Building Trust takes Time
Limits to Arbitrage for Blockchain-Based Assets
Voigt, Stefan; Hautsch, Nikolaus; Scheuch, Christoph

DOI:
10.2139/ssrn.3302159

Publication date:
2024

Document version
Early version, also known as pre-print

Citation for published version (APA):
Building Trust Takes Time: Limits to Arbitrage in Blockchain-Based Markets*

Nikolaus Hautsch  Christoph Scheuch  Stefan Voigt

January 29, 2021

Distributed ledger technologies replace central counterparties with time-consuming consensus protocols to record the transfer of ownership. This settlement latency slows down cross-market trading and exposes arbitrageurs to price risk. We theoretically derive arbitrage bounds induced by settlement latency. Using Bitcoin orderbook and network data, we estimate average arbitrage bounds of 121 basis points, explaining 91% of the cross-market price differences, and demonstrate that asset flows chase arbitrage opportunities. Controlling for inventory holdings as a measure of trust in exchanges does not affect our main results. Blockchain-based settlement without trusted intermediation thus introduces a non-trivial friction that impedes arbitrage activity.

JEL Codes: G00, G10, G14

Keywords: Arbitrage, Blockchain, Market Frictions

*Hautsch is at University of Vienna – Department of Statistics and Operations Research, Research Platform Data Science © Uni Vienna, Vienna Graduate School of Finance (VGSF) and Center for Financial Studies (CFS). Scheuch is at wikifolio Financial Technologies and Voigt is at University of Copenhagen – Department of Economics, and the Danish Finance Institute. Corresponding author: stefan.voigt@econ.ku.dk. We thank Bruno Biais, Sylvia Frühwirth-Schnatter, Sergey Ivliev, David Easley (discussant), Katya Malinova (discussant), Fahad Saleh (discussant), Peter Zimmerman (discussant), Viktor Todorov (discussant) as well as seminar participants at QFFE 2018, the 1st International Conference on Data Science in Finance with R, the 4th Konstanz-Lancaster Workshop on Finance and Econometrics, the Crypto Valley Blockchain Conference 2018, HFFE 2018, CFE 2018, CUNEF, University of Heidelberg, University of Vienna, University of Graz, the 2nd Toronto FinTech Conference, the 4th Vienna Workshop on High-Dimensional Time Series 2019, the Conference on Market Microstructure and High Frequency Data 2019, the 2019 FIRS Conference, the 12th Annual SoFiE Conference, the IMS at the National University of Singapore, the 3rd SAFE Microstructure Conference, the 2019 EFA Annual Meeting, the 2019 Vienna Congress on Mathematical Finance, the International Conference on Fintech & Financial Data Science 2019, the 4th International Workshop in Financial Econometrics, the 2019 CFM-Imperial Workshop, the Western Finance Association Conference 2020 and the NHH Young Scholars Finance Webinar series for helpful comments and suggestions. This paper replaces an earlier draft titled "Limits to Arbitrage in Markets with Stochastic Settlement Latency".
1 Introduction

Whenever two investors seek to agree on a financial contract, counterparty risk harms trading if either party may default on its contractual obligations. The mitigation of this risk requires trustworthy intermediation that guarantees the execution of the initial contract terms. Traditional stock markets overcome limitations due to counterparty risks by organizing the trading process around trusted intermediaries. Typically, central clearing counterparties bear all counterparty risks between transacting parties during the time span from trade agreement until the legal transfer of ownership (which typically takes two to three business days, e.g., SEC, 2017) through security depositories. Due to this temporal separation of the process of trading and the process of settlement, traders do not have to wait until the legal transfer of ownership is accomplished, but can continuously trade on unsettled positions. Market participants pay for this implied insurance against counterparty risks through fees and collateral deposits.

By contrast, distributed ledger technologies, such as blockchain, promise to mitigate counterparty risks and at the same time render trusted intermediaries obsolete. In such systems, decentralized validators interact with each other to establish consensus about transaction histories. Consensus protocols serve as the regulatory framework and specify how validators reach agreement, how fast this agreement is accomplished and how validators are incentivized to collaborate. The design of distributed consensus protocols can take different forms to ensure a reliable record of transaction histories. One of the best known and widely applied consensus mechanisms is proof-of-work, where the validation process involves substantial computational effort and is hence prohibitively costly to undermine (e.g., Biais et al., 2019).

Since proof-of-work systems typically achieve consensus within minutes, blockchain technology considerably speeds up the process of settlement compared to markets with central clearing. However, market participants cannot dispose of newly acquired positions until validators have recorded the transfer of ownership on the distributed ledger. In this way, trading and settlement are inextricably linked. We refer to the waiting time until validation of a transaction as settlement latency. This settlement latency is several magnitudes larger than the execution latency in traditional markets because the complexity in the validation process that is necessary to ensure the trustworthiness of the distributed ledger requires a critical amount of time.

The aim of this paper is to demonstrate that settlement latency, an inherent feature of proof-of-work consensus mechanisms, constitutes an economically significant friction.
We show that settlement latency leads to limits to arbitrage in the absence of trusted intermediaries since arbitrageurs cannot trade sufficiently fast in order to exploit cross-market price differences. As a consequence, violations of the law of one price can arise and impede the pricing of securities.

Empirically, we provide compelling evidence for such violations of the law of one price for Bitcoin, the most popular asset that is settled on a blockchain, against US Dollar. The average price differences across 120 exchange pairs through our sample (January 2018 until October 2019) are around 33 bp and thus economically highly significant. As Choi et al. (2018), Borri and Shakhnov (2019) and Makarov and Schoar (2020) also report, these price differences persist, are not traded away and (as we show in this paper) cannot be reconciled solely with transaction costs, capital controls or other exchange-specific frictions.

We show that the observed price differences are reconciled with arbitrage bounds that arise from the decision problem of a risk-averse arbitrageur. We model the trading decision of an arbitrageur who monitors prices on two different markets, but faces settlement latency, which limits her possibility to exploit concurrent price differences between the two markets. Whenever she buys on one market (at the current price), she has to wait until the transfer of the asset is validated before she can sell on the other market. The latency underlying this transfer thus exposes arbitrageurs to the risk of adverse price movements. Consequently, risk-averse arbitrageurs only exploit (concurrent) price differences if these price differences are sufficiently large to compensate for the price risk during the settlement period. Hence, price differences below these bounds may persist as it is rational for the arbitrageur to refrain from trading.

We provide closed-form expressions for the arbitrage bounds under fairly general assumptions and show that they increase with (i) the volatility on the sell-side market, (ii) the expected settlement latency, (iii) the variance of the settlement latency and (iv) the arbitrageur’s risk aversion. Our characterization of the arbitrage bounds also accounts for transaction costs and optimally chosen settlement fees which incentivize validators to enable faster validation (e.g., Easley et al., 2019).

Minute-level data from orderbooks of 16 exchanges that feature trading Bitcoin against US Dollar between January 2018 and October 2019 allows us to detect potential arbitrage opportunities between each exchange pair, taking transaction costs into account. Furthermore, we gather comprehensive information about the Bitcoin network, which includes the time it takes for every transaction from entering the Bitcoin network until its inclusion in the blockchain (i.e., its settlement). Using this information, the estimation
of arbitrage bounds rests on three ingredients. First, we construct estimates of exchange-
specific spot volatilities based on minute-level best bid quotes. Second, we parameterize
the latency distribution as a conditional Gamma distribution depending on network and
transaction-specific characteristics affecting the settlement latency. Third, in line with
the existing literature, we choose an isoelastic utility function with exogenously given
coefficient of relative risk aversion.

The average estimated arbitrage bound for a relative risk aversion of 2 amounts to
121 bp. This magnitude constitutes an economically highly significant friction and may
explain severe distortions of the law of one price. We find that 84% of all observed
concurrent price differences based on best bid and ask quotes across markets fall within
these bounds. Adjusting additionally for transaction costs, the bounds contain even
up to 91% of the observed price differences. Equivalently, we show that the average
implied relative risk aversion necessary to explain all observed concurrent cross-exchange
price differences amounts to 17. This estimate is high compared to existing estimates
of coefficients of relative risk aversion in the asset pricing literature (e.g., Hansen and
Singleton, 1982; Chetty, 2006) and suggests the existence of additional market frictions
as settlement latencies and transaction costs cannot fully explain all (apparent) arbitrage
opportunities.

