Liquidity Demand and BitCoin Transaction Fees

Alfred Lehar*
Haskayne School of Business
University of Calgary

Christine A. Parlour
Haas School of Business
UC Berkeley

May 2019
Preliminary and incomplete

*Corresponding author, Haskayne School of Business, University of Calgary, 2500 University Drive NW, Calgary, Alberta, Canada T2N 1N4. e-mail: alfred.lehar@haskayne.ucalgary.ca, Tel: (403) 220 4567.

Liquidity Demand and Bitcoin Transaction Fees

Abstract

We document a high variation of bitcoin fees not only over time but also within blocks. Statistics on average transaction fees can be misleading because they are driven by some severe outlyers. We find that transaction size in bytes is more important for determining fees than dollar or bitcoin value. Even though fees tend to be higher when locks are fuller, the blockchain rarely runs at capacity. Fees are higher when receipeints spend their finds sooner, when the payment is to and from an exchange, and during weekdays.

Keywords: Bitcoin, Transaction Costs, Blockchain

1 Introduction

Bitcoin is often portrayed as impractical for everyday use because of the high transaction fees. Indeed average fees as reported by blockchain analysis websites such as BitInfoCharts go up to $54.90 per transaction.[1] We argue that averages are misleading because there exist huge variations within the same block. Even at the peak times in Dec 2017 where somebody was willing to pay over $13.000 to have a transaction mined into a block, a similar transaction was included in the same block for a fee of 7 cent. A high dispersion in fees within the same block persists throughout the whole bitcoin blockchain.

In this paper we shed some light on the determinants of Bitcoin transaction fees. Analyzing a sample of over 350 million transactions we document huge variations in fees both over time and contemporaneously. We find that transaction size in bytes is far more important in determining fees than transaction amounts in BTC (or USD). Since the size of a bitcoin block is limited the space a transaction takes in the block determines a miner's opportunity cost.

While fees tend to be higher when blocks are fuller we find that the bitcoin blockchain rarely runs at capacity. Even in times of high transaction demand we observe empty or partially filled blocks which is not fully in line with the often floated explanation of 'congestion pricing'. We further document that fees are higher when the recipient spends the output faster, consistent with some liquidity need of the receiver. We find that transaction fees are lower over the weekend, and that over time fees ending in round numbers of Satoshis decline, indicating that algorithms instead of humans are becoming more prominent in setting fees. Using a sub-sample of identifiable wallets we find that payments to and from exchanges have substantially higher fees.

Our paper fits small and growing literature on transaction fees in blockchain systems. Easley, O'Hara, and Basu (2019) explain the observed shift from no-fee to fee paying transactions and model the interactions of fee payments and waiting times. While the focus of their empirical analysis is on the time series of average transaction fees our paper documents a huge variation of Bitcoin transaction within blocks analyses the cross section of transaction fees. Huberman, Leshno, and Moallemi (2017) compare Bitcoin to a traditional payment system and derive closed form solutions for equilibrium fees.

This research on fees fits into a larger body of literature that focuses on the economics and incentives

[1] see for example 'Big transaction fees are a problem for bitcoin — but there could be a solution' on CNBC, Dec 21, 2017

in blockchain ecosystems (among others Abadi and Brunnermeier (2018), Cong and He (2019), Budish (2018)) and the impact on financial markets (e.g. Malinova and Park (2017) or Brauneis, Mestel, Riordan, and Theissen (2018)). Cong, He, and Li (2019) analyze the incentives for miners to form pools. Other research focuses on the pricing of cyrypto-currencies in the market including frictions causing pricing differences (e.g. Hu, Parlour, and Rajan (2018), Makarov and Schoar (2018), Choi, Lehar, and Stauffer (2018)) .

2 The Bitcoin Protocol

A Bitcoin transaction comprises inputs and outputs. An output consists of a Bitcoin amount that is cryptogaphically locked up to a certain address, which is derived from a public key. Several outputs can be locked up under the same address but it is now best practice for a wallet to generate a new address for each transaction. By providing a matching unlocking signature which is derived from the rightful owner's private key the output can be unlocked and used as input to another transaction; each output can only be spent in its entirety or not at all.

Bitcoin transactions are only valid if they are included in a block. The transactions that are included in the block are chosen by the miners from the mempool. The latter consists of those transactions that are valid, i.e., that have been verified by nodes in the network but have not yet been mined. It is important to note that the mempool is not a centralized entity but rather is specific to each node (which typically has different capacity RAM). Transactions leave the mempool if they are chosen by miners to be worked on as part of the next block.

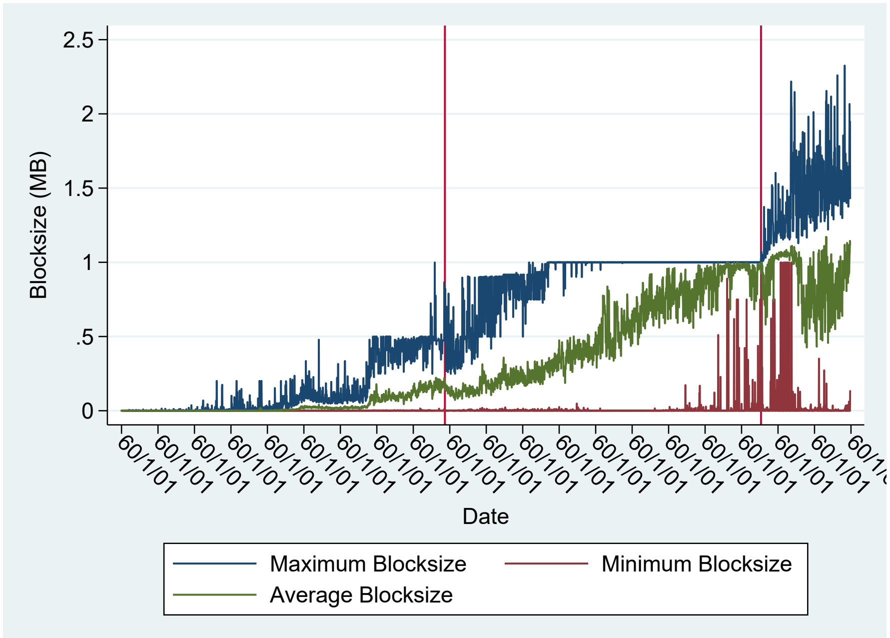
Miners create blocks by solving a computational puzzle and are compensated in two ways. First they can claim a mining reward, which was set at BTC 50 initially and is cut in half every 210,000 blocks (roughly every four years). The first transaction in each block is the called the coinbase transactions which awards the newly minted coins to the miner. Since coinbase transactions are part of the mining incentives we exclude them from much of the analysis in this paper. Second, and more germane to our analysis, participants who want their transaction to be included in a block can offer fees to miners. Fees are offered as the difference between inputs and outputs. A transaction might have, for example, a total of BTC 2.2 as inputs but only BTC 2.18 as outputs. Miners can claim this difference and pay it to themselves as part of the coinbase transaction.

