657	3.7929	6.6508	5.7965
		Median	5.3863	4.2279	6.6652	5.7219
		S.D.	1.5560	0.9737	0.7649	0.7748
obs.			676,544	262,786	177,464	262,608

This table shows the summary statistics of the control variables in the regression analysis. Definition of these variables can be found in subsection 4.5.1. All the independent variables are in log.

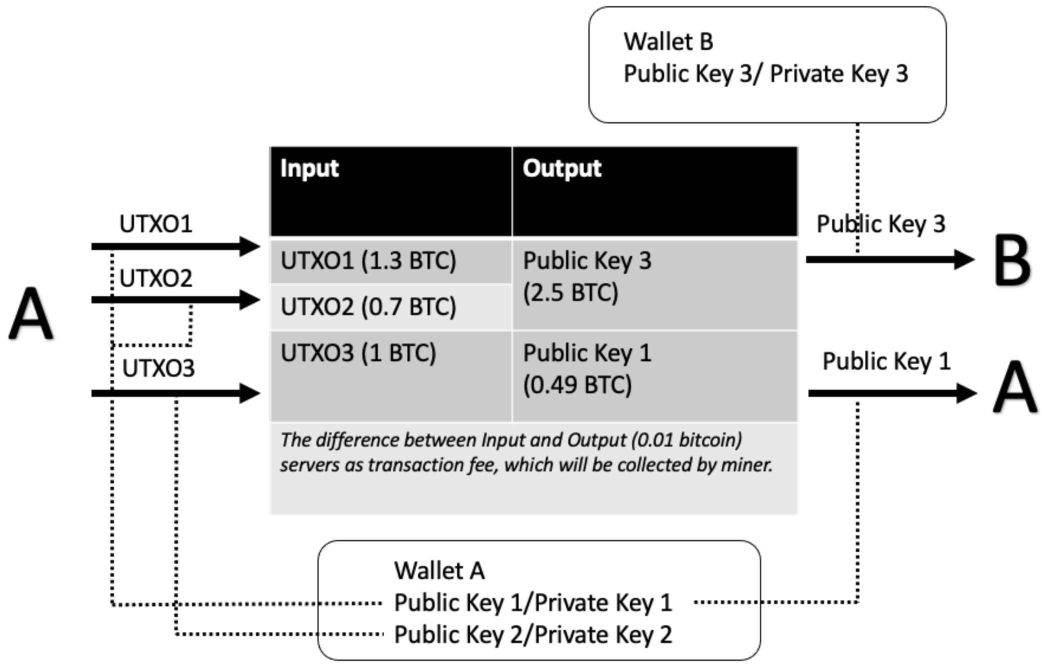
Table 4.10: Regression results for returns and centrality in the two most-connected groups

	Full Sample	First Period	Second Period	Third Period
Monthly Returns				
Degree	-0.0224	-0.1679	-0.5186	-0.0895
Centrality	(-1.27)	(-4.77)	(-7.42)	(-9.34)
Eigenvector	0.5251	0.4887	1.8362	0.1810
Centrality	(12.91)	(6.47)	(11.76)	(7.75)
Degree × Eigen	-0.0673	-0.0529	-0.2320	-0.0196
	(-13.96)	(-5.85)	(-12.11)	(-7.23)
$\bar{R}^2$	0.2336	0.2834	0.3369	0.3545
Half-monthly Returns				
Degree	-0.0737	-0.2836	-0.2964	-0.2027
Centrality	(-3.41)	(-7.14)	(-3.70)	(-14.79)
Eigenvector	0.6624	0.8449	2.1017	0.3679
Centrality	(13.43)	(9.81)	(11.82)	(11.13)
Degree × Eigen	-0.0878	-0.1010	-0.2761	-0.0402
	(-15.02)	(-9.77)	(-12.68)	(-10.49)
$\bar{R}^2$	0.2184	0.2703	0.3221	0.2406
Weekly Returns				
Degree	-0.0387	-0.2021	0.0116	-0.1422
Centrality	(-1.76)	(-4.54)	(0.14)	(-9.42)
Eigenvector	0.5314	0.5855	1.7357	0.2266
Centrality	(10.54)	(6.04)	(9.30)	(6.33)
Degree × Eigen	-0.0742	-0.0665	-0.2378	-0.0244
	(-12.41)	(-5.71)	(-10.41)	(-5.91)
$\bar{R}^2$	0.2038	0.2152	0.2394	0.1523
obs.	676,544	262,786	177,464	262,608

This table displays the results from the regression of monthly returns on addresses with different centrality in the two most-connected groups, controlling for log address-level trade volume, log address-level number of trades, log bitcoin price volatility, log market-level trade volume, log market-level number of trades (not reported here). Centrality is measured using either degree centrality or eigenvector centrality. All the results are reported with robust standard errors, and the numbers in the parentheses are the t-statistics.

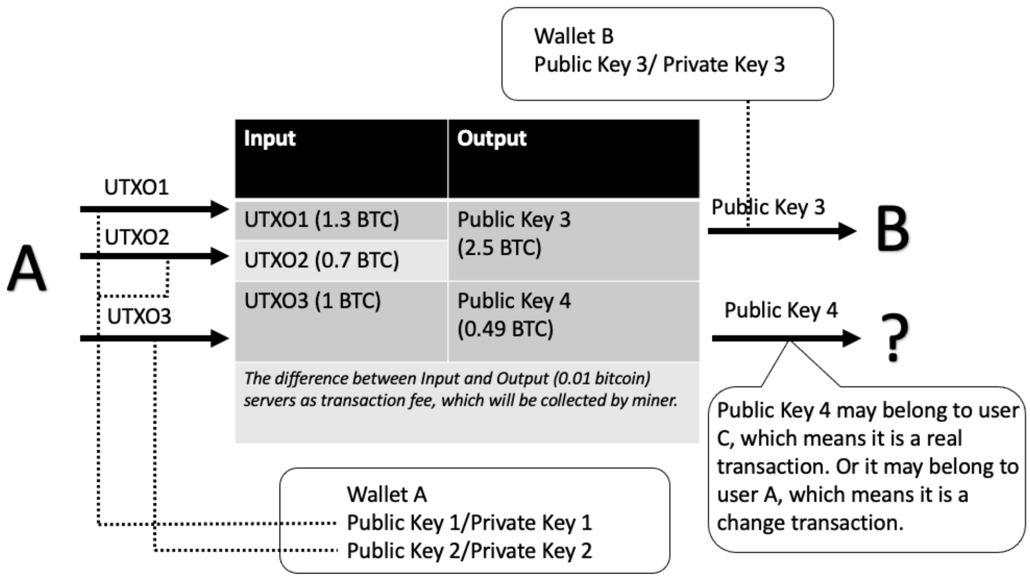
4.8 Figures

Figure 4.1: A simplified transaction framework I



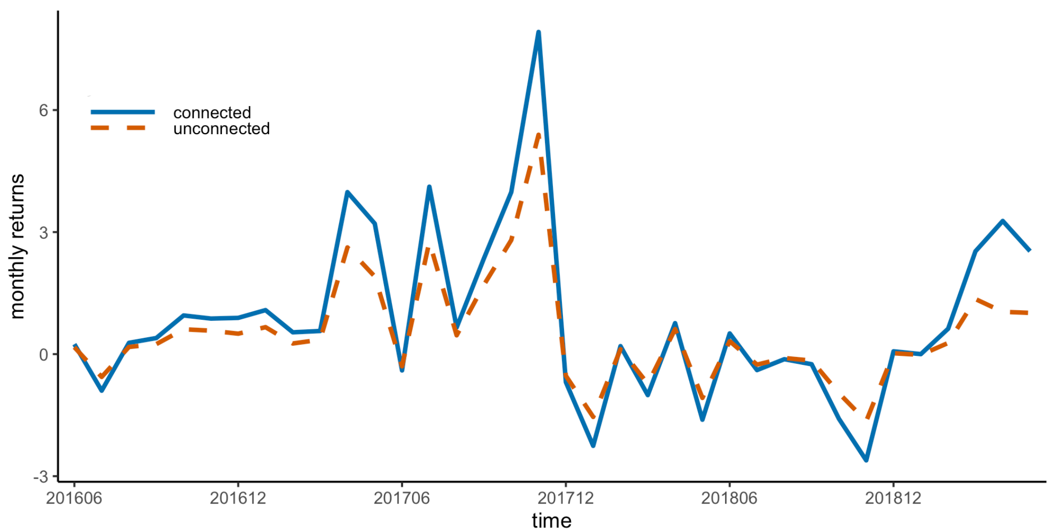
This figure shows a simple transaction. Private key 1 in wallet A controls UTXO1 and UTXO2, private key 2 controls UTXO3. User A takes out 3 bitcoins from her bitcoin wallet, and sends 2.5 bitcoins to user B, and sends back 0.49 bitcoins to herself. The 0.01 bitcoin difference serves as the transaction fee collected by the miner who successfully confirmed this transaction.

Figure 4.2: A simplified transaction framework II



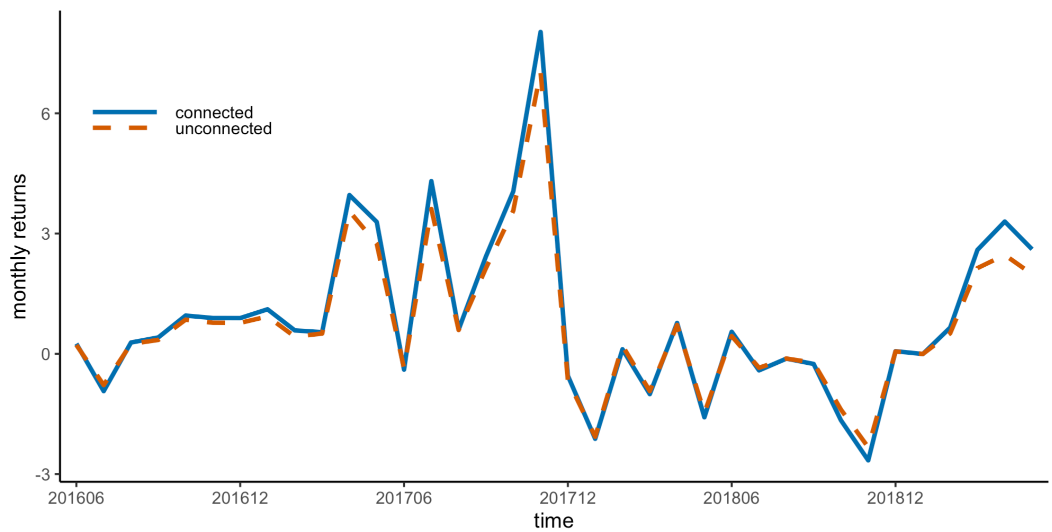
This figure shows a simple transaction. Private key 1 in wallet A controls UTXO1 and UTXO2, private key 2 controls UTXO3. User A takes out 3 bitcoins from her bitcoin wallet, and sends 2.5 bitcoins to user B. In this transaction, We do not know the identity of the person who receive the 0.49 bitcoins. It could be a new address created by user A to receive the change, or it could be owned by a third person. The 0.01 bitcoin difference serves as the transaction fee collected by the miner who successfully confirmed this transaction.