We derive the arbitrage bounds under the assumption that traders have no possibil-
ity to bypass the settlement latency. In this sense, we consider the benchmark case of
arbitrage bounds that occur in a purely decentralized system without any trusted inter-
mediation or whenever intermediation is prohibitively costly or risky. Therefore, these
bounds reflect the economic frictions – measured as the magnitudes of arbitrage bounds
in basis points – which arbitrageurs face in blockchain-based markets, that is, whenever
settlement is organized via decentralized consensus protocols without trusted third-party
intermediaries.

However, as the validation process is costly and time-consuming, cryptocurrency ex-
changes themselves exert substantial effort to enable bypassing the direct settlement of
transactions on the blockchain. On the one hand, they net executed transactions inter-
nally. Traders thus only face settlement latency for withdrawals or deposits. On the
other hand, trading venues established fast private inter-exchange settlement networks
that closely resemble clearing houses.\footnote{For instance, since October 2018, the company Blockstream runs a private side-chain to the main Bitcoin blockchain, which connects several exchanges and allows for transfer of assets between exchanges in less than 2 minutes.} In this sense, typical cryptocurrency exchanges
operate outside the premises of pure trustless intermediation, but serve as centralized intermediaries that enable market participants to overcome latency-related frictions in the settlement process for internally executed transactions. These practices open two main channels to exploit cross-market price differences without facing settlement latency: short-selling or simultaneous inventory holdings on multiple trading venues.

This form of centralization on the blockchain, however, clearly undermines the initial idea of a purely decentralized system and reintroduces counterparty risk: To render internal settlement feasible, exchanges need to serve as custodians of their customer’s funds and require full margin deposits before accepting orderbook transactions. As a result, it is impossible for cross-market arbitrageurs to dispose of their position before validators verify their deposit. Simultaneously, exchanges that perform internal settlement expose traders to exchange default risk. Therefore, considerable effort to (re-)establish the trustworthiness as a centralized intermediary is required, which is a prerequisite to overcome settlement latency related frictions. Examples for such trust-enhancing procedures include an increase of transparency with respect to actual cryptocurrency holdings under custody, the implementation of funds to repay damages due to security breaches or compliance with strict regulatory standards, as, e.g., the SEC BitLicense.

Notably, security standards at modern cryptocurrency exchanges are directly linked to settlement latency as most exchanges require not only one but several confirmations before they regard an incoming transaction as valid. This reduces the risk of so-called double spending attacks, where an attacker tries to spend his funds twice in different transactions. For this reason, some exchanges in our sample require up to 5 additional confirmations, which considerably increases the settlement latency and thus the implied arbitrage bounds. We estimate an increase of the resulting arbitrage bounds by on average 7 bp per additional confirmation. Therefore, in order to make internal settlement trustworthy, exchanges need to impose substantial additional costs on cross-market trading.

Our paper thus reveals an inherent dilemma of blockchain-based markets. While trusted intermediaries are able to overcome latency frictions, the essential idea of blockchain is to be trustless. Modern cryptocurrency exchanges hence struggle to build trust in a

---

2The efficiency gains for exchanges and their clients seem enormous: Reported trading volume of Bitcoin versus USD on October 31st 2019 was above 25 billion USD (according to the data provider coinmarketcap). At the same day, Bitcoin transactions worth only 913 million USD (excluding change) were verified on the blockchain.

3Biais et al. (2019) document more than 50 hacks and other losses on Bitcoin exchanges and find that expected returns of cryptocurrency investors reflect these risks.
world that promises to be trustless.\footnote{Changpeng Zhao, the CEO of Binance, the world’s largest cryptocurrency exchange expressed this concern as follows: “By focusing on delivering a superior user experience in tandem with top-notch security, centralized exchanges essentially live or die based on their ability to “create trust” among their users.” See https://www.coindesk.com/defi-why-the-days-of-centralized-exchanges-are-numbered.}

However, establishing centralized markets as trusted intermediaries is not the only feasible approach to facilitate trading in blockchain-based markets. Instead, in light of mainly unregulated cryptocurrency markets, decentralized finance (DeFi) protocols are worth noting. DeFi protocols use smart contracts on blockchains to enable trading on decentralized exchanges without any central financial intermediaries (e.g., Harvey et al., 2020). While such protocols render trusted intermediation obsolete because each bilateral transaction is directly settled on distributed ledgers only, trading will be affected by the frictions due to settlement latency highlighted in this paper. It remains an open question which platform characteristics (and hence which frictions) ultimately dominate.

To analyze the role of arbitrage bounds in light of these developments and to obtain a proxy for trust in an exchange, we use data from the platform glassnode to extract Bitcoin holdings under the custody of cryptocurrency exchanges from the publicly available blockchain. We document that at the end of our sample more than 12.4 Billion USD of Bitcoin were held at the custody of cryptocurrency exchanges, an increase of 25.8% relative to the beginning of 2018. We show that cross-market price differences tend to decline with inventory holdings, but latency-induced arbitrage bounds still remain significant drivers of both the magnitudes of observed cross-market price differences and their variation over time, even when controlling for inventory. This explanatory power persists if we additionally control for other well-known frictions that hamper arbitrage activity such as market liquidity (e.g., Roll et al., 2007; De Long et al., 1990) and the presence of intermediation facilities, such as margin trading (e.g., Pontiff, 1996; Lamont and Thaler, 2003a,b; De Jong et al., 2009).

We also study the relation between cross-exchange price differences and corresponding asset flows. To perform this analysis, we collect wallets that are under the control of the exchanges in our sample and compile a unique and novel data set of 3.9 million cross-exchange transactions with an average daily volume of 72 million US Dollar. We use the calibrated arbitrage bounds as instruments for cross-exchange price differences, to tackle the inherent endogeneity arising from the simultaneity between price differences and cross-exchange asset flows. We find that asset flows into an exchange significantly respond to variations in concurrent price differences, particularly those explained by variations in arbitrage bounds, while we also control for latency-induced price risks and
exchange-specific characteristics. Hence, markets perceives arbitrage opportunities and limits thereof. In this sense, we add to the literature on limits to arbitrage (e.g., De Long et al., 1990; Shleifer and Vishny, 1997; Gromb and Vayanos, 2010) by highlighting a friction that arises in blockchain-based markets and which impedes arbitrageurs’ abilities to exploit mispricing.

Our results contribute to a better understanding of the economic implications of distributed ledger technologies for trading on financial markets. In fact, the promise of fast and low-cost transaction settlement lead central banks and marketplaces to actively explore potential applications of such systems for transaction settlement (e.g., BIS, 2017; NASDAQ, 2017; ECB and BoJ, 2018; SIX, 2018). There are two main strands in the existing economic literature that we relate to: On the one hand, the incentive compatibility constraints of validators limit decentralization. For instance, Cong et al. (2019) show that risk-averse validators have incentives to pool their mining power that can result in inefficient accumulations of mining capacities. Hinzen et al. (2019) find that network security and fast settlement are mutually exclusive and Pagnotta (2018) finds miners’ competition for block rewards amplifies fundamental shock volatility and thus has severe pricing implications. On the other hand, other studies highlight the potential costs and welfare gains associated with the decentralized settlement process. In particular, Abadi and Brunnermeier (2018) point out a “Blockchain dilemma” in the sense that correctness, decentralization and cost efficiency cannot be achieved simultaneously in blockchain-based markets. Chiu and Koeppl (2019) illustrate that the benefits of faster blockchain-based settlement can nevertheless outweigh the implied frictions and lead to improvements relative to the legacy costs of the existing settlement system.

Our focus lies on the economic frictions that originate from the time-consuming effort necessary to establish trust in blockchain-based markets. In this sense, our results help to quantify the economic frictions arising from a lack of trust in a market without trusted intermediaries.

Overall, our paper sheds light on the important open question: to which extent can a trading system achieve trustworthiness while maintaining sufficiently fast trading and minimal third-party intermediation? Our main conclusion is that it is impossible to bypass the waiting time necessary to establish sufficient security and trust in a purely decentralized system without causing any other costs or frictions. The analysis and estimation of latency-induced arbitrage bounds is key to better understand the impact of these frictions.
2 Settlement Latency and Limits to Arbitrage

2.1 Arbitrage returns under settlement latency

We consider an economy containing a single asset that is traded on two different markets $b$ and $s$. The trading activity on these markets is exogenously given and we assume that agents can continuously monitor the quotes of the asset across all markets. We assume that market $i \in \{b, s\}$ continuously provides buy quotes (asks) $A_i^t$ and sell quotes (bids) $B_i^t$ (with $B_i^t \leq A_i^t$) for one marginal unit of the asset at time $t$. We address the possibility to trade more than one marginal unit of the asset as well as transaction costs in the next section and show that these generalizations do not affect our main insights.

