Liquidity and Price Informativeness in Blockchain-Based Markets

Stefan Voigt*

May 9, 2020

Blockchain-based markets impose substantial costs on cross-market trading due to the decentralized and time-consuming settlement process. I quantify the impact of the time-consuming settlement process in the market for Bitcoin on arbitrageurs activity. The estimation rests on a novel threshold error correction model that exploits the notion that arbitrageurs suspend trading activity when arbitrage costs exceed price differences. I estimate substantial arbitrage costs that explain 63% of the observed price differences, where more than 75% of these costs can be attributed to the settlement process. I also find that a 10 bp decrease in latency-related arbitrage costs simultaneously results in a 3 bp increase of the quoted bid-ask spreads. I reconcile this finding in a theoretical model in which liquidity providers set larger spreads to cope with high adverse selection risks imposed by increased arbitrage activity. Consequently, efforts to reduce the latency of blockchain-based settlement might have unintended consequences for liquidity provision. In markets with substantial adverse selection risk, faster settlement may even harm price informativeness.

JEL Codes: G00, G10, G14

Keywords: Arbitrage, Market Frictions, Distributed Ledger, Blockchain

---

*Stefan Voigt, VGSF (Vienna Graduate School of Finance) and WU (Vienna University of Economics and Business), Department of Finance, Accounting and Statistics, Welthandelsplatz 1, Building D4, 1020 Vienna, Austria. Correspondence to: stefan.voigt@vgsf.ac.at (www.voigtstefan.me). I thank Torben Andersen, Nikolaus Hautsch, Ravi Jagannathan, Robert Korajczyk, Gyöngyi Lóránth, Albert Menkveld, Christoph Scheuch, Günter Strobl, Viktor Todorov, Patrick Weiss, Christian Westheide, Kathy Yuan, Josef Zechner and seminar participants at Kellogg School of Management, VU Amsterdam, Amsterdam Business School, NHH Bergen, University of Copenhagen, Copenhagen Business School, University of Heidelberg, Brandeis International Business School, New Economic School, Higher School of Economics, the 4th Vienna Workshop on High-Dimensional Time Series in Macroeconomics and Finance 2019, the 5th Konstanz Lancaster Workshop on Finance and Econometrics and the VGSF Conference 2019 for helpful comments. I gratefully acknowledge financial support from the Austrian Science Fund (FWF project number DK W 1001-G16).

1 Introduction

In recent years, technological innovation and regulatory pressure spurred the emergence of many coexisting trading platforms in most asset classes. Market fragmentation may be beneficial for some participants, for instance due to increased competition between trading venues. However, it can also harm informational efficiency as long as limits to arbitrage prevent cross-market trading from equating prices across markets.[1] If frictions render arbitrage costly, arbitrageurs may refrain from exploiting price differences, thus giving up their pivotal role to restore the law of one price. As a result, the speed and extent to which arbitrageurs exploit price differences depends on the magnitude of arbitrage costs and contributes to the informational value of prices in fragmented markets.

One particular asset that simultaneously exhibits substantial market fragmentation and considerable arbitrage costs is Bitcoin, the most actively traded cryptocurrency to date. Bitcoin is traded worldwide at hundreds of exchanges, but persistent price differences across these exchanges repeatedly arise and cannot be reconciled solely with spreads or transaction costs.[2]

What distinguishes cross-market trading of Bitcoin from other assets (e.g., stocks or bonds) is the underlying decentralized settlement system. While most traditional security markets are organized around trusted intermediaries, in decentralized settlement procedures no central clearing counterparty guarantees the ultimate delivery of the asset. Cross-market arbitrageurs are thus not able to dispose of their position until validators reach consensus and publicly record the transaction on the blockchain. Hautsch et al. (2019) show that the associated settlement latency resembles a novel friction that exposes cross-market arbitrageurs to the risk of adverse price movements and thus imposes limits to arbitrage.

Settlement latency in blockchain-based systems can be of orders of magnitudes of a couple of minutes. Biais et al. (2019) show that settlement latency arises as a result of limited processing capacities of validators. This restriction induces competition among the originators of transactions. Originators have the option to offer higher fees to create

[1] Limits to arbitrage arise, for instance, as trading costs (Roll et al., 2007), holding costs (Pontiff (1996), Gagnon and Karolyi (2010)), constrained arbitrage capital (Shleifer and Vishny, 1997), or in the presence of short sale constraints (Ofek et al., 2004). I refer to Gromb and Vayanos (2010) for an extensive survey of this literature.

[2] Makarov and Schoar (2020) document average daily price difference ratios between exchanges based in the US and Korea of 15%. Brauneis et al. (2019) find that the market for Bitcoin against USD is highly liquid in terms of bid-ask spreads but also document that markets are crossed most of the time. Hautsch et al. (2019) investigate 120 exchange-pairs and document average cross-market price differences of 63 basis points.

incentives for validators to include their transactions in the next block, as illustrated by Easley et al. (2019). Arbitrageurs compete with other users of the Bitcoin network (e.g., motivated by consumption or private transactions) for limited settlement capacities. High demand for settlement activities increases arbitrage costs in terms of either longer expected settlement latencies or higher required offered fees. The adoption of blockchain-based markets therefore comes at substantial costs for cross-market arbitrageurs.

In this paper, I analyze how blockchain related settlement latency affects two main pillars of functioning financial markets: price informativeness and liquidity provisioning. First, I estimate the arbitrage costs for the Bitcoin market and show that they increase with the number of transactions waiting for verification and the associated higher settlement latency. Simultaneously, I document that faster settlement (and hence lower arbitrage costs) is associated with larger quoted spreads. I reconcile these two seemingly contradicting main findings in a theoretical model in which liquidity providers anticipate that arbitrageurs exploit price differences due to stale information across markets more frequently if settlement is fast and thus set wider spreads to cope with the adverse selection risk. As a result, the direct effect of faster settlement on arbitrageurs activity is partially offset by larger liquidity-related arbitrage costs in form of wider bid-ask spreads.

The main econometric challenge is to quantify the magnitude of arbitrage costs because, in general, cross-market trading activities are not observable. Instead, the estimation of arbitrage costs requires an econometric model.

I provide a dynamic estimation framework based on quoted prices of one asset traded simultaneously at two markets. The framework rests on the assumption that arbitrageurs actively monitor and exploit price differences across both markets. If price differences occur, arbitrageurs buy at the cheaper market and sell at the more expensive market. As a result, price pressure from arbitrage activity enforces price adjustments at both markets towards the law of one price. Thus, price pressure from arbitrage capital implies a cointegration relationship between the quotes at the two markets because price differences mean-revert towards the long-run equilibrium relationship.

This correction mechanism is suspended whenever limits to arbitrage prevent profitable cross-market trading. More specifically, arbitrage costs determine a no-trade regime during which arbitrageurs prefer to stay idle. In such a regime, price differences may persist and remain unexploited. As a result, the adjustment of quotes towards the law of one price is non-linear in price differences and depends on arbitrage costs. From an econometric perspective, these considerations imply a threshold error correction model

for quotes in which arbitrage costs determine the magnitude of the thresholds.[3] The estimated threshold defines the no-trade regime and thus allows to back out arbitrage costs from the dynamics of quoted prices.

The econometric framework incorporates two important novel features: First, I parametrize the threshold as a function of (observed) proxies for arbitrage costs such as quoted spreads and network activity. As a result, the no-trade regimes reflect time-varying arbitrage costs. Second, the parametrization allows me to decompose arbitrage costs into static latent (exchange-specific) and dynamic friction-specific features.[4] I use the econometric model to estimate the no-trade regime thresholds. Hereby, I exploit a large dataset of high-frequency orderbook snapshots of two of the largest cryptocurrency exchanges for trading Bitcoin versus US-Dollar from April 2018 until August 2019. The estimated arbitrage costs on average amount to 9 basis points. The estimated costs attributed to settlement latency constitute 75% of the estimated threshold. To quantify the contribution of settlement latency to total arbitrage costs, I use the number of transactions waiting for verification in the Bitcoin settlement network as a measure of network activity. Network activity varies considerably over time, where high activity increases the settlement latency of an individual transaction and therefore the risks and costs for arbitrageurs.

I find that a one percent increase in network activity raises arbitrage costs due to settlement latency by 1.2 USD, whereas exchange-specific risks seem to play minor roles in preventing arbitrageurs from exploiting price differences. The increase of arbitrage costs due to higher network activity, however, is partially offset by a simultaneous adjustment of the quoted spreads. I find that an increase in network activity by one percent leads to a 0.4 USD decrease in spreads. This result is robust to controlling for trading volume and volatility.

I reconcile this finding in a theoretical framework that builds on two markets with asynchronous arrival of fundamental information about the same asset. Market fragmentation manifests itself in restrictions for some market participants to actively monitor and trade on information from both markets at the same time. Cross-market price differences occur if quotes on one of the two markets are based on concurrently outdated information. Whenever arbitrageurs observe profitable arbitrage opportunities and trade, their activity implicitly transmits information between the two markets. Two frictions impose

[3] Incorporating nonlinear adjustment processes to cointegrated variables goes back to, among others, Balke and Fomby (1997) and Hansen and Seo (2002).

[4] Ters and Urban (2020) propose a 3-regime threshold model to estimate latent arbitrage costs which are constant over time. Theissen (2012) incorporates quoted spreads into a similar regime model but does not estimate latent arbitrage costs.

limits to arbitrage: (i) exogenous costs such as settlement latency may render arbitrage costly and (ii) liquidity in terms of spreads which market makers set in anticipation of cross-market trading activity.

I derive quoted equilibrium spreads as a function of exogenous arbitrage costs and show that higher arbitrage costs imply a lower adverse selection component in the spreads of the locally competitive market makers. Hereby, the adverse selection risk decreases because higher arbitrage costs reduce the likelihood of a profitable arbitrage opportunity. Consequently, quoted spreads are largest when costs due to settlement latency are absent. In line with the empirical findings, the magnitude of the adverse selection related arbitrage costs depends inversely on the latency-related arbitrage costs. The overall effect of changing exogenous arbitrage costs on arbitrageurs activity and thus price informativeness, measured as aggregate mispricing at both markets, is thus ambiguous: the change in spreads can even overcompensate the direct effect of latency-related costs such that a reduction (or even complete removal) of the exogenous friction *decreases* arbitrage activity and also harms price informativeness.

In the blockchain-based market under consideration, the economic magnitude of the adverse selection effect on arbitrage activity is substantial: If network activity increases, latency-related arbitrage costs increase, but the adverse selection component in the spreads decreases, hampering the network effect on overall arbitrage costs by almost 30%. Therefore, the variation of the estimated total arbitrage costs is much smaller than the variation of the individual components. However, the decomposition reveals that, during periods of narrow spreads, market participants who demand liquidity but are not arbitrageurs benefit from the presence of blockchain-related arbitrage costs.

Overall, I argue that blockchain technology imposes a novel and economically significant friction for cross-market trading which differs substantially from the well-documented limits to arbitrage in markets for equities. As blockchain-based settlement fundamentally differs from trading that involves (trusted) centralized clearing counterparties, there are still many unknowns when it comes to its microstructure implications. Abadi and Brunnermeier (2018) point out that blockchain-based settlement cannot simultaneously satisfy the demand for security, decentralization and cost efficiency and therefore centralized (trusted) intermediaries may dominate in some situations. For instance, Chiu and Koeppl (2019) estimate that the US-corporate debt market may benefit from blockchain-based settlement. I provide a novel trade-off that is particularly relevant in the context of trading cryptocurrencies: due to the nature of the decentralized settlement process, fragmented markets cannot simultaneously achieve price informativeness and narrow spreads.

Instead, reducing or entirely removing the friction associated with settlement latency may have unintended consequences in terms of local liquidity provision. Initiatives to reduce settlement latency should thus consider the extent to which fundamental information can become efficiently distributed across trading venues.

Apart from the implications for blockchain-based trading, this paper also speaks to the effect of market frictions on arbitrage activity and liquidity in other asset classes. Concerning the US equity markets, O'Hara and Ye (2011) conclude that Regulation National Market System Rule 611 fostered the coexistence of trading venues. At the same time, however, market consolidation fosters cross-market information dispersion. Current debates on intentional latency delays (speed bumps) address concerns which are in line with my empirical findings: crowding out cross-market liquidity takers provides market makers with the opportunity to set smaller spreads (see, e.g. Budish et al. (2015) and Brolley and Cimon (2019)).[5] Whereas it is well established that large spreads constitute limits to arbitrage (see, e.g. Stoll (1989)), the reverse direction – the interaction between technological frictions and adverse selection risks received less attention. The results of this paper highlight that it is important to understand this interaction to evaluate technological and regulatory changes that target cross-market trading activity, e.g., short-sale constraints, intentional latency delays or the adoption of blockchain technologies in financial markets.

The structure of this paper is as follows: In Section 2, I present the main frictions prevalent in cryptocurrency markets, Section 3 provides the econometric framework and main results of the estimated arbitrage costs. In Section 4, I provide a theoretical framework to reconcile the observed effect of latency-related arbitrage costs on liquidity providers and arbitrageurs activity. Section 5 concludes.

2 Arbitrage in Cryptocurrency Markets

Cryptocurrency markets provide an excellent framework to investigate arbitrage costs and the subsequent implications for price informativeness due to at least two reasons: First, hundreds of trading venues to exchange cryptocurrencies against fiat money exist around the globe but substantial obstacles to cross-market arbitrage seem to persist and hamper price informativeness. Persistent price differences across these trading venues are well-documented and can only partially be reconciled with frictions such as withdrawal

[5] The implications of intentional latency delays are still under debate, see, e.g., Aldrich and Friedman (2017), Hu (2018), Woodward (2018) and Aoyagi (2018).

restrictions, regulatory pressure or exchange risks. For instance, Choi et al. (2018) provide evidence for substantial mispricing of Bitcoin at Korean trading venues due to strict capital constraints. Makarov and Schoar (2020) document substantial violations of the law of one price at many more trading venues across but also within countries borders. Market frictions hamper arbitrageurs activities and therefore harm price informativeness. To address the economic relevance of particular frictions and to rationalize persistent price differences, I quantify arbitrage costs.

A second identifying feature of cryptocurrency markets is the settlement procedure of the underlying blockchain-based asset. Trustless verification of cryptocurrency transactions replaces fast but potentially inefficient intermediaries. However, it has been documented that the limited capacities of proof-of-work consensus protocols and the large intra-daily variation of transactions waiting for settlement results in costly and time-consuming competition for the service of verification (see, e.g. Biais et al. (2019) and Easley et al. (2019)). More specifically, in order to exchange ownership of units of a blockchain-based asset such as Bitcoin involves moving the asset across wallets controlled by the opposing counterparty. Settlement of a trade in blockchain-based markets entails that the network verifies a transaction and subsequently guarantees that only the party that controls the receiving wallet can pursue further transactions. The time it takes between announcing such a transaction until settlement is non-trivial and and usually in orders of magnitudes of a couple of minutes. As Hautsch et al. (2019) show, settlement latency implies arbitrage costs due to non-hedgeable price risk for cross-market arbitrageurs. After observing price differences across two trading venues, arbitrageurs are not able to dispose of their position before the legal change of ownership is accomplished and may thus be faced to adverse price movements during the settlement period. The absence of trusted intermediaries such as clearing houses therefore imposes direct costs on arbitrageurs.

I use data from the Bitcoin network, one of the most popular decentralized protocols since Nakamoto (2008) published the concept and the underlying code. As of 2020, Bitcoin can be traded continuously on more than 400 markets that differ substantially in terms of location, fee structure and investors access. I employ high-frequency orderbook information from the public application interfaces of the largest cryptocurrency exchanges that feature BTC versus USD trading.

For the analysis at hand I use minute-level orderbook data from the exchanges *Bitstamp* and *Gemini* which both do not allow any form of margin trading. Both exchanges comply with the virtual currency license of the New York State Department of Financial

Table 1: Summary statistics of the orderbook snapshots.

variable	Mean	SD	Min	5%	Median	95%	Max
z_t (USD)	0.22	7.20	-296.36	-7.65	-0.20	9.53	499.31
Spread (Bitstamp, USD)	3.28	3.14	0.01	0.01	2.31	9.47	53.65
Spread (Gemini, USD)	1.68	2.47	0.01	0.01	0.71	6.40	87.03
δ_t (USD)	0.22	6.08	-293.51	-5.05	-0.00	6.80	497.06
$\lvert z_t \rvert$ (bp)	5.90	8.78	0.00	0.39	4.38	15.80	951.86
Spreads (bp)	4.06	3.25	0.01	0.23	3.35	10.07	118.30
$\lvert\delta\rvert$ (bp)	3.15	8.35	0.00	0.00	0.90	12.20	947.57
% Fraction of Excess Price Differences	58	49	-	-	-	-	-
Volume (Gemini, million USD)	21.87	19.31	1.91	3.42	16.73	56.43	176.03
Volume (Bitstamp, million USD)	66.02	56.30	5.37	14.09	48.72	173.00	477.67

Notes: This table provides summary statistics for the quoted prices for Bitcoin in USD on *Bitstamp* and *Gemini*. The sample is based on minute level information starting from March 1st, 2018 until August 1st, 2019. z_t corresponds to the midquote price differences (Bitstamp - Gemini as of Equation (1)). δ_t is the spread-adjusted midquote price differential as of Equation (3). Basis points (*bp*) are always computed by scaling with the average midquote across both exchanges. *Spreads (bp)* denotes $S_t^{\text{Bitstamp}} + S_t^{\text{Gemini}}$ standardized by the midquote, thus the total spreads required to trade across the two markets. *% Fraction of Excess Price Differences* corresponds to minutes in which spread-adjusted price differences imply an arbitrage opportunity. Trading *volume* is computed in million USD per day.

Services (DFS) and therefore both venues are accessible for US investors. The sample ranges from March 1st, 2018 until August 1st, 2019. Issues with time differences do not arise because both markets are open every day without any trading pauses. Reported daily trading volume on the two exchanges varies considerably, ranging from 1.9 million USD to almost 180 million USD at *Gemini* and exceeding 450 million USD at *Bitstamp* during periods of particular high trading activity. The time-series of cross-market midquote price differences is computed each minute as follows

$$z_t := q_t^{\text{Bitstamp}} - q_t^{\text{Gemini}} = \frac{1}{2}\left(a_t^{\text{Bitstamp}} - a_t^{\text{Gemini}} + b_t^{\text{Bitstamp}} - b_t^{\text{Gemini}}\right) \tag{1}$$

where a_t^k and b_t^k are the quoted best ask and best bid, respectively, and q_t^k denotes the midquote at time t at market k. Spreads at the best level denote the difference between the ask and bid price on each market.

$$\begin{pmatrix} S_t^{\text{Bitstamp}} \\ S_t^{\text{Gemini}} \end{pmatrix} := \frac{1}{2}\begin{pmatrix} a_t^{\text{Bitstamp}} - b_t^{\text{Bitstamp}} \\ a_t^{\text{Gemini}} - b_t^{\text{Gemini}} \end{pmatrix}. \tag{2}$$

Table 1 provides summary statistics of the quoted price differences and spreads. Midquote price differences z_t are centred around zero but reveal substantial variation over time,

sometimes reaching almost 500 USD (1000 bp). Quoted (half-)spreads at the two exchanges are relatively small but exhibit substantial variation over time. During some periods, spreads at *Gemini* spike to almost 90 USD. However, average spreads are around 2 bp which indicates that *Gemini* and *Bitstamp* resemble rather liquid markets, also in comparison to US equity markets.[6] Price differences in excess of the spreads constitute potential arbitrage opportunities. The average spread-adjusted price difference is

$$\begin{aligned}\delta_t &:= \max\left\{0, |z_t| - \left(S_t^{\text{Bitstamp}} + S_t^{\text{Gemini}}\right)\right\} \\ &= \max\left\{0, b_t^{\text{Gemini}} - a_t^{\text{Bitstamp}}, b_t^{\text{Bitstamp}} - a_t^{\text{Gemini}}\right\}. \end{aligned} \tag{3}$$

Average cross-market midquote differences adjusted for spreads are around 3 bp and positive during 58% of all minutes in the sample period.