For our analysis it will be important that mined blocks are of finite capacity; or rather there is competition for transactions to be picked from the mempool and to be incorporated into the Blockchain. Transactions with more inputs and outputs are longer while the amount of Bitcoin transacted does not change the size of a transaction. For a miner, the direct cost of including a transaction in a block is space. Satoshi Nakamoto introduced a 1MB limit to Bitcoin blocks in 2010. However, due to its design Bitcoin's block size was limited by the number of database locks required to process it (at most 10000). This limit translated to around 500-750k in serialized bytes, and was forgotten until March 11, 2013, when an upgrade to V0.8.0 with a switch of databases caused a unplanned fork in the blockchain. After resolving the crisis, the community reached a consensus to remove this unknown limit and a hardfork was scheduled and cleanly activated on May 15, 2013. From this point forward, the 1 MB limit became the effective limiting factor of the block size for the first time.[2] After much debate in the Bitcoin community Segregated Witness (Segwit) was implemented on August 24, 2017 which allows to store signatures and redeem scripts in a data structure that is technically outside of the block, effectively increasing the block size to 4MB under ideal circumstances. The first block exceeding the 1MB limit was block 481.947 mined on Aug 25, 2017 with a size of 1032 KB. Yet, the overall adoption of Segwit was slow.

Figure 1 shows block size over time including the two important protocol updates mentioned above. Average block size increases steadily over time and it seems transaction capacity is scarce mid 2017 just before the segwit update. It is also interesting to note that many blocks are empty even at times when block space is in high demand and fees are high. Figure 2 shows the fraction of full and empty blocks. This is puzzling because this is money left on the table as it is technically not easier to mine an empty block. In fact for mining pools having non-empty blocks is essential to prevent free riding. The pool coordinator does not disclose a full list of all transactions that should be included in a candidate block to the pool participants in order to prevent them from claiming the whole mining reward for themselves. Instead pool participants get aggregate transaction information. Once a miner finds the right solution she has an incentive to report it to the pool as the the pool coordinator's information is required to claim the reward, which is then shared among all pool participants. Empty blocks do not allow the pool coordinator to withhold information and increase the moral hazard problem. One possible explanation

[2] see https://en.bitcoin.it/wiki/Block_size_limit_controversy, https://blog.bitmex.com/bitcoins-consensus-forks/

Figure 1. Block Size. Minimum, average, and maximum daily block size. The red bars mark the safe implementation of the 1MB block size limit and the Segwit update, respectively. Days are defined over UTC.



for empty blocks is the malicious intent to slow down transactions in Bitcoin, similar to the fight between two competing Bitcoin Cash developer groups in November 2018.[3]

The left panel in Figure 3 documents that from 2016 on fees gain importance as source of revenue for miners. We plot the fee revenue as fraction of the block reward with the red lines indicating cuts in the block reward. The block award started with 50 BTC and gets halved every 210,000 blocks. The right panel shows that users react to cuts in block rewards by offering higher fees immediately following the event date. The upper graph shows the median and quantiles of the distributions of the fee to blockreward ratio over 12 hour intervals surrounding the cut in block rewards. The lower graph is from the cut from 50 to 25 BTC and the upper graph is from the cut to 12.5 BTC.

[3]On November 15, 2008 the Bitcoin Cash blockchain forked in two parts as competing developer groups fought over the future development of the crypto-currency. Craig Wright, representing one group threatened to halt transactions in the competing fork by constantly adding empty blocks to the competitor's blockchain, therefore increasing processing times for transactions, thus making the other crypto-currency worthless.

Figure 2. Empty and full blocks. Number of blocks and fraction of empty and full blocks per day. Days are defined over UTC.

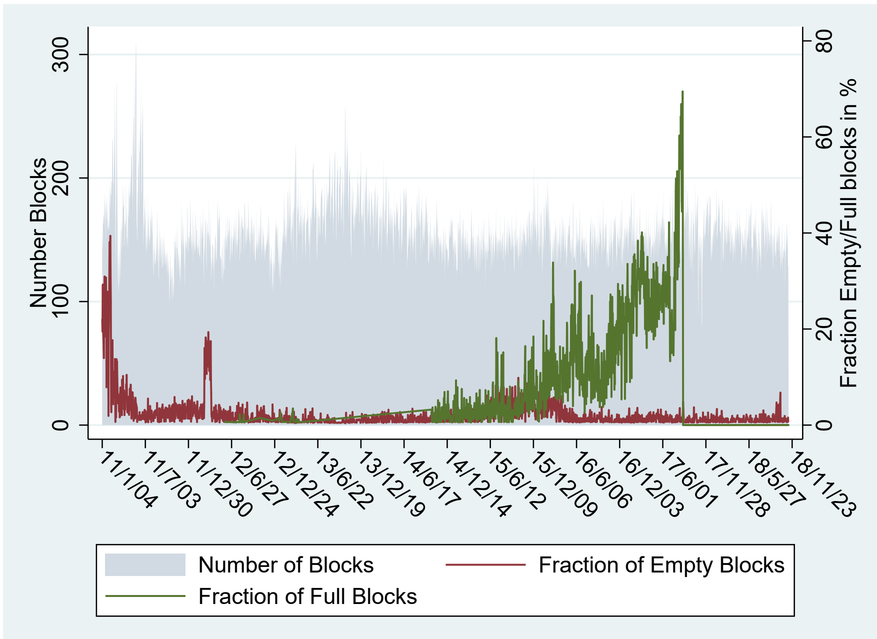


Figure 3. Fees and Block Rewards. Days are defined over UTC. Left panel: Fees per block as a fraction of the block reward. Median and 10%/90% quantile are of the daily distribution. The red lines show when the block reward was cut in half, the green lines indicate when the block capacity was increased. Right panel: Fees as a fraction over block reward around the events where the block reward was cut from 50 to 25 BTC (lower graph) and for the cut from 25 to 12.5 BTC (upper graph).

