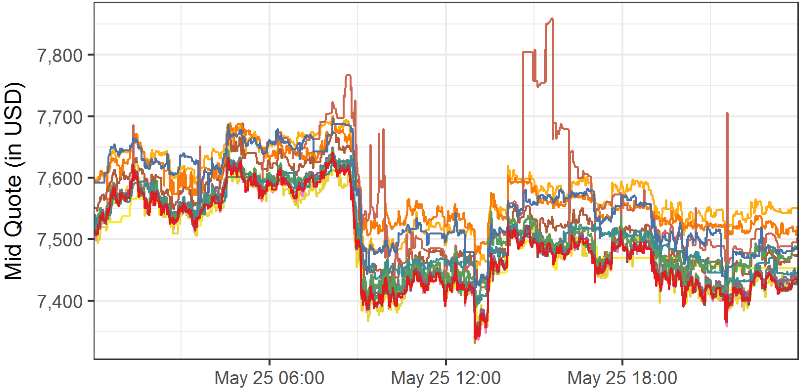
Limits to Arbitrage in Markets with Blockchain-Based Latency

Nikolaus Hautsch
University of Vienna
Christoph Scheuch
Vienna Graduate School of Finance
Stefan Voigt
Vienna Graduate School of Finance

Motivation

“Blockchain technology isn’t just a *more efficient* way to *settle securities*. It will *fundamentally change market structures*,[...]” – Abigail Johnson, CEO of Fidelity Investments



Stochastic latency and blockchain technology

Real-world application of blockchain

- Substantial price differences
- Persistence of price differences
- Are there inefficiencies?

Blockchain-based settlement

- Consensus algorithms introduce *stochastic latency*
- E.g. how long does it miners take to append a new block?

- **What are the implications of stochastic latency for price efficiency?**
- How inefficient are (current) blockchain-based markets?

How does stochastic latency affect price efficiency?

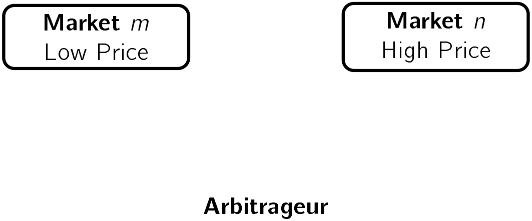
Our contribution

- Stochastic latency as a novel, but relevant microstructure friction
- Analytical limits to arbitrage for markets with stochastic latency
- In depth analysis of Bitcoin markets as a laboratory
- Stochastic latency alone accounts for 20% of Bitcoin price differences

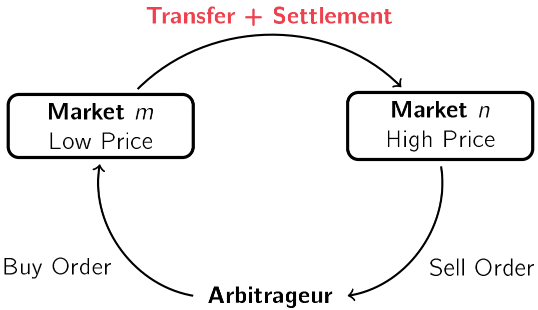
Relevance

- Understanding of design & impact of blockchain-based settlement
- Latency might affect pricing of blockchain-based products

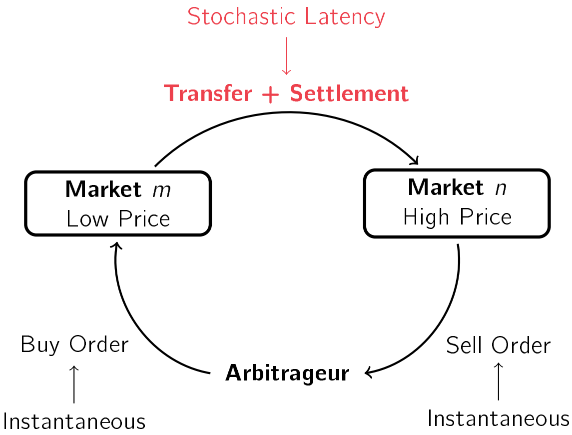
How does blockchain introduce stochastic latency?