Figure 1 visualizes price differences and quoted spreads at the two exchanges during the sample period. The black line corresponds to the midquote differences z_t in USD and the grey shaded area corresponds to the costs spanned by the minute-level sum of the quoted spreads $S_t^i + S_t^j$.[7] Consequently, the area between the two lines corresponds to cross-market spread-adjusted price differences δ_t. The figure suggests that arbitrage opportunities may exist as one can observe substantial price differences in excess of quoted spreads. Such an observation can only be reconciled with functioning financial markets if additional costs exceed the potential gains and therefore render trading unattractive. On longer time scales, deviations from the law of one price do not persist and instead, mean-reversion towards the law of one price seems to play a role.

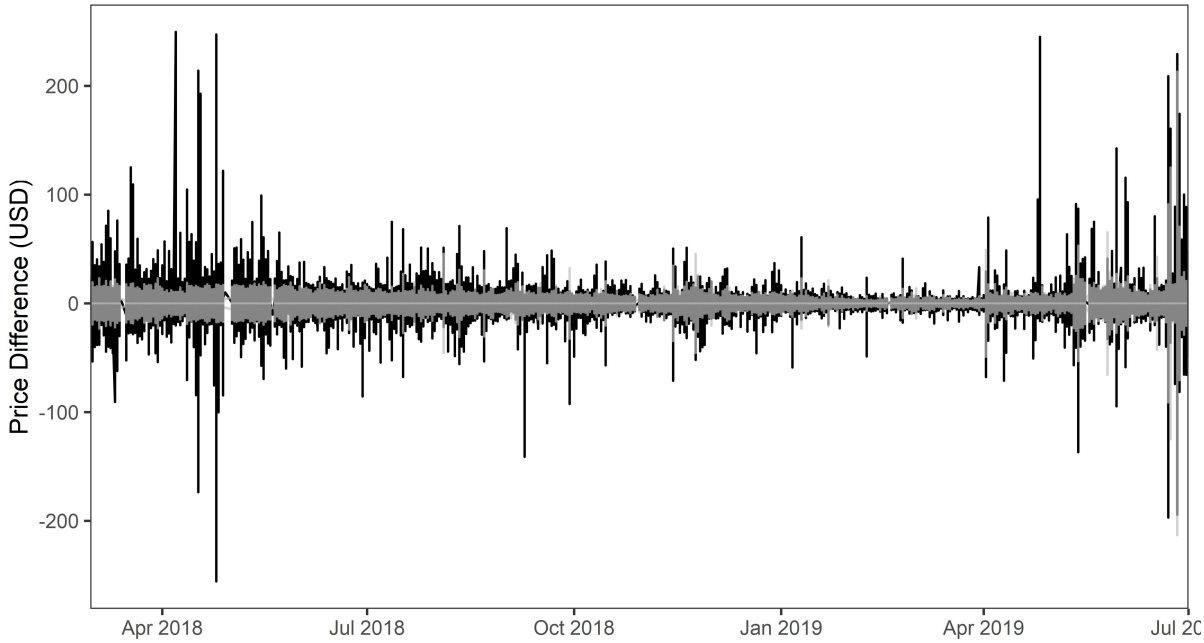
Therefore, the data reveals limits to arbitrage which prevent cross-market arbitrageurs from exploiting price differences $\delta_t > 0$. From an empirical perspective it is, however, challenging to identify the channels that impose arbitrage costs. Arguably, limits to arbitrage can arise in many different forms. As one source, I exploit variation in settlement latency as a proxy for price risks for arbitrageurs in order to quantify costs related to the decentralized settlement process in the market for cryptocurrencies.

To do so, I gather transaction-specific information from `blockchain.com`, a popular provider of Bitcoin network data and download all blocks verified during the sample period. I extract information about all verified transactions in this period. Each transaction contains a unique identifier, a timestamp of the initial announcement to the network, and

[6] Brogaard et al. (2014) documents relatve spreads from 6 basis points for large firms at NYSE.

[7] For the sake of readability I mirror the spreads around zero to visualize potential profitable arbitrage opportunities in both directions.

Figure 1: Quoted price differences and spreads.



Notes: This figure visualizes the minute-level price differences between *Bitstamp* and *Gemini*. The black line corresponds to the midquote price differences (z_t) in USD whereas a positive value indicates that *Bitstamp* quotes a higher price than *Gemini*. The grey area corresponds to the spreads required to buy and sell a marginal unit of Bitcoin at the two markets.

the fee (per byte) the initiator of the transaction offers validators to verify the transaction. Any transaction in the Bitcoin network, irrespective of its origin, has to go through the so-called mempool which is a collection of all unconfirmed transactions. These transactions wait in the mempool until they are picked up by validators and get verified. The Bitcoin protocol restricts the number of transactions that can enter a single block and therefore induces competition among the originators of transactions who can offer higher settlement fees to make it attractive for validators to include transactions in the next block.

Validators bundle transactions that wait for verification and try to solve a computationally expensive problem which involves numerous trials until the first validator finds the solution. For the Bitcoin protocol, validators successfully find a solution and append a block on average every 10 minutes. The number of transactions waiting for verification serves as a proxy for the activity of the Bitcoin network. The average number of transactions waiting for verification is above 8,400 and temporarily exceeds 39,000. As on average only around 1,000 transactions enter a single block, the queue of transactions

Table 2: Descriptive statistics of the Bitcoin network.

	Mean	SD	5 %	25 %	Median	75 %	95 %
Transaction Size	514.00	2169.86	192.00	225.00	248.00	372.00	962.00
Fee per Byte (Satoshi)	23.17	203.80	1.36	4.01	9.17	22.52	87.80
Fee per Transaction (USD)	0.60	8.08	0.02	0.07	0.15	0.40	2.00
Latency	30.54	165.31	0.73	3.58	8.85	20.28	90.87
# Waiting Transactions	8437.26	14438.03	324.00	1336.00	3429.50	8064.50	39415.00
% cross-exchange (daily)	2.62	1.92	0.47	1.40	2.11	3.45	6.56

Notes: This table reports descriptive statistics of our Bitcoin transaction data. The sample contains all transactions settled in the Bitcoin network from March 1st, 2018, until August 31, 2019. *Fee per Byte* is the total fee per transaction divided by the size of the transaction in bytes in Satoshi where 100,000,000 Satoshi are 1 Bitcoin. *Fee per Transaction* is the total settlement fee per transaction (in USD). I approximate the USD price by the average minute-level midquote across all exchanges in our sample. *Latency* is the time until the transaction is either validated or leaves the mempool without verification (in minutes). *Transaction Size* denotes the size of the transaction in bytes. *# Waiting transactions* is the number of transactions waiting for verification (per minute). % cross exchange (daily) corresponds to the (percentage) fraction of transactions that are associated with cross cryptocurrency-exchange transactions within a particular day.

waiting for verification implies settlement latency as the probability of being included in the next block decreases with the number of transactions that wait for settlement. Table 2 shows relevant summary statistics of the sample of Bitcoin transactions. The time until verification of a transaction in the Bitcoin network on average exceeds 30 minutes and exhibits substantial fluctuation. The costs of transferring Bitcoin from one wallet to another are on average 0.60 USD, irrespective of the trading size.

Panel A of Figure 2 illustrates the time-series of outstanding transactions during the sample period. Besides of regular intraday fluctuation patterns, periods of high network activity occurred particularly during December 2018 and since April 2019. The intra-daily variation of the Bitcoin network utilization is large. Panel B of Figure 2 illustrates the average number of transactions waiting for verification during the day, divided into intervals of 15 minutes in Central European Time. Network activity starts to spike at around 2pm CET which corresponds to 9am EST.

As postulated, for instance, by Easley et al. (2019) and Biais et al. (2019), the number of transactions waiting for verification increases the latency of all transaction waiting for verification. I illustrate this relationship in Figure 3. On days with many transactions waiting for verification, the average latency in minutes until settlement increases, reflecting the competition for settlement services in the decentralized network. Table 3 provides further evidence for the compelling relationship between network activity and

Table 3: Duration model.

Constant	α	Fee	Network activity	No flow	Full mempool
-0.0417	-1.5090	-0.0718	3.4773	0.6481	0.2798
(-1.2009)	(-56.2959)	(-2.5333)	(129.7267)	(29.2972)	(13.1538)

Notes: The table aggregates parameter estimates of a duration model of the latency of the observed transactions in the Bitcoin network. *Fee* is measured as fees per byte of transaction i, *Network acitvity* is measured by the (log) number of transactions waiting for verification. *No flow* is zero whenever a transaction can be identified as a cross-exchange transaction and one otherwise and *Full mempool* indicates a dummy variable that is one if the number of transactions waiting for verification exceeds the maximum allowed block size. Values in brackets denote $t - statistics$ based on the maximum likelihood estimation of the parameters in Equation (4).

the settlement latency in the Bitcoin network. The table contains parameter estimates of fitting the conditional probability density function of every transaction i with latency τ_i that has been verified by the Bitcoin network during the sample period. I model the latency as a Gamma distribution with rate parameter β_i and shape parameter α which implies a probability density function for τ_i of he form

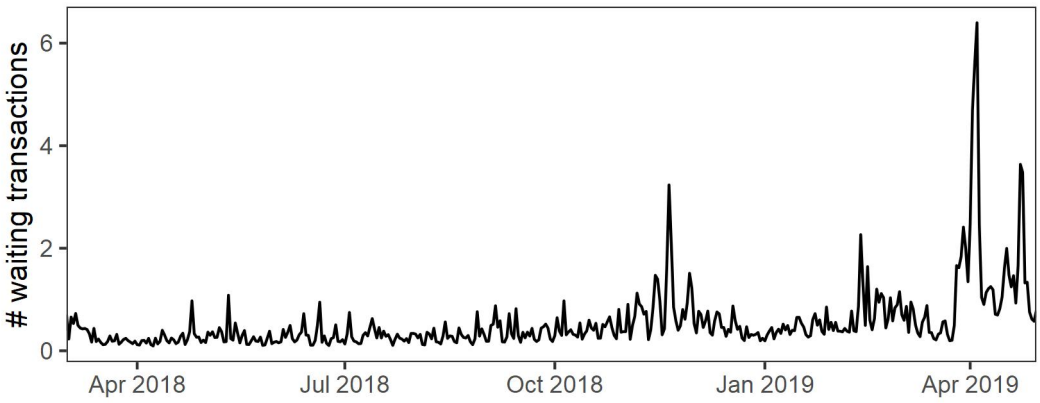
$$\pi(\tau_i|\theta_T) = \frac{\beta_i^\alpha}{\Gamma(\alpha)} \tau_i^{\alpha-1} e^{-\beta_i \tau_i}, \tag{4}$$

with

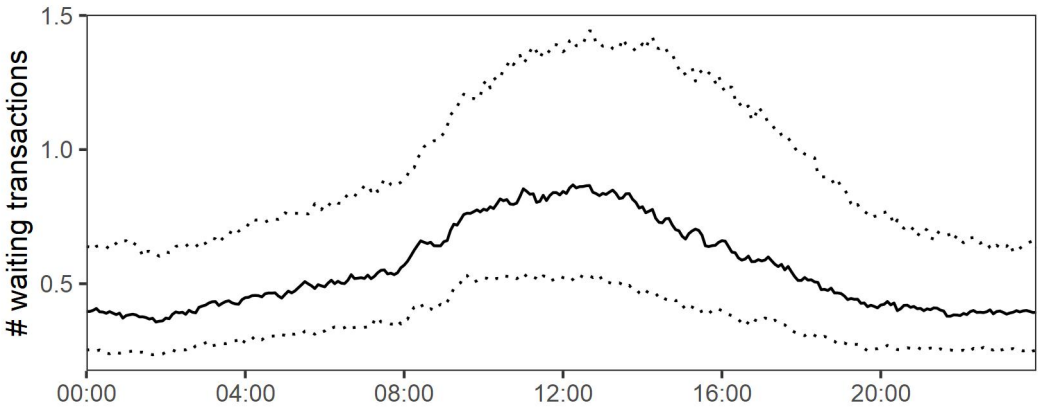
$$\theta_T := (\theta_T^\beta, \alpha)' \in \mathbb{R}^k \text{ and } \beta_i = \exp(-x_i' \theta_T^\beta), \alpha > 0. \tag{5}$$

Here, $\Gamma(\alpha)$ corresponds to the Gamma function. $x_i \in \mathbb{R}^K$ includes an intercept, the fee per byte, the number of transactions waiting for verification as a proxy for network activity, a dummy which is zero for cross-exchange transactions and one otherwise and a dummy which is one if the number of waiting transactions is larger than the hard-coded size constraint to adjust for potential non-linearities in the settlement latency. The Gamma distribution exhibits mean $\mathbb{E}(\tau_i) = \alpha \exp\left(x_i' \theta_T^\beta\right)$ and thus the specific parametrization allows to interpret estimate $\hat{\beta}_i$ as the sensitivity of the conditional mean with respect to (small) changes of variable x_i. The estimated parameters in Table 3 indicate that paying higher fees reduces the expected settlement latency whereas an increase in network activity, measured by the number of transactions waiting for verification, is associated with higher expected settlement latency. The effect is even more pronounced when the demand for settlement services by validators exceeds the network capacities in terms of the maximal blocksize. The relationship between the number of transactions waiting for

Figure 2: Network activity.



Notes Panel A: Number of transactions waiting for verification. This figure shows the time series of the daily average number of transactions waiting for verification (in 10,000 transactions waiting for verification).

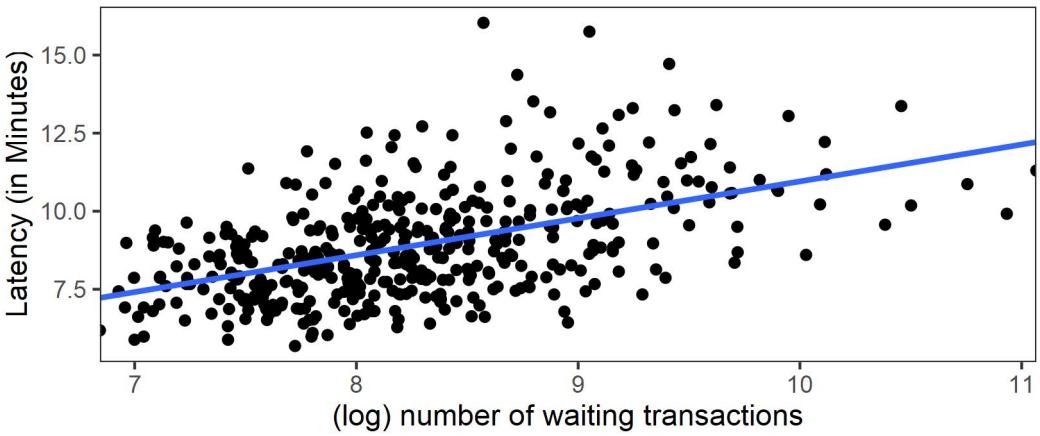


Notes Panel B: Intra-daily fluctuations.This figure shows the average number of transactions waiting for verification during 15 minute intervals over the day. The dotted lines correspond to the 5% and 95% quantiles (in 10,000 transactions waiting for verification).

verification and settlement latency implies non-hedgeable price risks. Therefore, if price risk due to settlement latency plays a role for the activity of arbitrageurs one would expect the no-trade region of arbitrageurs to widen during times of high network utilization.

For the analysis, two observations are important: First, the exchanges *Gemini* and *Bitstamp* both net trades internally, effectively circumventing settlement latency for quote updating activities. Therefore, the number of transactions waiting for verification is relevant only for cross-market arbitrageurs which are forced to use the blockchain to move funds between cryptocurrency-exchange controlled wallets. Second, a spike in network activity due to arbitrageurs transactions waiting for verification is unlikely to be of any

Figure 3: Network activity and settlement latency.



Notes: This figure visualizes the relationship between the number of transactions waiting for verification and the latency of subsequently verified transactions. The scatterplot shows with (log) number of transactions waiting for verification on the x-axis and the average daily waiting time (in minutes) of all transactions verified on that particular day. The blue line indicate the OLS estimator of the slope and constant of regressing waiting time on log of the number of waiting transactions.

concern. To merit the last point. a more careful discussion is required: The Bitcoin network is utilized for transactions of any purpose, including consumption or financing of illegal activities (see, e.g., Foley et al. (2019)). Cross-market trades with the purpose of exploiting price differences are of negligible relevance and therefore I do not detect any feedback effects from the presence of price differences on settlement latency. To illustrate that cross-market arbitrageurs do not induce inflated network activity, I examine a novel dataset that allows to identify potential arbitrage transactions and I find that such transactions only represent a minor fraction of all transactions waiting for verification and further do not differ with respect to the relevant summary statistics from transactions unrelated to cross-market arbitrage activity.

Settlement latency affects cross-market arbitrageurs because they cannot dispose of their position before the sell-side exchange accepts their Bitcoin deposit as valid. Whereas the Bitcoin blockchain is public, trades are usually not disclosed directly. However, to provide more information regarding the cross-market Bitcoin flows, I collect a list of wallets which are likely under the control of the exchanges in my sample.[8] Although Bitcoin transactions are pseudonymous in the sense that the transactions publicly reveal all addresses associated with a transaction, but it is hard to map these addresses to their

[8]The procedure to detect exchange-controlled wallets is laid out in Meiklejohn et al. (2013). I am grateful to Sergey Ivliev for his support in providing the data.

Table 4: Descriptive statistics of the cross-exchange transactions.

	Mean	SD	5 %	25 %	Median	75 %	95 %
Fee per Byte (Satoshi)	44.52	78.38	3.39	8.22	14.46	39.69	238.66
Fee per Transaction (USD)	4.91	27.84	0.06	0.17	0.74	2.97	9.86
Latency	12.39	24.70	0.50	3.00	7.27	14.70	36.53
Transaction Size (byte)	1708.47	4342.36	223.00	249.00	424.00	1172.00	5220.00

Notes: This table reports descriptive statistics of the Bitcoin transaction data. The sample contains all identified cross-exchange transactions settled in the Bitcoin network from March 1st, 2018, until August 31, 2019. *Fee per Byte* is the total fee per transaction divided by the size of the transaction in bytes in Satoshi where 100,000,000 Satoshi are 1 Bitcoin. *Fee per Transaction* is the total settlement fee per transaction (in USD). I approximate the USD price by the average minute-level midquote across all exchanges in our sample. *Latency* is the time until the transaction is either validated or leaves the mempool without verification (in minutes). *Transaction Size* denotes the size of the transaction in bytes. *Mempool Size* is the number of other transactions in the mempool at the time a transaction of our sample enters the mempool.

respective physical or legal owners. The wallet IDs allow to identify all transactions in which funds have presumably been moved between two exchange-controlled wallets. All cross-market arbitrage activity should therefore be a subset of the transactions included in this sample (in principle, agents could also move funds between trading platforms due to reasons which are not related to exploiting price differences). Table 4 provides comparative statistics of the underlying characteristics of the two subsets of transactions. *Standard* comprises of all transactions of the Bitcoin network and *Arbitrage* comprises of all transactions which are flagged as being potential cross-market fund flows. The table reveals that in general cross-exchange flows settle faster, the average latency in minutes only comprises about 50% of the latency of the entire network. However, this does not mean that arbitrageurs can move funds between two exchanges more efficiently than other market participants. The table reveals that fees paid for cross-exchange flows are on average about 5 USD, roughly 10 times as much as the average transaction fee in the entire network. Therefore, cross-exchange flows are settled faster because they are more valuable from the perspective of miners. However, these settlement fees resemble arbitrage costs for arbitrageurs which should reflect the trade-off between paying higher fees and getting faster settlement.