Figure 4.3: Average monthly returns of connected and unconnected groups, M=10



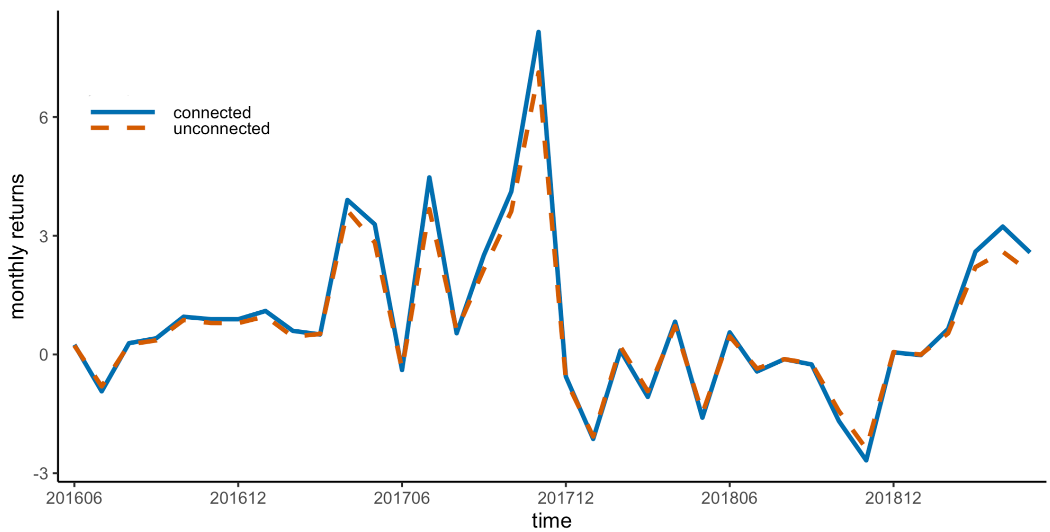
In this figure, addresses are divided into connected group and unconnected group based on the threshold $M = 10$, and then the average monthly returns of all the addresses in each group are calculated for each month.

Figure 4.4: Average monthly returns of connected and unconnected groups, M=30



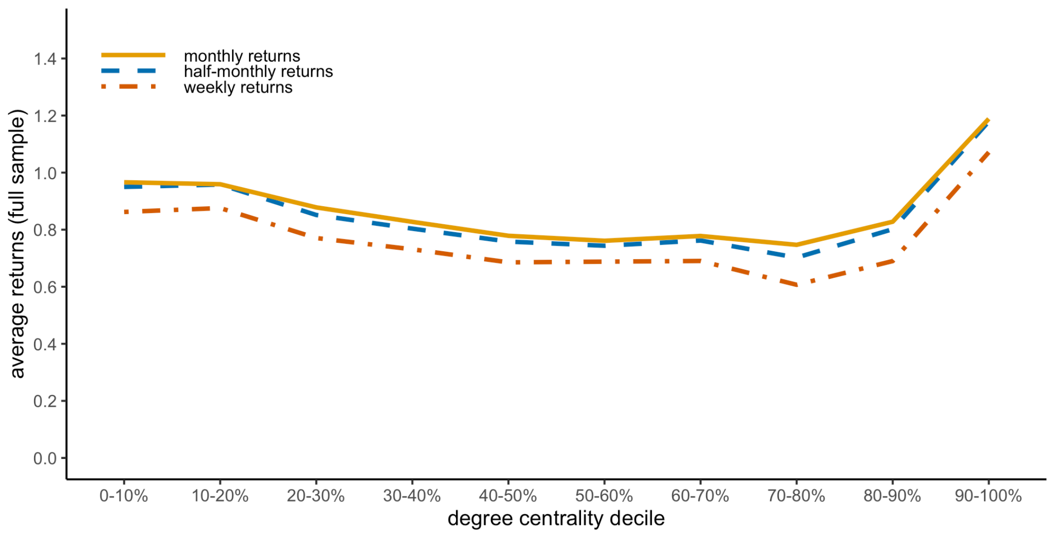
In this figure, addresses are divided into connected group and unconnected group based on the threshold $M = 30$, and then the average monthly returns of all the addresses in each group are calculated for each month.

Figure 4.5: Average monthly returns of connected and unconnected groups, M=50



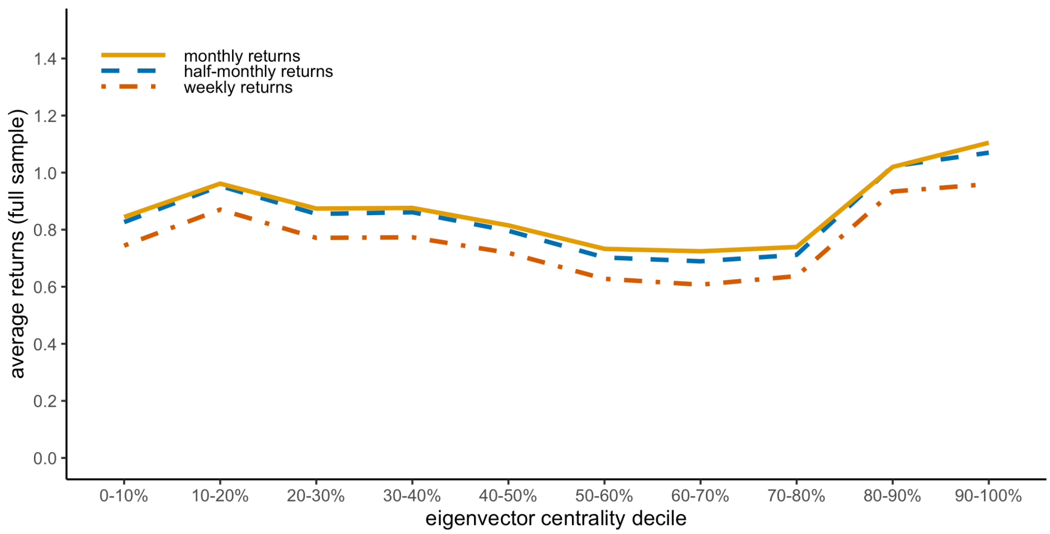
In this figure, addresses are divided into connected group and unconnected group based on threshold $M = 50$, and then the average monthly returns of all the addresses in each group are calculated for each month.

Figure 4.6: Average returns in deciles with different degree centrality



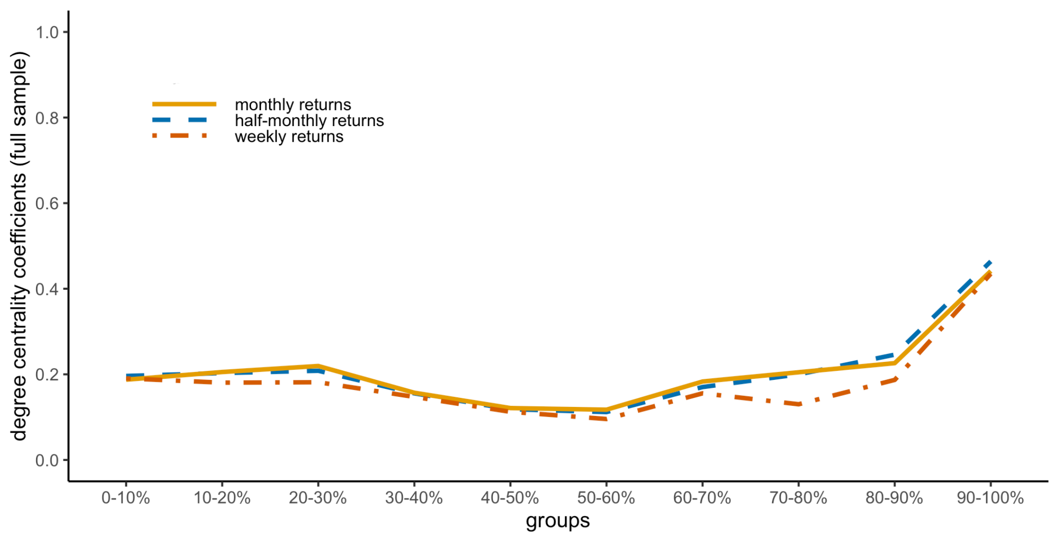
In this figure, we divide the addresses in the connected group into ten deciles based on their degree centrality. We calculate the average monthly returns of each group over the whole sample period and compare the return difference.

Figure 4.7: Average returns in deciles with different eigenvector centrality



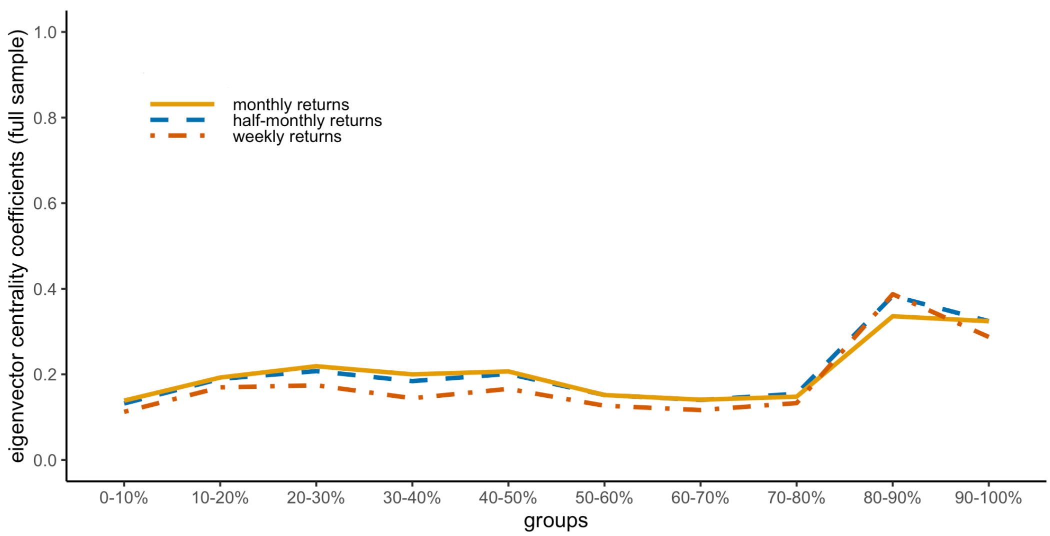
In this figure, we divide the addresses in the connected group into ten deciles based on their eigenvector centrality. We calculate the average monthly returns of each group over the whole sample period and compare the return difference.