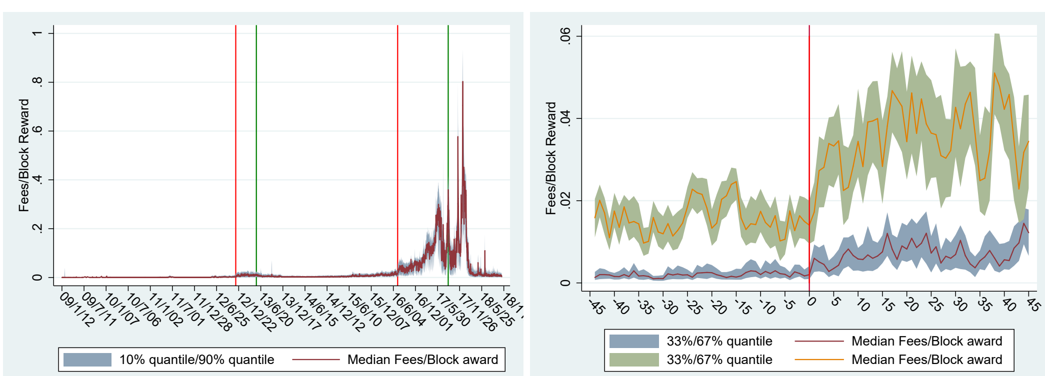
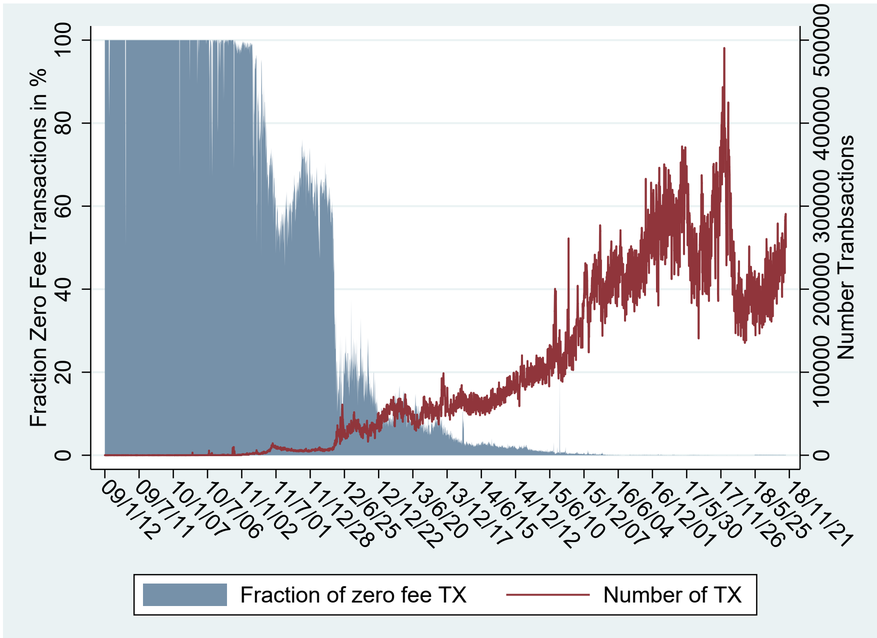


Figure 4. Number of transactions per day. Number of total transactions (red, right axis) and fraction of transaction without fees (blue area, left axis). Days are defined over UTC. Transactions exclude coinbase transactions



3 Summary statistics

Our sample consists of 353,830,421 transactions between Jan 3, 2009 and Nov 4, 2018 Our sample comprises all blocks from the Genesis block to block number 548,684 and includes 903,976,764 inputs and 961,308,921 outputs. Figure 4 shows the total number of transactions over time as well as the number of transactions with zero fees over time. The number of transactions increases over time and peaks in mid December 2017 coinciding with the peak in Bitcoin prices. The fraction of transactions without fees declines over time. On the last full month of our sample, October 2018 only 67 out of 7,570,407 non-coinbase transactions were mined without a fee. High variation in Bitcoin's early days of bitcoin can be attributed to the low transaction volume.

Figure 5 plots the time series of the median as well as the 25 and 75 percentiles of transaction fees starting in March 2015 when reliable price data exists.

Bitcoin fees are often portrayed as extremely expensive and therefore unusable in practice. Indeed we do see average as well as median transaction fees reaching very high levels in December 2017, when

Figure 5. Bitcoin fees and prices Bitcoin prices (gray area, left axis) and median, 25 and 75 percentile of daily fees (right axis) in BTC (upper panel), USD (medium panel) and KRW (lower panel). Days are defined over UTC. Transactions exclude coinbase transactions

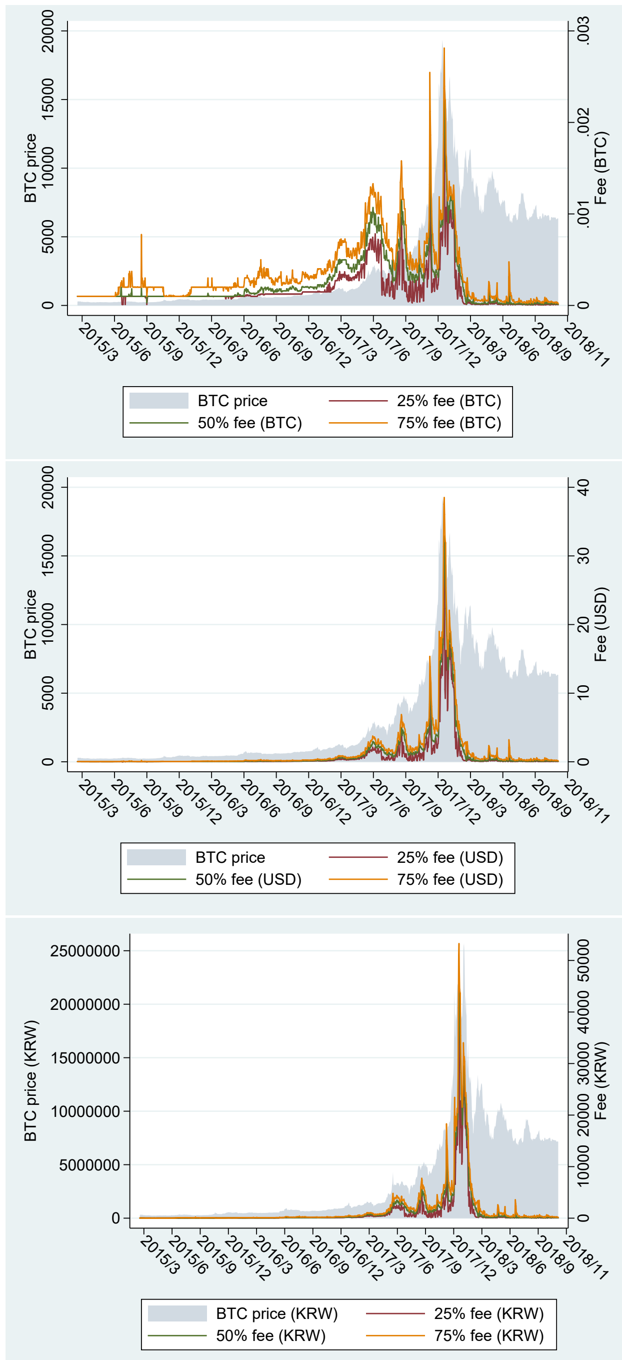
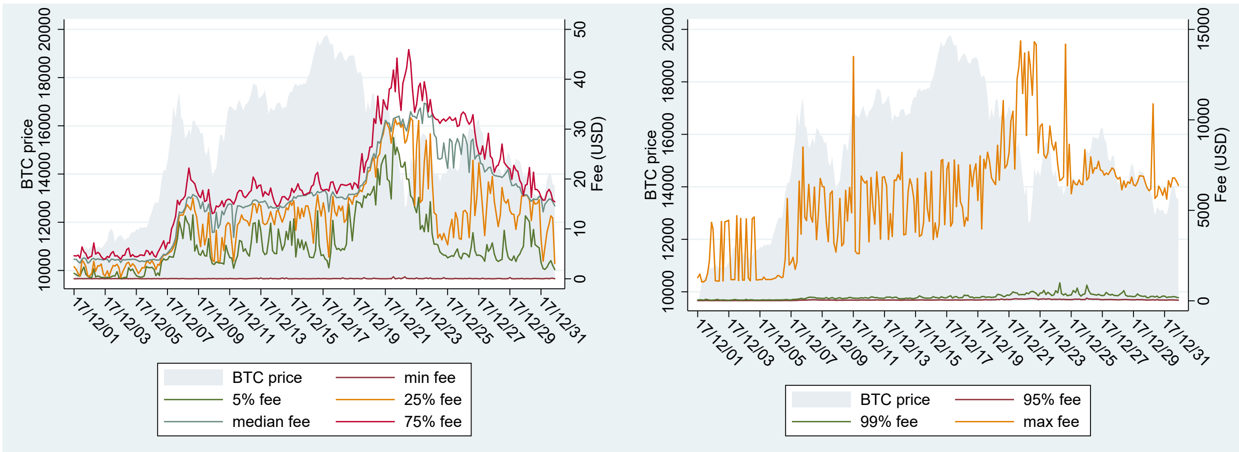