3 Estimating Arbitrage Costs

From an empirical perspective, it is a challenge to quantify arbitrage costs because they are generally not observable. Whereas potential impediments to arbitrage could be iden-

tified, e.g., the presence of short selling restrictions or settlement latency, the magnitude and economic relevance of these frictions remains opaque.

In principle, one could try to draw indirect inference about arbitrageurs' activity by evaluating potential portfolio holdings. Recent literature investigates the long side of arbitrage trading based on hedge fund stock holdings (Brunnermeier and Nagel, 2004), the short side by investigating short-selling activity on stocks (Hanson and Sunderam, 2014) or a combination of both trading directions to examine net arbitrage trading (Chen et al., 2019). However, in many applications and specifically in the context of cryptocurrency markets, asset holdings and trading activity may remain opaque. To overcome this measurement issue, I instead identify arbitrage trading activity based on quote dynamics to quantify the extent to which arbitrageurs refrain from exploiting price differences due to market frictions. The econometric framework rests on the notion of arbitrageurs actively enforcing the law of one price as long as arbitrage costs do not render the otherwise profitable trading activity unattractive.[9]

I derive a threshold vector error correction model for observed quoted prices based on the notion that cross-market arbitrage trading enforces the law of one price. Price pressure from arbitrage trades implies mean-reversion towards the efficient price. Arbitrage costs impose a threshold for price differences below which arbitrageurs prefer to stay idle instead of trading away price differences. Empirically, this strategic behaviour implies a three regime threshold model for the time series of price differences which allows to identify latent arbitrage costs. The underlying theory based on estimating threshold vector-error correction models is close in spirit to applications on exchange rates (Lo and Zivot, 2001), commodity markets (Park et al. (2007) and Stevens (2015)) and trading of futures (Dwyer et al. (1996) and Forbes et al. (1999)). As a result, the structural framework does rely solely on the dynamics of quoted prices at fragmented markets to detect cross-market arbitrage activity without relying on the identification of portfolio holdings of arbitrageurs.

The framework is flexible to allow for time varying arbitrage costs in response to shifts in potential proxies for the arbitrage costs. Given the focus to disentangle the effect of arbitrage costs due to settlement latency in blockchain-based markets on liquidity and price informativeness, time-variation of arbitrage costs is a particularly relevant feature, which however, may also apply in different market contexts. Time-varying thresholds constitute one of the main differences of my framework compared to, among others, Balke and

[9]The econometric model rests on a generalization of the theoretical framework presented in Section 4 which provides a micro-foundation for the underlying threshold vector-error correction model.

Fomby (1997) and Ters and Urban (2020). Whereas Theissen (2012) proposes to include time-varying liquidity in the regime specification, he abstracts from estimating latent arbitrage costs. Dwyer et al. (1996) instead, do not allow for asymmetries depending on the direction of the trade.

More specifically, I exploit time-variation of a proxy for a particular friction and attribute its contribution to the arbitrage costs. Then, the econometric model allows to back out the arbitrage costs and the sensitivity of the threshold magnitude with respect to the proxy using observed data.

3.1 Econometric model

I consider one risky asset, traded at two market i and j, where q_t^k denotes the midquote of the asset at time t at market $k \in \{i, j\}$.[10] As the asset at both markets provides a claim on otherwise identical cash flows, the central law of one price defines a cointegration relationship of the midquotes. Intuitively speaking, efficient markets imply that price differences

$$z_t := q_t^i - q_t^j \tag{6}$$

are stationary. Note, that price differences may occur due to asynchronous information revelation (Kotz et al., 2012) and asymmetric private valuation shocks (Foucault et al., 2017). Further, limits to arbitrage may render cross-market trading unattractive and therefore hamper the convergence of z_t towards the law of one price (Gromb and Vayanos, 2010).[11]

The distribution of the midquote price differences z_t evolves as a random walk within information arrivals but exhibits mean reversion whenever quotes are updated, either due to a news event or due to arbitrage trades.[12] The main assumption for the econometric model is that arbitrageurs exploit price differences $q_t^i - q_t^j$ only if they exceed the total arbitrage costs. Price pressure due to order flow arises as soon as prices deviate from their equilibrium relationship. Therefore, arbitrage activity reinforces the law of one price (see, e.g., Ross (1976)). The presence of arbitrage costs implies that price pressure is present only if price differences are large enough, which allows to identify the following three different regimes:

[10] In the estimation I always consider the log of midquotes as the relevant measure.

[11] In the theoretical framework in Section 4 price differences are caused by asynchronous information arrival. It turns out that even in presence of arbitrage costs, the equilibrium relationship in Equation 6 is stationary.

[12] See Chan (1993) for necessary stationary conditions of z_t.

Definition 1. *Arbitrage activity implies three possible regimes r_t of the economy at time t: either the arbitrageur prefers to stay idle (r_t = no trade), she sells at market i and buys at market j (r_t = pos), or vice versa (r_t = neg). Formally, given price difference $z_t = q_t^i - q_t^j$ at time t, the economy is in the following regime:*

$$r_t = \begin{cases} pos, & if\ z_t > c_t^{pos} \\ neg, & if\ -z_t > c_t^{neg} \\ no\ trade, & if\ c_t^{neg} \leq z_t \leq c_t^{pos} \end{cases} \tag{7}$$

where $c_t^{pos} \geq 0$ and $c_t^{neg} \geq 0$ correspond to the arbitrage costs at time t.

The regime r_t is central for the identification of the arbitrage costs. In a frictionless market the theoretical no-arbitrage condition requires price differences to be exploited as soon as they arise, price dynamics would collapse to a 1-regime error correction model, price differences would be stationary and price pressure would be present continuously. However, if arbitrage costs establish market fragmentation price differences may become sizeable which implies a non-linear adjustment process towards the long-run equilibrium given by the law of one price. The magnitude of the no-trade region determines if price differences persist because they do not correspond to a profitable arbitrage strategy. The boundaries of the no-trade region $[c_t^{\text{neg}}, c_t^{\text{pos}}]$ are defined as the minimum price differences required to make the arbitrageur indifferent between trading and staying idle (r_t = *no trade*). If price differences z_t exceed the threshold, however, one can expect price pressure to reinforce the equilibrium relationship represented by the law of one price. In other words, the law of one price may temporarily fail to hold whenever r_t = *no trade*, whereas quoted prices exhibit mean reversion if and only if the economy is in regime r_t = *pos* or r_t = *neg*. Two features in Definition 1 are of particular importance: First, arbitrage costs may be asymmetric in the sense that it can be costlier for the arbitrageur to perform cross-market trades in one direction rather than in the reverse direction ($c_t^{\text{pos}} \neq c_t^{\text{neg}}$). Potential reasons for asymmetries are, e.g., short-sale constraints or asymmetric buy and sell fees. In the specific case of cryptocurrency markets, asymmetric arbitrage costs play a role of particular importance: Regulatory differences across national borders and in particular capital controls have been shown to induce substantial premiums for the price of, e.g., Bitcoin, for instance on Korean exchanges (Choi et al., 2018). Further, as cryptocurrency exchanges serve as custodian of customer funds, substantial risks may be associated with holding assets on particular exchanges. As a result, price pressure stemming from arbitrageurs activity may response asymmetrically to price differences.

The second relevant feature of Definition 1 is that arbitrage costs can be time-varying such that during some periods the no-trade region widens. If $z_t > 0$, it is profitable to buy from market j and sell at market i. Consequently one can expect prices to adjust at least until $z_t = c_t^{\text{pos}}$. As a result, quotes at market j may increase and simultaneously quotes at market i may decrease. Further, price adjustments are presumably non-instantaneous. More specifically, the implied return dynamics $\Delta q_t^k := q_t^k - q_{t-1}^k$ are as follows:

$$\begin{pmatrix} \Delta q_t^i \\ \Delta q_t^j \end{pmatrix} = \begin{pmatrix} \mu_{i,t}^r \\ \mu_{j,t}^r \end{pmatrix} + \begin{pmatrix} \alpha_i^r \\ \alpha_j^r \end{pmatrix} \underbrace{\begin{pmatrix} 1 & -1 \end{pmatrix} \begin{pmatrix} q_{t-1}^i \\ q_{t-1}^j \end{pmatrix}}_{z_{t-1}} + \begin{pmatrix} u_t^i \\ u_t^j \end{pmatrix} \tag{8}$$

where $\mu_{k,t}^r$ corresponds to a potential time-varying mean specification, for instance in the presence of autoregressive dynamics and u_t^k denotes the innovation process. Equation (8) corresponds to a vector error correction model in the spirit of Engle and Granger (1987). The equation implies that price adjustments can occur due to idiosyncratic shocks u_t or due to the activity of arbitrageurs in response to profitable arbitrage opportunities. α_i^r and α_j^r correspond to the price adjustment in response to price pressure from arbitrage trades.

If, for instance, $r_t = pos$, the arbitrageur buys at exchange j and sells at the expensive exchange i. Consequently, prices at the buy side market should decrease ($\Delta q_t^i < 0$) and prices at the sell side market should increase ($\Delta q_t^j < 0$). For the adjustment terms this implies $\alpha_j^{\text{pos}} \geq 0$ and $\alpha_i^{\text{pos}} \leq 0$. The reverse case holds for $r_t = neg$, implying that $\Delta q_t^i > 0$ and $\Delta q_t^j < 0$. This is the case if, again, $\alpha_j^{\text{neg}} \geq 0$ and $\alpha_i^{\text{neg}} \leq 0$. However, although theoretically the sign of the adjustment terms should be identical for the regimes *pos* and *neg*, magnitudes of price adjustment may differ across the regimes. Most importantly, during the regime $r_t = no\ trade$, no price adjustment due to arbitrage trading should be present and thus price dynamics only follow idiosyncratic shocks such that $\Delta q_t^k = \mu_{k,t}^{\text{no trade}} + u_t$.

Instead of imposing the law of one price directly, one could also aim at estimating the cointegration relationship $(1, \quad \beta)\left(q_t^i, \quad q_t^j\right)'$ which would significantly increase the estimation uncertainty (see, e.g., Hansen and Seo (2002), Seo (2011), Ters and Urban (2020)). Latter approaches focus on uncertain or unstable cointegration relationships, for instance in applications related to statistical arbitrage. I, instead, impose the law of one price as an economically motivated relationship to mitigate identification issues (see, e.g., Martens et al. (1998) and Stevens (2015)).

3.2 Parametrization and estimation

Next, I provide a (Bayesian) framework to estimate the arbitrage costs c_t^r of the econometrics model given by Equation (7) and Equation (8). First, I parametrize the time-varying arbitrage costs c_t^r as a linear function of fixed costs and exposure to a potentially time-varying proxy of a source of exogenous arbitrage costs such as settlement latency. The estimation framework explicitly takes into account liquidity costs in form of bid-ask spreads.

Definition 2. *I parametrize the threshold c_t^r from Equation (7) as a function of observable proxy for arbitrage costs x_t with $\mathbb{V}(x_t) > 0$ such that*

$$c_t^r := \max(0, c^r + c_1 x_t) + S_t^i + S_t^j. \tag{9}$$

Here, c^r corresponds to (unobservable) fixed arbitrage costs which may depend on the current regime and reflect, for instance, exchange risks or constraints due to capital controls. x_t is an vector of observations that proxy a time-varying source of arbitrage costs for an arbitrageur. This could be, for instance, a time dummy to reflect changes in regulation (e.g., event fixed effect), shorting costs or, for blockchain-based markets, network activity as an instrument for settlement fees and latency. The parameter c_1 captures the effect of x_t on the magnitude of the no-trade region. The parametrization allows to derive the sensitivity of the threshold c_t^r with respect to changes in the chosen instrument for arbitrage costs as follows:

$$\frac{\partial c_t^r}{\partial x_t} = c_1 + \frac{\partial S_t^j}{\partial x_t} + \frac{\partial S_t^i}{\partial x_t}. \tag{10}$$

Here, $\frac{\partial S_t^k}{\partial x_t}$ captures the effect of exogenous arbitrage costs on quoted spreads and can be modelled directly because the spreads S_t^k and the proxy x_t are both observable. The econometric model therefore allows to infer the presence of arbitrageurs depending on the non-linear adjustment towards the law of one price. Further, the parametrization provides an intuitive decomposition of arbitrage costs into exchange-specific, proxy-related and liquidity-driven components.

For estimation purposes, the threshold vector error correction model in Equation (8) can be rewritten as a multivariate linear regression

$$\Delta V^r = X^r B^r + U^r \text{ where } r \in \{\textit{neg, pos, no trade}\} \tag{11}$$

with

$$\Delta V_{t_r}^r = \begin{pmatrix} \Delta v_{t_r}^i \\ \Delta v_{t_r}^j \end{pmatrix}' \text{ and } X_{t_r}^r = \begin{pmatrix} 1 \\ z_{t_r-1} \end{pmatrix}'. \tag{12}$$

Here, t_r corresponds to the stacked dates of all observations in regime r. The parameters to estimate[13] are

$$\theta = \left\{ \underbrace{\begin{pmatrix} \mu_1^{\text{neg}} & \mu_2^{\text{neg}} \\ \alpha_1^{\text{neg}} & \alpha_2^{\text{neg}} \end{pmatrix}}_{B^{\text{neg}}}, \begin{pmatrix} \mu_1^{\text{pos}} & \mu_2^{\text{pos}} \\ \alpha_1^{\text{pos}} & \alpha_2^{\text{pos}} \end{pmatrix}, \begin{pmatrix} \mu_1^{\text{no trade}} & \mu_2^{\text{no trade}} \\ \alpha_1^{\text{no trade}} & \alpha_2^{\text{no trade}} \end{pmatrix}, \begin{pmatrix} \Sigma^{\text{neg}} \\ \Sigma^{\text{pos}} \\ \Sigma^{\text{no trade}} \end{pmatrix}, \begin{pmatrix} c^{\text{neg}} \\ c^{\text{pos}} \\ c_1 \end{pmatrix} \right\}. \tag{13}$$

Under the assumption that the error terms U^r are zero-mean multivariate normal distributed with variance covariance matrix Σ^r, the likelihood of the data conditional on the parameters θ is given by

$$\mathcal{L}(\Delta V|\theta, X) \propto \prod_{\substack{r \in \{neg,\ pos,\\ no\ trade\}}} |\Sigma^r|^{-\frac{T^r}{2}} \exp\left(-\frac{1}{2}\text{tr}\left((\Sigma^r)^{-1} U_\theta^{r\prime} U_\theta^r\right)\right) \tag{14}$$

where T^r is the number of observations in regime r and $U_\theta^r = \Delta V_{t_r}^r - X^r B^r$.

The estimation can be performed either via concentrated maximum likelihood (see Tong (1983), Tsay (1998) and Hansen and Seo (2002)) or by means of Bayesian inference (see, e.g., Forbes et al. (1999) and Huber and Zörner (2019)). I employ a standard Bayesian approach that effectively circumvents issues related to the optimization of complex likelihood functions. The results are quantitatively similar for concentrated maximum likelihood methods.

I specify non-informative prior distributions for $\beta^i := \text{vec}\left(B^i\right), \Sigma^i$, c_0 and c_1 as follows:

$$p(\Sigma^r) \sim IW(V, v), p(\text{vec}(B^r)|\Sigma^r) \sim MN(0, C \otimes \Sigma^r), p(c) \sim U(-\infty, \infty) \tag{15}$$

where $IW(\cdot, v)$ corresponds to an inverse Wishart Distribution with v degrees of freedom and positive scale matrix V, $MN(\cdot)$ corresponds to a multivariate normal distribution and $U(\cdot)$ corresponds to a uniform distribution with unbounded support. In the empirical

[13]Extended models that capture potentially autoregressive components in the spot drift μ_i^r do provide qualitatively very similar results in the given setup. I therefore focus on a parsimonious structure with static drift term.

analysis, I use $v = 2$, $V = 10^{-5}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ and $C = V^{-1}$ as hyperparameters. I perform inference on θ using Monte Carlo Markov Chain methods and provide a detailed description of the sampling algorithm in Appendix D. Inference is drawn based on Gibbs sampling procedures with Metropolis Hasting steps. The sample of parameters is generated with 50 parallel chains of length 20.000 after discarding 20.000 burn-in iterations.

3.3 Estimation results

Table 5 summarises the posterior distribution of the parameters. Here, the estimation is based on setting $\mu^r_{i,t}$ to a constant such that the econometric model resembles Equation (11).

As a natural proxy for settlement fees and latency in the market for cryptocurrencies, I employ the (log) number of transactions waiting for verification in the Bitcoin network at minute t. High values of x_t are therefore associated with high network activity and larger competition among originators of transactions for mining services.

As of Panel A of Table 5, the estimated parameters $\hat{c}^r_0$ are between 3.7 USD and 5.05 USD. The estimated parameters reflect substantial costs for cost market trading which may be attributed to withdrawal fees and exchange risks. Further, sampling from the posterior distribution allows to evaluate the distribution of the difference $c_0^{\text{pos}} - c_0^{\text{neg}}$ which is positive, thus arbitrage costs are different depending on the direction of the trade. As a main result, the exposure of the arbitrage costs to the number of unconfirmed transactions, $\hat{c}_1$, is positive. Thus, higher blockchain activity imposes additional costs for cross-market arbitrageurs. Moreover, an increase in the number of outstanding transactions by 1% implies an average increase in the no-trade region by more than 2 bp. Therefore, fluctuations in the network activity, as for instance characterized by intra-daily variation as of Figure 2 imply considerable impediments for arbitrageurs and may impose severe effects on price informativeness.

Estimated total arbitrage costs in excess of quoted spreads, $\hat{c}^r := \max(0, \hat{c}^r_0 + \hat{c}^r_1 \bar{x}_t)$ where $\bar{x}_t$ is the time-series average of the (log) number of transactions waiting for verification amount to 15.11 USD in the case where *Bitstamp* is the sell-side market and 13.81 USD in the reverse direction, hinting at substantial arbitrage costs in excess of the spreads. Therefore, pushing forces back to to the law of one price due to arbitrageurs activity are present only, if price differences exceed these thresholds *after* already having adjusted for the spreads.

Table 5: Posterior parameter estimates.