Figure 4.8: Regression coefficients for different degree centrality



This figure shows the coefficients in the regression for returns on groups with different degree centralities. The full sample is used in this regression. Three different time horizon of returns: monthly returns, two-week returns and weekly returns are used to represent the return data.

Figure 4.9: Regression coefficients for different eigenvector centrality



This figure shows the coefficients in the regression for returns on groups with different eigenvector centralities. The full sample is used in this regression. Three different time horizon of returns: monthly returns, two-week returns and weekly returns are used to represent the return data.

Bibliography

[1] Kenneth R Ahern. Information networks: Evidence from illegal insider trading tips. *Journal of Financial Economics*, 125(1):26–47, 2017.

[2] Andreas M Antonopoulos. *The Internet of Money*. Merkle Bloom LLC, 2016.

[3] Andreas M Antonopoulos. *Mastering Bitcoin: programming the open blockchain*. O'Reilly Media, Inc., 2017.

[4] Raphael Auer. Beyond the doomsday economics of'proof-of-work'in cryptocurrencies. 2019.

[5] Brad M Barber, Yi-Tsung Lee, Yu-Jane Liu, and Terrance Odean. Just how much do individual investors lose by trading? *The Review of Financial Studies*, 22(2):609–632, 2009.

[6] Soumya Basu, David Easley, Maureen O'Hara, and Emin Sirer. Towards a functional fee market for cryptocurrencies. *Available at SSRN 3318327*, 2019.

[7] Ofir Beigel. The bitcoin mempool – a beginner's explanation. 2020. URL https://99bitcoins.com/bitcoin/mempool/.

[8] Morten Brandvold, Peter Molnár, Kristian Vagstad, and Ole Christian Andreas Valstad. Price discovery on bitcoin exchanges. *Journal of International Financial Markets, Institutions and Money*, 36:18–35, 2015.

[9] Ryan Browne. Big transaction fees are a problem for bitcoin — but there could be a solution. 2017. URL https://www.cnbc.com/2017/12/19/big-transactions-fees-are-a-problem-for-bitcoin.html.

[10] Daniel Cahill, Dirk G Baur, Zhangxin Frank Liu, and Joey W Yang. I am a blockchain too: How does the market respond to companies' interest in blockchain? *Journal of Banking & Finance*, 113:105740, 2020.

[11] Eng-Tuck Cheah, Tapas Mishra, Mamata Parhi, and Zhuang Zhang. Long memory interdependency and inefficiency in bitcoin markets. *Economics Letters*, 167:18–25, 2018.

[12] Weili Chen, Jun Wu, Zibin Zheng, Chuan Chen, and Yuren Zhou. Market manipulation of bitcoin: evidence from mining the mt. gox transaction network. In *IEEE INFOCOM 2019-IEEE Conference on Computer Communications*, pages 964–972. IEEE, 2019.

[13] Jonathan Chiu and Thorsten V Koeppl. The economics of cryptocurrencies–bitcoin and beyond. *Available at SSRN 3048124*, 2017.

[14] Shaen Corbet, Brian Lucey, Andrew Urquhart, and Larisa Yarovaya. Cryptocurrencies as a financial asset: A systematic analysis. *International Review of Financial Analysis*, 2018.

[15] Frank De Jong, Ronald Mahieu, Peter Schotman, and Irma Van Leeuwen. Price discovery on foreign exchange markets with differentially informed traders. Technical report, Tinbergen Institute Discussion Paper, 1999.

[16] Marco Del Negro and Giorgio E Primiceri. Time varying structural vector autoregressions and monetary policy: a corrigendum. *The review of economic studies*, 82(4):1342–1345, 2015.

[17] Francis X Diebold and Kamil Yilmaz. Measuring financial asset return and volatility spillovers, with application to global equity markets. *The Economic Journal*, 119(534):158–171, 2009.

[18] Nicola Dimitri. Transaction fees, block size limit, and auctions in bitcoin. *Ledger*, 4, 2019.

[19] David Easley, Maureen O'Hara, and Soumya Basu. From mining to markets: The evolution of bitcoin transaction fees. *Journal of Financial Economics*, 2019.

[20] Neil Gandal, JT Hamrick, Tyler Moore, and Tali Oberman. Price manipulation in the bitcoin ecosystem. *Journal of Monetary Economics*, 95:86–96, 2018.

[21] Rodney Garratt and Maarten RC van Oordt. Why fixed costs matter for proof-of-work based cryptocurrencies. *Available at SSRN*, 2019.

[22] Paolo Giudici and Iman Abu-Hashish. What determines bitcoin exchange prices? a network var approach. *Finance Research Letters*, 2018.

[23] Florian Glaser, Kai Zimmermann, Martin Haferkorn, Moritz Christian Weber, and Michael Siering. Bitcoin-asset or currency? revealing users' hidden intentions. *Revealing Users' Hidden Intentions (April 15, 2014). ECIS*, 2014.

[24] Gur Huberman, Jacob Leshno, and Ciamac C Moallemi. An economic analysis of the bitcoin payment system. *Columbia Business School Research Paper*, (17-92), 2019.

[25] Matthew O Jackson. *Social and economic networks*. Princeton university press, 2010.

[26] Harry Kalodner, Malte Möser, Kevin Lee, Steven Goldfeder, Martin Plattner, Alishah Chator, and Arvind Narayanan. Blocksci: Design and applications of a blockchain analysis platform. In *29th {USENIX} Security Symposium ({USENIX} Security 20)*, pages 2721–2738, 2020.

[27] Shoji Kasahara and Jun Kawahara. Effect of bitcoin fee on transaction-confirmation process. *arXiv preprint arXiv:1604.00103*, 2016.

[28] Yoshiaki Kawase and Shoji Kasahara. Transaction-confirmation time for bitcoin: a queueing analytical approach to blockchain mechanism. In *International Conference on Queueing Theory and Network Applications*, pages 75–88. Springer, 2017.

[29] Yoshiaki Kawase and Shoji Kasahara. Priority queueing analysis of transaction-confirmation time for bitcoin. *Journal of Industrial & Management Optimization*, 13(5):1, 2018.

[30] Stefan Klößner and Sven Wagner. Exploring all var orderings for calculating spillovers? yes, we can!—a note on diebold and yilmaz (2009). *Journal of Applied Econometrics*, 29(1):172–179, 2014.

[31] Olivier Kraaijeveld and Johannes De Smedt. The predictive power of public twitter sentiment for forecasting cryptocurrency prices. *Journal of International Financial Markets, Institutions and Money*, 65:101188, 2020.

[32] Ricardo Lagos and Randall Wright. A unified framework for monetary theory and policy analysis. *Journal of political Economy*, 113(3):463–484, 2005.

[33] Juanjuan Li, Yong Yuan, Shuai Wang, and Fei-Yue Wang. Transaction queuing game in bitcoin blockchain. In *2018 IEEE Intelligent Vehicles Symposium (IV)*, pages 114–119. IEEE, 2018.

[34] Yukun Liu and Aleh Tsyvinski. Risks and returns of cryptocurrency. Technical report, National Bureau of Economic Research, 2018.

[35] Yukun Liu and Aleh Tsyvinski. Risks and Returns of Cryptocurrency. *The Review of Financial Studies*, 09 2020. ISSN 0893-9454.

[36] Robert E Lucas. Macroeconomic priorities. *American economic review*, 93(1): 1–14, 2003.

[37] Nadim Mahomed et al. *Understanding consumer adoption of cryptocurrencies*. PhD thesis, University of Pretoria, 2018.

[38] Sarah Meiklejohn, Marjori Pomarole, Grant Jordan, Kirill Levchenko, Damon McCoy, Geoffrey M Voelker, and Stefan Savage. A fistful of bitcoins: characterizing payments among men with no names. In *Proceedings of the 2013 conference on Internet measurement conference*, pages 127–140, 2013.

[39] Saralees Nadarajah and Jeffrey Chu. On the inefficiency of bitcoin. *Economics Letters*, 150:6–9, 2017.

[40] Satoshi Nakamoto. Bitcoin: A peer-to-peer electronic cash system. 2008. URL https://bitcoin.org/bitcoin.pdf.

[41] Shunya Noda, Kyohei Okumura, and Yoshinori Hashimoto. A lucas critique to the difficulty adjustment algorithm of the bitcoin system. *Available at SSRN 3410460*, 2019.

[42] Han N Ozsoylev, Johan Walden, M Deniz Yavuz, and Recep Bildik. Investor networks in the stock market. *The Review of Financial Studies*, 27(5):1323–1366, 2014.

[43] Priyanka Prakash. Credit card processing fees: The complete guide, 2019. URL https://www.fundera.com/blog/credit-card-processing-fees.

[44] Giorgio E Primiceri. Time varying structural vector autoregressions and monetary policy. *The Review of Economic Studies*, 72(3):821–852, 2005.

[45] Alberto G Rossi, David Blake, Allan Timmermann, Ian Tonks, and Russ Wermers. Network centrality and delegated investment performance. *Journal of Financial Economics*, 128(1):183–206, 2018.

[46] Aviral Kumar Tiwari, RK Jana, Debojyoti Das, and David Roubaud. Informational efficiency of bitcoin—an extension. *Economics Letters*, 163:106–109, 2018.

[47] Kwok Ping Tsang and Zichao Yang. The market for bitcoin transactions. *Journal of International Financial Markets, Institutions and Money*, 71:101282, 2021.

[48] Kwok Ping Tsang and Zichao Yang. Do connections pay off in the bitcoin market? *Available at SSRN*, 2021.

[49] Andrew Urquhart. The inefficiency of bitcoin. *Economics Letters*, 148:80–82, 2016.

[50] Johan Walden. Trading, profits, and volatility in a dynamic information network model. *The Review of Economic Studies*, 86(5):2248–2283, 2019.

[51] Dingli Xi, Timothy Ian O'Brien, and Elnaz Irannezhad. Investigating the investment behaviors in cryptocurrency. *The Journal of Alternative Investments*, 2020.