Figure 6. Quantiles of bitcoin fees. Fees in USD (right axis) and Bitcoin price in USD (left axis) in December 2017.

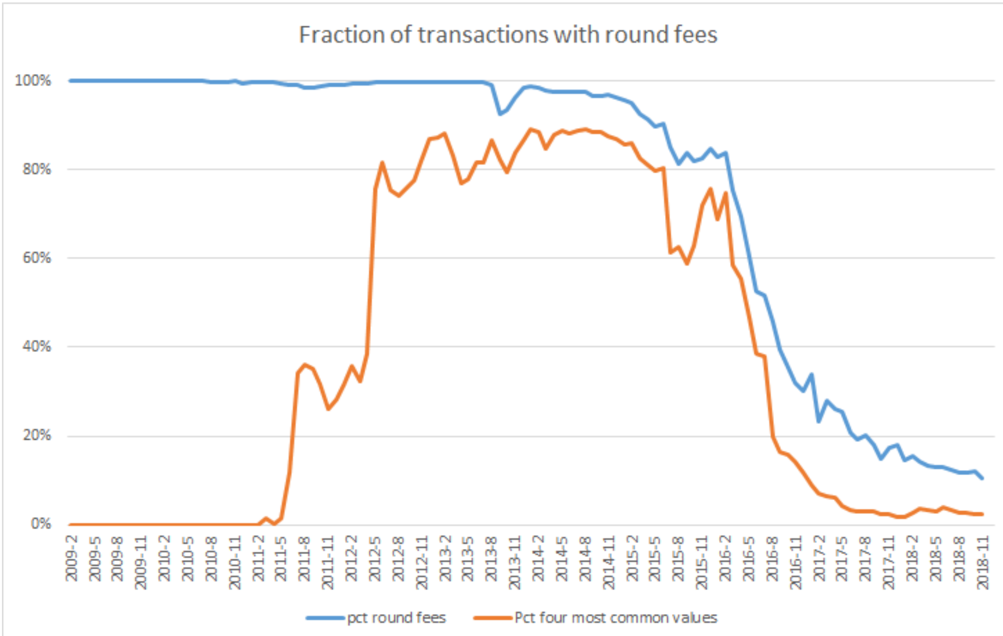


Bitcoin prices peaked in our sample. Indeed we find 80 transactions in our sample with fees greater than USD 10,000, out of which 51 occur between Dec 20^{th},2017 and Dec 24^{th},2017. Yet in these five days 1,674,141 transactions were processed out of which 752 had no fee and 16,191 transactions are mined with fees less than USD 5. Figure 6 visualizes the wide range of the Bitcoin fee distribution in December 2017. The left panel shows the lower end of the fee distribution documenting that a substantial fraction of transactions occurred at low or modest fees. The right panel details the upper end of the distribution documenting that as Bitcoin fell from its peak price some people were willing to pay enormous fees to have their transactions processed. On Dec 24^{th},2017, for example, the 99 percentile fee was USD 433.23 while the largest fee paid was USD 14,174.64.

4 Clustering

We looked for various patterns if Bitcoin fees cluster around round values. 19.68% of all transactions in our sample have a fee of 10,000 Satoshi. 32,47% of all transactions is conducted at a fee of 10k, 20k, 50k, or 100k Satoshi. 47.93% of all transactions are a multiple of 100 Satoshi. Yet these transactions are not evenly spread across time. Figure 7 shows the fraction of transactions with fees in a multiple of 100 Satoshis as well as fraction of transactions with fees of exactly 10k, 20k, 50k, or 100k Satoshis. The

Figure 7. Fraction of transactions with round fees. Fees in a multiple of 100 Satoshis aggregated per month (blue line) as well as fraction of transactions with fees of exactly 10k, 20k, 50k, or 100k (organge line). Transactions exclude coinbase transactions



former is steadily declining over time which would be consistent with a broader use of more sophisticated wallet software and with the arrival of websites that provide guidance to users on the current relationship between fees and transaction speeds. The second time series shows the fraction of transactions with fees equal to the four most popular absolute values. This fraction is close to zero towards the beginning of the sample as most transactions where mined without fees at the beginning.