Notes Panel A: Posterior estimates of c_t^r. Diff corresponds to the posterior distribution of the difference of the two fixed cost parameters, $c_0^{\text{pos}} - c_0^{\text{neg}}$.

r	Neg	Pos	$c_0^{\text{pos}} - c_0^{\text{neg}}$	
c_0	3.761	5.055	1.294	
	(3.45, 3.79)	(4.99, 5.62)	(1.21, 2.17)	
c_1				1.271
				(1.23, 1.67)

Notes Panel B: Posterior estimates of α_r and μ^r.

r	*Pos*	*No trade*	*Neg*
α_{bitstamp}	-0.789	-0.049	-0.655
	(-0.8, -0.77)	(-0.15, 0.14)	(-0.67, -0.62)
α_{gemini}	0.084	0.004	0.091
	(0.027, 0.091)	(-0.08, 0.09)	(0.081, 0.095)
μ_{bitstamp}	-0.77	0.011	0.36
	(- 0.78,-0.71)	(-0.01, 0.03)	(0.29, 0.38)
μ_{gemini}	0.847	-0.001	-0.848
	(0.74, 0.95)	(-0.02, 0.01)	(-0.94, -0.66)

Notes: This table summarises the posterior distribution of the main parameters of interest of the model as of Equation (7) and Equation (11). Panel A contains posterior means of the threshold parameters c_0^r and c_1^r. The values in brackets correspond to the 99% credible regions. Panel B contains posterior means and credible regions for the adjustment parameters α_k^r and the mean values μ from the different regimes r. Inference is drawn based on Gibbs sampling procedures with Metropolis Hasting steps as illustrated in the Appendix. The sample of parameters is generated with 50 parallel chains of length 20.000 after discarding 20.000 burn-in iterations.

In relation to the average absolute price difference, $|\delta_t|$, the magnitudes of $\hat{c}^r$ are substantial, covering almost 63% of all observed price differences in the sample. Relative to quoted spreads, latent arbitrage costs resemble more than 75% of the no-trade regions implied by the data.

Panel B of Table 5 contains summary statistics of the posterior distribution of the remaining parameters of interest, $\hat{\alpha}_k^r$ and $\hat{\mu}_k^r$. The adjustment parameters α_k^r exhibit the expected sign and reflect mean-reversion towards the law of one price whenever price differences indicate a regime different from r_t = no trade. Positive price differences (r_t = pos) imply that *Bitstamp* serves as a sell-side exchange and thus prices are expected to decrease at *Bitstamp* and to increase at the buy-side exchange *Gemini*. The credible regions of $\hat{\alpha}_k^0$ for a regime without arbitrage trading both contain zero, indicating that price differences evolve as random walks and are not actively exploited. The spot drift

terms μ_k^r reflect behaviour in line with the adjustment parameters: In periods where trading is profitable for arbitrageurs, price differences tend to decrease at both market.

The price dynamics further indicate that *Gemini* adjusts to price differences at a lower rate than its competitor *Bitstamp*. This could be due to at least two reasons: Either the price impact of arbitrageurs at *Gemini* is mostly absorbed or the rate of information arrival at *Gemini* leads the market with *Bitstamp* subsequently adjusting its current quotes.[14]

Next, I investigate the relationship between arbitrage costs and the liquidity component in the spreads. The empirical results indicate that market fragmentation due to constrained arbitrageurs activity plays a substantial role in cryptocurrency markets. As it has been noted before, market fragmentation may foster adverse selection risks that could harm liquidity provision. For instance, Foucault et al. (2017) notes that stale quote trading may impose adverse selection costs. However, it should be noted that cross-market liquidity providers may also act as arbitrageurs and benefit liquidity providers.

Based on the theoretical framework developed in Section 4, adverse selection should play a role in fragmented markets. Subsequently, shifts in arbitrage costs due to network activity should also be reflected in the dynamics of the quoted spreads. To estimate overall sensitivity of arbitrage costs due to network activity as of Equation (10), I turn to the effect of network activity on quoted spreads, $\partial S_t^k / \partial x_t$.

The following empirical analysis rests on the observed spread at the two markets in the sample. I estimate a vector autoregressive model of the form:

$$\begin{pmatrix} S_t^{\text{Bitstamp}} \\ S_t^{\text{Gemini}} \end{pmatrix} = \begin{pmatrix} \gamma_1 \\ \gamma_2 \end{pmatrix} + \begin{pmatrix} \omega_1 \\ \omega_2 \end{pmatrix} x_t + \begin{pmatrix} \rho_1^1 & \rho_2^1 \\ \rho_1^2 & \rho_2^2 \end{pmatrix} \begin{pmatrix} S_{t-1}^{\text{Bitstamp}} \\ S_{t-1}^{\text{Gemini}} \end{pmatrix} + \Gamma\beta + \begin{pmatrix} u_t^1 \\ u_t^2 \end{pmatrix} \tag{16}$$

where x_t is the (log) number of transactions waiting for verification at minute t as a proxy for the arbitrage costs and u_t^k are potentially correlated normally distributed error terms. Γ contains a host of control variables to explain quoted spreads. As controls, I use exchange trading volume V_t^k (see, e.g. Lin et al. (1995) and Stoll (1989)), the lagged cross-market average midquote and the cross-market average minute level spot volatility σ_t (see, e.g. Easley and O'Hara (1987)). I estimate the spot volatility using the procedure proposed by Kristensen (2010).[15]

Table 6 illustrates the parameter estimates of the vector autoregressive structure as of

[14] In Section 4, I provide a theoretical model of asynchronous information arrival at fragmented markets that incorporates different information arrival rates and could explain this result.

[15] I refer to Hautsch et al. (2019) for further information regarding estimating σ_t.

Table 6: Spread decomposition.

	Dependent variable:	
	S_t^{Bitstamp}	S_t^{Gemini}
$S_{t-1}^{\text{Bitstamp}}$	0.455***	0.028***
	(0.001)	(0.001)
S_{t-1}^{Gemini}	0.051***	0.552***
	(0.001)	(0.001)
γ	−13.593***	−2.683***
	(0.115)	(0.095)
x_t	**−0.114*** **	**−0.051*** **
	(0.003)	(0.002)
σ_t	5.818***	6.420***
	(0.136)	(0.112)
$\bar{q}_{t-1}$	1.824***	0.419***
	(0.013)	(0.011)
V_t^{Gemini}	0.0003	0.004***
	(0.0004)	(0.0003)
V_t^{Bitstamp}	0.005***	−0.003***
	(0.0002)	(0.0001)
Observations	588,716	

Parameter estimates of the Vector autoregressive model as of Equation (16). x_t denotes the (log) number of transactions waiting for verification, σ_t denotes the minute level spot volatility, $\bar{q}_t$ is the cross-market average midquote, V_t^k corresponds to trading volume (daily) at exchange k. Values in brackets denote p values of the estimated parameters.

Equation (16). The parameters ω_1 and ω_2 are both negative, indicating that an increase in the number of transactions waiting for verification decreases quoted spreads.

The empirical analysis reveals two different effects of a shift in the arbitrage costs for arbitrageurs on the arbitrageurs participation constraint as of Equation (7). First, Panel A of Table 5 illustrates a direct cost effect. More transactions waiting for verification increase the price risks cross-market arbitrageurs in the Bitcoin market are exposed to and consequently widen the no-trade region which constitutes the positive coefficient c_1. Second, an increase in network activity also narrows spreads.

In terms of arbitrage costs, the two effects above yield into different direction. To quantify the net effect of network activity on arbitrage costs in blockchain-based market, I return to Equation (10) which allows to decompose the two effects the effects.

A one percent increase of the estimated no-trade threshold c_t^r due to an increase in

network activity is associated with a 0.3 percent decrease of quoted spreads. Therefore, the adjusted spreads partially compensate for increased arbitrage costs and lower the burden of higher network activity for arbitrageurs. Overall, however, the effect of settlement latency in blockchain-based markets on arbitrage costs is large and harms price informativeness as it rationalizes persistent price differences across exchanges for otherwise identical assets. After adjusting for exchange-specific components reveals that network activity plays a major role beyond the documented market frictions in market for cryptocurrencies such as severe exchange default risks.

4 Arbitrage and Liquidity in Fragmented Markets

The following theoretical model rationalizes the empirical results and serves as a baseline framework to analyze the joint effect of a change of arbitrage costs, e.g. due to settlement latency in fragmented markets on the activity of arbitrageurs and liquidity providers.

The quote dynamics of the theoretical model are nested within the more general estimation framework presented in Section 3 to quantify arbitrage costs.

4.1 Market structure and participants

I assume there are two markets i and j. One risky asset is traded on both markets simultaneously. The terminal value of the asset v_T is uncertain and revealed to all market participants at time T.[16]

Assumption 1. *The value of the asset v_t is a random variable which follows a Brownian motion*

$$v_t = v_0 + \int_0^t \sigma dW_s \tag{17}$$

where σ corresponds to the volatility and W_t denotes a Wiener process.

Two groups of agents populate both markets: First, one arbitrageur who stands ready to exploit price differences between the two markets. The arbitrageur is the only participant able to trade at both platforms simultaneously. Second, at each market $k \in \{i, j\}$ there are competitive risk-neutral market makers specialized in trading the asset on their respective market.

[16] I assume that the expected time considered until the terminal payoff realizes is long relative to the units of time in the setup below, similar in spirit to Baldauf and Mollner (2019).

The arbitrageur is able to monitor quoted prices in real-time at both markets. Whenever quoted prices imply a profitable arbitrage opportunity, the arbitrageur buys at the market quoting the lower price, transfers the asset to the other market and sells. In order to keep the framework parsimonious I assume the arbitrageur trades one unit of the asset and thus the model abstracts from strategic choices of the order size. Two sources of arbitrage costs may render an arbitrage trade too costly to conduct: First, liquidity in form of bid and ask spreads that define the costs of buying or selling one unit of the asset at time t at market k:

$$a_t^k = q_t^k + S_t^k \quad \text{and} \quad b_t^k = q_t^k - S_t^k, \tag{18}$$

where S_t^k denotes the quoted bid-ask (half) spread and q_t^k corresponds to the midquote of market maker k at time t. An arbitrage opportunity occurs if markets are crossed, e.g., the bid price at one market exceeds the ask price of the other market.

The second source of arbitrage costs are, for instance, latency-related costs and denoted as $c \geq 0$. These costs are exogenous to the model. In the empirical analysis, these costs arise due to technological barriers with respect to the speed and settlement fees of the execution of the cross-market arbitrage trade.[17] In principle, however, risk aversion or capital constraints, for instance, can serve as valid examples for c as well, as long as c is determined outside the model. For that purpose, network activity as a proxy for settlement latency serves as a meaningful example of exogenous arbitrage costs as argued in Section 3.

Arbitrage costs $c \geq 0$ prevent the arbitrageur from trading if the cross-market difference between bid and ask is smaller than c. If instead $c = 0$, arbitrageurs exploit price differences as soon as markets are crossed. Therefore, the arbitrageur trades at time t if, for instance, the sell price on market i exceeds the buy price on market j such that $b_t^i - a_t^j > c$. The reverse case, $b_t^j - a_t^i \geq c$ can be handled analogously.

Definition 3. *The arbitrageur exploits any profitable cross-market price difference. I define a profitable arbitrage opportunity as any situation in which*

$$\left|q_t^i - q_t^j\right| > c + S_t^i + S_t^j. \tag{19}$$

Market makers determine the quotes that the arbitrageur faces. Each market maker

[17] Note for now that I assume that arbitrage costs arise as a fixed fee which is payable upfront the transaction instead of incorporating the risky nature of payoffs that arises due to uncertain settlement latency. Whereas this choice simplifies the model, I provide a direct mapping between arbitrage costs due to settlement latency and the limits to arbitrage due to quote changes during the settlement latency in Section 4.4.

$k \in \{i, j\}$ continuously commits to buy or sell one unit of the asset at the prices she quotes, $\{a_t^k, b_t^k\}$. a_t^k corresponds to the (ask) price at t at which market maker k is willing to sell, and b_t^k corresponds to the (bid) price at which she is willing to buy from sellers. Market makers determine quotes conditional on their (private) information regarding the asset value v_t.

The information set of market maker k is determined as follows: I assume that the starting value v_0 at the initial date $t = 0$ is publicly known. For $t > 0$, however, new information about the value of v_t is only observable at randomly sampled discrete time points, not necessarily simultaneous on both markets.

Market makers update their beliefs regarding the terminal payoff if they receive new information. If an information event occurs on market k and time t, the current state of the price process, v_t, is revealed to the market makers at the respective market. For market participants on the other market, however, this information is not available in real-time.[18] Instead, the arbitrageur is the only participant with the technology to monitor and act on both markets. Asynchronous information arrival resembles the core of market fragmentation in the theoretical model. It implies that informed investors are restricted in their access to multiple market venues and instead only act locally.[19] For the sake of simplicity the framework above implies a rather strict form of market fragmentation that is particularly severe for liquidity providers. I relax the restricted monitoring capacities of market makers in the Appendix B and I show that the results remain qualitatively similar when market makers are allowed to observe quote dynamics on the opposite market as long as some valuation uncertainty remains.

Information arrives on market k at times $\left\{0, t_1^k, \ldots, t_{n^k}^k\right\}$, where I denote the time between two information arrivals as $\tau_l^k := t_l^k - t_{l-1}^k$ for $l \in \{1, \ldots, n^k\}$. I put some structure on the sequence of random variables τ_l^k to obtain convenient analytical solutions in the following definition.

Assumption 2. *The sequence of information arrival times $\left\{0, t_1^k, \ldots, t_n^k\right\}$ follows a Poisson point process with parameter λ_k. Therefore, the inter-arrival times $\{\tau_l^k\}_{l=1,\ldots,n}$ are independent exponentially distributed variables with mean $\mathbb{E}\left(\tau^k\right) = \frac{1}{\lambda_k}$ and probability density function*

$$\pi\left(\tau^k\right) = \lambda_k \exp\left(-\lambda_k \tau^k\right). \tag{20}$$

[18] Evidence for such short-lived information asymmetries across markets is documented, for instance, by Kotz et al. (2012) and Budish et al. (2015).

[19] I do not provide a microfoundation for the actual process of information acquisition in this paper. Private information acquisition and how it is revealed through trading on local exchanges has been analyzed in depth (see, e.g., Grossman and Stiglitz (1980) and Verrecchia (1982)).

Intuitively, if $\lambda_j > \lambda_i$, news arrive more frequently on market j. An alternative but equivalent interpretation of the information arrival process is the following: new information about the current state of v_t is revealed to the economy with inter-arrival times $\{\tau_1, \tau_2, \ldots, \tau_{n^j+n^i}\}$ which are exponentially distributed with parameter $\lambda := \lambda_i + \lambda_j$. If new information arrives, it is revealed only either to market i with probability $\frac{\lambda_i}{\lambda_i+\lambda_j}$ or to market j with probability $\frac{\lambda_j}{\lambda_i+\lambda_j}$. I provide a formal proof of the equivalence of this statement and the information arrival processes in Assumption 2 in Appendix C.

After an information event at time t^k, market maker k considers the signal v_{t^k} to update her quotes. At time t^k, the best predictor of market maker k of the terminal payoff at T is $v^k_{t^k} = \mathbb{E}\left(v_T | v_{t^k}\right)$. Further, the valuation $v^k_{t^k}$ does not change until the next information event which takes places not before (random) time $t^k + \tau$.

Competitive pricing at both markets implies that the quotes on market k reflect the valuation $\mathbb{E}\left(v_T | v_{t^k}\right)$ at all times. Therefore, if market maker k received her last signal at t^k, her quotes at t (where $t^k \leq t < t^k + \tau$) are

$$a^k_t = v^k_{t^k} + S^k_t \quad \text{and} \quad b^k_t = v^k_{t^k} - S^k_t, \tag{21}$$

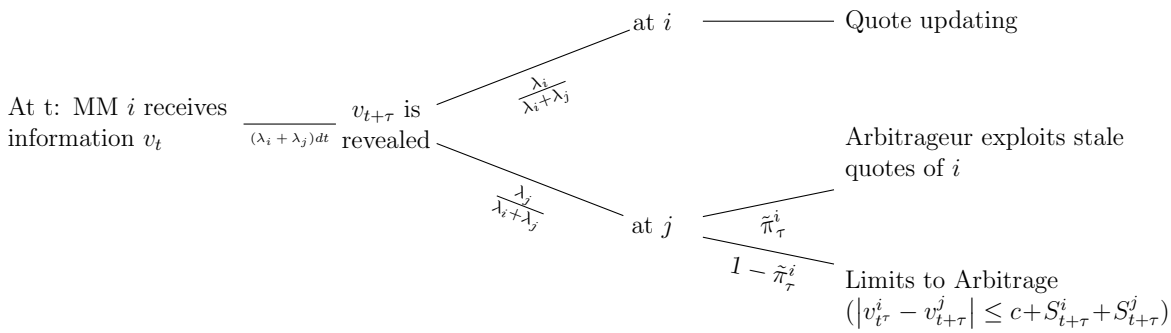
where S^k_t denotes the quoted bid-ask (half) spread.

Order flow stems from noise traders that arrive continuously on both markets and can be distinguished from the arbitrageur ex-post. I assume their expected arrival rate in a marginal unit of time is $2\lambda_L dt > 0$. Upon arrival, noise traders buy or sell a marginal unit of the asset at one of the two markets with equal probabilities, independent of the efficient price v_t.

Two assumptions deserve more attention: First, the purpose of liquidity traders is to provide a stream of order flow to market makers that is orthogonal to the actual value of the risky asset and thus does not resemble any form of adverse selection risk for the market maker. The presence of liquidity traders can be justified by (exogenously determined) private hedging or liquidation needs. Moreover, the assumption that the market-maker can ex-post distinguish arbitrageurs from liquidity traders clearly is a simplification. As in Easley and O'Hara (1987), a justification for this form of self-selection can be competition across arbitrageurs to exploit the potentially short-lived arbitrage opportunity which implies that it is optimal to maximize the trading volume. Note however, that market fragmentation implies that ex-ante the market makers cannot discriminate their prices between liquidity traders and arbitrageurs.

Absent any cross-market trading, market makers in expectation do not lose anything

Figure 4: Information revelation and decision making.



Notes: This figure illustrates the major elements of the theoretical framework: After market maker i receives information at time t, the next information takes place either at the same exchange or market makers i set of information becomes stale. The cross-market arbitrageur exploits the potential price difference if and only if the trade resembles a profit, which happens with probability $\tilde{\pi}^i_\tau$ and depends on the decision of the market maker and characterizes the equilibrium conditions.

and therefore do not require any compensation for providing liquidity. However, cross-market arbitrage leads to adverse selection risk due to the inability of market makers to cancel mispriced quotes before arbitrageurs exploit them (see, e.g. Budish et al., 2015).

Therefore, arbitrage opportunities arise if the quotes are *stale* in the sense that, for instance, at t, the quotes of market maker i reflect current information but market maker j still offers quotes based on her last signal received at $t^j < t^i \leq t$. Such an event exposes the market maker with the risk of selling to (buying from) the arbitrageur an asset at a price which is too low (high). I make the simplifying assumption that a trade by an arbitrageur resolves any information asymmetry and therefore triggers a new valuation by the market maker. This assumption can be justified by direct competition or fast information revelation in relation to the time it takes to monitor and trade on cross-market price differences.