Our sole agent is an arbitrageur who aims to exploit observed price differences across markets. The arbitrageur continuously monitors the quotes on markets $b$ and $s$ and considers the following strategy: if buying on one market and selling on the other market implies a profit, she intends to buy a marginal unit of the asset on 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.

Without loss of generality, we focus on a scenario where the arbitrageur buys on market $b$ and sells on market $s$. The converse case of selling on market $b$ and buying on market $s$ can be handled analogously. Hence, in case of frictionless trading and no settlement latency, the arbitrageur exploits observed price differences whenever

$$B_s^t > A_b^t,$$

as she can buy the asset on market $b$ at $A_b^t$, instantaneously transfer the asset to market $s$ and sell it again at price $B_s^t$.

To quantify economic frictions that arise in the benchmark scenario of decentralized settlement without any intermediation services, we assume that there is no cost-less possibility to bypass settlement latency. In particular, there are no short-selling functionalities provided by exchanges and the possibility to hold inventory of the underlying on several markets is too risky. Accordingly, in the absence of trusted intermediation an instantaneous transfer is not possible, whenever the settlement of the transaction is time-consuming. Such a (possibly random) latency constitutes a fundamental element of distributed ledger systems that do not rely on central clearing entities. It should not be confused, however, with latency in order execution, as heavily discussed in the context of high-frequency trading (e.g., Hasbrouck and Saar, 2013; Foucault et al., 2017). Such
latencies are in the order of milliseconds and thus of several magnitudes smaller than settlement latency. Therefore, without loss of generality, we refrain from latency in order execution and assume that markets process orders instantaneously.

Let the latency \( \tau \) denote the random waiting time until a transfer of the asset between markets is settled. If the buy transaction on market \( b \) takes place at time \( t \) and the transfer of the asset to market \( s \) is settled at \( t + \tau \), the arbitrageur faces the sell quote \( B_{t+\tau}^s \). The profit of the arbitrageur’s trading decision is thus at risk if the probability of losing money is non-zero, i.e., if

\[
P(B_{t+\tau}^s < A_t^b) > 0.
\]  

(2)

In this case, a risk averse arbitrageur faces limits to (statistical) arbitrage whenever the associated risk exceeds the expected return (see, e.g., Bondarenko, 2003). To formalize the trading decision of the arbitrageur, denote the log quotes by \( a_t^b := \log(A_t^b) \) and \( b_t^s := \log(B_t^s) \), respectively, to cast the payoff in log returns. The log return resulting from buying on market \( b \) at time \( t \) and selling on market \( s \) at time \( t + \tau \) is then given by

\[
r_{b,s}^{t:t+\tau} := b_{t+\tau}^s - a_t^b = \underbrace{\delta_{t}^{b,s}}_{\text{instantaneous return}} + \underbrace{b_{t+\tau}^s - b_t^s}_{\text{exposure to price risk}},
\]  

(3)

where \( \delta_{t}^{b,s} := b_t^s - a_t^b \) defines the return the arbitrageur would earn under instantaneous settlement, i.e., in the absence of any latency. The second part of the decomposition captures the risk of adverse price movements on the sell-side market. As the instantaneous return \( \delta_{t}^{b,s} \) is observable and thus known in \( t \), the arbitrageur faces only uncertainty about the evolution of prices on the sell-side market. The price process on the sell-side market is given as follows.

**Assumption 1.** For a given latency \( \tau \), we model the log price change on the sell-side \( b_{t+\tau}^s - b_t^s \) as a Brownian motion with drift \( \mu_t^s \) such that

\[
r_{b,s}^{t:t+\tau} = \delta_{t}^{b,s} + \tau \mu_t^s + \int_t^{t+\tau} \sigma_t^s dW_k^s,
\]  

(4)

where \( \sigma_t^s \) denotes the spot volatility of the bid quote process on market \( s \), and \( W_k^s \) denotes a Wiener process. We assume that \( \sigma_t^s \) is constant over the interval \( [t, t + \tau] \).\(^5\)

\(^5\)Time-varying and stochastic volatility can be incorporated by means of a change of the time-scale of the underlying Brownian motion. We provide the corresponding derivations in Appendix B. Both the time-variability of \( \sigma_t^s \) and the presence of jumps would further increase the price risk the arbitrageur is
The dynamics of the sell price thus expose the arbitrageur to uncertainty about her profits. We require only weak assumptions regarding the stochastic nature of the latency.

Assumption 2. The stochastic latency $\tau \in \mathbb{R}_+$ is a random variable equipped with a conditional probability distribution $\pi_t(\tau) := \pi(\tau | I_t)$, where $I_t$ denotes the set of available information at time $t$. We assume that the moment-generating function of $\pi_t(\tau)$, defined as $m_\tau(u) := \mathbb{E}_t(e^{u\tau})$ for $u \in \mathbb{R}$, is finite on an interval around zero.

Assumptions 1 and 2 allow us to fully characterize the return distribution $\pi_t(I_{(t+\tau)})$ through the interval of random length from $t$ to $t + \tau$ for a wide range of latency distributions.

**Lemma 1.** Under Assumptions 1 and 2, the returns follow a normal variance-mean mixture with probability distribution

$$
\pi_t(I_{(t+\tau)}) = \int_{\mathbb{R}_+} \pi_t(I_{(t+\tau)} | \tau) \pi_t(\tau) d\tau,
$$

and corresponding characteristic function$^6$

$$
\varphi_{\pi_t(I_{(t+\tau)})}(u) = e^{iu\mu_s} m_\tau \left( iu\mu_s - \frac{1}{2} u^2 (\sigma_s)^2 \right).
$$

**Proof.** See Appendix A. \qed

For any valid distribution $\pi_t(\tau)$, Lemma 1 characterizes the impact of settlement latency on the return distribution. In Appendix C, we illustrate the special case where $\pi_t(\tau)$ follows an exponential distribution and show that the resulting return distribution follows an asymmetric Laplace distribution.

### 2.2 Arbitrage bounds for risk averse arbitrageurs

Settlement latency in blockchain-based markets implies that riskless cross-market arbitrage is infeasible. To quantify the arbitrageur’s assessment of risk, we have to equip her with a corresponding utility function.

---

$^6$The characteristic function fully describes the behavior and properties of a probability distribution. For a random variable $X$, $\varphi_X(u)$ is defined as $\varphi_X(u) = \mathbb{E}(e^{iuX})$, where $i$ is the imaginary unit and $u \in \mathbb{R}$ is the argument of the characteristic function.
Assumption 3. The arbitrageur has an utility function $U_\gamma(r)$ with risk aversion parameter $\gamma$, where $r$ are the log returns implied by her trading decision. Furthermore, we assume $U'_\gamma(r) > 0$ and $U''_\gamma(r) < 0$.

The arbitrageur maximizes the expected utility $E_t(U_\gamma(r))$, which we express in terms of the certainty equivalent (CE). We derive the CE of exploiting concurrent cross-market price differences in the following theorem.

**Theorem 1.** Under Assumptions 1 - 3, the certainty equivalent (CE) resulting from the arbitrage trade is given by

$$CE = \delta^b_s + \mu^s + \sum_{k=2}^{\infty} \frac{U^{(k)}_\gamma}{k!U'_\gamma} \left( \delta^b_s + \mu^s \right)^k, \quad (7)$$

where $U^{(k)}_\gamma(r) := \frac{\partial^k}{\partial r^k} U_\gamma(r)$.

**Proof.** See Appendix A.

Theorem 1 allows us to compare the expected utility of executing the arbitrage trade versus staying idle (which yields a riskless return of zero). The arbitrageur is willing to exploit cross-market price differences if and only if the CE of trading given by Equation (7) is positive. A positive CE corresponds to a statistical arbitrage opportunity in the sense of positive expected risk-adjusted profits. Whenever the observed price differences $\delta^b_s$ are positive, but $CE$ is negative, the arbitrageur does not trade. In this case, although the trade would be profitable under the possibility of instantaneous settlement, limits to (statistical) arbitrage arise due to stochastic latency. Hence, the arbitrageur is indifferent between trading and staying idle if the observed price differences $\delta^b_s$ imply $CE = 0$.