How does blockchain introduce stochastic latency?



How does blockchain introduce stochastic latency?



Theoretical Framework

Market $i \in \{1, \ldots, K\}$ continuously provides buy quotes (ask) A_t^i and sell quotes (bid) B_t^i for the asset.

No short selling, margin trading or derivatives

Arbitrageur continuously monitors the quotes on markets m and n and acts on the following strategy: if buying and selling quotes across markets imply a profit, buy one unit of the asset at the market with the lower buy quote, transfer the asset to the market with a higher sell quote and sell it as soon as the transfer is settled.

Instantaneous trading: Arbitrageur exploits price differences if

$$\delta_t^{m,n} := \log\left(B_t^n\right) - \log\left(A_t^m\right) > 0 \quad (1)$$

Trading decision with deterministic latency τ

Log return of arbitrageur's strategy

$$r^{m,n}_{(t:t+\tau)} = \underbrace{\delta^{m,n}_t}_{\text{Instantaneous Return}} + \underbrace{\int_t^{t+\tau} \sigma^n_t dW^n_s}_{\text{Price Risk on Sell-Side Market}} \quad (2)$$

Trading decision with deterministic latency τ

Log return of arbitrageur's strategy

$$r^{m,n}_{(t:t+\tau)} = \underbrace{\delta^{m,n}_t}_{\text{Instantaneous Return}} + \underbrace{\int_t^{t+\tau} \sigma^n_t dW^n_s}_{\text{Price Risk on Sell-Side Market}} \tag{2}$$

Risk-averse arbitrageur with exponential utility and risk aversion parameter γ:

$$U_\gamma(r) = \frac{1 - e^{-\gamma r}}{\gamma} \tag{3}$$

Arbitrageur maximizes expected-utility and exploits price differences for given latency τ if

$$\delta^{m,n}_t > \gamma(\sigma^n_t)^2\tau \tag{4}$$

Trading decision with stochastic latency $\tilde{\tau}$

Assume the **probability distribution of the stochastic latency** is exponential:

$$\pi(\tau|\mathcal{I}_t) = \lambda_t e^{-\lambda_t \tau}, \tag{5}$$

Trading decision with stochastic latency $\tilde{\tau}$

Assume the **probability distribution of the stochastic latency** is exponential:

$$\pi\left(\tau|\mathcal{I}_t\right) = \lambda_t e^{-\lambda_t \tau}, \tag{5}$$

We show that the $\mathcal{I}_t$-conditional distribution of the returns

$$\pi\left(r^{m,n}_{(t:t+\tilde{\tau})}|\mathcal{I}_t\right) = \int_{\mathbb{R}_+} \pi\left(r^{m,n}_{(t:t+\tau)}|\tau\right)\pi\left(\tau|\mathcal{I}_t\right) d\tau \tag{6}$$

corresponds to a Laplace distribution with $E\left(r^{m,n}_{(t:t+\tilde{\tau})}|\mathcal{I}_t\right) = \delta^{m,n}_t$ and $Var\left(r^{m,n}_{(t:t+\tilde{\tau})}|\mathcal{I}_t\right) = \frac{\left(\sigma^{m,n}_t\right)^2}{\lambda_t}$.

Trading decision with stochastic latency $\tilde{\tau}$

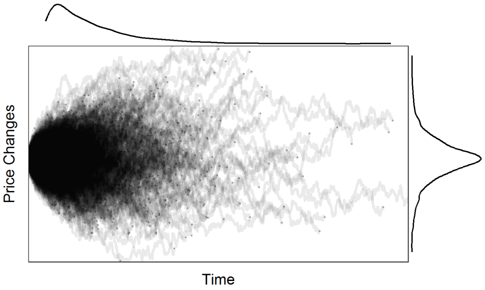
Assume the **probability distribution of the stochastic latency** is exponential:

$$\pi\left(\tau|\mathcal{I}_t\right) = \lambda_t e^{-\lambda_t \tau}, \tag{5}$$

We show that the $\mathcal{I}_t$-conditional distribution of the returns

$$\pi\left(r^{m,n}_{(t:t+\tilde{\tau})}|\mathcal{I}_t\right) = \int_{\mathbb{R}_+} \pi\left(r^{m,n}_{(t:t+\tau)}|\tau\right)\pi\left(\tau|\mathcal{I}_t\right) d\tau \tag{6}$$

corresponds to a Laplace distribution with $E\left(r^{m,n}_{(t:t+\tilde{\tau})}|\mathcal{I}_t\right) = \delta^{m,n}_t$ and $Var\left(r^{m,n}_{(t:t+\tilde{\tau})}|\mathcal{I}_t\right) = \frac{\left(\sigma^{m,n}_t\right)^2}{\lambda_t}$.



Implied costs of stochastic latency: limits to arbitrage

4th-degree Taylor-expansion of the utility function yields

$$CE = \delta_t^{m,n} - \frac{\gamma}{2}\frac{\left(\sigma_t^n\right)^2}{\lambda_t} - \frac{\gamma^3}{24}\frac{\left(\sigma_t^n\right)^4}{\lambda_t^2}K \tag{7}$$

where K is the kurtosis of the returns. $\Rightarrow$ The arbitrageur exploits price differences if

$$\delta_t^{m,n} > \underbrace{\frac{1}{2}\gamma\left(\sigma_t^n\right)^2\mathbb{E}\left(\tilde{\tau}|\mathcal{I}_t\right)}_{\text{Exptected Latency}} + \underbrace{\frac{1}{4}\gamma^3\left(\sigma_t^n\right)^4\mathbb{V}\left(\tilde{\tau}|\mathcal{I}_t\right)}_{\text{Uncertainty in Latency}}. \tag{8}$$

Stochastic latency implies a no-trade region which increases if

- spot volatility is high
- expected latency is large
- latency uncertainty is high
- risk aversion is high

What about trading fees & price impact?

Consider transaction costs of the form

$$P_t^{i,B}(q) = P_t^{i,B}\left(1 - \rho_t^{i,B}(q)\right) \tag{9}$$

$$P_t^{i,A}(q) = P_t^{i,A}\left(1 + \rho_t^{i,A}(q)\right) \tag{10}$$

Arbitrageur exploits price differences if

$$\delta_t^{m,n} > d_t^n + \underbrace{\log\left(\frac{1 + \rho_t^{m,A}(q^*)}{1 - \rho_t^{n,B}(q^*)}\right)}_{\text{Transaction Costs}} \tag{11}$$

for a given trading quantity q^* where

$$d_t^n := \frac{1}{2}\gamma\left(\sigma_t^n\right)^2 \mathbb{E}\left(\tilde{\tau}|\mathcal{I}_t\right) + \frac{1}{4}\gamma^3\left(\sigma_t^n\right)^4 \mathbb{V}\left(\tilde{\tau}|\mathcal{I}_t\right). \tag{12}$$

Some more extensions...

What about different utility functions?

- Analytically less tractable, but basic intuition is the same

More general latency distributions?

- Inverse-Gamma distributed latency implies Student-t-distribution

What if price process has a drift?

- We get skewness through an asymmetric Laplace distribution

What if the volatility of price process is stochastic?

- Integrated volatility enters the equation but intuition remains the same

Aggregating price differences to market-wide inefficiency

To construct an aggregated measure of price inefficiency, we define efficiency boundaries for all K markets by

$$d_t = \begin{pmatrix} d_t^1 \\ \vdots \\ d_t^K \end{pmatrix}. \tag{13}$$

and let Δ_t be the matrix of observed price differences,

$$\Delta_t = \begin{pmatrix} 0 & \cdots & B_t^1 - A_t^K \\ \vdots & \ddots & \vdots \\ B_t^K - A_t^1 & \cdots & 0 \end{pmatrix}. \tag{14}$$

Then, the aggregated price differences in excess of efficiency boundaries are

$$\hat{\mathcal{E}}_t = \left\| \left(\Delta_t - \hat{d}_t \iota' \right) \odot \mathbf{\Psi}_t \right\|, \tag{15}$$

where ι is a vector of ones and the (n, m)-th element of $\mathbf{\Psi}_t$ is defined as $\mathbf{\Psi}_{t,n,m} = \mathbb{1}\{B_t^n - A_t^A > \hat{d}_t^n\}$.