4.2 Equilibrium spreads

Next, I derive the equilibrium spreads at the individual markets in presence of an arbitrageur as a function of the arbitrage costs c and the information arrival rates λ_i and λ_j. Figure 4 illustrates the relevant elements of the theoretical framework. Assume that a news event occurs on market i at time $t = t^i$. By Assumption 2, the time until the next information event, τ, is exponentially distributed with parameter $\lambda := \lambda_i + \lambda_j$ and expected inter-arrival time $\mathbb{E}(\tau) = \frac{1}{\lambda}$. At time $t + \tau$, new information arrives on one of the two markets and the corresponding market makers update their quotes.

I define the change of the signal regarding the terminal payoff of the asset during the time period $[t, t+\tau]$ as $\delta_{t,\tau} := v_{t+\tau} - v_t$. If information arrives for example on market j at time $t+\tau$ and $\delta_{t,\tau} > 0$, the mid-quotes on market i will be too low. As a result, the law of one price is violated.

However, Definition 3 implies that the arbitrageur trades only if the payoffs also exceed the arbitrage costs, thus if $|\delta_{t,\tau}| > c + S^i_{t+\tau} + S^j_{t+\tau}$ as in Equation (19). In the particular example, the arbitrageur buys on market i, transfers the asset and sells on market j. The payoffs and the trading strategy are reversed if, instead, $\delta_{t,\tau} < 0$. Upon arrival of news at market j, adverse selection does only affect the opposite market maker i. From the perspective of the individual market maker, there is no threat of quoting outdated prices at the time when she receives information. Subsequently, at t^k, competitive spreads of market maker k are zero. Only the adverse selection component in the spreads of their own markets is relevant for the market makers when it comes to their equilibrium spreads.

In the following Lemma, I derive the probability of an arbitrage trade from the perspective of market maker $k \in \{i, j\}$, τ units of time after she updated her quotes for the last time.

Lemma 1. *Given Assumptions 1 and 2, the probability of an arbitrage trade at time* $t^k + \tau$, $\pi^k_\tau(S_{t+\tau}, c, \sigma) := \mathbb{P}\left(|v_{t^k+\tau} - v_t| > S_{t+\tau} + c \mid v_t = v^k_t\right)$ *is*

$$\pi^k_\tau(S_{t+\tau}, c, \sigma) = 1 - \frac{2}{\sqrt{\pi}} \int_0^{\frac{c+S_{t+\tau}}{\sigma\sqrt{2\tau}}} e^{-z^2} dz. \tag{22}$$

Further, $\pi^k_\tau(S_{t+\tau}, c, \sigma)$ *exhibits the following characteristics*

$$\frac{\partial \pi^k_\tau}{\partial \tau} > 0, \quad \frac{\partial \pi^k_\tau}{\partial S_{t+\tau}} < 0, \quad \frac{\partial \pi^k_\tau}{\partial c} < 0, \quad \frac{\partial \pi^k_\tau}{\partial \sigma} > 0. \tag{23}$$

Proof. See Appendix. □

Large price changes $|\delta_{t,\tau}|$ are more likely if the time interval since last arrival of information, τ, is long, or if the volatility of the price process, σ is high. In particular, Assumption 1 and the random arrival rate of new information both imply that the unconditional volatility of the change of the value, $\mathbb{V}(|\delta_{t,\tau}|)$ is $\sigma_v := \sqrt{\frac{\sigma^2}{2\lambda}}$. As a consequence, the likelihood of a profitable arbitrage opportunity that exceeds the threshold $S_{t+\tau} + c$ increases with volatility σ and the expected waiting time, λ^{-1}.

Lemma 1 shows that the total arbitrage costs, $S_{t+\tau} + c$, affect the activity of the arbitrageur in a straightforward manner: The probability that the valuation difference

$|\delta_{t,\tau}|$ exceeds the threshold that makes the arbitrage opportunity profitable decreases with the quoted spread $S_{t+\tau}$ and c. Whereas c is exogenous, market makers control the spread based on their knowledge about the total arbitrage costs paid by the arbitrageur at both markets. A higher spread makes it less likely that the change of v_t during the time interval $[t, t+\tau]$ exceeds the boundaries implied by the arbitrageurs participation constraint as of Definition 3. The extreme case of the spread $S_{t+\tau}$ approaching infinity corresponds to a zero probability event of arbitrage activity. At the other extreme, even if the market marker quotes a zero spread, the presence of arbitrage costs, $c > 0$, implies a positive probability that the absolute differences in valuation, $|\delta_{t,\tau}|$ do not exceed arbitrage costs c.

If an information event occurs at time $t+\tau$, the price process is revealed with probability $\frac{\lambda_i}{\lambda_i+\lambda_j}$ on market i or with probability $\frac{\lambda_j}{\lambda_i+\lambda_j}$ on market j. In the first case, market maker i updates her quotes and shifts v_t^i correspondingly. As discussed above, this scenario does not leave any uncertainty for market maker i, and trading against the arbitrageur does not expose her to any adverse selection risk. In the second case, market maker j updates her quotes whereas market maker i is not fast enough to react. If the arbitrageur does not get active because the differences in valuation do not offset the quoted spreads, neither market maker i nor j earn or lose anything and price differences persist. Conditional on the set of information of the market maker k, $\pi_\tau^i\left(S_{t+\tau}, c, \sigma\right)$ denotes the probability that the arbitrageur exploits an occurring price difference. Market maker i earns the spread but trades against the arbitrageur at stale quotes and her expected losses conditional on an arbitrage trade amount to $S_{t+\tau}^i - \mathbb{E}\left(|\delta_{t,\tau}| \middle| S_{t+\tau}^i + c \leq |\delta_{t,\tau}|\right)$. The following lemma summarizes the expected profits of market maker i during the period of time $[t^i+\tau, t^i+\tau+dt)$ for small dt.

Lemma 2. *Under assumptions 1 and 2, the expected profits of market maker i at $d(t^i+\tau)$ with spread S are*

$$\mathbb{E}\left(\Pi_{i,t+\tau}\left(S\right)\right) = \underbrace{\lambda_L S}_{\substack{\textit{Profit from} \\ \textit{noise trading}}} + \frac{\lambda_j}{\lambda_i+\lambda_j}\left(\pi_\tau S - \underbrace{\sigma\sqrt{\frac{2\tau}{\pi}}\exp\left(-\frac{\left(S+c\right)^2}{2\tau\sigma^2}\right)}_{\textit{Expected loss from stale quote}}\right). \tag{24}$$

The expected profit of market maker i at $d(t+\tau)$ exhibit the following characteristics:

$$\frac{\partial \mathbb{E}\left(\Pi_{i,t+\tau}(S)\right)}{\partial \lambda_L} > 0, \qquad \frac{\partial \mathbb{E}\left(\Pi_{i,t+\tau}(S)\right)}{\partial \lambda_i} > 0, \quad \frac{\partial \mathbb{E}\left(\Pi_{i,t+\tau}(S)\right)}{\partial \lambda_j} < 0,$$
$$\frac{\partial \mathbb{E}\left(\Pi_{i,t+\tau}(S)\right)}{\partial \tau} < 0, \qquad \frac{\partial \mathbb{E}\left(\Pi_{i,t+\tau}(S)\right)}{\partial \sigma} < 0, \quad \frac{\partial \mathbb{E}\left(\Pi_{i,t+\tau}(S)\right)}{\partial c} > 0.$$

Proof. See Appendix. □

The expected gains from trading against the arbitrageur are always negative. As trading takes place only if the difference in the valuation exceeds at least the spreads, the losses conditional on this event always exceed the spread because

$$S^k_{t+\tau} + c \leq \mathbb{E}\left(|\delta_{t,\tau}| | S^k_{t+\tau} + c < |\delta_{t,\tau}|\right). \tag{25}$$

Expected profits can only become non-negative, if $\lambda_L > 0$ because noise trades compensate the market maker for providing liquidity even in anticipation of trading against the arbitrageur.

A higher spread S increases expected profits due to higher expected gains from trading with liquidity traders and it reduces the likelihood of arbitrage activity as shown in Lemma 1. Lemma 2 reflects that if information arrives on market k, profits would be strictly positive for $S^k_{t+\tau} > 0$ at that point in time ($\tau = 0$), due to the absence of any asymmetric information. Therefore, local competition forces market makers to set their spreads to zero at the time of information arrival.

From the perspective of market maker k at time $t^k + \tau$ it is uncertain if the next information event will occur at her market or if she will be exposed to stale quote trading. Consequently, large λ_k decreases the likelihood of an (adverse) information event and therefore decreases expected losses of market maker k. Conversely, if the probability of information arrival on the opposite market increases, the expected profits decrease.

Next, I characterize the equilibrium spreads $\tilde{S}^k_{t+\tau}$. Equilibrium is characterized by market makers at both markets setting their spreads such that at every point in time, the expected profits at $d(t^k+\tau)$ are zero due to local competition. Further, in equilibrium, arbitrageurs mechanically exploit price differences whenever profitable according to their participation constraint in Definition 3.

Definition 4. *In equilibrium for $\lambda_k > 0$, after an information arrival on market k at time t^k, the competitive spread of market maker k at time $t^k + \tau$ is the maximum of zero*

and the unique root of the equation

$$\mathbb{E}\left(\Pi_{k,t^k+\tau}\left(\tilde{S}_{t^k+\tau}\right)\right) = 0. \tag{26}$$

Equilibrium spreads $\tilde{S}^k_{t^k+\tau}$ are the minimum required spreads such that market makers earn zero expected profits. Requiring higher spreads is not a feasible solution due to competition among market makers on market k. Reversely, less compensation than $\tilde{S}^k_{t^k+\tau}$, both increases the likelihood of arbitrage trading and reduces profits from uninformed order flow and thus would render losses for the market maker.

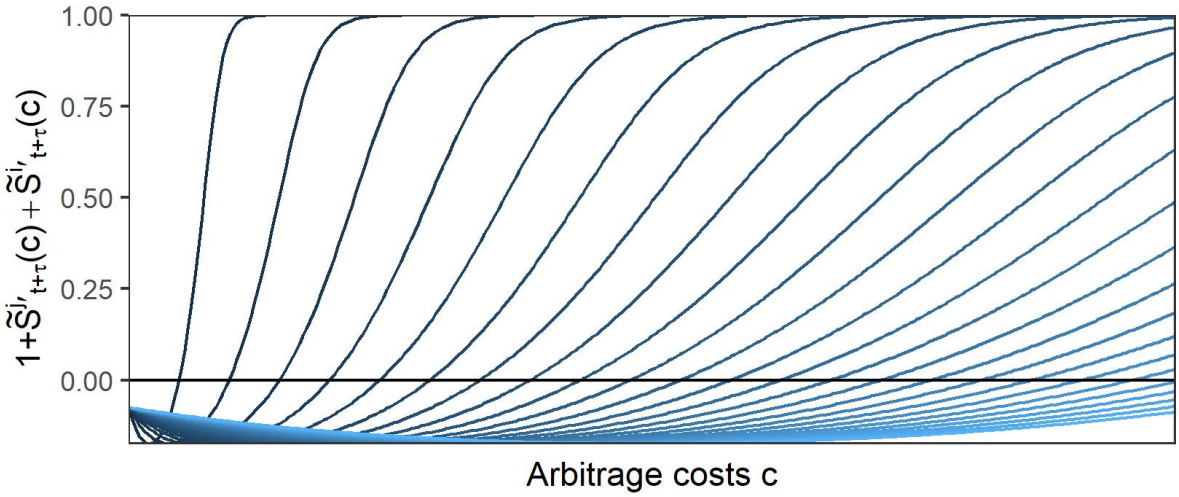
By Lemma 2, the equilibrium spreads are strictly positive but diverge to infinity for $\lambda_L \to 0$ which is in line with the common notion of the non-existence of an equilibrium in absence of belief dispersion (Grossman and Stiglitz (1980)). The arrival rate of noise traders determines the relation between losses due to trading activity from arbitrageurs and gains from trade without any subsequent price adjustment. For $\lambda_L \to \infty$, instead, the market maker is able to compensate her losses due to the high arrival rate of liquidity traders and equilibrium spreads $\tilde{S}^k_{t^k+\tau}$ converge towards zero (similar to vanishing price impact as described by Kyle (1985)). Further and in line with well-established results, equilibrium spreads increase with volatility σ (Easley and O'Hara (1987)) which renders information asymmetries more costly for the market maker. Similarly, Table 6 indicates that spreads at cryptocurrency exchanges increase with spot volatility as a measure of uncertainty.

The arrival rate of information, $\lambda = \lambda_i + \lambda_j$, constitutes an important parameter, both for the equilibrium spreads but also for the informativeness of quoted prices in general. Faster information arrival decreases the volatility σ_v and therefore reduces the adverse selection risk. However, from the perspective of the individual market maker, the probability of information arrival at her own market, $\frac{\lambda_k}{\lambda}$, is the relevant measure of adverse selection risk.

Keeping everything else equal, higher costs c reduce the likelihood of an arbitrage event $\tilde{\pi}^k_\tau$ and reduce the threat of stale quote trading. However, conditional on an arbitrage event, higher costs c also increase the expected losses due to trading for the market maker. As characterized in Lemma 1, an increase of the arbitrage costs c increases the expected profits of the market maker (and therefore the equilibrium spreads become smaller).

Figure 5 illustrates the trade-off between a shift of technology-related arbitrage costs, c, and the total arbitrage costs, $\tilde{S}^i_t(c) + \tilde{S}^j_t(c) + c$. The figure shows the gradient of the total costs for the arbitrageur as a function of the exogenous arbitrage costs, c. Increas-

Figure 5: Gradient of aggregated arbitrage costs.



Notes: This figure illustrates the effect of a marginal increase of the technology-related arbitrage costs, c on the total arbitrage costs $\tilde{S}_t^i(c) + \tilde{S}_t^j(c) + c$. The different lines correspond to shifted values of the volatility σ of the efficient price process. Brighter lines denote larger volatility.

ing the arbitrage costs c has two effects on the arbitrageurs' activity: first, it directly increases total arbitrage costs as it shifts the minimum price difference required before the arbitrageurs prefers to trade. Second, the endogenous component of the costs decreases due to the reduced spreads in response to lower threat of adverse selection. As postulated before, for large enough values of c, the spreads converge towards zero and therefore the aggregate arbitrage costs increase linear with c. The figure illustrates that regions exist for which the total costs of the arbitrageur even decrease if c increases. Intuitively, this seemingly puzzling result occurs when a marginal increase in c substantially decreases the probability of trading and thus simultaneously relaxes the zero expected profit constraints of the individual market makers.

As a main result, the theoretical model reconciles the empirical findings that higher settlement-latency related costs due to network activity are associated with wider spreads due to limits to arbitrage. Further, focussing on the Bitcoin network indicates that that an increase in network activity is associated with large no-trade regions, thus there is no evidence that the gradient of aggregated arbitrage costs is negative and thus the endogenous component of the arbitrage costs (spreads) seems to play a minor role relative to the high costs related to settlement fees and latency.

The results visualized in Figure 5 indicate, however that the reduction of arbitrage costs (for instance by improving the throughput of blockchain consensus systems), does

i) comes at the cost of increased adverse selection component in the spreads and ii) may even be harmful for arbitrageurs in the sense that for low values of c the adjustment in the spreads more than overcompensates the initial gains.

The figure further illustrates that adverse selection costs from the perspective of market makers play a major role when volatility σ is large. Increasing σ corresponds to increasing the relevance of the endogenous component in the arbitrage costs. Increasing c may thus be beneficial for both, arbitrageurs and liquidity providers (depicted by the negative gradient of aggregated arbitrage costs). The focus on volatility is especially relevant in light of the findings of, for instance, Pagnotta and Philippon (2018) and Zimmerman (2020), arguing that decentralized consensus mechanisms and the usage of cryptocucrencies for both, payment and investment purposes may induce excess volatility, rendering Bitcoin less attractive as a stable payment vehicle.

4.3 Price informativeness

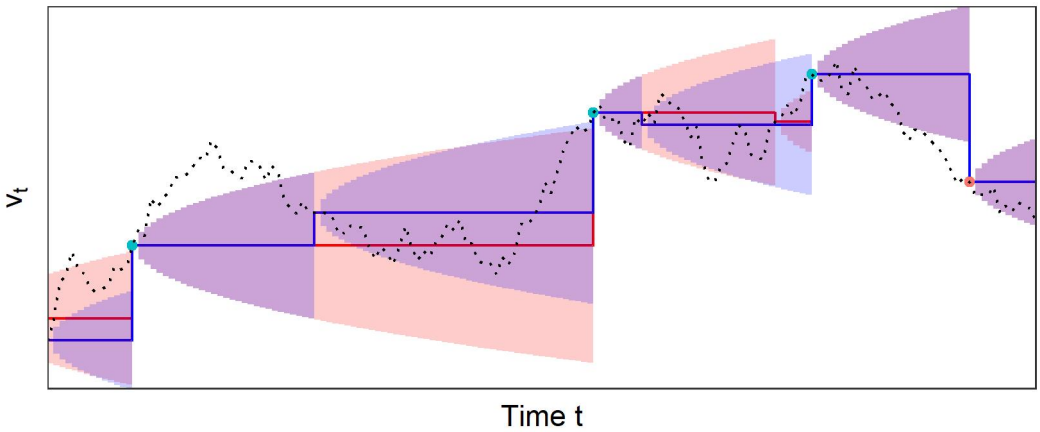
In the theoretical model, quoted prices can deviate from the efficient price for two reasons. On the one hand, the efficient price process v_t is observable only at infrequent points in time. On the other hand, information is revealed asynchronous due to absence of (profitable) arbitrage opportunities.

The frequency of information arrival events, $(\lambda_i + \lambda_j)^{-1}$ determines the aggregate deviation of quoted prices from the underlying efficient price process. Arbitrage activity, which depends on c and $\tilde{S}_t^k$, in fact only facilitates cross-market information aggregation and enforces price informativeness by updating prices across markets. Figure 6 illustrates both determinants of mispricing. All three panels are based on a simulated time series. The grey line corresponds to the latent efficient price process and the red and blue lines denote the midquotes at both markets. The efficient price process evolves as a Brownian motion in line with Assumption 1 and information is revealed with independent exponential distributed waiting times as in Assumption 2. During the remaining time no additional information is available regarding the current value of v_t, constituting one source of mispricing. Jumps in the lines correspond to information arrivals and illustrate updated information sets of the market makers. The shaded areas correspond to the aggregate arbitrage costs (the sum of quoted spreads and exogenous costs c).

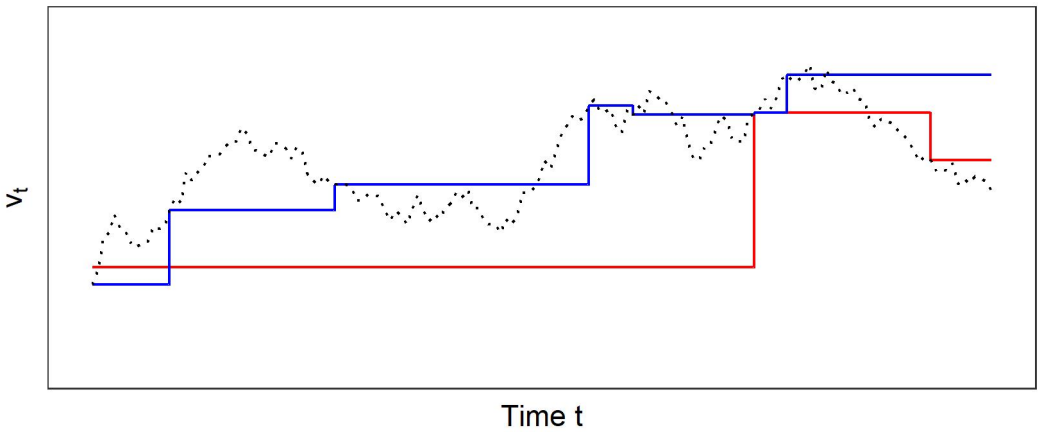
Panel A illustrates the case without any exogenous arbitrage costs, i.e., $c = 0$. Whereas spreads increase with the waiting time τ since the last information arrival event, quoted price differences are bounded within narrow intervals as the arbitrageur continu-

Figure 6: Simulated Price Paths.