5 Paying for Information

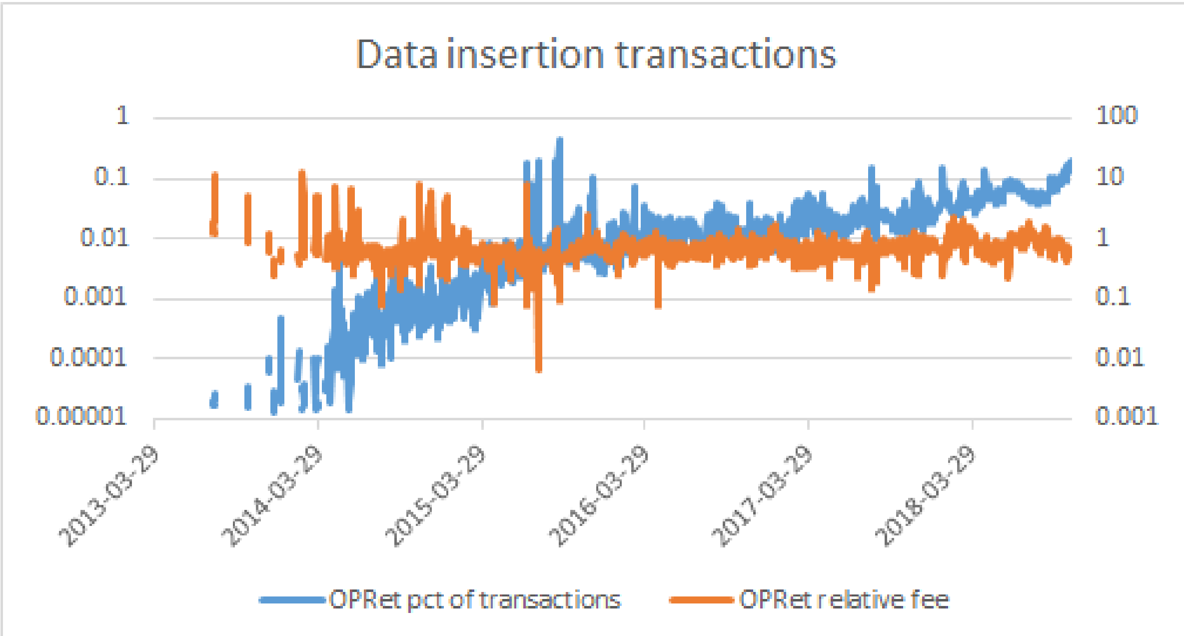
Not all transactions on the Bitcoin blockchain are used to record financial transactions, i.e. the transfer of Bitcoin from one wallet to another. Some transactions' only purpose is to record data on the blockchain. Less serious examples include declarations of love, e.g. 'Charles loves Heidi in block 308570, a picture of Nelson Mandela in Block 415336, or a copy of the original Bitcoin paper in Block 230009. More serious applications include notary services where a hash of a document, for example, a will is published. Other uses include digital asset management or meta-layers running on the Bitcoin Blockchain such as the Omni Layer.

In early blocks data insertion is implemented by sending money to Bitcoin addresses that do not exist (i.e. where presumably nobody owns the corresponding private key). Combining several of these addresses then constitutes the payload, i.e. the data that is recorded in the blockchain. Any Bitcoin sent to these addresses are unrecoverably lost. This practice of using unrecoverable addresses raised many concerns and is heavily criticized in the developer community for its resource use. For faster validation of transactions Bitcoin clients keep a table of all unspent transaction outputs (UTXO) in memory. Since it is impossible to determine whether an address is a true address or part of a binary file the unrecoverable addresses and their associated outputs must be kept in the UTXO table and thus waste resources for all clients.

To overcome this problem OP_RETURN was introduces with bitcoin core release 0.9.0 in March 2014. OP_RETURN is a command in the bitcoin script language that identifies the following bit-sequence as data and hence it must not be kept in the UTXO table in memory. The release update provided a resource friendly way of inserting data into the bitcoin blockchain. Empirical studies in the computer science literature like e.g. Matzutt, Hiller, Henze, Ziegeldorf, Müllmann, Hohlfeld, and Wehrle (2018) find that 99.92% of all data-insertion in non-coinbase transactions are using the OP_RETURN, which is why we will focus on this method in this paper.

Figure 8 plots the fraction of data insertion transactions (blue line) which is increasing over time and constituted 5318 transactions per day or 10% of all transactions in the last 50 days of our sample period. Those transactions also pay on average a slightly lower fee than normal transactions as is illustrated by the orange line in the graph, in the last 50 days they paid 72.2% of the fees of an average transaction. Data insertion transactions are on average larger in terms of bytes than financial transactions and move very little BTC around. Due to the different objective of those transactions we control for them in our analysis.

Figure 8. Data Insertion transactions. Number of transactions and fees paid by transactions without financial transactions, having output under the OP_RETURN code. Fraction of OP_RETURN transactions as percentage of all transactions (blue line, left axis, log scale) ad fees paid by OP_RETURN transactions relative to fees paid by other transactions in the same block (orange line, right axis, log scale)



6 Regressions

To identify drivers of transaction fees we regress the fee per transaction on some explanatory variables from the blockchain. We define minimum outtime as the time (in blocks) until the first output of this transaction is spent again. Recipient wallet owners that spend their funds very quickly have a more immediate need for funds and might thus put pressure on their debtors to send funds quickly. The Opret dummy is set to one for all data insertion transactions as discussed in Section 5. Because their main purpose is record-keeping rather than financial transactions services posting such transactions might have a different sense of urgency for getting their transaction included in a block.

The Sum Inputs variable adds up all input values to a transaction. We choose inputs rather than outputs because some transactions, most of them data-insertion transactions, have almost all their inputs dedicated to fees and only have negligible output. Transaction size is the physical size of all inputs and outputs in bytes. The transaction size represents an opportunity cost for miners as block space is

Table 1. Summary statistics

	Nobs	Mean	Std.Dev.	Median	Min	Max
Fee (Satoshi)	353,306,421	52,406	2,826,555	18,802	0	29,124,090,000
Fee (USD)	291,702,649	2.90	36.86	.20	0	137,186.10
Sumin (Satoshi)	353,306,421	1,370,000,000	40,500,000,000	19,800,000	0	55,000,000,000,000
Sumin(USD)	291,702,649	25,690.01	556,020.30	256.64	0	947,000,000.00
Blocksize (bytes)	353,306,421	834,072	325,063	998,126	267	2,324,736
Tx-Size (bytes)	355,725,559	537	2,210	250	62	999,657
dum-OPRet	353,306,421	.02	.15	.00	.00	1.00
Minouttime (blocks)	350,031,634	547.54	4,950.34	4.00	.00	474,195.00
Priceusd (USD)	291,702,649	3,604.66	4,162.71	1,188.28	204.84	19,828.76

limited. Blocksize is the absolute size of the block in bytes. Larger blocks have less space for additional transactions and are an indication for transaction demand.

Table 1 contains summary statistics. Some transactions contain unusually high fees which are either errors or payments miners make to themselves to consolidate their bitcoin balances together with the block award they earned under a new address. For example on April 26, 2016 someone paid 291.241 BTC to an output with 0.0001 BTC, leaving fees of 291.2409 (or USD 137,186.07 at the time) for the miner.[4] We found overall 80 transactions with fees over USD 10,000. These transactions have an average input of 140.89 BTC, and an average fee of 7.55 BTC or (USD 16,974.42 at the time). These transactions can be found between April 2016 and October 2018, but most of them are concentrated from Dec 21-14 in 2017. Bitcoin prices peaked on Dec 16 2017 at USD 19,200 and feel back to USD 15,191 on Dec 21. One potential explanation for these high fees is that investors wanted to move Bitcoin from their private wallet to an exchange quickly in a market panic. Transactions from the earlier days of Bitcoin may also show unusually high fees. In December 2011 a transaction can be found with inputs of over 207 BTC and outputs of 36 BTC, leaving a huge fee for the miner. At that time bitcoin were worth very little, yet since there was no shortage of transaction space in the blockchain, there is no obvious rationale for paying such a high fee.