Empirical Analyis - Roadmap

Quantify limits to arbitrage by estimating (given γ)

$$\hat{d}_n^t = \gamma \frac{1}{2} \left(\hat{\sigma}_t^n\right)^2 \hat{\lambda}_t^{-1} + \frac{1}{4} \gamma^3 \left(\hat{\sigma}_t^n\right)^4 \hat{\lambda}_t^{-2}, \tag{16}$$

1. Quantify price differences Δ_t
2. Estimate spot volatility $\hat{\sigma}^2_{n,t}$
3. Parametrize stochastic latency $\hat{\lambda}_t$
4. Compute efficiency boundary estimator $\hat{d}_n^t$

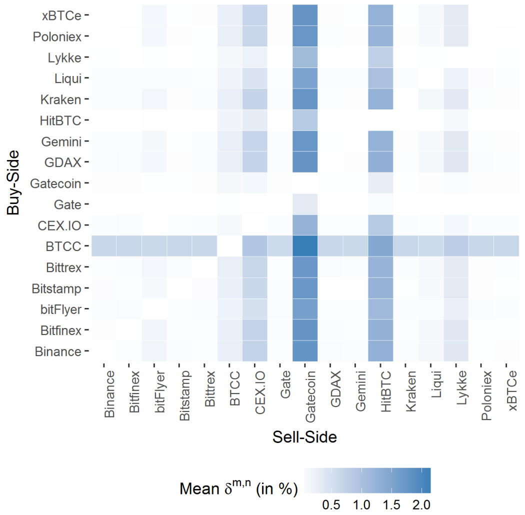
Bitcoin as laboratory for markets with stochastic latency

- Bitcoin can be traded on more than 400 exchanges
- Daily trading volume for Bitcoin/Dollar exceeds 1 billion Dollar
- We collect minute-level Bitcoin/Dollar orderbooks from 18 exchanges since December 2017 ($\approx$ 95% of trading volume)

Bitcoin market structure - Summary statistics

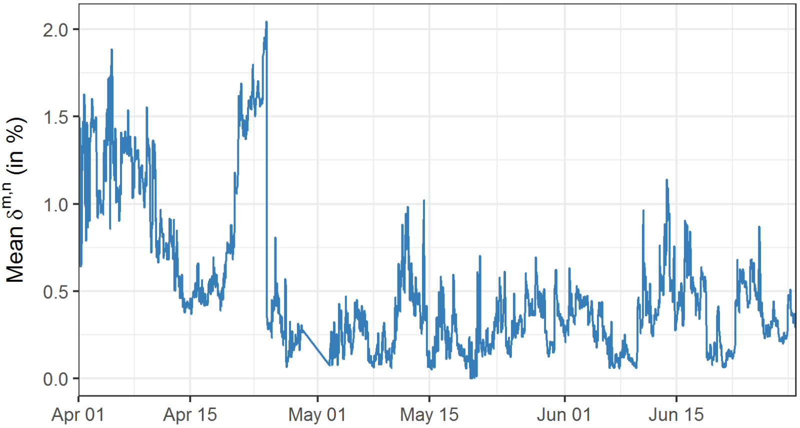
	N	Spread (USD)	Spread (%)	Depth (Ask)	Depth (Bid)
Binance	122,872	2.98	0.04	186,153	186,193
Bitfinex	122,560	0.28	0.00	431,910	459,571
bitFlyer	122,639	18.86	0.24	385,788	258,952
Bitstamp	122,217	5.58	0.07	487,181	517,634
Bittrex	122,878	13.37	0.17	159,462	162,656
BTCC	108,871	110.87	1.46	133,869	75,811
CEX.IO	122,275	13.75	0.18	406,944	426,745
Gate	122,685	219.79	2.84	123,234	129,820
Gatecoin	121,539	163.41	1.92	61,722	67,358
GDAX	122,672	0.04	0.00	249,977	259,122
Gemini	121,678	2.12	0.03	441,173	486,447
HitBTC	122,497	4.22	0.05	128,729	110,010
Kraken	122,929	3.00	0.04	420,540	399,207
Liqui	122,211	34.47	0.45	131,535	182,248
Lykke	124,355	51.52	0.65	110,695	128,994
Poloniex	123,910	8.65	0.11	200,314	183,140
xBTCe	119,808	7.56	0.10	451,848	473,457