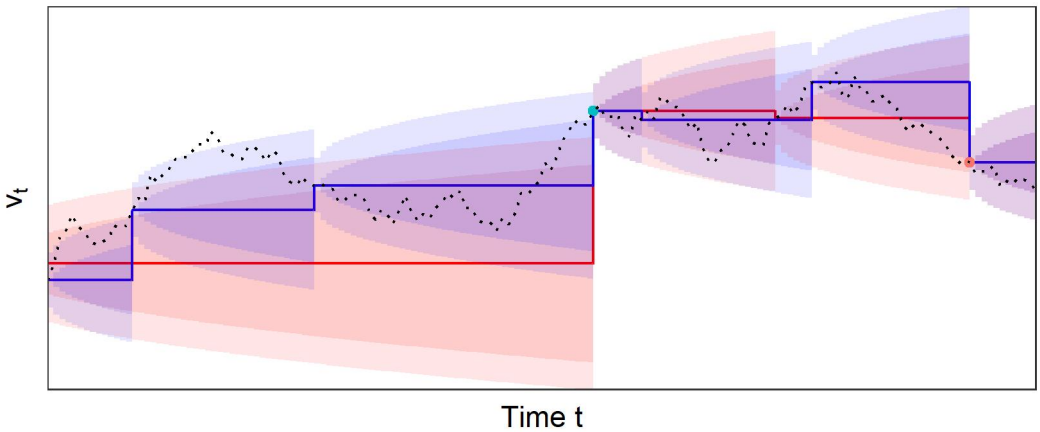
Panel A: No technology-related arbitrage costs ($c = 0$).



Panel B: Absence of arbitrageur ($c \to \infty$).



Panel C: Intermediate technology-related arbitrage costs ($0 < c < \infty$).



Notes: This figure shows three outcomes based on one simulated Wiener process v_t (black dots). Time is plotted on the x-axis. Information inter-arrival times are exponentially distributed. Information is revealed at one of the two markets with the equal probabilities ($\lambda_i = \lambda_j$). The blue (red) line correspond to the quoted mid-prices of the two markets. The shaded area corresponds to the corresponding (equilibrium) spreads. Arbitrage trades are indicated with green dots.

ously monitors and eventually performs cross-market trades. The green dot corresponds to a arbitrage transaction and subsequent adjustment of quotes at both markets.

On the other extreme, Panel B of Figure 6 corresponds to the case with prohibitively high exogenous arbitrage costs, thus $c \to \infty$. In this case the adverse selection component in the spreads is negligible because the probability of an arbitrage trade $\tilde{\pi}_\tau\left(\tilde{S}^k_{t^k+\tau}, c, \sigma\right)$ is close to zero. Therefore, quoted spreads at both markets are zero. Price differences, however, can persist within wider bands. In fact, in the most extreme case both markets are entirely decoupled and quotes are only updated with (market-specific) intensities λ_i and λ_j. Panel C illustrates the intermediate case where c constitutes one source of limits to arbitrage but does not fully prevent arbitrageurs activity. As a result, spreads are positive but smaller in magnitudes than in Panel A.

I derive the aggregate level of price informativeness as the unconditional expected mispricing prevalent in the aggregated market which I define in the following proposition.

Proposition 1. *Under Assumption 1 and Assumption 2 the expected (L_1-) error of the aggregated quoted prices is*

$$\mathbb{E}\left(\left|v_t - q_t^i\right| + \left|v_t - q_t^j\right|\right) = \sqrt{\frac{\sigma^2}{2}\mathbb{E}\left(\tau\right)}\Psi\left(c\right) \tag{27}$$

where $\mathbb{E}\left(\tau\right) = \frac{1}{\lambda_i+\lambda_j}$ *and* $\Psi\left(c\right) := \sqrt{\frac{1+\lambda_j/\lambda_i}{1+\mathbb{E}\left(\tilde{\pi}^i_\tau\left(\tilde{S}^i_{t^i+\tau},c,\sigma\right)\right)\frac{\lambda_j}{\lambda_i}}} + \sqrt{\frac{1+\lambda_i/\lambda_j}{1+\mathbb{E}\left(\tilde{\pi}^j_\tau\left(\tilde{S}^j_{t^j+\tau},c,\sigma\right)\right)\frac{\lambda_i}{\lambda_j}}}$.

Proof. See Appendix. □

Proposition 1 illustrates the fundamental trade-off between price informativeness and endogenous spreads in a setting with exogenous arbitrage costs: First, the expected pricing error is always positive and increases with the volatility of the efficient price process σ and the expected inter-arrival times $\mathbb{E}\left(\tau\right) = \frac{1}{\lambda}$. Both components increase the uncertainty with respect to the true value. The role of the arbitrageur can be understood as increasing the speed with which information is reflected at the individual markets. In the case of prohibitive high arbitrage costs ($c \to \infty$) arbitrageurs never trade ($\tilde{\pi}^k = 0$) and the arrival rates of information at both market do not change. In that case $\lim_{c\to\infty}\Psi(c) = \left(\sqrt{\frac{\lambda_i+\lambda_j}{\lambda_i}} + \sqrt{\frac{\lambda_i+\lambda_j}{\lambda_j}}\right)$ and the expected pricing errors are

$$\lim_{c\to\infty}\mathbb{E}\left(\left|v_t - v_t^i\right| + \left|v_t - v_t^j\right|\right) = \sqrt{\frac{\sigma^2}{2}}\left(\sqrt{\mathbb{E}\left(\tau_i\right)} + \sqrt{\mathbb{E}\left(\tau_j\right)}\right). \tag{28}$$

Intuitively, Equation (28) resembles the mispricing of two independent markets. On the contrary, if exogenous arbitrage costs c are small, the probability of an arbitrage event on market i given information arrived at the opposite market may become positive ($\tilde{\pi}_\tau^i > 0$). In that case, the expected arrival rate of information on market i increases to $\tilde{\lambda}_i := \lambda_i + \mathbb{E}\left(\tilde{\pi}^i\right)\lambda_j$. The increased information arrival rate can be interpreted as information spillover across the markets, whereby arbitrageurs extract the corresponding rent from market makers. The similar effect holds for market j such that $\tilde{\lambda}_j := \lambda_j + \mathbb{E}\left(\tilde{\pi}^j\right)\lambda_i$. Due to the adverse selection component in the spreads the probability of an arbitrage event $\tilde{\pi}_\tau^k$ will never reach one even if arbitrage costs c are 0 and is thus bounded from above. The resulting difference can be interpreted as the (informational) friction that arises due to the presence of market fragmentation.

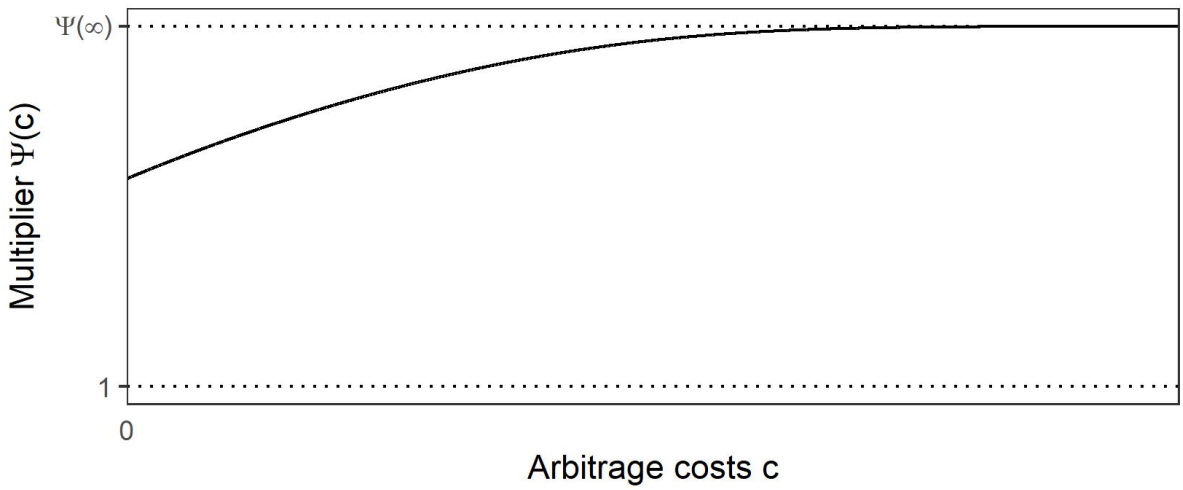
Removing the friction of fragmentation and instead allowing all participants to trade on a consolidated orderbook would make cross-market trading superfluous and could be interpreted as $\mathbb{E}\left(\tilde{\pi}_\tau^k\right) = 1$ which would result in the highest attainable value of price informativeness with

$$\mathbb{E}\left(\left|v_t - q_t^i\right| + \left|v_t - q_t^j\right|\right) \overset{\mathbb{E}\left(\tilde{\pi}_\tau^k\right)=1}{=} \sqrt{2\sigma^2\mathbb{E}\left(\tau\right)}. \tag{29}$$

In case of Equation (29), the only source of mispricing that prevails is due to the asynchronous arrival of information in the economy. Subsequently, for increasing λ, pricing errors converge to zero. Figure 7 shows the value of the multiplier $\Psi(c)$ for intermediate values of c, resembling the case with exogenous arbitrage costs and endogenous adverse selection component in the spreads. Without exogenous arbitrage costs, endogenous liquidity-related arbitrage costs induce persistent mispricing. For increasing values of c, $\Psi(c)$ gradually converges towards entirely fragmented markets as of Equation (28).

As a final step of the analysis, the theoretical framework allows to compare the joint equilibrium outcomes of liquidity (measured in the endogenous adverse selection component in the spreads) and price informativeness (in terms of the inverse of expected aggregate pricing errors). As discussed in Lemma 2, an increase in latency related arbitrage costs, c, reduces the adverse selection component of the spreads because the lower probability of arbitrage trading relaxes the zero expected profits constraint of the market makers. On the other hand, price differences remain unexploited if arbitrage costs are large and reduce the speed of information revelation across markets. As spreads increase with vanishing arbitrage costs, market markers' anticipation of the arrival of arbitrageurs is already sufficient to generate limits to arbitrage in terms of higher total

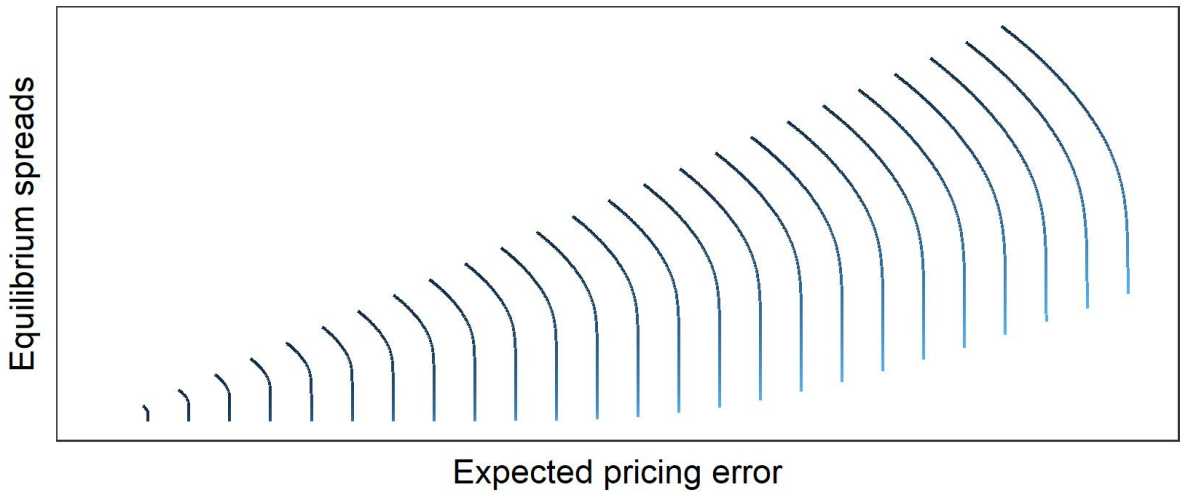
Figure 7: Mispricing Multiplier.



Notes: This figure illustrates the relation between the mispricing multiplier $\Psi(c)$ and the exogenous arbitrage costs c as of Proposition 1.

arbitrage costs, $c + \tilde{S}_t^i + \tilde{S}_t^j$. Therefore, market fragmentation harms informational efficiency because the rate of information arrival is smaller than the hypothetical equivalent of integrated markets, i.e., $\tilde{\lambda}_k < \lambda_i + \lambda_j$. Figure 8 shows equilibrium spreads and pricing errors as a function of arbitrage costs c and shows that price informativeness (in terms of lower expected aggregate pricing errors) can only be achieved if arbitrage activity is high. The figure shows expected pricing errors on the x-axis and equilibrium spreads on the y-axis. Each line corresponds to the equilibrium values based on different values of the exogenous arbitrage costs, c, keeping everything else equal. Bright color corresponds to high technology-related arbitrage costs c, dark color indicates low values of c. The figure shows that for high arbitrage costs (at the lower right hand side) the expected pricing error is generally higher than for lower values of c because the expected arrival rate of new information reaches its minimum, $\frac{1}{\lambda_i} + \frac{1}{\lambda_j}$. The corresponding equilibrium spreads, however, are lowest for high values of c due to the reduced threat of adverse selection. Decreasing c (at the upper left hand side) increases arbitrage activity, thus widens the quoted spreads of the market makers. Further, the expected pricing error decreases because $\Psi(c)$ converges to its lower limit. Note, that the end points of the lines at the left tails correspond to the extreme case with no exogenous arbitrage costs c. Here, total arbitrage costs, $\tilde{S}_t^i + \tilde{S}_t^j$ arise entirely endogenous and therefore limit the expected pricing error from below. The different lines in the figure correspond to shifts in the volatility σ of the efficient price process. The lines to the left represent small volatilities which have two

Figure 8: Equilibrium Price Informativeness and Liquidity.



Notes: This figure summarizes the (sub)-space of equilibrium outcomes for the expected pricing error and the equilibrium spread as functions of arbitrage costs c. The different lines correspond to shifted values of the volatility σ of the efficient price process. Colors denote the exogenous arbitrage costs, c, whereas brighter colors correspond to larger values. Pricing error denotes the expected L_1 norm of aggregate mispricing according to Proposition 1 and the equilibrium spreads are according to Definition 4.

effects: first, smaller uncertainty regarding the efficient price corresponds to less adverse selection and smaller spreads and second, the expected pricing error decreases.

The analysis with respect to price informativeness as a response to arbitrage costs c is closely aligned with the empirical findings related to cryptocurrency markets. First, note that arbitrage costs c provide the fundamental justification for persistent deviations from the law of one price and are the cause for aggregate mispricing. The theoretical framework thus does not only rationalize the effect of network activity on quoted spreads, but instead also suggests that settlement latency hampers price informativeness. Although the empirical analysis does not allow to identify the fundamental value v_t of Bitcoin it nevertheless provides an identification strategy to estimate the otherwise latent arbitrage costs c which depends on the implied quote dynamics of the theoretical framework.

More specifically, note that information revelation at the two fragmented markets implies that the midquotes at time t reflect the (possibly) stale belief of the market makers about the current value of q_t. Midquote price differences between market i and market j, however, are stationary even in absence of any arbitrage trading because as

long as information arrival times are stationary the distribution of $t^i - t^j$ is stationary:

$$z_t := q_t^i - q_t^j = \left(q_t^i - q_t\right) - \left(q_t^j - q_t\right) = \int_{t^i}^{t} \sigma dW_s - \int_{t^j}^{t} \sigma dW_s = \pm \int_{t^i}^{t^j} \sigma dW_s \sim I(0). \quad (30)$$

Equation (30) confirms that the law of one price exhibits a valid cointegration relationship.

Also, Equation (8) nests the price dynamics implied by the theoretical framework. First, quote adjustments do not occur on market k continuously but at irregular frequencies τ_l^k. Therefore, the noise process $u_{t^k}^k$ exhibits volatility $\mathbb{V}\left(u_t^k\right) = \sqrt{2\sigma^2\lambda_k^{-1}}$. Second, setting $\alpha_i^{\text{pos}} = 0, \alpha_j^{\text{pos}} = 1$ and $\alpha_i^{\text{neg}} = -1, \alpha_j^{\text{pos}} = 0$ ensures corresponding price pressure if news occurred either on market i or on market j. The econometric framework is less strict in the sense that it does not require to reveal all information from arbitrage trades instantaneously and instead allows gradual adjustments of the quoted prices at both, the buy and the sell market. The resulting dynamics fully recover the dynamics of the theoretical model. In this case, if, for instance, information arrived on market i at time t_l^i and triggered an arbitrage trade, it holds that

$$\begin{pmatrix} v_{t_{l-1}^i}^i \\ v_{t_{l-1}^i}^j \end{pmatrix} + \begin{pmatrix} \Delta v_{t_l^i}^i \\ \Delta v_{t_l^i}^j \end{pmatrix} = \begin{pmatrix} v_{t_{l-1}^i}^i \\ v_{t_{l-1}^i}^j \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \end{pmatrix} z_{t-1} = \begin{pmatrix} v_{t_{l-1}^i}^i \\ v_{t_{l-1}^i}^i \end{pmatrix}. \quad (31)$$

4.4 Settlement latency as a specific example for c

The empirical analysis rests on latency-related arbitrage costs in terms of risky arbitrage payoffs due to settlement latency. As shown, direct costs do arise in form of settlement fees paid to provide incentives to miners to pick up the arbitrageurs transaction. Costs, however, also do arise simply by the time-consuming settlement process which may render seemingly profitable arbitrage strategies undesirable to exploit. In this section, I show that the time it takes to transfer an asset between two markets, implies costs for the arbitrageur which resemble technology-related arbitrage costs c in the framework above.

Settlement latency is a market friction which arises, for instance, for any asset where legal change of ownership is recorded on a blockchain (Hautsch et al., 2019). Settlement latency prevents the arbitrageur from selling instantaneously, as the transfer of assets to the more expensive market and subsequent sale is only possible with a certain delay. Latency $\tau_a > 0$ is the (possibly random) waiting time until settlement occurs and affects the relative speed of the arbitrageur. The longer the waiting time, the higher the likeli-

hood that the market maker updates her quotes and therefore the arbitrage opportunity disappears.

