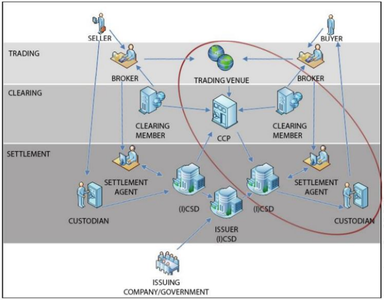
Market making and stochastic settlement latency

Stefan Voigt
Vienna Graduate School of Finance

"Traditional" settlement versus distributed ledgers



? Security without intermediaries

\+ (Legal) settlement process much faster

? On a public blockchain, transactions are (semi-)transparent

? ... (probably many more things)

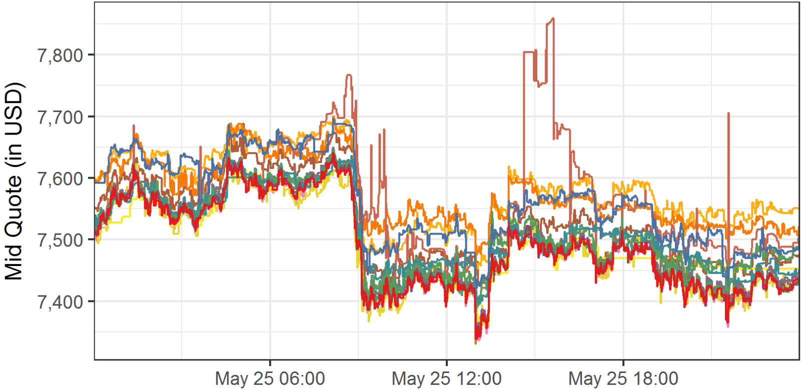
Obvious open economic questions ...

- Market transparency and privacy?
- Control and surveillance?
- **What about blockchain-induced market frictions?**

Distributed ledgers as alternative settlement systems?

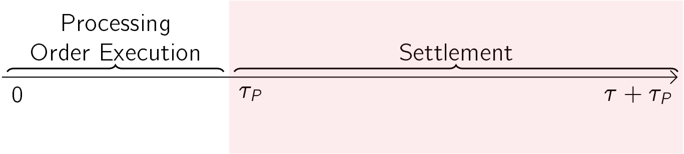
"Blockchain technology isn't just a *more efficient* way to *settle securities*. It will *fundamentally change market structures*, [...]"

Abigail Johnson, CEO of Fidelity Investments



Research focus: latency and trading decisions

- (Relative) speed affects decisions of market participants



- Usual focus: time of information processing and order execution
- **With DLT, no intermediaries** step in to secure ultimate payoffs.
- Traders are exposed to additional source of latency due to decentralized settlement.

Blockchain technology and stochastic latency

Stylized Facts of blockchain-based settlement (Hautsch et al, 2019)

- Consensus algorithms introduce *stochastic latency*
- (Stochastic) settlement latency implies limits to arbitrage
- Quantitatively important friction in Bitcoin markets (on average, 88% of all observed price differences fall within latency related limits to arbitrage)

Follow-up research questions

- **What are the implications of limits to arbitrage (due to stochastic latency) for financial markets?**
- How does the design of the consensus mechanism affect liquidity?
- Can market makers/investors benefit from latency monitoring?

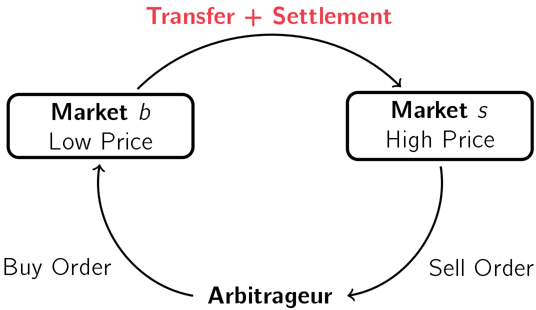
How does blockchain introduce limits to arbitrage?