Many variables show some extreme observations. The average Bitcoin transaction was for 13.7 BTC (one bitcoin equals 100 million Satoshi) while the largest transaction was for 550,000 BTC on Nov 16, 2011 at a zero fee.[5] The largest transaction in dollar terms occurred on Dec 17th, 2017 at the peak of

[4]this is the highest fee transaction in BTC and USD terms in our sample, see transaction cc455ae816e6cdafdb58d54e35d4f46d860047458eacf1c7405dc634631c570d.

[5]see transaction 29a3efd3ef04f9153d47a990bd7b048a4b2d213daaa5fb8ed670fb85f13bdbcf

the bitcoin price when 48,500 BTC valued at over 946 million USD changed wallet for a fee of 80,908 Satoshi or USD 15.8.[6] In our sample we find 54,907 transactions with a value of more than USD 10 million. Out of those 20,956 were processed with a fee of less than USD 5 and the average fee of those transactions was USD 90.54.

Most transactions are small in size with a mean of 537 and a median of 250 bytes. The largest transaction in our sample consumes the entire block 364,292 with a size of 999,657 bytes. The transaction was mined on July 7th, 2015 and consolidates 5,569 inputs of 1,000 Satoshi each in a single output.[7] Minouttime measures the time (in number of blocks) until the first output of a transaction gets spent again. In some cases a transaction output gets spent in the same block that it is received, showing a minouttime of zero. The mining difficulty is set such that on average one block is mined every 10 minutes. We chose blocks and not calendar time as measure to make our findings robust to smaller variations in the time between blocks especially when the output is spent shortly after it is being received. For longer timespans the difference between calendar time and block time is irrelevant as blocks are mined every 10 minutes.

To purge our sample from extreme outliers we windsorize our fee data, the transaction size, the inputs, and the outtime at the 99.9% level. We find our results to be robust to different levels of windsorizing. Bitcoin fees also vary tremendously over time. To control for this time variation we include day-fixed effects in our regression analysis. To control for variation of fees across miners we cluster standard errors per block.

Table 2 presents our findings for the fees in term of Satoshis (1 BTC is 100 million Satoshis) and Table 3 for fees in USD. In line with our intuition fees are higher when transactions space is scarce (larger blocksize) and when the transaction is larger in terms of bytes (TXSize) and value (Sum Inputs). Fees are also higher when the funds are spent sooner (lower min. outtime). One possible explanation is that receivers who are in need for liquidity pressure senders to accelerate payments by offering higher

[6]see transaction 261d69b25896034325d8ad3e0668f963346fd79baefb6a73b4eabd68c58c81ff

[7]According to some forum posts an unknown spammer created all these wallets to test the limits of the UTXO database. Somebody created many small outputs locked up under different addresses. To speed up the time it takes to verify the validity of a transaction each node holds in RAM memory a complete lust of all unspent BTC amounts per address. This list is called the Unspent Transaction Output (UTXO) database. The attack of the spammer forced each of the nodes to expand the UTXO list thereby increasing the memory requirement for each node. The Intentions of the spammer are unclear as all the transactions were protected by very weak private keys such as *'passowrd'* or *'cat'*. Somebody eventually guessed the private keys and consolidated all the small inputs in block 364,292, freeing up space in the UTXO list.

Table 2. Regression results of fees in Satoshis (1 BTC is 100 million Satoshis). Regressions include day fixed effects. Standard errors are clustered by block. One, two, and three stars indicate significance at the 10%, 5%, and 1% level, respectively.

Blocksize	0.0024***					0.0012***
	(0.0001)					(0.0001)
TXSize		49.4070***				49.5693***
		(0.1380)				(0.1382)
Sum Inputs			1.04e-06***			5.75e-07***
			(8.65e-09)			(5.65e-09)
OP-Ret				-10371.4000***		-502.8570**
				(232.5050)		(243.3099)
Min. Outtime					-1.2260***	-2.1106***
					(0.0144)	(0.0199)
constant	44750.4000***	21403.3000***	45713.6000***	46993.6000***	47288.5000***	20521.8000***
	(74.5277)	(67.0622)	(39.8744)	(40.5487)	(40.2871)	(115.3693)
R^2	0.0951	0.3688	0.0973	0.0952	0.0952	0.3708
N	353,306,421	353,281,736	353,306,421	353,306,421	350,031,634	350,008,404

fees. Data insertion transactions pay on average lower fees in BTC other users of the blockchain. While being statistically significant some results are economically insignificant. The value of the transaction, for example, only has a small impact on the fees, which is consistent with the fact that the miner's opportunity cost for including a transaction in a block of limited size is determined by the transaction size in bytes and independent of the value. In Appendix A we provide more details on the determinants of transaction size. Transaction size has therefore an economically significant impact on fees. Adding one input (with an average size of 180 bytes) to a transaction (the average transaction has 2.5 inputs in our sample) increases the fee by 8,893 Satoshi or USD 0.41. Given that the median fee over the whole sample is USD 0.20 this effect is economically large. Outtime is measured in blocks, which are mined every 10 minutes on average. People spending their funds earlier pay on average USD 0.014 more in fees per day.

Results are mixed for the data-insertion (OP-Ret) dummy. In terms of BTC we see that data-insertion transactions post lower fees, yet when controlling for other variables we see that they pay a higher fee in USD. We attribute that to the increasing volume of data insertion transactions towards the end of our sample, where BTC prices were generally higher than at the beginning.

The results in Table 4 show that transaction fees are lower on weekends by almost USD 0.196, which is close to median transaction fee for the whole sample. Looking at individual days of the week it seems

Table 3. Regression results of fees in USD. Regressions include day fixed effects. Standard errors are clustered by block. One, two, and three stars indicate significance at the 10%, 5%, and 1% level, respectively.

Blocksize	9.06e-08***					5.03e-08***
	(5.89e-09)					(6.87e-09)
TXSize		0.0023***				0.0023***
		(9.90e-06)				(9.76e-06)
Sum Inputs			7.69e-06***			6.15e-06***
			(6.65e-08)			(5.85e-08)
OP-Ret				-0.3681***		0.1247***
				(0.0168)		(0.0157)
Min. Outtime					-0.0001***	-0.0001***
					(1.30e-06)	(1.45e-06)
constant	2.3626***	1.2527***	2.2953***	2.4579***	2.4814***	1.1323***
	(0.0068)	(0.0054)	(0.0034)	(0.0035)	(0.0035)	(0.0088)
R^2	0.2741	0.3997	0.2858	0.2741	0.2747	0.4079
N	291,702,649	291,678,054	291,702,649	291,702,649	288,555,299	288,532,152

that fees gradually increase during the work week with the highest observed fees on Fridays.