Assumption 3. *Settlement latency $\tau_a > 0$ is the time it takes to transfer an asset between the two markets. An asset can be sold at a market only, if the arbitrageur is in possession of this asset, thus, only if the legal change of ownership has been completed.*

Consider for now the following example: new information arrives on market i at time t. Settlement latency then implies that during the time period $[t + \tau_a)$ the market maker on market i is not at risk of quoting stale quotes and trading against an arbitrageur at outdated prices. Furthermore, at time $t + \tau$ where $\tau > \tau_a$, the market maker knows that an arbitrageur may have exploited price differences which occurred $\tau - \tau_a$ periods after she updated her quotes last. Therefore, the implied volatility of price changes from the perspective of the market maker, is $\mathbb{V}\left(|v_{t+\tau_a} - v_t|\right) = \sqrt{\sigma^2 (\tau - \tau_a)}$ which is lower than it would have been without settlement latency ($\tau_a = 0$). The following lemma shows that from the perspective of the market maker, settlement latency can be interpreted as giving her a time advantage relative to the arbitrageur:

Lemma 3. *Under assumptions 1, 2 and 3 at $\tau > \tau_a$, the expected profits of market maker i at $d(t + \tau)$ are*

$$\mathbb{E}\left(\Pi^s_{i,t+\tau}(S)\right) = \lambda_L S + \frac{\lambda_j}{\lambda_i + \lambda_j}\left(\tilde{\pi}^s_{\tau-\tau_a} S - \sigma\sqrt{\frac{2(\tau - \tau_a)}{\pi}} \exp\left(-\frac{S^2}{2(\tau - \tau_a)\sigma^2}\right)\right). \quad (32)$$

Here, $\tilde{\pi}^s_{\tau-\tau_a} = 1 - \frac{2}{\sqrt{\pi}} \int_0^{\frac{S}{\sigma\sqrt{2(\tau-\tau_a)}}} e^{-z^2} dz$. For $\tau \leq \tau_a$,

$$\mathbb{E}\left(\Pi^s_{i,t+\tau}(S)\right) = 0 \;\Leftrightarrow S = 0. \quad (33)$$

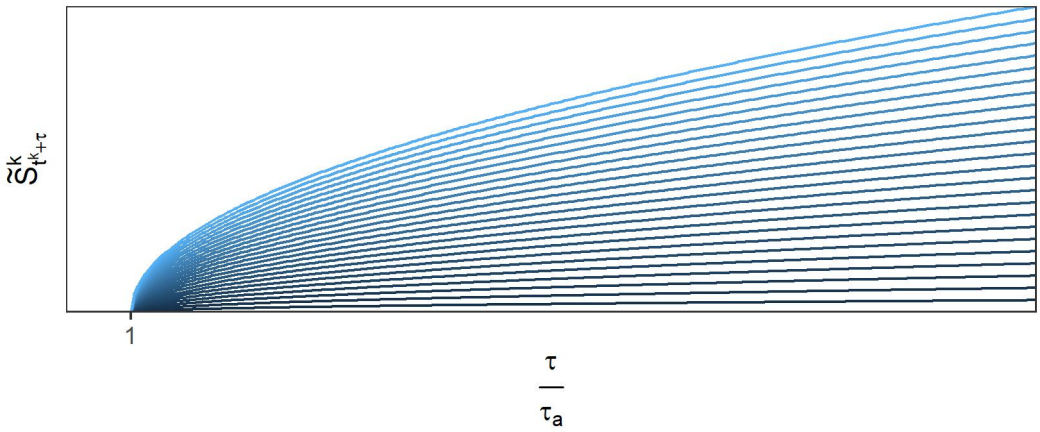
Further, it holds that

$$\frac{\partial \mathbb{E}\left(\Pi^s_{i,t+\tau}(S)\right)}{\partial \tau_a} > 0. \quad (34)$$

Proof. See Appendix. □

Lemma 3 takes into account that the (risk-neutral) arbitrageur upon observing a price difference at time t does only trade if expected profits are positive. This is the case, if $|\delta_{t,\tau}| > \mathbb{E}\left(S_{t+\tau+\tau_a}\right) \geq S_{t+\tau}$. Abstracting from any other costs for the arbitrageur, settlement latency imposes limits to arbitrage in the sense that price differences are exploited if and only if the hypothetical instantaneous returns exceed the (latency-adjusted) spreads.

Figure 9: Settlement Latency and Equilibrium Spreads.



Notes: This figure shows the effect of settlement latency τ_a on the equilibrium spreads. The case with no latency is equivalent to the benchmark case in Lemma 2 with $c = 0$. Brighter lines denote larger volatility.

Figure 9 illustrates the effects of settlement latency on the quoted spreads: the blue line shows the equilibrium spreads for different values of τ_a. Higher values of τ_a indicate longer expected waiting times for the arbitrageur and correspond to higher price risk for the arbitrageur.

Therefore, the main result of Lemma 3 justifies the modelling choice of arbitrage costs c which increase in settlement latency and thus in network activity. Also note, that in contrast to the modelling framework of Hautsch et al. (2019), risk aversion does not play a role to determine arbitrage costs. Instead, in the framework above, price differences are mean-reverting due to the error correction mechanism of both, arbitrageurs activity and information revelation. Therefore, even for risk-neutral arbitrageurs settlement latency may impose limits to arbitrage.

5 Conclusions

Fragmented trading characterizes nowadays financial market infrastructure - similar assets are traded on multiple venues that differ with respect to transparency, order types and, more generally, access for investors. Price informativeness, however, requires that cross-market arbitrageurs monitor and exploit price differences such that quoted prices at all markets ultimately incorporate information regarding the fundamental value of the underlying asset. This pivotal role of arbitrage activity evaporates when market frictions

render cross-market trading costly and impose limits to arbitrage.

Cryptocurrencies are one particular asset class that exhibits substantial market fragmentation and simultaneously imposes considerable arbitrage cost related to blockchain-based settlement. In this paper, I analyze the implications of blockchain-related settlement latency (see, e.g. Hautsch et al. (2019)) on arbitrage activity and liquidity providers. I find that faster settlement reduces latency-related arbitrage costs but at the same time increases quoted bid-ask spreads. Thus, the effect of reducing the latency-related arbitrage costs is partially offset by an increase in costs related to liquidity. In extreme cases, this substitution effect can even predominate and therefore *harm* price informativeness.

The main econometric challenge hereby is to estimate arbitrage costs, which are generally not observable. When arbitrageurs enter the market and start to exploit price differences, however, associated price pressure towards the law of one price should reveal their activity. Arbitrage trading therefore implies a cointegration relationship between markets. Whenever arbitrage costs do not render arbitrage trades profitable (e.g., price differences fall within a no-arbitrage regime) the correction mechanism should evaporate.

I exploit high-frequency orderbook data of two of the largest cryptocurrency exchanges to estimate the no-arbitrage regimes. More specifically, I provide a (Bayesian) estimation procedure to parametrize the thresholds as functions of latent exchange-specific arbitrage costs and time-varying observable proxies for arbitrage costs. The parametrization allows to decompose arbitrage costs into liquidity- and latency-related components. The number of transactions waiting for verification serves as a measure of network activity that increases the price risks of arbitrageurs.

I show that faster settlement reduces arbitrage costs. More specifically, an increase of one percent in the number of transactions waiting for verification implies around 2 basis points wider no-trade regions for arbitrageurs. Most importantly, however: This finding does not necessarily imply that developing faster consensus protocols ultimately benefits price informativeness. Instead, I also document that faster settlement is associated with larger spreads. In fact, my findings suggest that a 10 basis point decrease in technology-related arbitrage costs is associated with a 3 basis point *increase* in spreads. The partially offsetting effects suggest that efforts to reduce the latency of blockchain-based settlement might have unintended consequences for liquidity provision which in extreme cases could even *harm* informational efficiency.

I back the empirical evidence by theoretical reasoning and embed my results in a model in which liquidity providers anticipate that arbitrageurs exploit stale quotes more frequently if settlement is fast and thus set wider spreads to cope with the adverse se-

lection risk. As a result, the direct effect of faster settlement on price informativeness is offset by larger liquidity-related arbitrage costs.

Ultimately, the empirical analysis of cryptocurrency data addresses the fundamental question how settlement procedures that rely on distributed ledger technologies affect market efficiency. In the recent past, for instance, the number of cryptocurrency exchanges across the globe has grown significantly. At the same time, frictions due to the time-consuming settlement latency impose limits to arbitrage that cause deviations from the law of one price to persist. The limited capacities of distributed consensus protocols served as motivation to decrease blockchain related frictions (e.g. Segregated Witness, an implemented soft fork change in the transaction format for Bitcoin was also intended to mitigate the transaction speed problem). However, the findings of this paper suggest that price informativeness *and* narrow spreads are hard to achieve if market makers fear adverse selection.

Reversely, the consolidation of the highly fragmented US equity market landscape (Regulation National Market System Rule 611) seems to be a way to foster price informativeness by reducing latency-related frictions for cross-market trading. However, the growing debate on intentional access delays, as a means to hamper cross-market liquidity taking activities sheds light from a different perspective on a very similar trade-off between price informativeness and adverse selection risks. In this particular example, however, liquidity providers criticize the absence of latency-related costs and claim that adverse selection risks impose disproportionally high costs on all market participants.

This paper provides a first step on discovering the implications and potential drawbacks of blockchain-based settlement in financial markets. Certainly, frictions due to settlement latency do not only affect liquidity providers and arbitrageurs but could arguably affect incentives for informatoin acquisition in general and, more generally, may impose spoll-over effects to overall market efficiency by harming liquidity-taking activities (see, e.g. Kyle and Xiong (2001)). Such an analysis would certainly be interesting, but is left for future research.

References

Abadi, J. and M. Brunnermeier (2018). Blockchain Economics. Working paper.

Aldrich, E. M. and D. Friedman (2017). Order Protection through Delayed Messaging. Working paper.

Aoyagi, J. (2018). Strategic Speed Choice by High-Frequency Traders Under Speed Bumps. Working paper.

Baldauf, M. and J. Mollner (2019). Trading in Fragmented Markets. *Journal of Financial and Quantitative Analysis (forthcoming).*

Balke, N. S. and T. B. Fomby (1997). Threshold Cointegration. *International Economic Review 38*(3), 627–645.

Biais, B., C. Bisiere, M. Bouvard, and C. Casamatta (2019). The Blockchain Folk Theorem. *The Review of Financial Studies 32*(5), 1662–1715.

Brauneis, A., R. Mestel, R. Riordan, and E. Theissen (2019). A High-Frequency Analysis of Bitcoin Markets. Working paper.

Brogaard, J., T. Hendershott, and R. Riordan (2014). High-Frequency Trading and Price Discovery. *The Review of Financial Studies 27*(8), 2267–2306.

Brolley, M. and D. Cimon (2019). Order Flow Segmentation, Liquidity and Price Discovery: The Role of Latency Delays. *Journal of Financial and Quantitative Analysis (forthcoming).*

Brunnermeier, M. and S. Nagel (2004). Hedge Funds and the Technology Bubble. *The Journal of Finance 59*(5), 2013–2040.

Budish, E., P. Cramton, and J. Shim (2015). The High-Frequency Trading Arms Race: Frequent Batch Auctions as a Market Design Response. *Quarterly Journal of Economics 130*(4), 1547–1621.

Chan, K. S. (1993). Consistency and Limiting Distribution of the Least Squares Estimator of a Threshold Autoregressive Model. *The Annals of Statistics 21*(1), 520–533.

Chen, Y., Z. Da, and D. Huang (2019). Arbitrage Trading: The Long and the Short of It. *The Review of Financial Studies 32*(4), 1608–1646.

Chiu, J. and T. V. Koeppl (2019). Blockchain-based Settlement for Asset Trading. *The Review of Financial Studies 32*(5), 1716–1753.

Choi, K. J., A. Lehar, and R. Stauffer (2018). Bitcoin Microstructure and the Kimchi Premium. Working paper.

Dwyer, G. P., P. Locke, and W. Yu (1996). Index Arbitrage and Nonlinear dynamics Between the S&P 500 Futures and Cash. *The Review of Financial Studies 9*(1), 301–332.

Easley, D. and M. O'Hara (1987). Price, Trade Size, and Information in Securities Markets. *Journal of Financial Economics 19*(1), 69–90.

Easley, D., M. O'Hara, and S. Basu (2019). From Mining to Markets: The Evolution of Bitcoin Transaction Fees. *Journal of Financial Economics 134*(1), 91–109.

Engle, R. F. and C. Granger (1987). Co-Integration and Error Correction: Representation, Estimation, and Testing. *Econometrica 55*(2), 251.

Foley, S., J. Karlsen, and T. J. Putninš (2019). Sex, Drugs, and Bitcoin: How Much Illegal Activity is Financed Through Cryptocurrencies? *Review of Financial Studies 32*(5), 1798–1853.

Forbes, C. S., G. R. J. Kalb, and P. Kofhian (1999). Bayesian Arbitrage Threshold Analysis. *Journal of Business & Economic Statistics 17*(3), 364–372.

Foucault, T., R. Kozhan, and W. W. Tham (2017). Toxic Arbitrage. *Review of Financial Studies 30*(4), 1053–1094.

Gagnon, L. and A. G. Karolyi (2010). Multi-market Trading and Arbitrage. *Journal of Financial Economics 97*(1), 53–80.

Gromb, D. and D. Vayanos (2010). Limits of Arbitrage. *Annual Review of Financial Economics 2*(1), 251–275.

Grossman, S. J. and J. E. Stiglitz (1980). On the Impossibility of Informationally Efficient Markets. *The American Economic Review 70*(3), 393–408.

Hansen, B. E. and B. Seo (2002). Testing for two-regime Threshold Cointegration in Vector Error-Correction Models. *Journal of Econometrics 110*(2), 293–318.

Hanson, S. G. and A. Sunderam (2014). The Growth and Limits of Arbitrage: Evidence from Short Interest. *The Review of Financial Studies 27*(4), 1238–1286.

Hautsch, N., C. Scheuch, and S. Voigt (2019). Building Trust takes Time: Limits to Arbitrage in Blockchain-Based Markets. Working paper.

Hu, E. (2018). Intentional Access Delays, Market Quality, and Price Discovery: Evidence from IEX Becoming an Exchange. Working paper.

Huber, F. and T. O. Zörner (2019). Threshold Cointegration in International Exchange Rates: A Bayesian approach. *International Journal of Forecasting 35*(2), 458–473.

Kotz, S., T. Kozubowski, and K. Podgorski (2012). *The Laplace Distribution and Generalizations: a Revisit with Applications to Communications, Economics, Engineering, and Finance*. Springer Science & Business Media.

Kristensen, D. (2010). Nonparametric Filtering of the Realized Spot Volatility: A Kernel-Based Approach. *Econometric Theory 26*(1), 60–93.

Kyle, A. S. (1985). Continuous Auctions and Insider Trading. *Econometrica 53*(6), 1315.

Kyle, A. S. and W. Xiong (2001). Contagion as a Wealth Effect. *The Journal of Finance 56*(4), 1401–1440.

Lin, J.-C., G. C. Sanger, and G. G. Booth (1995). Trade Size and Components of the Bid-Ask Spread. *The Review of Financial Studies 8*(4), 1153–1183.

Lo, M. C. and E. Zivot (2001). Threshold Cointegration and Nonlinear Adjustment to the Law of One Price. *Macroeconomic Dynamics 5*(4), 533–576.

Makarov, I. and A. Schoar (2020). Trading and Arbitrage in Cryptocurrency Markets. *Journal of Financial Economics 135*(2), 293–319.

Martens, M., P. Kofman, and T. C. F. Vorst (1998). A Threshold Error-Correction Model for Intraday Futures and Index Returns. *Journal of Applied Econometrics 13*(3), 245–263.

Meiklejohn, S., M. Pomarole, G. Jordan, K. Levchenko, D. McCoy, G. M. Voelker, and S. Savage (2013). A Fistful of Bitcoins: Characterizing Payments among Men with no Names. In *Proceedings of the 2013 conference on Internet measurement conference*, pp. 127–140.

Nakamoto, S. (2008). Bitcoin: A Peer-to-Peer Electronic Cash System: Working Paper. Working paper.

Ofek, E., M. Richardson, and R. F. Whitelaw (2004). Limited Arbitrage and Short Sales Restrictions: Evidence from the Options Markets. *Journal of Financial Economics 74*(2), 305–342.

O'Hara, M. and M. Ye (2011). Is Market Fragmentation harming Market Quality? *Journal of Financial Economics 100*(3), 459–474.

Pagnotta, E. S. and T. Philippon (2018). Competing on Speed. *Econometrica 86*(3), 1067–1115.

Park, H., J. W. Mjelde, and D. A. Bessler (2007). Time-varying Threshold Cointegration and the Law of one Price. *Applied Economics 39*(9), 1091–1105.

Pontiff, J. (1996). Costly Arbitrage: Evidence from Closed-End Funds. *Quarterly Journal of Economics 111*(4), 1135–1152.

Roll, R., E. Schwartz, and A. Subrahmanyam (2007). Liquidity and the Law of one Price: the Case of the Futures–Cash Basis. *The Journal of Finance 62*(5), 2201–2234.

Ross, S. A. (1976). The Arbitrage Theory of Capital Asset Pricing. *Journal of Economic Theory 13*(3), 341–360.

Seo, B. (2011). Nonparametric Testing for Linearity in Cointegrated Error-Correction Models. *Studies in Nonlinear Dynamics & Econometrics 15*(2).

Shleifer, A. and R. W. Vishny (1997). The Limits of Arbitrage. *The Journal of Finance 52*(1), 35–55.

Stevens, J. (2015). Do Transaction Costs prevent Arbitrage in the Market for Crude Oil? Evidence from a Threshold Autoregression. *Applied Economics Letters 22*(3), 169–172.

Stoll, H. R. (1989). Inferring the Components of the Bid-Ask Spread: Theory and Empirical Tests. *The Journal of Finance 44*(1), 115.

Ters, K. and J. Urban (2020). Estimating unknown Arbitrage Costs: Evidence from a 3-regime Threshold Vector Error Correction Model. *Journal of Financial Markets 47*, 100503.

Theissen, E. (2012). Price Discovery in Spot and Futures Markets: a Reconsideration. *The European Journal of Finance 18*(10), 969–987.

Tong, H. (Ed.) (1983). *Threshold Models in Non-linear Time Series Analysis*, Volume 21 of *Lecture Notes in Statistics*. New York, NY: Springer New York.

Tsay, R. S. (1998). Testing and Modeling Multivariate Threshold Models. *Journal of the American Statistical Association 93*(443), 1188–1202.

Verrecchia, R. E. (1982). Information Acquisition in a Noisy Rational Expectations Economy. *Econometrica 50*(6), 1415–1430.

Woodward, M. (2018). Bumping Up the Competition: The Influence of IEX's Speed Bump in US Financial Markets. Working paper.

Zimmerman, P. (2020). Blockchain Structure and Cryptocurrency Prices. Working paper.