Substantial price differences across markets



Median price differences (in %)

Average Price Differences over Time



Estimating spot volatilities $\sigma^2_{n,t}$

- Current volatility affects price risk of arbitrageur

$$dB^n_t = \sigma_{n,t} dW_t \tag{17}$$

- Nonparametric filtering of the realized spot volatility (Kristensen, 2010)
- For each market n and time t, we estimate $(\sigma^n_t)^2$ by

$$\widehat{(\sigma^n_t)}^2(h) = \sum_{s=1}^{t} K(s-t,h)\left(B^n_s - B^n_{s-1}\right)^2, \tag{18}$$

where $K(s-t,h)$ denotes a one-sided Gaussian kernel smoother with bandwidth h.
- Bandwidth h chosen by minimizing the Integrated Squared Error (ISE)

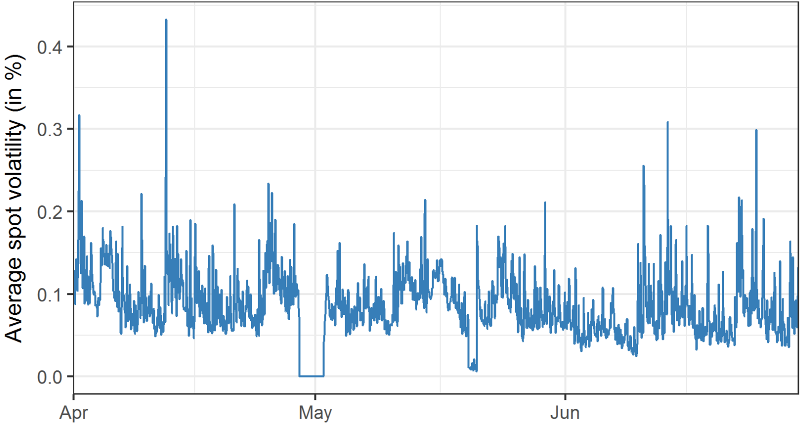
$$\widehat{\text{ISE}}_{T-1}(h) = \sum_{i=1}^{I}\left[\left(B^n_i - B^n_{i-1}\right)^2 - \widehat{(\sigma^n_i)}^2(h)\right]^2, \tag{19}$$

where $i = 1, \ldots, I$ refers to the observations on day $T-1$ and $\widehat{(\sigma^n_t)}^2(h)$ is the spot volatility estimator based on bandwidth h. The optimal bandwidth on day T is thus chosen

$$h = \arg\min_{h>0} \widehat{\text{ISE}}_{T-1}(h) \tag{20}$$

.

Estimated exchange-specific spot volatilities



Parameterizing waiting times $\lambda(\mathcal{I}_t)$

Collect transaction-level data from Bitcoin network

- Unique ID, size, fee, *waiting time* until included in a block

Latency of a transaction depends on

- Current state of the system (i.e. no. of unconfirmed transactions)
- Transaction fee relative to other transactions

Variable	Mean	SD	5 %	25 %	Median	75 %	95 %
Fee per Byte (Satoshi)	31.27	106.58	2.80	5.07	10.07	25.56	133.14
Fee per Transaction (USD)	1.01	10.02	0.05	0.09	0.20	0.62	3.63
Waiting Time (Minutes)	17.57	41.68	0.85	3.45	8.08	17.08	55.93
Mempool Size	2703.01	3322.76	171.00	687.00	1571.00	3320.50	9305.10