Bitcoin is a pseudo-anonymous system, all wallet addresses can be identified and payments can be tracked moving from one wallet to another but the identity of the wallet owner is usually unknown. In some cases, e.g. voluntary disclosure, court proceedings, the owner of some addresses gets known. Gambling sites often use vanity addresses, which start with a suggestive word like '1dice....' or '1Lucky....' which are easily identified. Such vanity addresses have to be found by trial and error in a time consuming process and therefore get often reused.[8] Once an address is known other addresses controlled by the same walled can be inferred in a process commonly known as address clustering see e.g. Reid and Harrigan (2013) or Foley, Karlsen, and Putniņš (2018). The idea is that if multiple addresses are used as inputs in the same transaction these addresses most likely belong to the same person because the private key has to be used to sign the transaction.[9]

We use data from lists of known addresses such as wallet explorer and are able to identify the identity of the sender for 18,123,498 transactions, out of which 7,250,374 (or 2.05% of all transactions) were initiated by an exchange and 10,873,124 (or 3.07% of all transactions) were initiated by a gambling site.

[8] Addresses are encoded in a Base58 alphabet (i.e. there are 58 possible 'letters' consisting of upper case, lower case letters and numbers with some combinations dropped that are often mixed up when printed on paper, e.g. capital i and lower case L) and start with 1. To get an address with '1Lucky' one has to try $58^5 \approx 656\textit{million}$ combinations. Vanity address companies offer computing resources for custom bitcoin addresses.

[9] One notable exception are anonymizing services that for a fee combine transactions of several users into one large transaction so that it is not that clear who paid whom. See e.g. Möser and Böhme (2017).

Table 4. Regression results of fees in USD - day of the week. Regressions include week fixed effects. Days are defined in UTC. Standard errors are clustered by block. One, two, and three stars indicate significance at the 10%, 5%, and 1% level, respectively.

Blocksize	1.88e-08**	3.90e-08***
	(8.34e-09)	(8.13e-09)
TXSize	0.0023***	0.0022***
	(9.76e-06)	(9.76e-06)
Sum Inputs	6.13e-06***	6.13e-06***
	(5.91e-08)	(5.91e-08)
OP-Ret dummy	0.0832***	0.0848***
	(0.01870)	(0.01859)
Min. Outtime	-0.0001***	-0.0001***
	(1.51e-06)	(1.51e-06)
Monday	0.2683***	
	(0.02050)	
Tuesday	0.3749***	
	(0.01963)	
Wednesday	0.4388***	
	(0.01873)	
Thursday	0.5236***	
	(0.01961)	
Friday	0.6379***	
	(0.02209)	
Saturday	0.4847***	
	(0.02119)	
Weekend		-0.1962***
		(0.01197)
Constant	0.7662***	1.1966***
	(0.01649)	(0.01089)
R^2	0.3952	0.3949
N	288,532,152	288,532,152

Similarly we are able to identify 32,771.960 payments to an exchange and 23,099,021 payments to a gambling site. Table 5 presents our findings for fees in USD. Flows to and from exchanges transact at higher than average fees. Since we control for day fixed effects our results cannot be driven by more exchange flows occurring on days when fees are generally higher. Our findings are also not driven by outliers since the data is windsorized. Transactions flowing out of exchanges pay USD 3.15 more, which is remarkable given that the median fee for the whole sample is USD 0.19. Fees into exchanges are paying USD 1.24 more than average. Possible explanations include the desire of arbitrageurs to move finds quickly between markets.

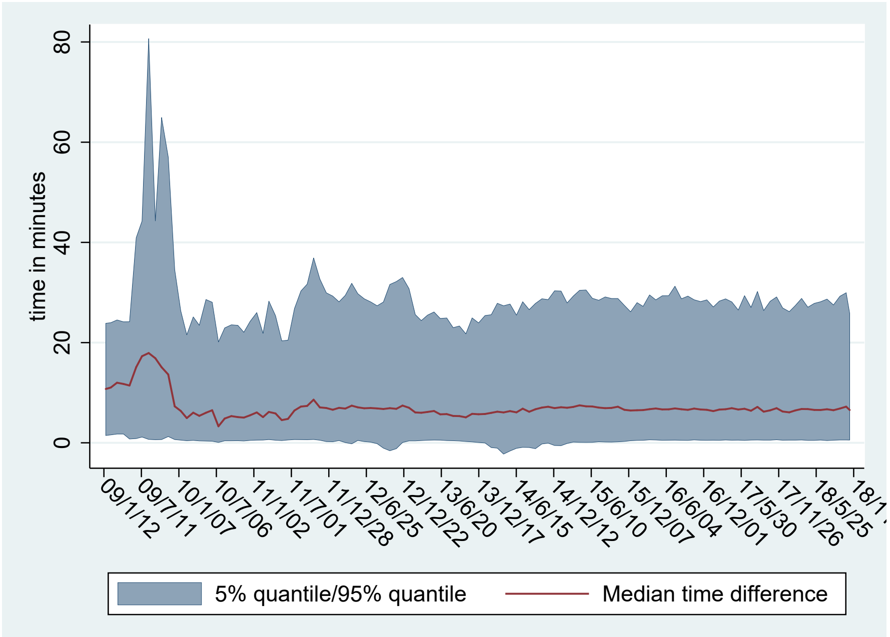
We examine two measures how transaction costs are driven by congestion in transaction processing on the blockchain. We identify two potential channels of congestion. First we examine how many

Table 5. Regression results of fees in USD - day of the week. Regressions include week fixed effects. Days are defined in UTC. Standard errors are clustered by block. One, two, and three stars indicate significance at the 10%, 5%, and 1% level, respectively.