Appendix

A Proofs

Proof of Lemma 1. First, I derive the distribution of $|v_{t+\tau} - v_t|$ given information available at time t. Assumption 1 implies that given v_t and τ, the distribution of $\delta_{t,\tau} = v_{t+\tau} - v_t$ is Gaussian with mean 0 and variance $\sqrt{\tau}\sigma$. Then, $|\delta_{t,\tau}|$ corresponds to a half-normal distribution with probability density function

$$\pi(|\delta_{t,\tau}|) = \frac{\sqrt{2}}{\sigma\sqrt{\tau\pi}} \exp\left(-\frac{|\delta_{t,\tau}|^2}{2\tau\sigma^2}\right). \tag{A1}$$

The half-normal distribution has mean $\sqrt{\frac{2\tau}{\pi}}\sigma$. The probability of $|\delta_{t,\tau}| > z$ for $z > 0$ is

$$\pi\left(|\delta_{t,\tau}| > z\right) = 1 - \operatorname{erf}\left(\frac{z}{\sqrt{2\tau}\sigma}\right) \tag{A2}$$

where $\operatorname{erf}(\cdot)$ corresponds to the Gauss error function defined as

$$\operatorname{erf}\left(\frac{z}{\sqrt{2\tau}\sigma}\right) := \frac{2}{\sqrt{\pi}} \int_0^{\frac{z}{\sqrt{2\tau}\sigma}} e^{-x^2} dx. \tag{A3}$$

The derivatives follow immediately from

$$\frac{\partial \operatorname{erf}(x)}{\partial x} = \frac{2}{\sqrt{\pi}} \exp^{-x^2}. \tag{A4}$$

□

Proof of Lemma 2. First, I derive the expected losses of the market maker k *conditional on an information event* at the opposite market. After an information event occurs, the arbitrageur trades if the difference in valuation, $|\delta_{t,\tau}|$ exceeds the limits to arbitrage given by the threshold $c + S^k_{t+\tau}$. Therefore, it holds that

$$E\left(|\delta_{t,\tau}| \mid c + S_{t+\tau} < |\delta_{t,\tau}|\right) = \frac{\int_{c+S^k_{t+\tau}}^{\infty} |\delta_{t,\tau}| \pi(|\delta_{t,\tau}|) d|\delta_{t,\tau}|}{1 - P(|\delta_{t,\tau}| < c + S^k_{t+\tau})}. \tag{A5}$$

From Lemma 1, $|\delta_{t,\tau}|$ follows a half-normal distribution. Thus, the denominator of Equation (A5) can easily be derived as $\tilde{\pi}_t$, whereas for the nominator I make use of the fact that $\pi(|\delta_{t,\tau}|) = 2\phi\left(\frac{\delta_{t,\tau}}{\sigma\sqrt{\tau}}\right)$ for $\delta_{t,\tau} \geq 0$, where $\phi(\cdot)$ is the probability density function of

the standard normal distribution. Then, I get

$$\int_{c+S_{t+\tau}^k}^{\infty} |\delta_{t,\tau}|\pi(|\delta_{t,\tau}|)d|\delta_{t,\tau}| = \sqrt{\frac{2}{\tau\pi\sigma^2}}\int_{c+S_{t+\tau}^k}^{\infty} \delta \exp\left(-\frac{\delta^2}{2\sigma^2\tau}\right) d\delta = \sqrt{\frac{2\tau\sigma^2}{\pi}} \exp\left(-\frac{\left(S_{t+\tau}^k + c\right)^2}{2\sigma^2\tau}\right). \tag{A6}$$

Therefore, the expected loss from trading against arbitrageurs is

$$\tilde{\pi}_\tau \left(S_{t+\tau}^k - \frac{1}{\tilde{\pi}_\tau}\sqrt{\frac{2\tau\sigma^2}{\pi}} \exp\left(-\frac{\left(S_{t+\tau}^k + c\right)^2}{2\sigma^2\tau}\right)\right). \tag{A7}$$

Note, that for $\frac{\partial \mathbb{E}(\Pi_{i,t+\tau}(S))}{\partial c}$ the following holds (based on Lemma 1):

$$\frac{\partial \tilde{\pi}_{i,\tau}(S)}{\partial c} = -\sqrt{\frac{2}{\pi\tau\sigma^2}} \exp\left(-\frac{\left(S_{t+\tau}^k + c\right)^2}{2\sigma^2\tau}\right). \tag{A8}$$

Then, it holds that

$$\frac{\partial \mathbb{E}\left(\Pi_{i,t+\tau}(S)\right)}{\partial c} = \sqrt{\frac{2}{\pi\tau\sigma^2}} \exp\left(-\frac{\left(S_{t+\tau}^k + c\right)^2}{2\sigma^2\tau}\right)\left(S_{t+\tau} + c - S_{t+\tau}\right) > 0 \tag{A9}$$

□

Proof of Proposition 1. First, I derive the distribution of the pricing error $|z_t^k| := |v_t^k - v_t|$. Assumption 1 implies that given $v_t^k = v_{t-\tau}$ for known τ, the distribution of $z_t^k = v_t - v_{t+\tau}$ is Gaussian with mean 0 and variance $\sigma^2\tau$. However, τ is a random variable which follows an exponential distribution with parameter λ_k. Therefore, the characteristic function of the stopped Wiener process z_t^k is

$$\varphi_{z_t^k} : \mathbb{R} \to \mathbb{C} \tag{A10}$$

$$\varphi_{z_t^k}(s) = E\left(e^{isz_t^k}\right) = E\left(E\left(e^{isz_t^k} \mid \tau\right)\right) \tag{A11}$$

$$= E\left(e^{-\tau\sigma^2 s^2/2}\right) \tag{A12}$$

$$= \lambda_k \int_0^\infty e^{-(\lambda_k + \sigma^2 s^2/2)t}\, dt \tag{A13}$$

$$= \frac{1}{1 + \frac{\sigma^2 s^2}{2\lambda_k}}. \tag{A14}$$

Equation (A14) corresponds to the characteristic function of a Laplace distribution with

expected value $\mathbb{E}\left(z1k_t\right) = 0$, scale parameter $\sqrt{\frac{\sigma^2}{2\lambda_k}}$ and corresponding probability density function

$$\pi\left(z_t^k\right) = \sqrt{\frac{\lambda_k}{2\sigma^2}} \exp\left(-\sqrt{\frac{2\lambda_k}{\sigma^2}}\left|z_t^k\right|\right). \tag{A15}$$

By equation (A15), the distribution of $|z_t^k| \propto \exp\left(-\sqrt{\frac{2\lambda_k}{\sigma^2}}\left|z_t^k\right|\right)$ corresponds to the kernel of an exponential distribution with rate parameter $\sqrt{\frac{2\lambda_k}{\sigma^2}}$. Therefore, the expected (L_1) difference between quoted price v_t^k and efficient price v_t is

$$\mathbb{E}\left(|z_t|\right) = \sqrt{\frac{\sigma^2}{2}\mathbb{E}\left(\tau\right)}. \tag{A16}$$

Then, for independent information arrivals ($\tilde{\pi}_t = 0$) the expected pricing error is additive:

$$\mathbb{E}\left(\left|v_t - v_t^i\right| + \left|v_t - v_t^j\right|\right) = \frac{\sigma}{\sqrt{2}}\left(\frac{1}{\sqrt{\lambda_i}} + \frac{1}{\sqrt{\lambda_j}}\right) \tag{A17}$$

$$= \sqrt{\frac{\sigma^2}{2}\mathbb{E}\left(\tilde{\tau}\right)}\left(\sqrt{\frac{\lambda_i + \lambda_j}{\lambda_i}} + \sqrt{\frac{\lambda_i + \lambda_j}{\lambda_j}}\right). \tag{A18}$$

Here, $\mathbb{E}\left(\tilde{\tau}\right) := \frac{1}{\lambda_i + \lambda_j}$ is the expected inter-arrival of information at the economy and the adjustment terms $\left(\sqrt{\frac{\lambda_i+\lambda_j}{\lambda_i}} + \sqrt{\frac{\lambda_i+\lambda_j}{\lambda_j}}\right) > 1$ are related to the odds of information arriving on market$k$. Next, arbitrage activity increases the arrival rate of information at the individual markets. More specifically, recall that Assumption 2 implies that information arrives at the economy with rate $\tilde{\lambda} := \lambda_i + \lambda_j$ and is then revealed on market$k$ with probability $\frac{\lambda_k}{\lambda_i + \lambda_j}$. Given an information event on market$j$, the probability of a trade is characterized by $\tilde{\pi}_\tau^i$ as of Lemma 1 which depends on the arbitrage costs c and the equilibrium spreads $\tilde{S}_\tau^i$.

Therefore, by the properties of the exponential distribution, the (expected) arrival rate of information on market$i$ is $\tilde{\lambda}_i = \lambda_i + \lambda_j \mathbb{E}\left(\tilde{\pi}^i\right) > \lambda_i$ and similar $\tilde{\lambda}_j = \lambda_j + \lambda_i \mathbb{E}\left(\tilde{\pi}^j\right) > \lambda_j$. Replacing λ_i and λ_j in the adjustment term of Equation (A18) concludes the proof. □

Proof of Lemma 3. The proof of Lemma 3 follows immediately from Lemma 2 in combination with Assumption 3. □

B Market makers with cross-market monitoring capacities

Fragmentation in my theoretical framework resembles a strict separation of the market participants at the two markets which can be overcome only by arbitrageurs. More specifically, I restrict local market makers to observe quotes at the other market, respectively. Arguably, a more realistic trading infrastructure presumably imposes less severe restrictions on the monitoring capacities of local market makers. Instead, assume that market makers can observe their own quotes and the quotes of the competing market at all times but they are restricted from providing liquidity at both markets. Then, market maker $k \in \{i, j\}$ instantaneously reacts to a shift in the quotes of market $k' \neq k$ and limited price informativeness due to market fragmentation does not play a role. However, if I instead impose noisy information revelation in the spirit of Foucault et al. (2017), price differences may persist and the thread of adverse selection remains active. More specifically, assume that at random times $\tilde{\tau}_l^k$, market makers at k receive a private valuation shock of magnitude $\gamma \sim N(0, \sigma_\gamma)$. Then from the perspective of market maker k, quoted prices at the opposite market only reveal information about the aggregate valuation $v_t^{k'} + \gamma$. Further, from the perspective of market maker k' the uncertainty with respect to the efficient price process v_t remains present and causes her to set spreads accordingly. In line with Foucault et al. (2017) such a setup does not reveal the full information set of market maker k' to market maker k and therefore price differences remain until they are either dissolved by new information or by an arbitrage trade.

C Details regarding assumption 2

Proposition 2. *Assumption 2 implies an equivalent distribution of arrival times at the individual markets as assuming that information arrives at times* $\{0, t_1, \ldots, t_n\}$ *and is then revealed on marketk* $\in \{i, j\}$ *with probability* $\lambda_k/(\lambda_i + \lambda_j)$.

Proof. In the following I show that the probability density function of the event $\tau^k = x$, $x > 0$ is equivalent to an exponential distribution with parameter λ_k. Define $\lambda; = \lambda_i + \lambda_j$. First $\tau^k = x$ holds if and only if all information arrivals until τ^k occurred at the opposite market. Further, the sum of s independent and identical exponential distributed variables with scale parameter λ follows the Erlang distribution with probability density function

$\pi(x) = \frac{\lambda^s x^{s-1} \exp(-\lambda x)}{(s-1)!}$. Then it holds that

$$\pi\left(\tau^k = x\right) = \sum_{s=1}^{\infty} \pi\left(s \text{ trials}\right) \pi\left(\sum_{i=1}^{s} \tau_i = x\right) \tag{C19}$$

$$= \sum_{s=1}^{\infty} \left(1 - \frac{\lambda_k}{\lambda}\right)^{s-1} \frac{\lambda_k}{\lambda} \frac{\lambda^s}{(s-1)!} x^{s-1} \exp^{-\lambda x} \tag{C20}$$

$$= \lambda_k \exp^{-\lambda x} \sum_{s=1}^{\infty} \left(\lambda - \lambda_k\right)^{s-1} x^{s-1} \tag{C21}$$

$$= \lambda_k \exp^{-\lambda x} \exp^{(\lambda - \lambda_k)x} \tag{C22}$$

$$= \lambda_k \exp^{-\lambda_k x}. \tag{C23}$$

Therefore, $\pi\left(\tau^k = x\right)$ corresponds to an exponential distributed random variable with scale parameter λ_k which concludes the proof. □

D MCMC Algorithm

The following section illustrates the Monte Carlo Markov Chain Algorithm to calibrate the three-regime threshold vector error correction model. In line with the notation of Equation (11) the model can be rewritten as a (stationary) multivariate linear regression

$$\Delta V^r = X^r B^r + U^r \tag{D24}$$

where

$$\Delta V^r_{t_r} = \begin{pmatrix} \Delta v^i_{t_r} \\ \Delta v^j_{t_r} \end{pmatrix}' \text{ and } X^r_{t_r} = \begin{pmatrix} 1 \\ z_{t_r - 1} \end{pmatrix}'. \tag{D25}$$

Here, t_r corresponds to the stacked dates of all observations in regime r. Therefore, the data is of the form $\Delta V \in \mathbb{R}^{T \times 2}$ and $X \in \mathbb{R}^{T \times 2}$. The data is separated into three regimes by the thresholds c_t^{neg} and c_t^{pos} such that

$$\Delta V^{\text{neg}} := \{Y_t : z_{t-1} < c_t^{\text{neg}}\} \text{ and } X^{\text{neg}} := \{X_t : z_{t-1} < c_t^{\text{neg}}\} \tag{D26}$$

$$\Delta V^{\text{pos}} := \{Y_t : z_{t-1} > c_t^{\text{pos}}\} \text{ and } X^{\text{pos}} := \{X_t : z_{t-1} > c_t^{\text{pos}}\} \tag{D27}$$

$$\Delta V^{0} := \{Y_t : c_t^{\text{neg}} < z_{t-1} < c_t^{\text{pos}}\} \text{ and } X^{0} := \{X_t : c_t^{\text{neg}} < z_{t-1} < c_t^{\text{neg}}\} \tag{D28}$$

The respective size of the partitioned matrices is $\Delta V^0 \in \mathbb{R}^{T^0 \times 2}$, $\Delta V^{\text{pos}} \in \mathbb{R}^{T^{\text{pos}} \times 2}$ and $\Delta V^{\text{neg}} \in \mathbb{R}^{T^{\text{neg}} \times 2}$ with $T^0 + T^{\text{pos}} + T^{\text{neg}} = T$. Thus, for $i \in \{\text{neg}, 0, \text{pos}\}$, the underlying data-generating process takes the form:

$$\Delta V^r = X^r \beta^r + U^r \text{ with } U_t^R \sim MN(0, \Sigma^r) \tag{D29}$$

where $MN(\cdot)$ corresponds to the multivariate normal distribution with probability density function with zero mean

$$\pi\left(x\right) = \pi \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = (2\pi)^{-1} \det\left(\Sigma^r\right)^{-\frac{1}{2}} \exp\left(-\left('\Sigma^{r-1} x\right)\right). \tag{D30}$$

Conditional on the parameters $c_0^{\text{pos}}, c_0^{\text{neg}}$ and c_1, standard inference from Bayesian multivariate linear regression models applies. First, the likelihood takes the form

$$\mathcal{L}(\Delta V|\theta, X) \propto \prod_{r \in \{\text{neg, 0, pos}\}} |\Sigma^r|^{-T^r/2} \exp\left(-\frac{1}{2}\text{tr}\left(\left(\Sigma^r\right)^{-1} U_\theta^{r\prime} U_\theta^r\right)\right) \tag{D31}$$

Then, conditional conjugate priors for $\beta^r = \text{vec}\left(B^r\right)$ and Σ^r are chosen such that for suitable hyperparameters $\beta = \text{vec}(B), \Psi$, and ν I have:

$$\pi(\Sigma^r) \sim IW(\Psi, \nu) = \frac{|\Psi|^{\nu/2}}{2^{\nu p/2} c_p(\frac{\nu}{2})} |\Sigma^r|^{-(\nu+p+1)/2} \exp\left(-\frac{1}{2}\text{tr}(\Psi\left(\Sigma^r\right)^{-1})\right) \tag{D32}$$

$$\pi(\beta^r|\Sigma^r) \sim MN(\beta, \Sigma^r \otimes \Lambda^{-1}) \tag{D33}$$

The priors for c_0^r and c_1 are uniform such that

$$\pi(c_1) \sim \pi(c_0^r) \sim U\left(-\infty, \infty\right). \tag{D34}$$

A standard Gibbs sampling scheme applies for the posterior when conditioning on the threshold variables c_t^{pos} and c_t^{neg}. Given initial (or sampled) values of $\{c_t^{\text{pos}}, c_t^{\text{neg}}\}$ the algorithm works as follows:

- Separate the data ΔV and X into $\Delta V^{\text{neg}}, \Delta V^0, \Delta V^{\text{pos}}$ and $X^{\text{neg}}, X^0, X^{\text{pos}}$.
- For each of the three regimes $r \in \{\text{neg, 0, pos}\}$ generate a draw Σ^r from the Inverse Wishart distribution based on the posterior distribution

$$\pi(\Sigma^r|\Delta V, X) \sim IW(\tilde{\Psi}, \tilde{\nu}) \tag{D35}$$

where

$$\tilde{\Psi} = \Psi + (\Delta V - X\tilde{B})'(\Delta V - X\tilde{B}) + (\tilde{B} - B)'\Lambda(\tilde{B} - B) \tag{D36}$$

$$\tilde{\nu} = \nu + T^r \tag{D37}$$

$$\tilde{B} = (X'X + \Lambda)^{-1}(X'\Delta V + \Lambda B) \tag{D38}$$

$$\tag{D39}$$

- For each of the three regimes $r \in \{\text{neg, 0, pos}\}$ generate a draw β from the multivariate normal distribution based on the (conditional) posterior distribution

$$\pi\left(\beta^r|\Sigma^r\right) \propto MVN\left(\text{vec}\left(\tilde{B}\right), \Sigma^r \otimes \left(X^{r\prime}X^r + \Lambda\right)^{-1}\right) \tag{D40}$$

- Random walk Metropolis Hastings step within Gibbs: Sample $\tilde{c}_0^{\text{neg}}, \tilde{c}_0^{\text{pos}}$ and $\tilde{c}_1$ independently from a normal distribution with means $c_0^{\text{neg}}, c_0^{\text{pos}}$ and c_1 and sampling variances σ_0 and σ^1. Compute the acceptance ratio

$$\alpha(\{\tilde{c}_0^{\text{neg}}, \tilde{c}_0^{\text{pos}}, \tilde{c}_1\}|\{c_0^{\text{neg}}, c_0^{\text{pos}}, c_1\}) = \min\left(1, \frac{\mathcal{L}(\Delta V|\tilde{c}_t, \beta^0, \beta^{\text{pos}}, \beta^{\text{neg}}, \Sigma^0, \Sigma^{\text{neg}}, \Sigma^{\text{pos}}, X)}{\mathcal{L}(\Delta V|c_t, \beta^0, \beta^{\text{pos}}, \beta^{\text{neg}}, \Sigma^0, \Sigma^{\text{neg}}, \Sigma^{\text{pos}}, X)}\right). \tag{D41}$$

Letting $\tilde{T}^r$ the sample size of ΔV based on the new segmentation due to $\tilde{c}_t$ and $\tilde{U}^r$ the corresponding residuals $\Delta\tilde{V}^r - \tilde{X}^r\beta^r$, I get:

$$\log\left(\alpha(\{\tilde{c}_0^{\text{neg}}, \tilde{c}_0^{\text{pos}}, \tilde{c}_1\}\right) = \frac{1}{2}\sum_r \left(\left(T^r - \tilde{T}^r\right)\log|\Sigma^r| + \text{tr}(\Sigma^{r-1}(U^{r\prime}U^r - \tilde{U}^{r\prime}\tilde{U}^r))\right). \tag{D42}$$