Market b
Low Price

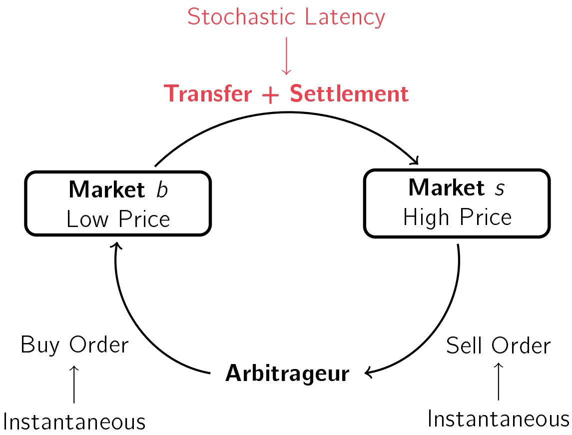
Market s
High Price

Arbitrageur

How does blockchain introduce limits to arbitrage?



How does blockchain introduce limits to arbitrage?



Theoretical framework (without equations)

Market continuously provides buy quotes (ask) and sell quotes (bid) for the asset.

No short selling, margin trading or derivatives

Arbitrageur continuously monitors the quotes on markets b and s.

Stochastic latency τ is the random waiting time until a transfer of the asset between markets is settled.

Profit of arbitrageur's trading decision is at risk if the sell price at $t + \tau$ is below the buy price at t.

Our main theoretical result

If the arbitrageur is risk-averse (or capital constrained), price differences are exploited only if they exceed a certain threshold:

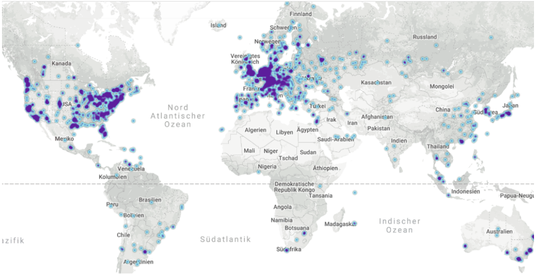
$$b_t^s - a_t^b > d_t^s$$

Stochastic latency implies limits to arbitrage d_t^s which increase if

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

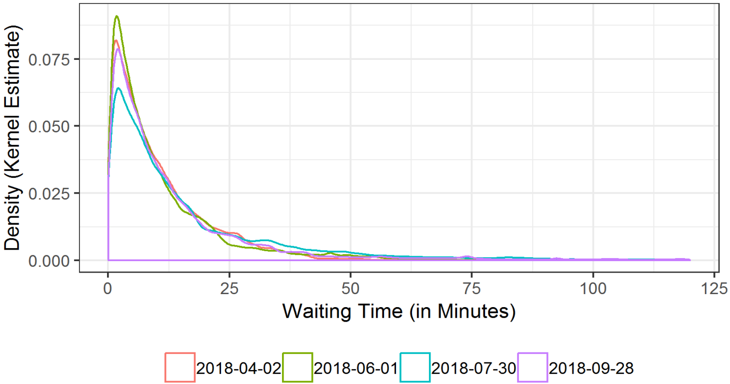
Bitcoin: a settlement system with stochastic latency

- We run a full Bitcoin Node and monitor the mempool
- For every transaction we collect
  - unique ID, size, fee, waiting time until included in a block
- In total, we tracked > 22.000.000 transactions since April

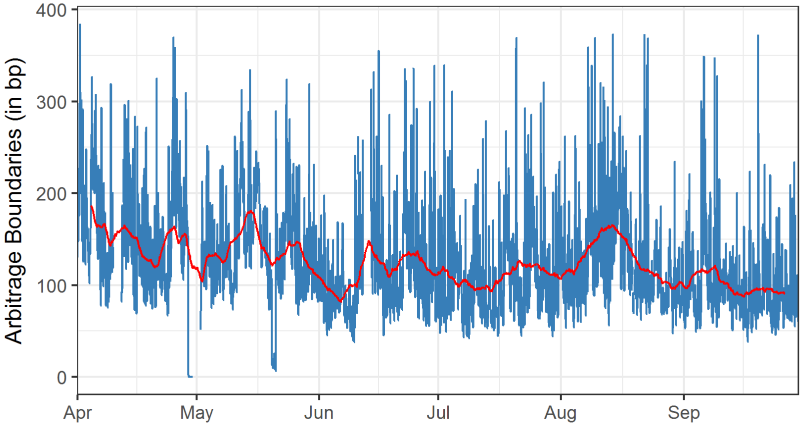


Bitcoin blockchain mempool

	Mean	SD	5 %	Median	95 %
Fee per Byte (Satoshi)	24.16	106.00	2.31	8.31	102.64
Fee per Transaction (USD)	0.73	7.26	0.04	0.15	2.78
Latency (Minutes)	**18.13**	41.97	0.88	8.28	58.30
Transaction Size (Byte)	474.93	2087.21	142.00	225.00	933.00
Mempool Size	2938.35	3474.47	193.00	1765.00	9878.25
Block Mining Time (Minutes)	9.53	9.33	0.55	6.70	28.13