**Definition 1.** We define the arbitrage bound $d^*_{t}$ as the minimum price difference necessary such that the arbitrageur prefers to trade. Formally, $d^*_{t}$ is the maximum of zero and the unique root of

$$F(d) = d + \mu^s + \sum_{k=2}^{\infty} \frac{U^{(k)}_\gamma}{k!U'_\gamma} \left( d + \mu^s \right)^k, \quad (8)$$

By definition of the CE, we have $F(d) = U^{-1}_\gamma \left( \mu^s \right)$. Since $U'_\gamma(r) > 0$, the expectation is increasing in $d$. Moreover, since $U''_\gamma(r) < 0$, the inverse $U^{-1}_\gamma(r)$ is also strictly concave. Thus, $F(d)$ is strictly increasing and has a unique root.
Price differences below the arbitrage bound $d_t^s$ might persist as the arbitrageur prefers not to trade in such a scenario. More intuitive representations of the arbitrage bound can be derived by assuming that the arbitrageur is equipped with constant absolute or relative risk aversion. In particular, we follow Schneider (2015) in ignoring the impact of higher order moments above the fourth degree of the Taylor representation in Equation (8) and assume that the price process has a drift of $\mu_t^s$ of zero. These two additional assumptions yield an analytically tractable formulation of the arbitrage bound. The following lemma gives the analytical closed-form expression for $d_t^s$ under the assumption of a power utility function.

**Lemma 2.** If, in addition to Assumptions 1 and 2, the arbitrageur has an isoelastic utility function $U_\gamma(r) := \frac{(1+r)^{1-\gamma}}{1-\gamma}$ with risk aversion parameter $\gamma > 1$, the arbitrage bound for $\mu_t^s = 0$ is given by

$$d_t^s = \frac{1}{2} \sigma_t^s \sqrt{\gamma E_t(\tau) + \sqrt{\gamma^2 E_t(\tau)^2 + 2\gamma(\gamma + 1)(\gamma + 2)(V_t(\tau) + E_t(\tau)^2)}}.$$  \hfill (9)

**Proof.** See Appendix A. \hfill $\square$

Hence, $d_t^s$ positively depends on (i) the arbitrageur’s risk aversion, $\gamma$, (ii) the local volatility on the sell-side market, $\sigma_t^s$, (iii) the (conditionally) expected waiting time until settlement, $E_t(\tau)$, and (iv) the conditional variance of the waiting time, $V_t(\tau)$. We show in Appendix D that the special case of an exponential utility function with constant absolute risk aversion $\gamma$ yields a similarly tractable expression.

The arbitrage bound obviously depends on the arbitrageur’s risk aversion $\gamma$. Riskless arbitrage profits are not possible in the presence of settlement latency and therefore the arbitrage bound depends on the arbitrageur’s risk aversion $\gamma$. Accordingly, for a risk neutral arbitrageur, we have $d_t^* = 0$ and she would exploit any positive price difference $\delta_t^{b,s} > 0$. In this case, any price differences between the two markets should be absorbed immediately. Hence, in the absence of any other frictions, the existence of persistent price differences between two markets (which are not traded away) indicates that the markets are populated by risk averse arbitrageurs who do not exploit price differences below the threshold $d_t^s$. We thus denote the interval $[0, d_t^s]$ as a no-trade region, in which price differences between markets $b$ and $s$ are not exploited.\(^8\)

\(^8\)The risk aversion is associated with the arbitrageur’s attitude towards the risk of a single trade. Theoretically, repeatedly exploiting price differences may lead to a vanishing variance of the arbitrageurs’ aggregate returns which is equivalent to a contraction of the relevant bounds. From an empirical per-
3 Transaction Costs and Settlement Fees

Most markets demand trading fees that agents pay upon the execution of a trade. For instance, traders frequently pay fees as a percentage of the trading volume when they execute trades on centralized exchanges. Similarly, broker-dealers usually charge markups for the execution of trades in over-the-counter markets. Moreover, markets typically exhibit limited supply in the form of price-quantity schedules that agents are willing to trade, possibly leading to substantial price impacts for large trading quantities. To incorporate trading fees and liquidity effects into our framework, we make the following assumption.

**Assumption 4.** Trading the quantity \( q \geq 0 \) on market \( i \) exhibits proportional transaction costs such that the average per unit sell and buy quotes are

\[
B^i_t(q) = B^i_t \left(1 - \rho^{i,B}(q)\right), \\
A^i_t(q) = A^i_t \left(1 + \rho^{i,A}(q)\right),
\]

with \( \rho^{i,B}(q) \geq 0 \) and \( \rho^{i,A}(q) \geq 0 \), both monotonically increasing in \( q \).

The presence of transaction costs changes the objective function of the arbitrageur who focuses on maximizing returns net of transaction costs defined as

\[
\tilde{r}_{(t:t+\tau)}^{b,s} = b^{s}_{t+\tau} - b^{s}_t + \delta^{b,s}_t - \log \left( \frac{1 + \rho^{b,A}(q)}{1 - \rho^{s,B}(q)} \right) \\
= r^{b,s}_{(t:t+\tau)} - \log \left( \frac{1 + \rho^{b,A}(q)}{1 - \rho^{s,B}(q)} \right).
\]

From this expression immediately follows that transaction costs decrease the expected utility of the arbitrageur. A different interpretation of Equation (12) is that transaction costs only increase the instantaneous return required to make the arbitrageur indifferent between trading and staying idle. The following lemma summarizes the arbitrageur’s decision problem in the presence of transaction costs.

**Lemma 3.** Under assumptions 1 - 4, the arbitrageur prefers to trade a quantity \( q > 0 \)

spective, however, high autocorrelation in the resulting individual returns due to the latency questions the feasibility of such a law of large numbers. Moreover, competition among arbitrageurs or information transmission across markets can imply \( \mu^{s}_t < 0 \), which induces arbitrage bounds even for risk-neutral arbitrageurs (e.g., Voigt (2020)).
over staying idle if
\[ \delta_i^{b,s} - \log \left( \frac{1 + \rho_b^A(q)}{1 - \rho_s^B(q)} \right) > d_i^s. \] (13)

Proof. See Appendix A.

In addition, we introduce settlement fees that play a pivotal role in distributed ledger systems. Validators typically receive a reward for confirming transactions which (at least partly) comprises of fees that originators of transactions offer to potential validators. Since the information that can be added to the ledger at any point in time is usually limited, such fees aim to provide validators with incentives to prioritize the settlement of transactions that include a higher fee (e.g., Easley et al., 2019). By offering a higher fee, arbitrageurs can thus decrease the expected settlement latency they face. We extend our framework to incorporate such latency-reducing settlement fees as follows.

Assumption 5. A settlement fee \(f > 0\) implies a latency distribution \(\pi_t(\tau|f)\) that can be ordered in the sense that for \(\tilde{f} > f\), \(\pi_t(\tau|f)\) first-order stochastically dominates \(\pi_t(\tau|\tilde{f})\), i.e., \(P(\tau \leq x|\tilde{f}) > P(\tau \leq x|f)\) for all \(x \in \mathbb{R}_+\).

The ordering of latency distributions in Assumption 5 implies a lower CE of trading for \(\tilde{f} > f\).\(^9\) Denote by \(d_i^s(f)\) the arbitrage bound associated with the latency distribution \(\pi_t(\tau|f)\). Theorem 1 then implies that \(d_i^s(f) > d_i^s(\tilde{f})\), i.e., by paying a higher settlement fee, the arbitrageur can reduce the risk associated with settlement latency and becomes more likely to trade. For simplicity, we assume that \(d_i^s(f)\) is differentiable such that Assumption 5 implies \(\frac{\partial}{\partial f} d_i^s(f) < 0\).

While settlement fees reduce the latency, they are costly for the arbitrageur. Since the arbitrageur does not hold inventory of the asset on the buy-side market, she has to acquire the additional quantity \(f\) to spend it in the settlement process. In line with practical implementation in most systems, where cryptocurrencies are transferred, we assume that the arbitrageur has to pay the settlement fee in terms of the underlying asset. Given the transaction costs from above, the choice of \(f\) thus also affects the trading quantity \(q\). The following lemma characterizes the arbitrageur’s decision problem in the presence of transaction costs and settlement fees.

\(^9\)We refer to Hadar and Russell (1969) and Levy (1992) for an explicit analysis of the relation between stochastic dominance and expected utility.
Lemma 4. Under assumptions 1 - 5, the arbitrageur prefers to trade a quantity $q > 0$ and pay a settlement fee $f > 0$ over staying idle if

$$\delta_t^{b,s} - \log \left( \frac{1 + \rho^{b,A}(q + f)}{1 - \rho^{s,B}(q)} \right) > d_t^s(f).$$

(14)

Proof. See Appendix A. 