	Identity sender		Identity Receiver	
Blocksize		7.00e-08***		5.76e-08***
		(0.0000)		(0.0000)
TXSize		0.0023***		0.0024***
		(0.0000)		(0.0000)
Sum Inputs		5.98e-06***		7.75e-06***
		(0.0000)		(0.0000)
OP-Ret dummy		-0.0874***		0.0884***
		(0.0154)		(0.0158)
Minimum Outtime		-0.0001***		-0.0001***
		(0.0000)		(0.0000)
Exchange	5.4745***	3.1496***	2.1978***	1.2460***
	(0.0478)	(0.0350)	(0.0140)	(0.0099)
Gambling	0.3528***	0.1914***	0.7724***	0.5259***
	(0.0039)	(0.0069)	(0.0060)	(0.0050)
constant	2.3485***	1.0723***	2.3281***	0.9826***
	(0.0033)	(0.0089)	(0.0036)	(0.0098)
R^2	0.2789	0.4095	0.2637	0.4062
N	291,702,649	288,532,152	294,586,994	291,416,436

blocks have been recently mined with the idea that fewer recently mined blocks cause more unprocessed transactions to pile up in the mempool which increases users' willingness to pay higher fees to get their transaction processed. The Bitcoin protocol calibrates the difficulty, i.e. the number of leading zeros that a block hash has to have to qualify as valid, in such a way that on average a new block is added every ten minutes. Yet the times between blocks as they are recorded on the blockchain wary widely as illustrated in Figure 9. This is because mining a successful block is purely random and so sometimes blocks are found very quickly and sometimes it takes a long time. Another reason is that the time a block was mined is self reported by the miner and miners can have their local clocks not aligned with the world time. These clock mis-alignments can be substantial. Out of the 548,648 blocks in our sample we find 13,848 cases in which a block has an earlier time-stamp than its predecessor. This is technically impossible. Each block contains information from the previous block, which links the blocks together in a blockchain. The only rational explanations for the inconsistency in time-stamps is improper alignment of miners' clocks. To accommodate potential synchronization problems in miners' clocks the bitcoin protocol allows a block to have time-stamp up to two hours earlier than the previously mined block. Despite these problem cases for the vast majority of the sample the time-stamps seem to be properly recorded. Second we measure the capacity utilization of the recently mined blocks, which we define as the fraction of available space

Figure 9. Time between blocks. Time between blocks as recorded on the blockchain. Quantiles are computed from the monthly distribution.



used by the miners.

Table 6 includes our measures of congestion in the fee regression. In line with our intuition we find that fees are lower when more blocks have been mined recently. Contrary to expectation we find that fees are lower when previous blocks are more full. To understand this counter-intuitive finding note that we focus on the very short term variation in fees. Any longer term trends would be absorbed by the day fixed effects. In the short term when more pending transactions get picked up from the mempool fees fall because congestion reduces and because previous miners picked up the high fee transactions.

Table 6. Regression results of fees in USD including measures of congestion. Regressions include day fixed effects. Days are defined in UTC. Standard errors are clustered by block. One, two, and three stars indicate significance at the 10%, 5%, and 1% level, respectively.

Blocksize		-3.51e-07***
		(0.0000)
TXSize		0.0023***
		(0.0000)
Sum Inputs		6.11e-06***
		(0.0000)
OP-Ret dummy		0.1416***
		(0.0152)
Minimum Outtime		-0.0001***
		(0.0000)
Num blocks 1h	-0.0957***	-0.1343***
	(0.0018)	(0.0018)
Num Blocks 24h		-0.0107***
		(0.0006)
Capacity use 1h	-0.6112***	-0.8866***
	(0.0160)	(0.0172)
Capacity use 24h		-0.9620***
		(0.0999)
constant	3.3813***	5.0140***
	(0.0193)	(0.1433)
R^2	0.2789	0.4088
N	291,702,649	288,532,152

References

Abadi, Joseph, and Markus Brunnermeier, 2018, Blockchain economics, Working paper, National Bureau of Economic Research.

Brauneis, Alexander, Roland Mestel, Ryan Riordan, and Erik Theissen, 2018, A high-frequency analysis of bitcoin liquidity, .

Budish, Eric, 2018, The economic limits of bitcoin and the blockchain, Working paper, National Bureau of Economic Research.

Choi, Kyoung Jin, Alfred Lehar, and Ryan Stauffer, 2018, Bitcoin Microstructure and the Kimchi premium, .

Cong, Lin William, and Zhiguo He, 2019, Blockchain disruption and smart contracts, *The Review of Financial Studies* 32, 1754–1797.

Cong, Lin William, Zhiguo He, and Jiasun Li, 2019, Decentralized mining in centralized pools, Working paper, National Bureau of Economic Research.

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

Foley, Sean, Jonathan R Karlsen, and Tālis J Putniņš, 2018, Sex, drugs, and bitcoin: how much illegal activity is financed through cryptocurrencies?, *Review of Financial Studies, Forthcoming*.

Hu, Albert S., Christine A. Parlour, and Uday Rajan, 2018, Cryptocurrencies: stylized facts on a new investible instrument, working paper.

Huberman, Gur, Jacob Leshno, and Ciamac C Moallemi, 2017, Monopoly without a monopolist: An economic analysis of the bitcoin payment system, .

Makarov, Igor, and Antoinette Schoar, 2018, Trading and arbitrage in cryptocurrency markets, working paper.

Malinova, Katya, and Andreas Park, 2017, Market design with blockchain technology, *Available at SSRN 2785626*.

Matzutt, Roman, Jens Hiller, Martin Henze, Jan Henrik Ziegeldorf, Dirk Müllmann, Oliver Hohlfeld, and Klaus Wehrle, 2018, A Quantitative Analysis of the Impact of Arbitrary Blockchain Content on Bitcoin, in *Proceedings of the 22nd International Conference on Financial Cryptography and Data Security (FC). Springer*.

Möser, Malte, and Rainer Böhme, 2017, The price of anonymity: empirical evidence from a market for Bitcoin anonymization, *Journal of Cybersecurity* 3, 127–135.

Reid, Fergal, and Martin Harrigan, 2013, An analysis of anonymity in the bitcoin system, in *Security and privacy in social networks* (Springer,).

Table 7. Regression of transaction size in bytes on the number of inputs, number of outputs, and the value of the inputs. One, two, and three stars indicate significance at the 10%, 5%, and 1% level, respectively.

Numin	164.74***
	(0.06414)
Numout	37.781***
	(0.02734)
Sum Inputs	9.23e-12***
	(1.07e-12)
constant	19.617***
	(0.16118)
R^2	0.9394
N	350,008,404

A Determinants of Btcoin Transaction Sizes

Transactions sizes are influenced by the number of inputs and outputs as well as by the type of the transaction. The Bitcoin eco-system allows several types of transactions from the simple pay to an address, to multi-signature transactions where, for example, two out of three people need to digitally sign a transaction for validation, to pay-to-script-hash transactions where the precise rules of redemption are not known at the time an output is locked but only the hash of the redemption script is recorded. Transaction size increases in the number of inputs and outputs with inputs contributing more since they have to hold the transaction signature and the public key. For basic transactions with i inputs and o outputs a common rule of thumb is that the transaction size in bytes is $180i + 34o \pm i$.[10]. To confirm this idea and to show that value does not drive transaction size we regress transaction size in bytes on the number of in, and outputs (numin and numout, respectively) as well as on the value of all inputs (sumin). The results in Table 7 document that the rough rule mentioned above is approximately true for our sample as well. We also find that that the value has no economically significant impact on the size of a transaction.

[10]see e.g. https://bitcoin.stackexchange.com/questions/1195/how-to-calculate-transaction-size-before-sending-legacy-non-segwit-p2pkh-p2sh