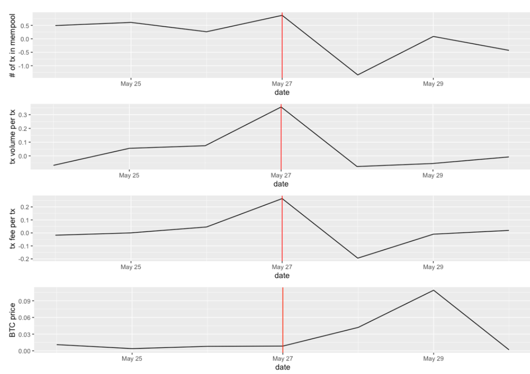
[52] Aaron Yelowitz and Matthew Wilson. Characteristics of bitcoin users: an analysis of google search data. *Applied Economics Letters*, 22(13):1030–1036, 2015.

Appendices

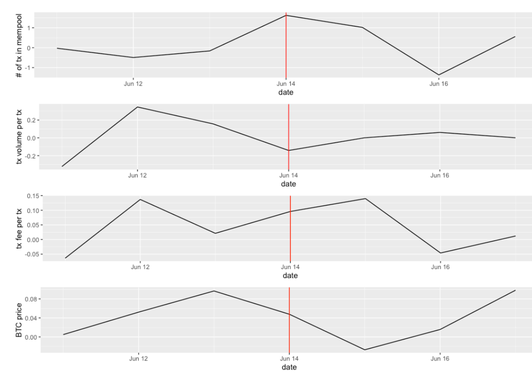
Appendix A

Chapter 1: Event study

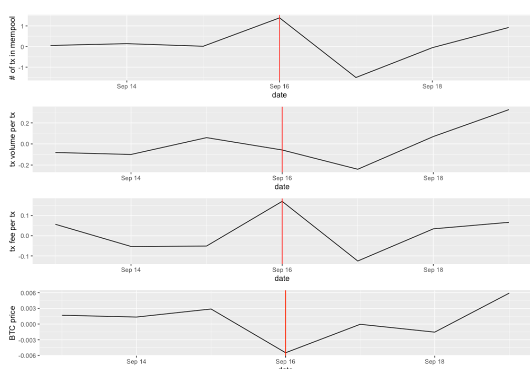
Figure A.1: Event studies in the early period



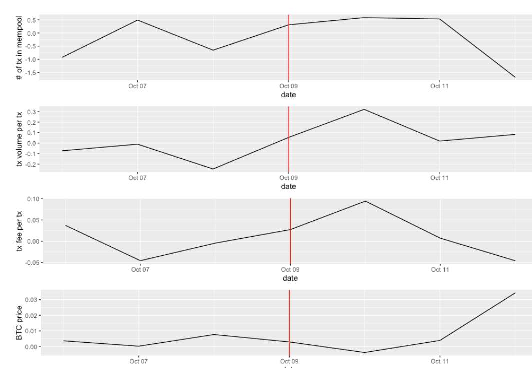
(a) Event date: 2016-05-27



(b) Event date: 2016-06-14



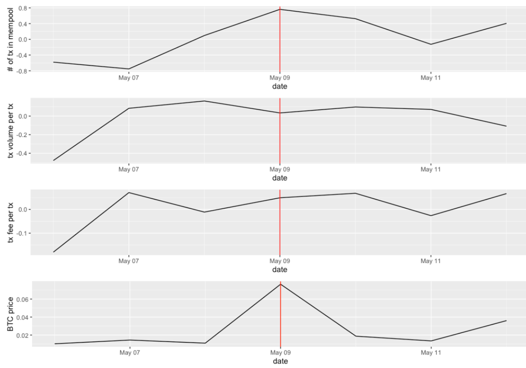
(c) Event date: 2016-09-16



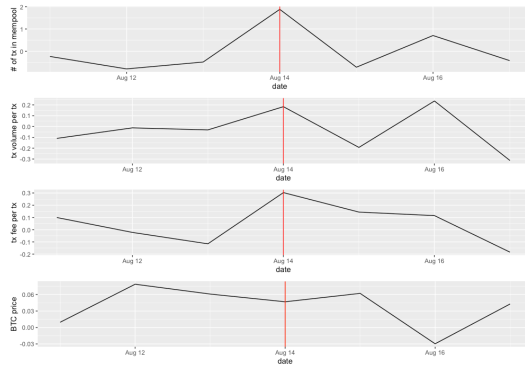
(d) Event date: 2016-10-09

(a) 2016-05-27, bitcoin price reached 6-month high. (b) 2016-06-14, bitcoin price reached 2-years high, $700. (c) 2016-09-16, Russia government blocked finnish exchange. (d) 2016-10-09, China yuan fell to its lowest level in 6 years.

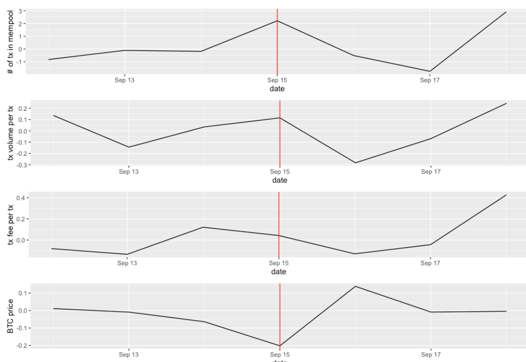
Figure A.2: Event studies in the volatile period



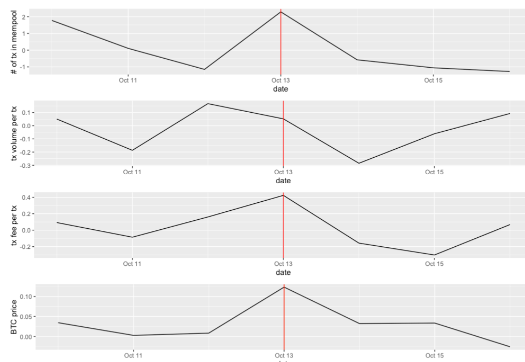
(a) Event date: 2017-05-09



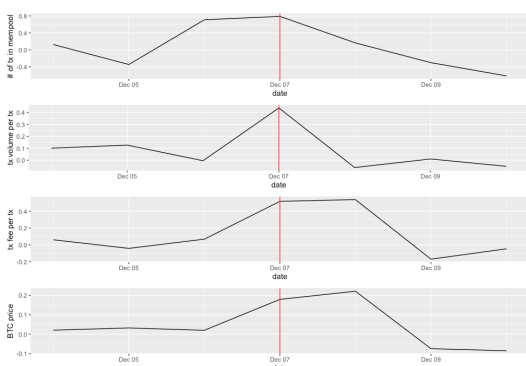
(b) Event date: 2017-08-14



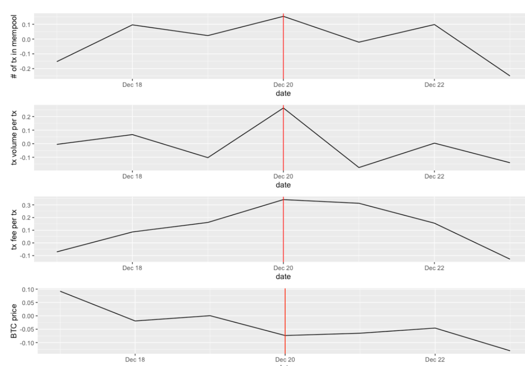
(c) Event date: 2017-09-15



(d) Event date: 2017-10-13



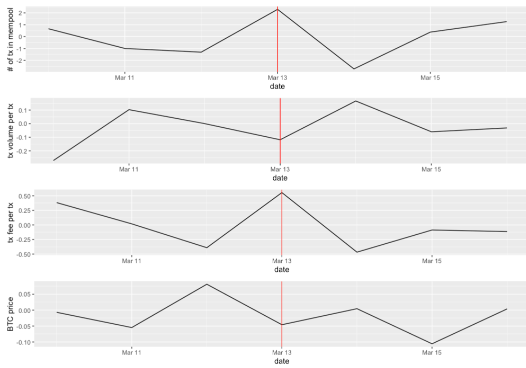
(e) Event date: 2017-12-07



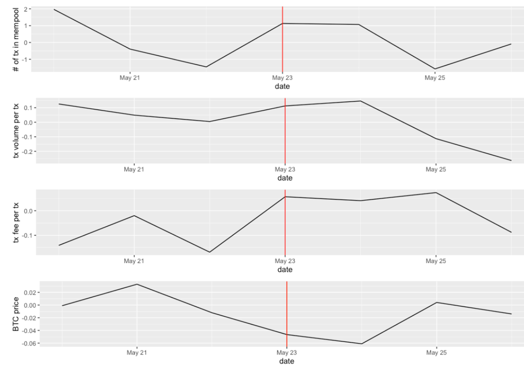
(f) Event date: 2017-12-20

(a) 2017-05-09, bitcoin price jumped from $1600 to $1700 in one day and investors think the price can reach $4000 soon. (b) 2017-08-14, bitcoin price reached $4000. (c) 2017-09-15, China's second-largest digital-currency exchange was shut down by government. (d) 2017-10-13, bitcoin price reached $5000. (e) 2017-12-07, bitcoin price passed $15000. (f) 2017-12-20, bitcoin price dropped from historical high $19000.