Trading a larger quantity might deliver higher total returns, but it comes at the cost of higher transaction costs on both the buy-side and sell-side market. Moreover, paying higher settlement fees leads to lower arbitrage bounds, but at the cost of additional transaction costs on the buy-side market. The arbitrageur’s trading decision thus features a trade-off between $q$ and $f$ with endogenous arbitrage bounds. Formally, the arbitrageur aims to maximize total returns

$$\max_{(q,f) \in \mathbb{R}_+^2} B_t^s (1 - \rho^{s,B}(q)) q - A_t^b (1 + \rho^{b,A}(q + f))(q + f)$$

(15)

subject to the constraint

$$\delta_t^{b,s} - \log \left( \frac{1 + \rho^{b,A}(q + f)}{1 - \rho^{s,B}(q)} \right) \geq d_t^s(f).$$

(16)

We characterize the arbitrageur’s optimal choice of trading quantities and settlement fees in the following lemma.

Lemma 5. A total return maximizing arbitrageur only pays a settlement fee $f^* > 0$ to trade a quantity $q^* > 0$ if the following necessary conditions are met:

$$\frac{1 - \rho^{s,B}(q^*)}{q^*} > \frac{\partial}{\partial q} \rho^{s,B}(q^*)$$

(17)

$$- \frac{\partial}{\partial f} d_t^s(f^*) > \frac{\partial}{\partial q} \rho^{s,B}(q^*) \frac{1}{1 + \rho^{s,B}(q^*)}.$$

(18)

Otherwise, the arbitrageur optimally sets $f^* = 0$. Moreover, a total return maximizing arbitrageur chooses trading quantities $q^* > 0$ and settlement fees $f^* \geq 0$ such that

$$\delta_t^{b,s} - \log \left( \frac{1 + \rho^{b,A}(q^* + f^*)}{1 - \rho^{s,B}(q^*)} \right) = d_t^s(f^*).$$

(19)

Proof. See Appendix A.
The first part of the lemma provides conditions for the choice of the settlement fee. According to Equation (17), the arbitrageur chooses a positive settlement fee as long as the marginal price impact for the trading quantity is below the average price impact. However, Equation (18) shows that the reduction of the arbitrage bound through a higher settlement fee must exceed the implied opportunity costs, i.e., the possible gain in selling a higher quantity. As a consequence, the arbitrageur tends to pay a higher settlement fee if the sell-side market is very liquid (keeping the marginal price impact low) and the settlement fee has a high impact on the arbitrage bound (i.e., reducing the latency and thus risk). If any of these two conditions is violated, the arbitrageur optimally chooses not to pay any settlement fee, but might still decide to trade.

The second part of the lemma states that the arbitrageur always chooses trading quantities and settlement fees such that the constraint in Equation (16) binds. If the constraint would not be binding, the arbitrageur could trade a larger quantity to increase her total returns at the expense of higher transaction costs.

4 Bitcoin Orderbook and Network Data

4.1 Bitcoin orderbook data

We gather orderbook information from the public application programming interfaces (APIs) of the 16 largest cryptocurrency exchanges that feature BTC versus USD trading.\textsuperscript{10} We retrieve all open buy and sell orders for the first 25 levels on a minute interval from January 1, 2018, to October 31, 2019. The granularity of our data yields detailed information on orderbook depth.\textsuperscript{11} Table 1 gives the corresponding exchanges and provides summary statistics of the underlying orderbook data of our sample period. We observe a strong heterogeneity of exchange-specific liquidity. For instance, whereas investors could have traded BTC versus USD at Coinbase Pro with an average spread of 0.45 USD, the average quoted spread at Gatecoin has been about 337 USD since January 2018. For most exchanges, however, the relative bid-ask spreads are comparable to those from equity markets such as Nasdaq or NYSE, where relative spreads range from 5 basis points (bp) for large firms to 38 bp for small firms (e.g., Brogaard et al., 2014).

\textsuperscript{10}Some exchanges do not feature fiat currencies. However, they allow trading BTC against Tether, a token that is backed by one USD for each token and trading close to par with USD.

\textsuperscript{11}To the best of our knowledge, none of these exchanges offers the opportunity to place hidden orders. Our data set thus reflects a real-time image of the available liquidity on each exchange.
Table 1: Descriptive Statistics of the Orderbook Sample

<table>
<thead>
<tr>
<th>Exchange</th>
<th>Orderbooks</th>
<th>Spread (USD)</th>
<th>Spread (bp)</th>
<th>Taker Fee</th>
<th>With. Fee</th>
<th>Conf.</th>
<th>Margin</th>
<th>Business</th>
</tr>
</thead>
<tbody>
<tr>
<td>Binance</td>
<td>941,399</td>
<td>2.61</td>
<td>3.29</td>
<td>0.10</td>
<td>0.00100</td>
<td>2</td>
<td>✓</td>
<td>x</td>
</tr>
<tr>
<td>Bitfinex</td>
<td>938,703</td>
<td>0.62</td>
<td>0.74</td>
<td>0.20</td>
<td>0.00080</td>
<td>3</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>bitFlyer</td>
<td>919,182</td>
<td>15.13</td>
<td>20.52</td>
<td>0.15</td>
<td>0.00080</td>
<td></td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Bitstamp</td>
<td>938,483</td>
<td>5.11</td>
<td>6.33</td>
<td>0.25</td>
<td>0.00000</td>
<td>3</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Bittrex</td>
<td>940,523</td>
<td>9.07</td>
<td>13.20</td>
<td>0.25</td>
<td>0.00000</td>
<td>2</td>
<td>x</td>
<td>✓</td>
</tr>
<tr>
<td>CEX.IO</td>
<td>936,378</td>
<td>11.73</td>
<td>15.07</td>
<td>0.25</td>
<td>0.00100</td>
<td>3</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Gate</td>
<td>907,874</td>
<td>81.24</td>
<td>90.92</td>
<td>0.20</td>
<td>0.00200</td>
<td>2</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Gatecoin</td>
<td>560,111</td>
<td>336.52</td>
<td>515.87</td>
<td>0.35</td>
<td>0.00060</td>
<td>6</td>
<td>x</td>
<td>✓</td>
</tr>
<tr>
<td>Coinbase Pro</td>
<td>941,539</td>
<td>0.45</td>
<td>0.54</td>
<td>0.30</td>
<td>0.00000</td>
<td>3</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Gemini</td>
<td>912,944</td>
<td>2.57</td>
<td>3.40</td>
<td>1.00</td>
<td>0.00200</td>
<td>3</td>
<td>x</td>
<td>✓</td>
</tr>
<tr>
<td>HitBTC</td>
<td>919,686</td>
<td>2.96</td>
<td>3.68</td>
<td>0.10</td>
<td>0.00085</td>
<td>2</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Kraken</td>
<td>936,970</td>
<td>2.63</td>
<td>3.24</td>
<td>0.26</td>
<td>0.00100</td>
<td>6</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Liqui</td>
<td>491,516</td>
<td>30.15</td>
<td>45.13</td>
<td>0.25</td>
<td></td>
<td></td>
<td>✓</td>
<td>x</td>
</tr>
<tr>
<td>Lykke</td>
<td>918,768</td>
<td>44.04</td>
<td>57.95</td>
<td>0.00</td>
<td>0.00050</td>
<td>3</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Poloniex</td>
<td>916,876</td>
<td>5.38</td>
<td>7.51</td>
<td>0.20</td>
<td></td>
<td>1</td>
<td>✓</td>
<td>x</td>
</tr>
<tr>
<td>xBTCe</td>
<td>887,289</td>
<td>13.34</td>
<td>17.87</td>
<td>0.25</td>
<td>0.00300</td>
<td>3</td>
<td>✓</td>
<td>x</td>
</tr>
</tbody>
</table>

Notes: This table reports descriptive statistics of orderbook data used in our study. We gather high-frequency orderbook information of 16 exchanges by accessing the public application programming interfaces (APIs) every minute. Orderbooks denotes the number of successfully retrieved orderbook snapshots between January 1, 2018 and October 31, 2019. Spread (USD) is the average quoted spread in USD, Spread (bp) is the average spread relative to the quoted best ask price (in basis points). Taker Fee are the associated trading fees in percentage points relative to the trading volume. With. Fee are the withdrawal fees in BTC. Conf. refers to the number of blocks that the exchange requires to consider incoming transactions as being valid. Empty cells indicate missing values. Margin refers to the existence of BTC shorting instruments at the exchange. Business indicates whether the exchange allows business accounts and hence access for institutional investors.