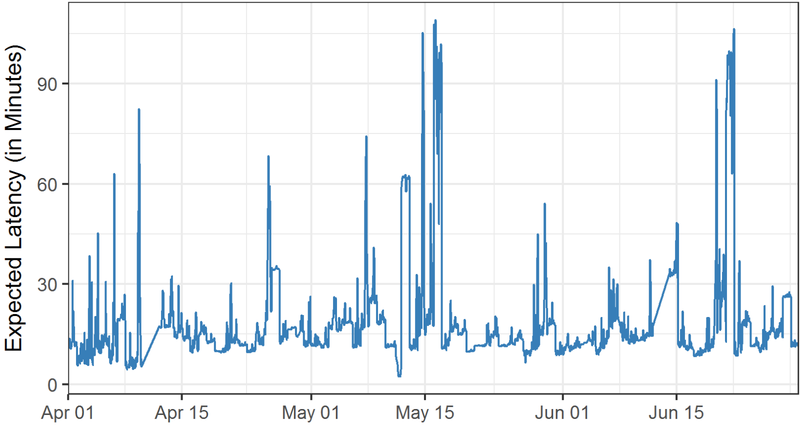
The waiting time τ_t can be written as an accelerated failure time model

$$\ln \tau_t = x_t'\gamma + \varepsilon_t, \tag{21}$$

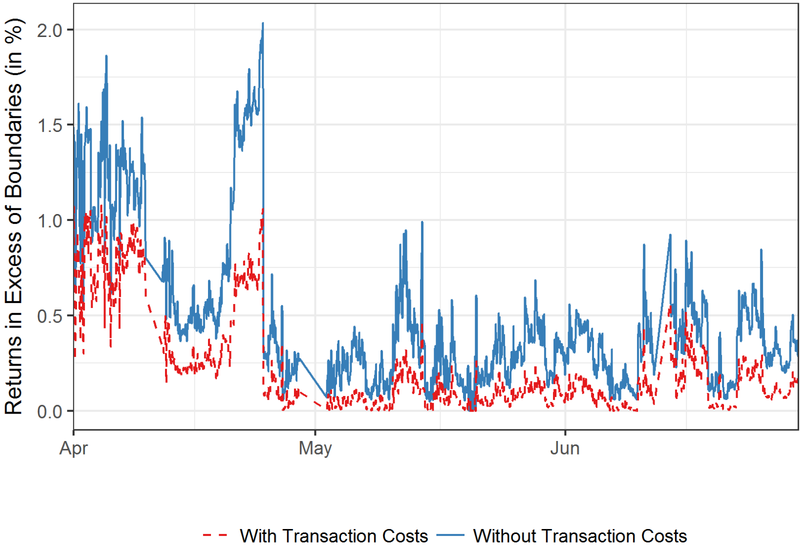
with x_t denoting covariates driving τ_t and γ being a vector of parameters. ε_t follows an extreme value distribution with probability density function

$$\pi(\varepsilon_t) = \exp(\varepsilon_t + \exp(\varepsilon_t)). \tag{22}$$

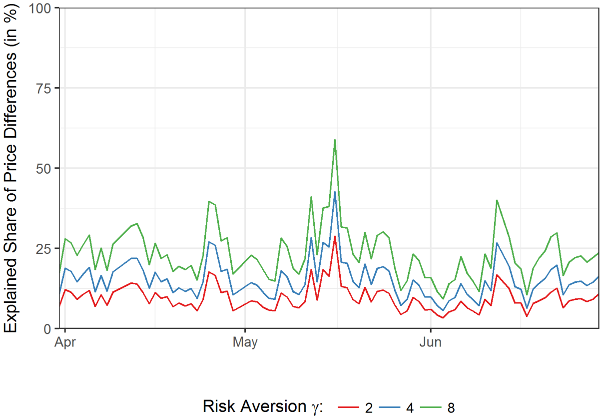
Rolling window estimation of accelerated failure time model