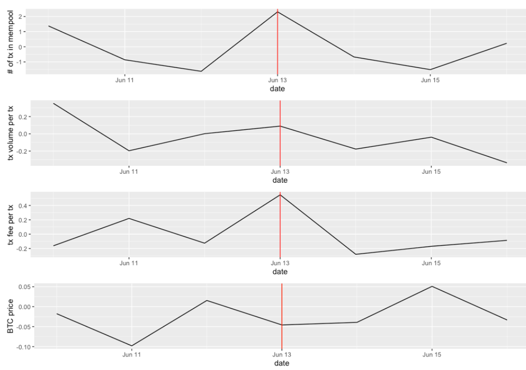
Figure A.3: Event studies in the recent period



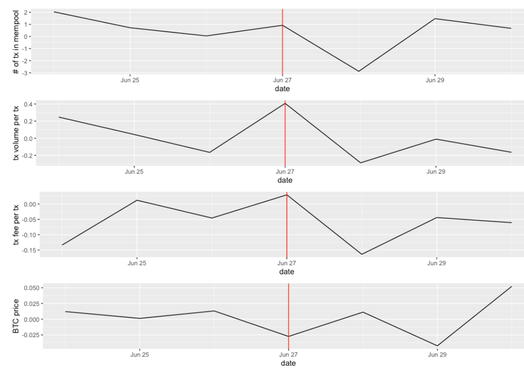
(a) Event date: 2018-03-13



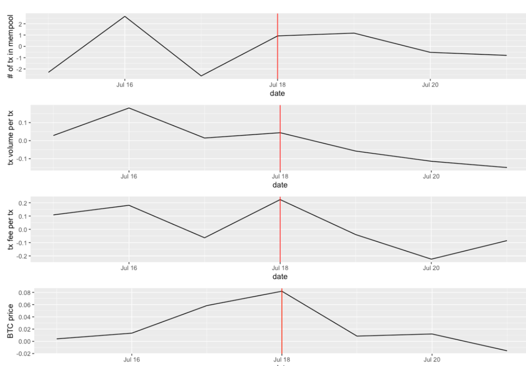
(b) Event date: 2018-05-23



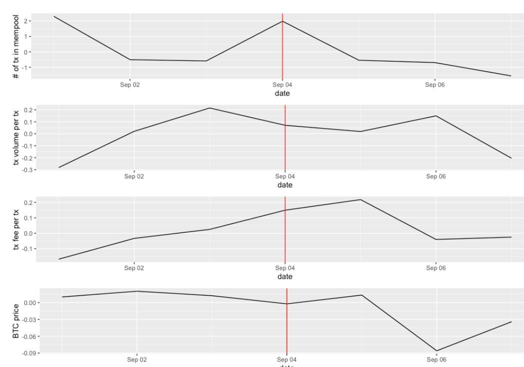
(c) Event date: 2018-06-13



(d) Event date: 2018-06-27



(e) Event date: 2018-07-18



(f) Event date: 2018-09-04

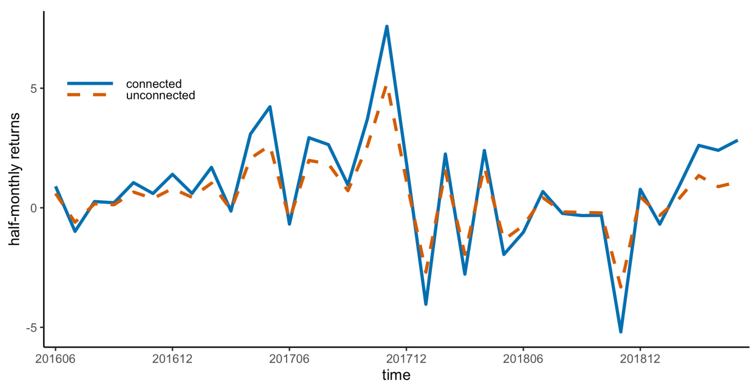
(a) 2018-03-13, IMF Managing Director Christine Lagarde calls for bitcoin crackdown. (b) 2018-05-23, U.S. Department of Justice probed bitcoin price manipulation. (c) 2018-06-13, South Korean exchange CoinRail got hacked. (d) 2018-06-27, U.S. authorities seized over \$20M in crypto in massive Darknet crackdown. (e) 2018-07-18, bitcoin price surged 10% in 30 minutes. (f) 2018-09-04, Iran officially recognized cryptocurrency mining.

Appendix B

Chapter 3: Robustness Check

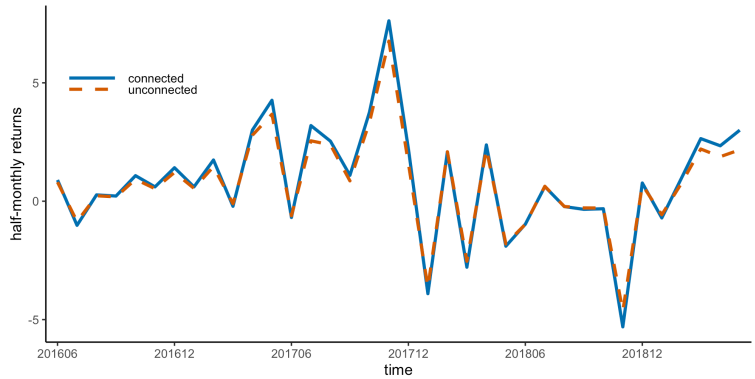
This appendix includes a number of results that serve as a series of robustness checks. One set of results use $M = 10, 30$ as the threshold to establish addresses connections in the bitcoin investor network and verify if the results will be consistent with the main results. The other set of results use 2-week returns and weekly returns, instead of monthly returns, to verify if the conclusions in the main results still hold. We find that all the results are consistent with the main results reported in the paper.

Figure B.1: Average half-monthly returns in connected and unconnected groups, M=10



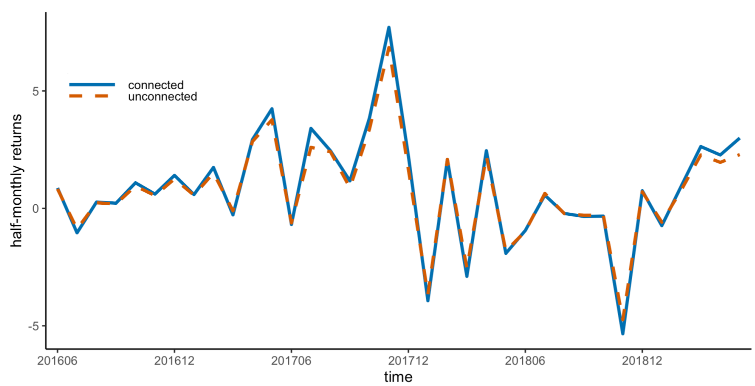
In this figure, addresses are divided into connected group and unconnected group based on threshold $M = 10$, and then the average half-monthly returns of all the addresses in each group are calculated for each month.

Figure B.2: Average half-monthly returns in connected and unconnected groups, M=30



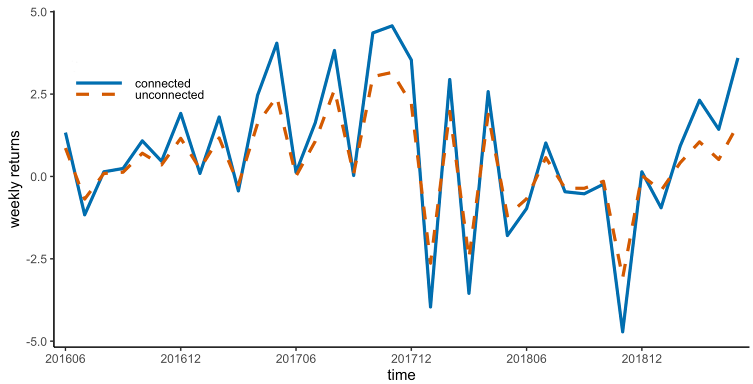
In this figure, addresses are divided into connected group and unconnected group based on threshold $M = 30$, and then the average half-monthly returns of all the addresses in each group are calculated for each month.

Figure B.3: Average half-monthly returns in connected and unconnected groups, M=50



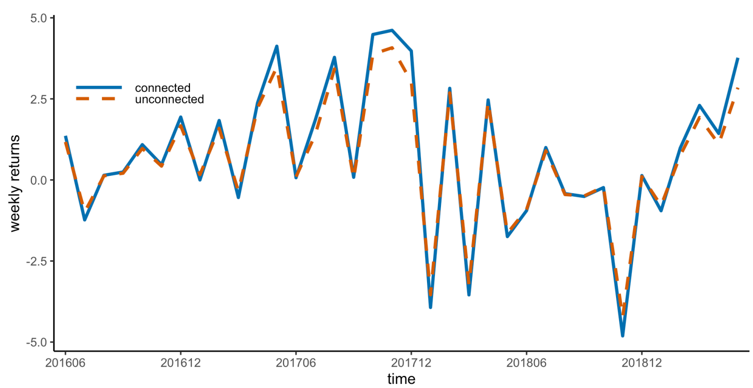
In this figure, addresses are divided into connected group and unconnected group based on threshold $M = 50$, and then the average half-monthly returns of all the addresses in each group are calculated for each month.

Figure B.4: Average weekly returns in connected and unconnected groups, M=10



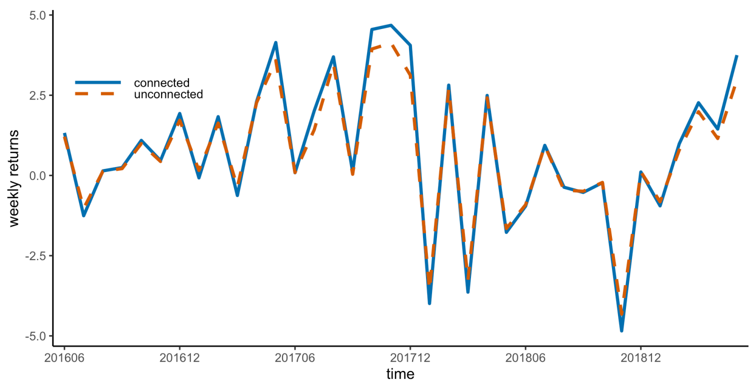
In this figure, addresses are divided into connected group and unconnected group based on threshold $M = 10$, and then the average weekly returns of all the addresses in each group are calculated for each month.

Figure B.5: Average weekly returns in connected and unconnected groups, M=30



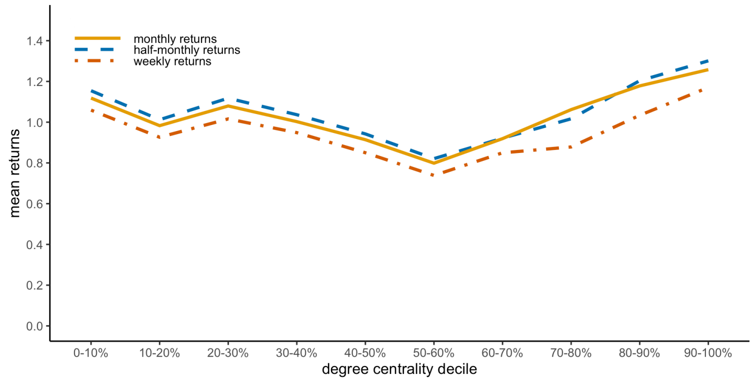
In this figure, addresses are divided into connected group and unconnected group based on threshold $M = 30$, and then the average weekly returns of all the addresses in each group are calculated for each month.