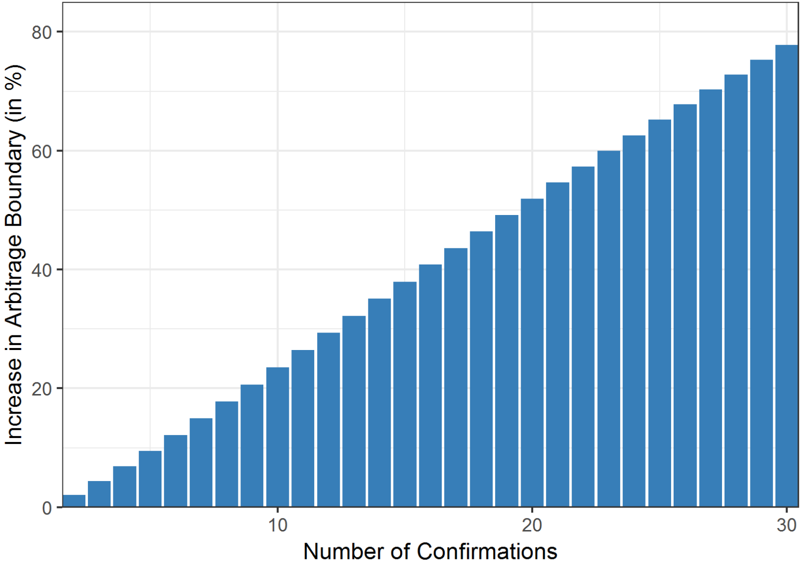


Estimated limits to arbitrage over time (CRRA, $\gamma = 2$)



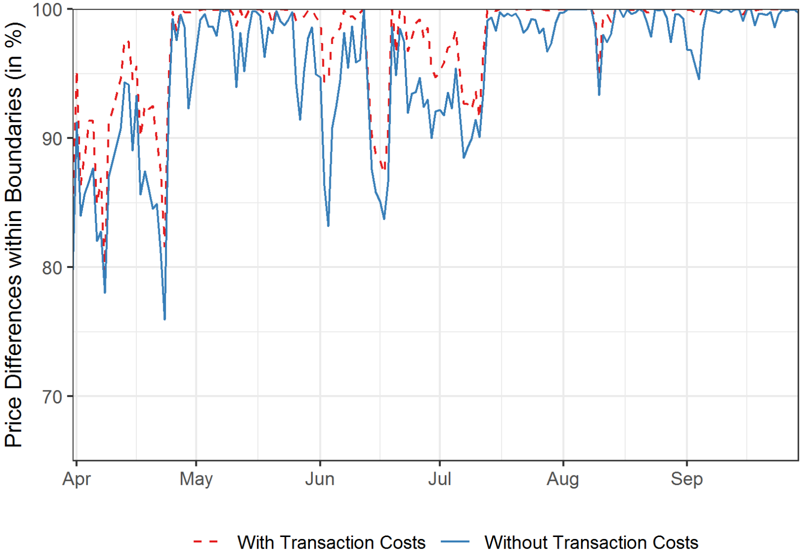
- Large fluctuations due to time-variation in volatility *and* network utilization

Double spending attacks and arbitrage boundaries



Are limits to arbitrage reflected in price differences?

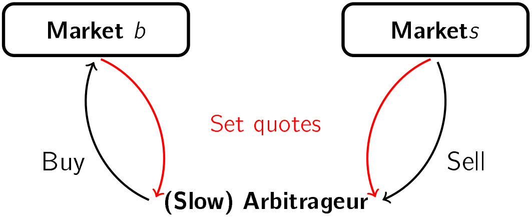
Proportion $\mathbf{min}\{1, \hat{d}_t^n/\delta_t^{b,s}\}$ with and without transaction costs