Returns in excess of efficiency boundaries

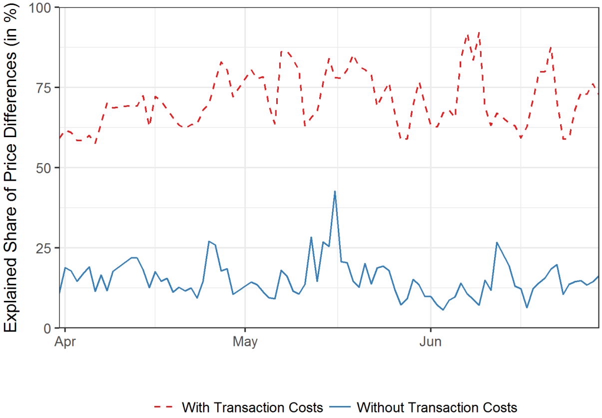


How much do efficiency boundaries explain?



- **Main finding 1:** Boundaries to efficiency, $\hat{D}_t$, explain 20 percent of variation in quoted price differences (for conservative $\gamma = 4$)
- **Main finding 2:** Bitcoin prices exhibit substantial price differences beyond our derived boundaries to price efficiency

How much do transaction costs add?



Exchange Characteristics and Excess Price Differences

	Only Crypto	Taker Fee	Margin Trading	US Citizens	Confirmations	Company location
Binance	✓	0.10	✓	✓	2	Tokyo, Japan
Bitfinex	✗	0.20	✓	✗	3	Central, Hong Kong
bitFlyer	✗	0.15	✓	✓		Tokyo, Japan
Bitstamp	✗	0.25	✗	✓	3	London, United Kingdom
Bittrex	✓	0.25	✗	✓	2	Las Vegas NV, United States
BTCC	✗	0.10	✗		2	Shanghai, China
CEX.IO	✗	0.16	✓	✓	3	London, United Kingdom
Gatecoin	✗	0.35	✗		6	Wanchai, Hong Kong
Gate	✓	0.20	✗		2	Sparta NJ, United States
Gemini	✗	0.25	✗	✓	3	New York NY, United States
GDAX	✗	0.25	✓	✓	3	San Francisco CA, United States
HitBTC	✓	0.10	✗	✓	2	Hong Kong
Kraken	✗	0.26	✓	✓	6	San Francisco CA, United States
Liqui	✓	0.25	✓	✗		Kiev, Ukraine
Lykke	✗	0.14	✗	✗	10	Zug, Switzerland
Poloniex	✓	0.25	✓	✓	1	Wilmington DE, USA
xBTCe	✗	0.25	✓	✗	3	Charlestown, Nevis

- Price differences seem to persist even after adjusting for derived boundaries stochastic latency
- Potential remaining frictions include exchange-specific constraints, withdrawal restrictions, country regulations, HFT-access, ...
- We investigate the effects in a large-scale panel regression

	$\pi(\delta_t^{m,n}(q^*) > d_t^n)$		
	(1)	(2)	(3)
Both Margin Trading	-0.365***	-0.356***	-0.058***
	(-27.49)	(-26.99)	(-7.48)
Both Only Crypto	0.138***	0.138***	-0.012*
	(10.17)	(9.70)	(-1.90)
Both US Citizens	-0.182***	-0.172***	-0.059***
	(-14.59)	(-13.28)	(-7.94)
Constant	Yes	No	No
Timestamp FE	No	Yes	Yes
Buy-Side FE	No	No	Yes
Sell Side FE	No	No	Yes
N	11,158,476	11,158,476	11,158,476
Adjusted R^2	0.13	0.20	0.50

Conclusion

Stochastic latency exposes arbitrageurs to price risk
$\Rightarrow$ imposes limits to arbitrage

Key friction of blockchain-based settlement systems

Quantitatively important friction in Bitcoin markets

- Stochastic latency explains 20% of price differences
- Latency & transaction costs explain 75% of price differences
- Results suggest unexploited arbitrage opportunities