Figure B.6: Average weekly returns in connected and unconnected groups, M=50



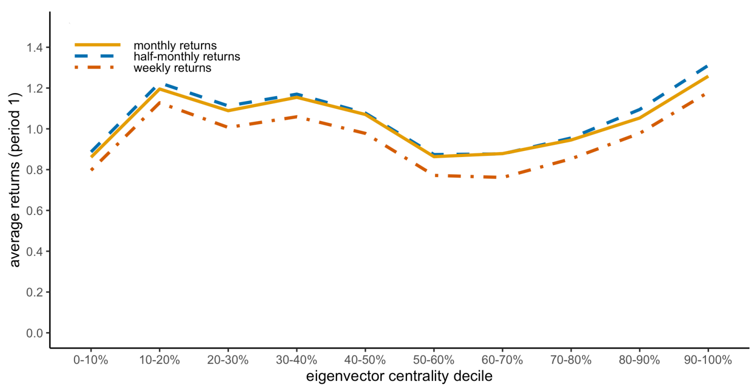
In this figure, addresses are divided into connected group and unconnected group based on threshold $M = 50$, and then the average weekly returns of all the addresses in each group are calculated for each month.

Figure B.7: Average returns in deciles with degree centrality, period 1



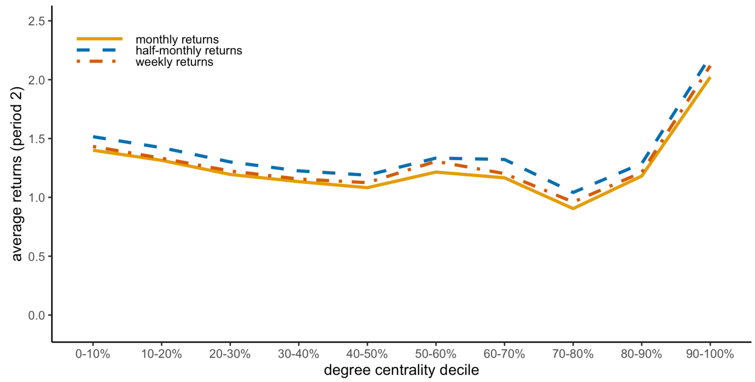
In this figure, we divide the addresses in the connected group into ten deciles based on their degree centrality. We calculate the average monthly returns of each group in period 1 and compare the return difference.

Figure B.8: Average returns in deciles with eigenvector centrality, period 1



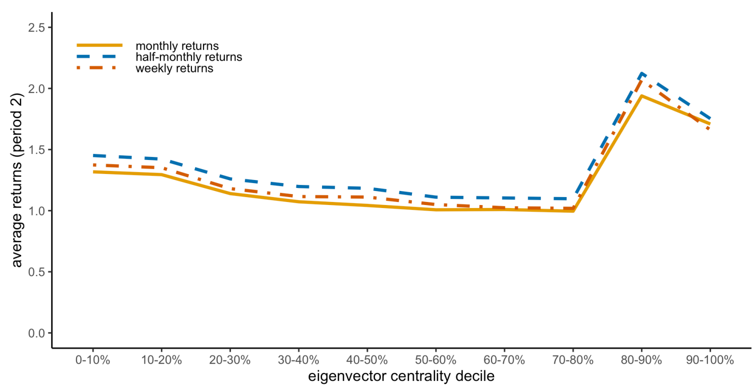
In this figure, we divide the addresses in the connected group into ten deciles based on their eigenvector centrality. We calculate the average monthly returns of each group in period 1 and compare the return difference.

Figure B.9: Average returns in deciles with degree centrality, period 2



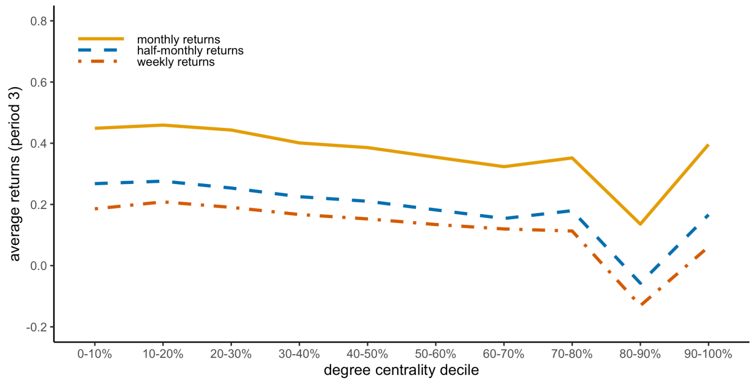
In this figure, we divide the addresses in the connected group into ten deciles based on their degree centrality. We calculate the average monthly returns of each group in period 2 and compare the return difference.

Figure B.10: Average returns in deciles with eigenvector centrality, period 2



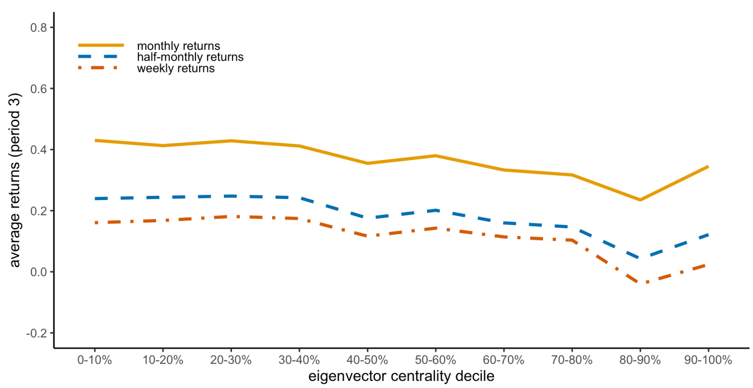
In this figure, we divide the addresses in the connected group into ten deciles based on their eigenvector centrality. We calculate the average monthly returns of each group in period 2 and compare the return difference.

Figure B.11: Average returns in deciles with degree centrality, period 3



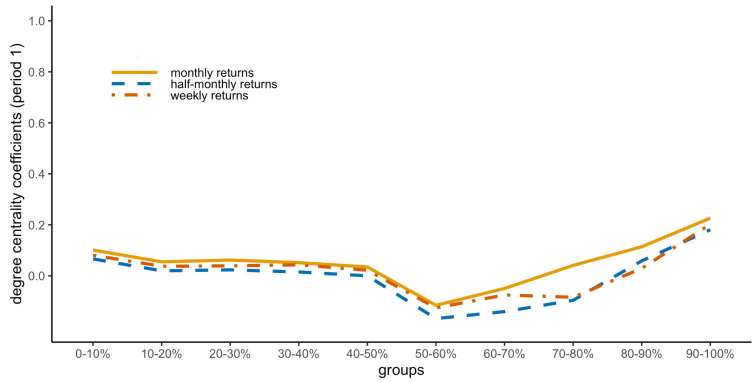
In this figure, we divide the addresses in the connected group into ten deciles based on their degree centrality. We calculate the average monthly returns of each group in period 3 and compare the return difference.

Figure B.12: Average returns in deciles with eigenvector centrality, period 3



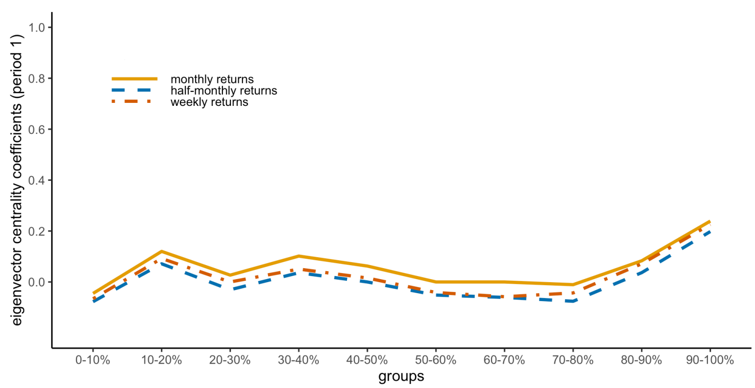
In this figure, we divide the addresses in the connected group into ten deciles based on their eigenvector centrality. We calculate the average monthly returns of each group in period 3 and compare the return difference.

Figure B.13: Regression coefficients for returns on groups with different degree centralities, period 1



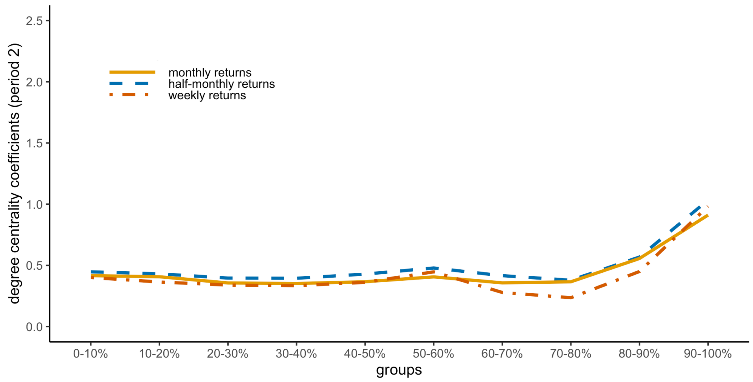
This figure shows the coefficients in the regression for returns on groups with different degree centralities. This regression focus on the first period. Three different returns: monthly returns, two-week returns and weekly returns are used to represent the return data.

Figure B.14: Regression coefficients for returns on groups with different eigenvector centralities, period 1



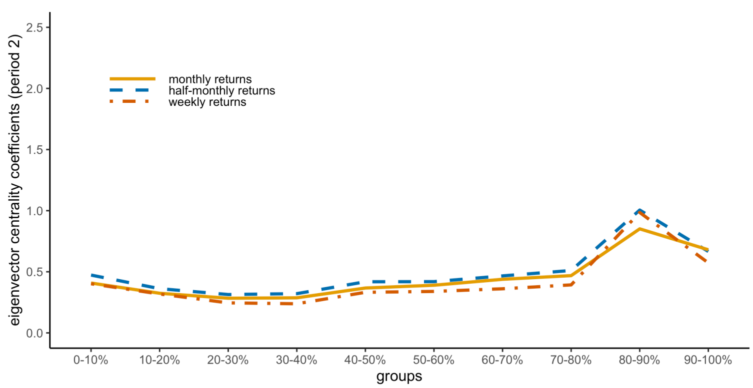
This figure shows the coefficients in the regression for returns on groups with different eigenvector centralities. This regression focus the first period. Three different returns: monthly returns, two-week returns and weekly returns are used to represent the return data.