Next step: implications of limits to arbitrage

- How informative are prices?
  - Role of arbitrageurs: correct mispricing across markets
- Pricing of derivatives?
  - The law of one price constitutes a fundamental pillar of asset pricing
- What are the implied costs of switching to a distributed ledger?
  - Depending on the application, limits to arbitrage may have more/less severe consequences
  - If the volatility of an asset is low relative to the settlement latency, limits to arbitrage are less relevant
  - For equity, however, the picture may look different
- **Implications for High-Frequency trading?**
  - HF Traders/market participants fulfill manifold different roles in financial markets
  - Slowing down trading flow due to technological frictions (or security considerations) entirely changes the market microstructure

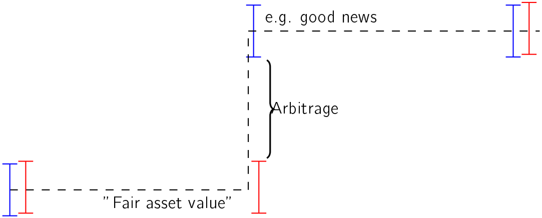
What is the effect of slowing-down arbitrageurs on liquidity?



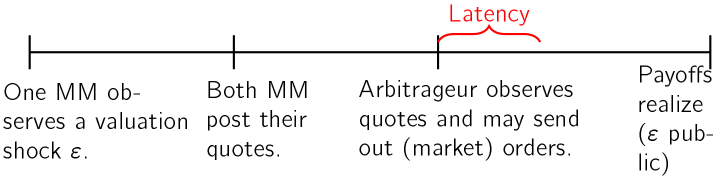
In this ongoing project, I analyze

- (Theory): Effect of limits to arbitrage (due to latency delay) on liquidity provision
  - Less toxic arbitrage (?)
  - Impediments to inventory control (?)
  - Should we strive to build ever faster consensus protocols/ trading platforms?
- (Empirics): Relationship between latency and liquidity in novel blockchain-market HF-orderbook dataset

A (very) stylized model



- Asset traded at two (segmented) markets X and Y
- 3 types of participants
  - Two risk neutral (specialized) market makers
  - One arbitrageur
  - Liquidity traders who buy or sell one share on X or Y with equal probabilities



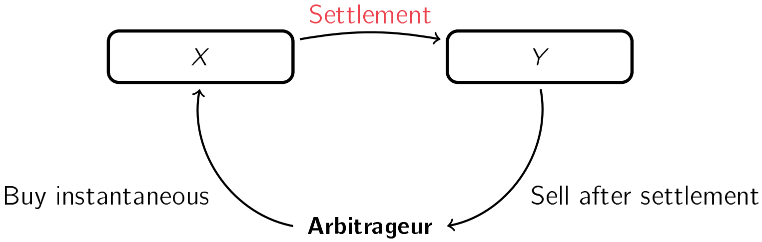
The role of market makers' quotes

- At date $t = 1$, market makers simultaneously post an ask price a_j and a bid price b_j, which consists of two elements: their private valuation v_j and the spread S_j:

$$a_j = v_j + \frac{S_j}{2} \qquad \text{and } b_j = v_j - \frac{S_j}{2}$$

- The spread compensates for the risk of trading against an arbitrageur
- Quoted liquidity attracts investors and is actively monitored (see Christophs' talk)
- In a (competitive) equilibrium, quotes are such that X and Y earn zero expected profit

Latency, limits to arbitrage and optimal quotes

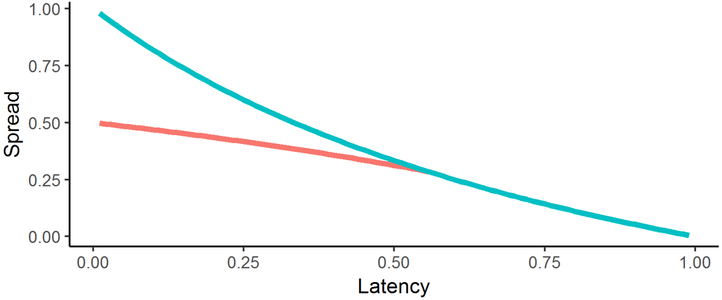


- At $t = 2$, arbitrageur has two choices
  1. Stay idle - no payoffs at all
  2. Enter arbitrage trade: Buy asset for $v \pm S_j$, transfer to $i \neq j$ and sell after settlement at $t + \tau$.
- Quote updating does not require settlement
- Sell-side market maker may update her quotes during settlement period
- Decision to trade depends on the spread!