The exchanges also exhibit substantial heterogeneity in terms of trading-related characteristics. Taker fees range from 0% on Lykke to 1% on Gemini. Other potential transaction costs are withdrawal fees that have to be paid upon the transfer of BTC from the exchange to any other exchange or private wallet address. Exchanges charge up to 0.003 BTC for withdrawal requests, which corresponds to roughly 30 USD in prices as of April 2018, irrespective of the withdrawn amount. Furthermore, exchanges have different requirements with respect to the number of block confirmations before they proceed to process BTC deposits. For instance, Kraken requires that incoming transactions must be included in at least 6 blocks. The objective of these requirements is to reduce the possibility of an attack that aims at revoking previous transactions, i.e., a so-called ‘double-spending attack’. In such a scenario, a potential attacker has to alter all blocks containing the corresponding transaction. The probability that an attacker catches up with the honest chain decreases exponentially with the number of blocks the attacker has
to alter. For instance, in the case of a confirmation requirement of 10 blocks, the probability of a successful attack is less than 0.01% (5%), if the attacker has a share of 30% (10%) of the total available computing power (Nakamoto, 2008). As we discuss below, these requirements confront arbitrageurs with a mechanical increase in the settlement latency.

Finally, we collect information about two exchange characteristics that might help arbitrageurs to circumvent the exposure to settlement latency. On the one hand, some exchanges offer margin trading instruments which allow traders to short BTC and avoid settlement latency. However, such margin trading always comes at the cost of substantial collateral deposits which the exchanges control. On the other hand, some exchanges allow businesses to open an account which provides institutional investors who might have lower risk aversion with the opportunity to hold inventories and exploit price differences. Holding inventories at exchanges is costly though, since it is associated with continuous exposure to fluctuations in prices and exchange-specific default or hacking risks. However, as we demonstrate in Section 6, the mere presence of margin trading instruments or access for institutional investors is not a sufficient condition to offset the impact of settlement latency.

4.2 Bitcoin network data

We gather transaction-specific information from blockchain.com, a popular provider of Bitcoin network data. We download all blocks verified between January 1, 2018 and October 31, 2019 and 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, among other details, the fee (per byte) the initiator of the transaction offers validators to verify the transaction.\footnote{The fee per byte is more relevant than the total fee associated with a transaction as block sizes are limited in terms of bytes. In principle, a transaction can have multiple inputs and outputs, i.e., several addresses that are involved as senders or recipients of a transaction, which increases the number of bytes.}

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 until they are picked up by validators and get verified. The size of the mempool thus reflects the number of transactions that wait for confirmation. By design, the Bitcoin protocol restricts the number of transactions that can enter a single block. This restriction 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. Consequently, transactions with no or very low settlement fees may not attract validators and thus stay in the mempool until they become verified eventually.\textsuperscript{13}

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. By design of the Bitcoin protocol, validators successfully find a solution and append a block on average every 10 minutes (during our sample period, new blocks are announced to the network on average every 9.7 minutes). The time until verification, however, should not be confused with the time it takes until a new block is mined. Even though the expected block validation time is 10 minutes, it is ex-ante uncertain when a transaction is included in a block for the first time. The number of outstanding transactions serves as a proxy for fluctuations in congestion of the Bitcoin network. Whereas on average 1,644 transactions have been included per block in our sample period, the average number of transactions in the mempool is above 10,000 with temporarily more than 41,000 transactions waiting for verification. For any transaction this induces stochastic settlement latency. The probability of being included in the next block decreases with the number of transactions that wait for settlement and increases with the settlement fee the investor is willing to pay.

Table 2 provides summary statistics of the recorded transactions. The average settlement fee per transaction is about 2 USD. The distribution of fees exhibits a strong positive skewness with a median of 0.28 USD. The average waiting time until the verification of a transaction is about 41 minutes, while the median is about 8.8 minutes.

4.3 Price differences across markets

To provide systematic empirical evidence on the extent of potential arbitrage opportunities and thus violations of the law of one price, we compute the observed instantaneous cross-market price differences, adjusted for transaction costs, of all 120 exchange pairs (with the total number of exchanges \( N = 16 \)), defined as

\[
\tilde{\Delta}_t := \begin{pmatrix}
0 & \ldots & \tilde{\delta}_t^{N,1} \\
\vdots & \ddots & \vdots \\
\tilde{\delta}_t^{1,N} & \ldots & 0 \\
\end{pmatrix}
= \begin{pmatrix}
0 & \ldots & \tilde{b}_t^1 \left( q_t^{N,1} \right) - \tilde{a}_t^N \left( q_t^{N,1} \right) \\
\vdots & \ddots & \vdots \\
\tilde{b}_t^N \left( q_t^{1,N} \right) - \tilde{a}_t^1 \left( q_t^{1,N} \right) & \ldots & 0 \\
\end{pmatrix}, \quad (20)
\]

\textsuperscript{13}\text{Relaxing this artificial supply constraint might reduce issues pertaining to settlement latency but at the cost of reduced network security (see, e.g., Hinzen et al., 2019).}
Table 2: Descriptive Statistics of Transactions in the Bitcoin Network

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>SD</th>
<th>5 %</th>
<th>25 %</th>
<th>Median</th>
<th>75 %</th>
<th>95 %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fee per Byte (in Satoshi)</td>
<td>47.41</td>
<td>183.08</td>
<td>1.21</td>
<td>5.00</td>
<td>14.06</td>
<td>45.52</td>
<td>200.25</td>
</tr>
<tr>
<td>Fee per Transaction (in USD)</td>
<td>1.98</td>
<td>24.19</td>
<td>0.02</td>
<td>0.09</td>
<td>0.28</td>
<td>1.12</td>
<td>7.54</td>
</tr>
<tr>
<td>Latency (in Min)</td>
<td>41.03</td>
<td>289.26</td>
<td>0.73</td>
<td>3.55</td>
<td>8.82</td>
<td>20.75</td>
<td>109.52</td>
</tr>
<tr>
<td>Mempool Size (in Number)</td>
<td>10,018.74</td>
<td>14,876.52</td>
<td>432.00</td>
<td>1,812.00</td>
<td>4,503.50</td>
<td>11,057.50</td>
<td>41,884.50</td>
</tr>
<tr>
<td>Transaction Size (in Bytes)</td>
<td>507.28</td>
<td>2174.13</td>
<td>192.00</td>
<td>225.00</td>
<td>248.00</td>
<td>372.00</td>
<td>958.00</td>
</tr>
</tbody>
</table>

Notes: This table reports descriptive statistics of the Bitcoin transaction data used in our study. The sample contains all transactions settled in the Bitcoin network from January 1, 2018, to October 31, 2019. Our sample comprises 139,704,737 transactions that are verified in 99,129 blocks. 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). We 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 a 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.

where $\tilde{b}_i(q_{i,j}^t)$ is the transaction cost adjusted (log) sell price of $q_{i,j}^t$ units of the asset on exchange $i$ at time $t$ and $\tilde{a}_i(q_{i,j}^t)$ is the transaction cost adjusted (log) buy price of $q_{i,j}^t$ units of the asset. For analyses in which we abstract from transaction costs, we use the best bid and ask quotes of each exchange.

In line with our definition in Section 3, transaction costs are proportional to the trading quantity. We choose $q_{i,j}^t$ as the quantity that maximizes the resulting return for the exchange pair $i$ and $j$ given the prevailing orderbooks at time $t$, the taker fees of exchanges $i$ and $j$ and withdrawal fees of exchange $j$. Accordingly, we account for proportional exchange-specific taker fees (as reported in Table 1), which increase the average buy price and decrease the average sell price. We then use the resulting transaction cost adjusted orderbook queues and apply a grid search algorithm to identify the trading quantity that maximizes the total return for each exchange pair. As a last step, we check if the resulting trading quantity exceeds the withdrawal fee that the buy-side exchange charges for outgoing transactions (see Table 1). If the optimal trading quantity is below the withdrawal fee, we set the trading quantity to zero. This data-driven approach thus mimics the strategy of an arbitrageur who aims to maximize profits by optimally accounting for the prevailing orderbook depth and other trading-related fees. As price differences obviously can only be positive in one trading direction, we set negative price differences to zero as such scenarios (even without latency) do not correspond to arbitrage opportunities. The resulting matrix of price differences thus contains only non-negative values.
Figure 1: Price Differences over Time

Notes: This figure shows the daily average of price differences adjusted for transaction costs $\tilde{\delta}_{b,s,t}$, across all exchange pairs from January 1, 2018, to October 31, 2019. Price differences are based on minute-level transaction cost adjusted bid and ask quotes for each exchange according to Equation (20). We account for exchange-specific taker fees according to Table 1 and compute the quantity which maximizes the return for each exchange pair using a grid search algorithm. The shaded area indicates the 10% and 90% quantiles of price differences on a given day.

Figure 1 depicts the daily average of minute-level price differences based on optimal trading quantities according to Equation (20) across all exchange pairs. We observe a substantial variation over time. The average daily price difference across all exchange pairs is on average 33 bp. The 90% quantile is on average 129 bp, indicating a large dispersion of price differences through our sample period.