Figure B.15: Regression coefficients for returns on groups with different degree centralities, period 2



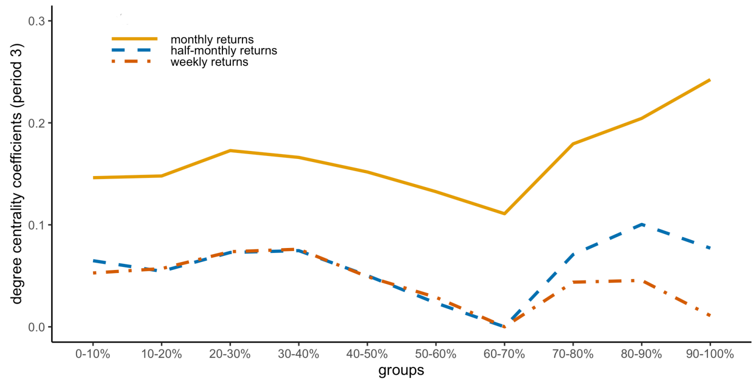
This figure shows the coefficients in the regression for returns on groups with different degree centralities. This regression focus on the second period. Three different returns: monthly returns, two-week returns and weekly returns are used to represent the return data.

Figure B.16: Regression coefficients for returns on groups with different eigenvector centralities, period 2



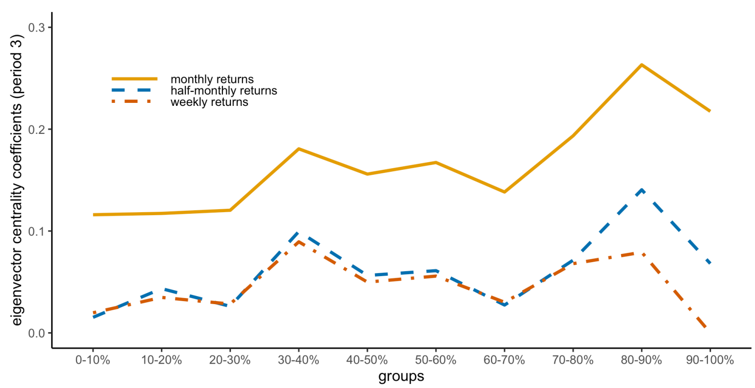
This figure shows the coefficients in the regression for returns on groups with different eigenvector centralities. This regression focus the second period. Three different returns: monthly returns, two-week returns and weekly returns are used to represent the return data.

Figure B.17: Regression coefficients for returns on groups with different degree centralities, period 3



This figure shows the coefficients in the regression for returns on groups with different degree centralities. This regression focus on the third period. Three different returns: monthly returns, two-week returns and weekly returns are used to represent the return data.

Figure B.18: Regression coefficients for returns on groups with different eigenvector centralities, period 3



This figure shows the coefficients in the regression for returns on groups with different eigenvector centralities. This regression focus the third period. Three different returns: monthly returns, two-week returns and weekly returns are used to represent the return data.

Table B.1: Stability of the bitcoin investor network, M=10

Compared periods	1-2	2-3	1-3
Number of links in period 1, k_1	702,360,012		702,360,012
Number of links in period 2, k_2	742,308,301	742,308,301	
Number of links in period 3, k_3		770,470,599	770,470,599
Overlap links, y	48,022,189	64,121,487	6,268,211
$E_{random}[y]$	31,306	34,342	32,494
$y/E_{random}[y]$	1533.95	1867.13	192.90
$E_{degree-adjusted}[y]$	4,377,357	2,700,984	4,543,429
$y/E_{degree-adjusted}[y]$	10.97	23.74	1.38

The table shows the stability of the bitcoin investor network across the three periods. The network is constructed when the connection threshold is set as $M = 10$. Period 1 is from Jun 2016 to May 2017, period 2 is from Jun 2017 to May 2018, and period 3 is from Jun 2018 to May 2019. Two agents are linked if they trade in the same direction for multiple time M. Overlap links, y, shows the number of intersecting links between two different periods. $E_{random}[y]$ is the expected number of intersecting links between two periods if the network is random. $E_{degree-adjusted}[y]$ is another measure of expected number of intersection links between two periods, but it takes the degree distributions of the original networks into consideration.

Table B.2: Stability of the bitcoin investor network, M=30

Compared periods	1-2	2-3	1-3
Number of links in period 1, k_1	209,602,037		209,602,037
Number of links in period 2, k_2	214,921,858	214,921,858	
Number of links in period 3, k_3		262,626,789	262,626,789
Overlap links, y	16,303,616	22,031,201	5,876,037
$E_{random}[y]$	2,705	3,389	3,305
$y/E_{random}[y]$	6027.26	6500.26	1777.72
$E_{degree-adjusted}[y]$	425,191	319,101	519,569
$y/E_{degree-adjusted}[y]$	38.34	69.04	11.31

The table shows the stability of the bitcoin investor network across the three periods. The network is constructed when the connection threshold is set as $M = 30$. Period 1 is from Jun 2016 to May 2017, period 2 is from Jun 2017 to May 2018, and period 3 is from Jun 2018 to May 2019. Two agents are linked if they trade in the same direction for multiple time M. Overlap links, y, shows the number of intersecting links between two different periods. $E_{random}[y]$ is the expected number of intersecting links between two periods if the network is random. $E_{degree-adjusted}[y]$ is another measure of expected number of intersection links between two periods, but it takes the degree distributions of the original networks into consideration.

Table B.3: Regression results for connectedness and monthly returns, M=30

	Full Sample	First Period	Second Period	Third Period
Address Level Variables				
Connectedness	0.2401	0.1150	0.4169	0.1959
	(>20)	(>20)	(>20)	(>20)
Trade Volume	-0.0533	-0.0749	-0.0985	0.0001
	(<-20)	(<-20)	(<-20)	(0.4553)
# of Trades	-0.0625	0.0691	-0.0787	-0.1194
	(<-20)	(>20)	(<-20)	(<-20)
Market Level Variables				
Price Volatility	0.0535	0.1645	0.8576	-0.4944
	(>20)	(>20)	(>20)	(<-20)
Trade Volume	1.7645	0.3756	7.0990	0.6902
	(>20)	(>20)	(>20)	(>20)
# of Trades	2.2483	4.8171	-8.2662	0.5964
	(>20)	(>20)	(<-20)	(>20)
$\bar{R}^2$	0.1207	0.1405	0.2059	0.2467
obs.	26,181,341	8,026,523	9,296,250	8,858,550

This table display the results from the regression of monthly returns on connectedness, and other control variables (log-form) . All the results are reported with robust standard errors. The numbers in the parentheses represent the t-statistics.

Table B.4: Regression results for connectedness and monthly returns, M=10

	Full Sample	First Period	Second Period	Third Period
Address Level Variables				
Connectedness	0.4114	0.3683	0.4739	0.4132
	(>20)	(>20)	(>20)	(>20)
Trade Volume	-0.0507	-0.0686	-0.0978	0.0017
	(<-20)	(<-20)	(<-20)	(6.788)
# of Trades	-0.0580	0.0541	-0.0590	-0.1155
	(<-20)	(>20)	(<-20)	(<-20)
Market Level Variables				
Price Volatility	0.0502	0.1665	0.8589	-0.4939
	(>20)	(>20)	(>20)	(<-20)
Trade Volume	1.7690	0.3760	7.0990	0.6902
	(>20)	(>20)	(>20)	(>20)
# of Trades	2.2602	4.7901	-8.2644	0.6226
	(>20)	(>20)	(<-20)	(>20)
$\bar{R}^2$	0.1223	0.1441	0.2058	0.2498
obs.	25,924,373	7,954,317	9,199,468	8,770,535

This table display the results from the regression of monthly returns on connectedness, and other control variables (log-form) . All the results are reported with robust standard errors. The numbers in the parentheses represent the t-statistics.

Table B.5: Regression results for connectedness and 2-week returns, M=50

	Full Sample	First Period	Second Period	Third Period
Address Level Variables				
Connectedness	0.2082	0.0074	0.4934	0.0610
	(>20)	(3.295)	(>20)	(>20)
Trade Volume	-0.0585	-0.0807	-0.1129	-0.0019
	(<-20)	(<-20)	(<-20)	(-5.429)
# of Trades	-0.0536	0.1097	-0.0887	-0.0610
	(<-20)	(>20)	(<-20)	(<-20)
Market Level Variables				
Price Volatility	0.1030	0.5761	2.0104	-0.9339
	(>20)	(>20)	(>20)	(<-20)
Trade Volume	1.9569	0.0848	9.1081	6.0760
	(>20)	(>20)	(>20)	(>20)
# of Trades	2.3324	2.7412	-11.5520	0.8139
	(>20)	(>20)	(<-20)	(>20)
$\bar{R}^2$	0.0906	0.1276	0.1684	0.1536
obs.	26,242,261	8,046,594	9,314,410	8,881,202

This table display the results from the regression of 2-week returns on connectedness, and other control variables (log-form) . All the results are reported with robust standard errors. The numbers in the parentheses represent the t-statistics.

Table B.6: Regression results for connectedness and 2-week returns, M=30

	Full Sample	First Period	Second Period	Third Period
Address Level Variables				
Connectedness	0.2390	0.0769	0.4647	0.0984
	(>20)	(>20)	(>20)	(>20)
Trade Volume	-0.0575	-0.0792	-0.1118	-0.0017
	(<-20)	(<-20)	(<-20)	(-4.926)
# of Trades	-0.0671	0.0959	-0.1039	-0.0707
	(<-20)	(>20)	(<-20)	(<-20)
Market Level Variables				
Price Volatility	0.1048	0.5787	2.0086	-0.9324
	(>20)	(>20)	(>20)	(<-20)
Trade Volume	1.9643	0.0888	9.1059	6.0769
	(>20)	(>20)	(>20)	(>20)
# of Trades	2.3225	2.7158	-11.5408	0.8127
	(>20)	(>20)	(<-20)	(>20)
$\bar{R}^2$	0.0908	0.1274	0.1685	0.1534
obs.	26,181,341	8,026,523	9,296,250	8,858,550

This table display the results from the regression of 2-week returns on connectedness, and other control variables (log-form) . All the results are reported with robust standard errors. The numbers in the parentheses represent the t-statistics.

Table B.7: Regression results for connectedness and 2-week returns, M=10

	Full Sample	First Period	Second Period	Third Period
Address Level Variables				
Connectedness	0.4105	0.3628	0.5330	0.3055
	(>20)	(>20)	(>20)	(>20)
Trade Volume	-0.0550	-0.0723	-0.1113	-0.0005
	(<-20)	(<-20)	(<-20)	(-1.346)
# of Trades	-0.0624	0.0736	-0.0791	-0.0761
	(<-20)	(>20)	(<-20)	(<-20)
Market Level Variables				
Price Volatility	0.1020	0.5817	2.0102	-0.9300
	(>20)	(>20)	(>20)	(<-20)
Trade Volume	1.9695	0.0932	9.1099	6.0760
	(>20)	(>20)	(>20)	(>20)
# of Trades	2.3344	2.6783	-11.5440	0.8327
	(>20)	(>20)	(<-20)	(>20)
$\bar{R}^2$	0.0917	0.1297	0.1683	0.1540
obs.	25,924,373	7,954,317	9,199,468	8,770,535

This table display the results from the regression of 2-week returns on connectedness, and other control variables (log-form). All the results are reported with robust standard errors. The numbers in the parentheses represent the t-statistics.