Two effects of settlement latency

- If arbitrageur trades, her payoff is at risk with prob. $1 - \pi^*$ (if market makers update quotes faster than settlement occurs)
- Shorter latency (fast settlement) decreases risk of arbitrageur
- Limits to arbitrage reduce threat of *toxic arbitrage* for market makers (market makers can reduce their spreads)

Adverse selection and latency



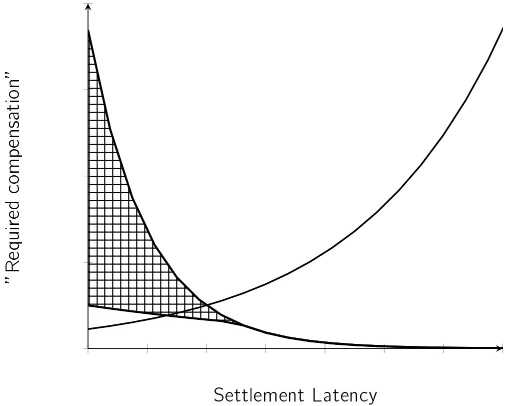
- Limits to arbitrage reduce adverse selection component
- Latency implies smaller spread but less price discovery

Implications for market makers

- Spread should be a function of the settlement latency
- Number of required block confirmations is an (exchange-specific) tool to control limits to arbitrage

Is there an "optimal" settlement latency?

- Faster consensus protocol increases risk of stale quotes
- Slower settlement switches off arbitrageurs activity as cross-market liquidity providers



Conclusion and outlook

Main Take-aways:

- Settlement latency constitutes a fundamental technological friction of every decentralized ledger
- In the BTC market, limits to arbitrage due to settlement latency are large
- Actions to impede threat of double-spending attacks increase these limits even further
- Settlement latency opens path for liquidity providers to narrow down spreads (and to attract more order flow from non-arbitrageurs)

Outlook:

- Impact of liquidity on arbitrage is well-established, way less results for the reverse case
- Empirical Analysis rests on novel dataset with High-frequency data for BTC, LTC, ETH, XRP
- Running full nodes allows to provide real-time latency monitoring tools

Theoretical framework

Market $i \in \{1, \dots, N\}$ continuously provides (log) 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 b and s.

Instantaneous trading: Arbitrageur exploits price differences if

$$\delta_t^{b,s} := b_t^s - a_t^b > 0$$

Stochastic latency τ is the random waiting time until a transfer of the asset between markets is settled.

Profit of arbitrageur's trading decision is at risk if

$$\mathbb{P}\left(b_{t+\tau}^s \leq a_t^b\right) > 0$$

Log return of arbitrageur's strategy

$$r^{b,s}_{(t:t+\tau)} := b^s_{t+\tau} - a^b_t = \underbrace{\delta^{b,s}_t}_{\substack{\text{instantaneous}\\ \text{return}}} + \underbrace{b^s_{t+\tau} - b^s_t}_{\substack{\text{exposure to}\\ \text{price risk}}}, \quad \text{where } \delta^{b,s}_t := b^s_t - a^b_t.$$

Assumption 1. For given latency τ, we model the log price changes on the sell-side $b^s_{t+\tau} - b^s_t$ as a Brownian motion with drift μ^s_t such that

$$r^{b,s}_{(t:t+\tau)} = \delta^{b,s}_t + \tau\mu^s_t + \int_t^{t+\tau} \sigma^s_t dW^s_k,$$

We assume σ^s_t and μ^s_t are locally constant over the interval $[t, t+\tau]$.

Assumption 2. Stochastic latency $\tau \in \mathbb{R}_+$ is a random variable equipped with a (conditional) probability distribution $\pi_t(\tau) := \pi(\tau|\mathcal{I}_t)$. We assume that the moment-generating function of $\pi_t(\tau)$ is finite on an interval around zero.

Stochastic latency and limits to arbitrage

Assumption 3. Arbitrageur has power utility function

$$U_\gamma(r) := \frac{r^{1-\gamma} - 1}{1-\gamma},$$

with relative risk aversion parameter $\gamma > 1$.

Lemma. Under the above assumptions with $\mu_t^s = 0$, price differences are exploited if

$$\delta_t^{b,s} > d_t^s = \frac{1}{2}\sigma_t^s \sqrt{\gamma \mathbb{E}_t(\tau) + \sqrt{\gamma^2 \mathbb{E}_t(\tau)^2 + 2\gamma(\gamma+1)(\gamma+2)\mathbb{E}_t(\tau^2)}}.$$

Stochastic latency implies limits to arbitrage which increase if

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