Figure 2 shows the average price differences for each exchange pair. The heatmap shows that some exchanges exhibit quotes that tend to deviate quite systematically from (nearly) all other exchanges. For instance, Bitfinex, CEX.IO, Gatecoin and HitBTC quote on average higher bid prices than most other exchanges and thus exhibit large price differences when used as a sell-side market. Conversely, other exchange pairs do not feature large average price differences. For instance, there are hardly any price differences whenever Coinbase Pro or Kraken serve as sell-side markets.
Figure 2: Price Differences between Exchanges

Notes: The heatmap shows the average price differences, adjusted for transaction costs, $\delta_{t,s}^{b,s}$, across time for each exchange pair in our sample. Price differences are based on minute-level transaction cost adjusted bids and asks for each exchange according to Equation (20). We account for exchange-specific taker fees according to Table 1 and compute the quantity which maximizes the return for each exchange pair using a grid search algorithm. The darker the color, the higher the average price difference through our sample period in the specific exchange pair. White or very light colors indicate that there are on average no or few price differences for a specific exchange pair.

5 Quantifying Arbitrage Bounds

5.1 Spot volatility estimation

To estimate the spot volatility, we follow the approach of Kristensen (2010). For each market $s$ and minute $t$, we estimate $(\sigma_t^s)^2$ by

$$\hat{\sigma}_t^2 \left( h_T \right) = \sum_{l=1}^{\infty} K \left( l - t, h_T \right) \left( b_l^s - b_{l-1}^s \right)^2,$$

where $K \left( l - t, h_T \right)$ is a one-sided Gaussian kernel smoother with bandwidth $h_T$ and $b_l^s$ corresponds to the quoted bid price on market $s$ at minute $l$. The choice of the bandwidth $h_T$ involves a trade-off between the variance and the bias of the estimator. Considering too many observations introduces a bias if the volatility is time-varying, whereas shrinking the estimation window through a lower bandwidth results in a higher variance of the
Kristensen (2010) thus proposes to choose \( h_T \) such that information on day \( T - 1 \) is used for the estimation on day \( T \). Formally, the bandwidth on any day of our sample is the result of minimizing the Integrated Squared Error (ISE) of estimates on the previous day, i.e.,

\[
h_T = \arg\min_{h>0} \sum_{l=1}^{1440} \left[ (b^*_l - b^*_T)^2 - (\widehat{\sigma}_T^2(h))^2 \right],
\]

where \( l \) refers to the minutes on day \( T - 1 \) and \( (\widehat{\sigma}_T^2(h)) \) is the spot variance estimator for minute \( l \) on day \( T - 1 \) based on bandwidth \( h \).

For each exchange, we trim the distribution of all estimates at 1% on both tails to eliminate outliers (e.g., due to flickering quotes). Figure 3 displays the cross-market average of spot volatility estimates on a daily basis. Since the underlying asset is identical, the resulting estimates—as expected—do not differ substantially across exchanges. The average minute-level volatility across exchanges is about 0.09%, which translates into a daily volatility of about 3.4%, significantly higher than the average daily volatility of the S&P 500 index during the same period, which yields roughly 0.65%.

### 5.2 Latency prediction

We use all verified transactions to parametrize the latency in the settlement process of the Bitcoin blockchain. In line with Chiu and Koeppl (2019) and Easley et al. (2019) we expect that transaction fees and mempool congestion play an important role in the determination of the expected time until verification. Accordingly, we employ a Gamma regression, where the conditional probability density function of latency \( \tau_i \) with rate parameter \( \beta_i \) and shape parameter \( \alpha_T \) is given by

\[
\pi(\tau_i|\theta_T) = \frac{\beta_i^{\alpha_T}}{\Gamma(\alpha_T)} \tau_i^{\alpha_T-1} e^{-\beta_i \tau_i},
\]

where

\[
\theta_T := (\theta_T^\beta, \alpha_T) \in \mathbb{R}^k \text{ and } \beta_i = \exp(-x_i^T \theta_T^\beta), \alpha_T > 0.
\]

---

14 We convert minute-level estimates to the daily level by multiplying it with the square root of the number of minutes on any given trading day, i.e., \( \sqrt{1440} \).
Notes: This figure shows the daily cross-market average of minute-level spot volatility estimates from January 1, 2018, to October 31, 2019. For each exchange, we estimate minute-level spot volatilities as the time-weighted average of squared bid price changes with a one-sided Gaussian kernel (Kristensen, 2010). For each day, we compute the average volatility across all exchanges. The shaded area corresponds to the range of average daily exchange-specific volatility estimates.

Here, $x_i \in \mathbb{R}^K$ includes an intercept and denotes (pre-determined) covariates driving $\tau_i$, $\theta^\beta_T \in \mathbb{R}^K$ denotes the corresponding vector of parameters and $\Gamma(x) := \int_{\mathbb{R}_+} z^{x-1} e^{-z} dz$ is the Gamma function. The Gamma distribution collapses to an exponential distribution for $\alpha_T = 1$. We estimate the parameter vector $\theta_T$ using all verified transactions on day $T - 1$ via maximum likelihood, both with and without covariates. In addition, we estimate an exponential model by fixing $\alpha_T = 1$. As covariates $x_i$ we include settlement fees and the (log) size of the mempool. The settlement fees enter as fees per byte as the relevant metric for validators who face a restriction in terms of the maximum size of a block in bytes. The number of transactions waiting for verification at the time when a transaction is announced serves as a proxy for competition among transactions.

In Table 3, we provide summary statistics of the estimated parameters. The numbers in the brackets denote the 5% and 95% quantiles of the time series of estimated parameters. The marginal effect of settlement fees is statistically significant and has the expected sign for nearly all days, i.e., higher fees predict a lower latency. The mempool size exhibits a positive impact on latencies through our sample period, i.e., congestion of the mempool decreases the probability of inclusion of a transaction in the next block.
(see, e.g., Huberman et al., 2017; Easley et al., 2019). A likelihood ratio test against a model without covariates indicates that the regressors are jointly highly significant. We therefore find clear evidence that the waiting time until a transaction enters the next block of the blockchain is predictable. We moreover find that the exponential distribution is rejected in favor of the more general Gamma distribution in nearly 93% of all days.

To predict the (conditional) moments of the latency distribution, while avoiding any look-ahead bias, we use the estimated parameter $\hat{\theta}_T$ based on transactions from day $T - 1$ to parameterize the latency distribution for every minute $t$ of day $T$. We provide further direct evidence for the predictability of settlement latency by computing the in-sample as well as out-of-sample root mean square prediction errors (MSPEs). In particular, for the in-sample MSPE, we use all transactions that feed into the estimation of $\hat{\theta}_T$ (i.e., all transactions verified on day $T - 1$). The out-of-sample MSPE is based on predictions for all transactions verified on day $T$ using the estimated parameter vector $\hat{\theta}_T$. We find that the in-sample MSPE is on average smaller for the unrestricted model specifications and that the unrestricted models exhibit on average a lower out-of-sample MSPE compared to their restricted counterparts. As a consequence, we predict the latency using the unrestricted Gamma model.

Accordingly, the conditional mean and variance of the latency $\tau$, induced by a transaction at minute $t$ on day $T$ with characteristics $x_t$, is given by

$$
\hat{\mathbb{E}}_t(\tau) = \hat{\alpha}_T \exp (x_t' \hat{\theta}_T^{\alpha}), \quad \text{and} \quad \hat{\mathbb{V}}_t(\tau) = \hat{\alpha}_T \exp (2x_t' \hat{\theta}_T^{\alpha}),
$$

(25)

where $x_t$ consists of the mempool size and the fee an arbitrageur is willing to pay at time $t$. While the mempool size is observable at any point in time, we use the optimal fee as derived in Lemma 5 as a proxy for the individually chosen settlement fees.