Table B.8: Regression results for connectedness and weekly returns, M=50

	Full Sample	First Period	Second Period	Third Period
Address Level Variables				
Connectedness	0.1827	0.0277	0.4180	0.0472
	(>20)	(10.53)	(>20)	(15.11)
Trade Volume	-0.0609	-0.0698	-0.1242	-0.0065
	(<-20)	(<-20)	(<-20)	(-15.13)
# of Trades	-0.0345	0.0619	-0.0488	-0.0348
	(<-20)	(>20)	(<-20)	(<-20)
Market Level Variables				
Price Volatility	0.1529	0.7357	1.5492	-0.4930
	(>20)	(>20)	(>20)	(<-20)
Trade Volume	1.7971	0.1113	6.7777	3.4794
	(>20)	(>20)	(>20)	(>20)
# of Trades	2.4357	1.0374	-7.3683	2.2461
	(>20)	(>20)	(<-20)	(>20)
$\bar{R}^2$	0.0645	0.0844	0.0899	0.0749
obs.	26,242,261	8,046,594	9,314,410	8,881,202

This table display the results from the regression of weekly returns on connectedness, and other control variables (log-form). All the results are reported with robust standard errors. The numbers in the parentheses represent the t-statistics.

Table B.9: Regression results for connectedness and weekly returns, M=30

	Full Sample	First Period	Second Period	Third Period
Address Level Variables				
Connectedness	0.2137	0.0873	0.4147	0.0789
	(>20)	(>20)	(>20)	(>20)
Trade Volume	-0.0600	-0.0684	-0.1230	-0.0064
	(<-20)	(<-20)	(<-20)	(-14.80)
# of Trades	-0.0469	0.0493	-0.0656	-0.0428
	(<-20)	(>20)	(<-20)	(<-20)
Market Level Variables				
Price Volatility	0.1544	0.7377	1.5470	-0.4920
	(>20)	(>20)	(>20)	(<-20)
Trade Volume	1.8034	0.1144	6.7752	3.4814
	(>20)	(>20)	(>20)	(>20)
# of Trades	2.4275	1.0163	-7.3555	2.2434
	(>20)	(>20)	(<-20)	(>20)
$\bar{R}^2$	0.0646	0.0843	0.0901	0.0748
obs.	26,181,341	8,026,523	9,296,250	8,858,550

This table display the results from the regression of weekly returns on connectedness, and other control variables (log-form). All the results are reported with robust standard errors. The numbers in the parentheses represent the t-statistics.

Table B.10: Regression results for connectedness and weekly returns, M=10

	Full Sample	First Period	Second Period	Third Period
Address Level Variables				
Connectedness	0.3715	0.3409	0.5002	0.2537
	(>20)	(>20)	(>20)	(>20)
Trade Volume	-0.0579	-0.0623	-0.1226	-0.0054
	(<-20)	(<-20)	(<-20)	(-12.44)
# of Trades	-0.0428	0.0307	-0.0450	-0.0472
	(<-20)	(>20)	(<-20)	(<-20)
Market Level Variables				
Price Volatility	0.1517	0.7389	1.5486	-0.4905
	(>20)	(>20)	(>20)	(<-20)
Trade Volume	1.8077	0.1147	6.7830	3.4866
	(>20)	(>20)	(>20)	(>20)
# of Trades	2.4391	0.9927	-7.3640	2.2547
	(>20)	(>20)	(<-20)	(>20)
$\bar{R}^2$	0.0651	0.0857	0.0901	0.0749
obs.	25,924,373	7,954,317	9,199,468	8,770,535

This table display the results from the regression of weekly returns on connectedness, and other control variables (log-form). All the results are reported with robust standard errors. The numbers in the parentheses represent the t-statistics.

Table B.11: Regression results for groups with different centrality and half-monthly returns, M=50

	Degree Centrality				Eigenvector Centrality			
	Full Sample	First Period	Second Period	Third Period	Full Sample	First Period	Second Period	Third Period
Group 1	0.1962	0.0663	0.4474	0.0648	0.1320	-0.0774	0.4731	0.0153
(0-10%)	(>20)	(14.00)	(>20)	(10.89)	(>20)	(-12.35)	(>20)	(2.25)
Group 2	0.2029	0.0196	0.4308	0.0545	0.1894	0.0713	0.3618	0.0435
(10-20%)	(>20)	(2.85)	(>20)	(7.05)	(>20)	(12.28)	(>20)	(6.30)
Group 3	0.2083	0.0234	0.3962	0.0729	0.2075	-0.0304	0.3138	0.0261
(20-30%)	(>20)	(4.00)	(>20)	(11.29)	(>20)	(-5.09)	(>20)	(3.84)
Group 4	0.1552	0.0150	0.3946	0.0747	0.1842	0.0362	0.3210	0.0995
(30-40%)	(>20)	(2.68)	(>20)	(11.01)	(>20)	(6.14)	(>20)	(14.49)
Group 5	0.1186	0.0020	0.4288	0.0503	0.2010	-0.0093	0.4184	0.0564
(40-50%)	(>20)	(0.392)	(>20)	(7.56)	(>20)	(-1.61)	(>20)	(8.19)
Group 6	0.1119	-0.1677	0.4786	0.0234	0.1517	-0.0515	0.4193	0.0611
(50-60%)	(>20)	(<-20)	(>20)	(3.44)	(>20)	(-9.48)	(>20)	(9.20)
Group 7	0.1703	-0.1402	0.4158	-0.0019	0.1401	-0.0606	0.4663	0.0274
(60-70%)	(>20)	(<-20)	(>20)	(-0.29)	(>20)	(-11.88)	(>20)	(4.17)
Group 8	0.1999	-0.0961	0.3789	0.0788	0.1547	-0.0758	0.5106	0.0714
(70-80%)	(>20)	(-16.27)	(>20)	(10.89)	(>20)	(-14.78)	(>20)	(11.48)
Group 9	0.2460	0.0587	0.5684	0.1004	0.3834	0.0366	1.0049	0.1405
(80-90%)	(>20)	(11.58)	(>20)	(15.18)	(>20)	(7.80)	(>20)	(>20)
Group 10	0.4639	0.1808	1.0336	0.0771	0.3242	0.1980	0.6664	0.0681
(Top 10%)	(>20)	(>20)	(>20)	(11.33)	(>20)	(>20)	(>20)	(9.65)
$\bar{R}^2$	0.0906	0.1275	0.1684	0.1533	0.0906	0.1274	0.1684	0.1534
obs.	26,242,261	8,046,594	9,314,410	8,881,202	26,242,261	8,046,594	9,314,410	8,881,202

This table display the results from the regression of half-monthly returns on groups with different centrality, controlling log address-level trade volume, log address-level number of trades, log bitcoin price volatility, log market-level trade volume, log market-level number of trade (not reported here). Centrality is measured using either degree centrality or eigenvector centrality. All the results are reported with robust standard errors. The numbers in the parentheses represent the t-statistics.

Table B.12: Regression results for groups with different centrality and weekly returns, M=50

	Degree Centrality				Eigenvector Centrality			
	Full Sample	First Period	Second Period	Third Period	Full Sample	First Period	Second Period	Third Period
Group 1	0.1905	0.0805	0.4007	0.0528	0.1124	-0.0659	0.4019	0.0197
(0-10%)	(>20)	(14.66)	(>20)	(7.51)	(>20)	(-9.15)	(>20)	(2.55)
Group 2	0.1806	0.0376	0.3643	0.0571	0.1696	0.0922	0.3182	0.0348
(10-20%)	(>20)	(4.71)	(>20)	(6.29)	(>20)	(13.75)	(>20)	(4.32)
Group 3	0.1816	0.0391	0.3390	0.0737	0.1743	-0.0137	0.2457	0.0284
(20-30%)	(>20)	(5.86)	(>20)	(9.74)	(>20)	(-1.98)	(18.95)	(3.57)
Group 4	0.1473	0.0431	0.3342	0.0761	0.1442	0.0505	0.2383	0.0894
(30-40%)	(>20)	(6.62)	(>20)	(9.71)	(>20)	(7.47)	(18.26)	(11.16)
Group 5	0.1124	0.0218	0.3613	0.0492	0.1660	0.0155	0.3322	0.0499
(40-50%)	(>20)	(3.58)	(>20)	(6.40)	(>20)	(2.32)	(>20)	(6.31)
Group 6	0.0956	-0.1258	0.4471	0.0290	0.1267	-0.0414	0.3393	0.0558
(50-60%)	(19.00)	(<-20)	(>20)	(3.72)	(>20)	(-6.41)	(>20)	(7.31)
Group 7	0.1557	-0.0751	0.2786	0.0095	0.1165	-0.0579	0.3610	0.0301
(60-70%)	(>20)	(-11.68)	(>20)	(1.27)	(>20)	(-9.64)	(>20)	(4.06)
Group 8	0.1301	-0.0840	0.2357	0.0438	0.1328	-0.0431	0.3928	0.0680
(70-80%)	(>20)	(-12.15)	(18.21)	(6.09)	(>20)	(-7.23)	(>20)	(9.63)
Group 9	0.1870	0.0292	0.4500	0.0455	0.3873	0.0714	0.9850	0.0792
(80-90%)	(>20)	(4.86)	(>20)	(6.26)	(>20)	(13.23)	(>20)	(11.46)
Group 10	0.4350	0.2023	0.9827	0.0109	0.2873	0.2258	0.5731	0.0067
(Top 10%)	(>20)	(>20)	(>20)	(1.48)	(>20)	(>20)	(>20)	(0.87)
$\bar{R}^2$	0.0645	0.0843	0.0899	0.0748	0.0645	0.0843	0.0900	0.0748
obs.	26,242,261	8,046,594	9,314,410	8,881,202	26,242,261	8,046,594	9,314,410	8,881,202

This table display the results from the regression of weekly returns on groups with different centrality, controlling log address-level trade volume, log address-level number of trades, log bitcoin price volatility, log market-level trade volume, log market-level number of trade (not reported here). Centrality is measured using either degree centrality or eigenvector centrality. All the results are reported with robust standard errors. The numbers in the parentheses represent the t-statistics.