### 5.3 Estimation of arbitrage bounds

Based on the empirically relevant CRRA case of Lemma 2, the estimated arbitrage bounds $\hat{d}_t^s$ at minute $t$ are given by

$$
\hat{d}_t^s = \frac{1}{2} \hat{\sigma}_t^s \sqrt{\gamma m_1 + \sqrt{\gamma^2 m_1^2 + 2\gamma (\gamma + 1)(\gamma + 2)m_2}},
$$

(26)
## Table 3: Parameter Estimates for the Duration Models

<table>
<thead>
<tr>
<th></th>
<th>Exponential W/o Covariates</th>
<th>Exponential W/ Covariates</th>
<th>Gamma W/o Covariates</th>
<th>Gamma W/ Covariates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>3.31</td>
<td>1.41</td>
<td>3.86</td>
<td>1.19</td>
</tr>
<tr>
<td></td>
<td>[2.510, 4.246]</td>
<td>[-0.070, 3.675]</td>
<td>[2.626, 5.250]</td>
<td>[0.013, 2.596]</td>
</tr>
<tr>
<td>α</td>
<td>0.62</td>
<td>0.63</td>
<td>0.358</td>
<td>0.365</td>
</tr>
<tr>
<td>Fee per Byte</td>
<td>-0.22</td>
<td>-0.22</td>
<td>-0.486, -0.031</td>
<td>-0.501, -0.031</td>
</tr>
<tr>
<td>Mempool Size</td>
<td>0.23</td>
<td>0.31</td>
<td>[0.043, 0.452]</td>
<td>[0.059, 0.530]</td>
</tr>
</tbody>
</table>

|                      |                             |                           |                      |
| LR (Covariates)      | 91.33                       | 74.59                     |                      |
| LR (Gamma vs. Exponential) | 92.68                     |                           |                      |
| MSPE (In-Sample)     | 65.67                       | 65.74                     | 65.67                | 66.02               |
| MSPE (Out-of-Sample) | 70.97                       | 70.81                     | 70.97                | 70.55               |

**Notes:** This table reports summary statistics for the estimated parameters of the Gamma duration model given by Equation (23). Fee denotes fee per byte and Mempool Size refers to the number of unconfirmed transactions in the mempool. We estimate each model for each day in our sample, where we consider all transactions confirmed on a particular day. We report the time series averages of the estimated parameters. Values in brackets correspond to the 5% and 95% percent quantiles of the estimated parameters. LR (Covariates) summarizes likelihood ratio tests of the corresponding unrestricted duration model with covariates against the restricted model without covariates. LR (Gamma vs. Exponential) summarizes likelihood ratio tests of the Gamma duration model against the exponential specification. The reported values denote the percentage of days where the null hypothesis that the likelihood of the more general model equals the likelihood of the restricted model is rejected at the 95% significance level. MSPE refers to the mean squared prediction error for out-of-sample and in-sample tests, respectively.

with

\[
m_1 = \hat{\sigma}_t^s \cdot (B^s - 1),
\]

\[
m_2 = \hat{\sigma}_t^s \cdot (B^s - 1)^2 + \left( \hat{\sigma}_t^s \cdot (B^s - 1) + \hat{\sigma}_t^s \right)^2,
\]

where \( \hat{\sigma}_t^s \) denotes the square-root of the estimated spot volatility on the sell-side exchange, and \( \hat{\sigma}_t^s \) and \( \hat{\sigma}_t^s \) denote the estimated conditional mean and variance of the latency distribution, respectively. Moreover, \( B^s \) refers to the number of blocks that the sell-side exchange \( s \) requires to consider incoming transactions as valid (see Table 1). This exchange-specific security requirement thus further increases the settlement latency beyond the waiting time until a transaction’s validation in the first block.\(^{15}\)

\(^{15}\)bitFlyer and Liqui do not report a minimum number of confirmations. They rather use a discretionary system depending on the individual transaction and the state of the network. In this case, we assume the number of confirmations to be equal to the median across all exchanges that provide such information.
We thus decompose the latency into two components: the time it takes until a transaction is included in the blockchain (i.e., the first block), \( \tau \), and the additional time until exchanges accept the transaction as de facto being immutable. While \( \tau \) is partially under the control of the arbitrageur, the validation time of subsequent blocks is exogenous. In fact, we do not find evidence against non-zero autocorrelation in waiting times and constant volatility in the block validation time. This evidence supports the notion that the validation times of blocks are partially under control of the Bitcoin network and are internally impaired by the computational complexity of the underlying cryptographic problem. As a result, we can safely assume that the waiting times between subsequent blocks after the first one, which includes the current transaction, are independently and identically distributed. As validators append a new block on average every 9.7 minutes in our sample, we use this magnitude as the best-possible prediction of the time between two subsequent blocks, \( \bar{E}_t(\tau_B) \). Accordingly, \( \bar{V}_t(\tau_B) \) denotes the (sample) variance of the time between two consecutive blocks.

We fix the coefficient of risk aversion to \( \gamma = 2 \) and estimate \( \hat{d}_t \) for each exchange on a minute level.\(^{16}\) As shown by Figure 4, we observe substantial variation of these bounds over time. Arbitrage bounds are large especially during phases of high price volatility. We cannot reject the null hypothesis that the correlation between volatility and expected latency is significantly different from zero, which suggests that settlement latency constitutes a source of risk which is not captured by price fluctuations.

Table 4 gives summary statistics of the resulting time series of arbitrage bounds due to settlement latency. We observe that these bounds range, on average, between 91 bp and 197 bp.

While the conditional moments of the latency distribution affect the time series variation of the bounds, the cross-sectional variation is driven by the exchange-specific spot volatilities and the required number of confirmations, \( B^s \). For instance, Gatecoin and Kraken require \( B^s = 6 \) confirmations and produce on average the highest bounds, while Poloniex requires only \( B^s = 1 \) confirmation yielding the smallest median bound. To quantify the effect of the exchange-specific security component \( B^s \), we decompose the arbitrage bounds into the component resulting from the latency until a transaction is included in a block for the first time, \( \tau \), and the component resulting from the waiting time until a transaction fulfills exchange-specific security requirements, \( (B^s - 1)\tau_B \). The second

\(^{16}\)Our calibration follows Conine et al. (2017), who estimate an average coefficient of relative risk aversion of about 2 over an extensive sample period.
Figure 4: Estimated Arbitrage Bounds over Time

Notes: This figure shows the daily average estimated arbitrage bound based on a CRRA utility function with risk aversion parameter $\gamma = 2$ from January 1, 2018, to October 31, 2019. We estimate the bounds using spot volatility estimates following Kristensen (2010) and out-of-sample predictions of the conditional moments of the latency based on a Gamma duration model. The solid blue line shows the daily averages (in basis points) across all exchanges. The shaded area corresponds to the range of daily exchange-specific averages.

to last column in Table 4 gives the increase in the median arbitrage bound when we take the exchange-specific number of confirmations into account. The values correspond to the (percentage) difference between the median arbitrage bound as of Equation (26) and the respective bounds based on the assumption $B^s = 1$ for all exchanges. We observe that the impact of exchange-specific security components on arbitrage bounds is substantial and accounts on average for 23% of the bounds.

To shed more light on the implied costs of decentralized settlement under sufficiently high security standards, we quantify the relation between the level of security and the resulting latency. For each exchange, we compute arbitrage bounds for a given hypothetical number of confirmations and compare it to the baseline case of no additional security requirements (i.e., whenever the inclusion in the first upcoming block is sufficient). As a result, we find that requiring 10 confirmations at all exchanges (high security) would yield average arbitrage bounds of 175 bp, whereas the average bound when requiring only 1 confirmation (low security) would be 101 bp, an increase by more than 73%. This analysis thus shows how security in a distributed ledger translates into settlement latency.
Table 4: Summary of Exchange-Specific Arbitrage Bounds

<table>
<thead>
<tr>
<th>Security</th>
<th>Mean</th>
<th>SD</th>
<th>5%</th>
<th>25%</th>
<th>Median</th>
<th>75%</th>
<th>95%</th>
<th>Security</th>
<th>Uncertainty</th>
</tr>
</thead>
<tbody>
<tr>
<td>Binance</td>
<td>114.75</td>
<td>318.76</td>
<td>24.35</td>
<td>42.10</td>
<td>68.92</td>
<td>125.59</td>
<td>320.28</td>
<td>13.54</td>
<td>41.53</td>
</tr>
<tr>
<td>Bitfinex</td>
<td>117.22</td>
<td>299.25</td>
<td>18.89</td>
<td>42.47</td>
<td>73.26</td>
<td>136.19</td>
<td>324.19</td>
<td>23.98</td>
<td>40.85</td>
</tr>
<tr>
<td>bitFlyer</td>
<td>130.85</td>
<td>317.68</td>
<td>33.02</td>
<td>57.07</td>
<td>89.41</td>
<td>143.51</td>
<td>333.88</td>
<td>14.32</td>
<td>41.63</td>
</tr>
<tr>
<td>Bitstamp</td>
<td>126.34</td>
<td>294.72</td>
<td>28.45</td>
<td>50.53</td>
<td>80.46</td>
<td>145.61</td>
<td>341.72</td>
<td>23.69</td>
<td>40.79</td>
</tr>
<tr>
<td>Bittrex</td>
<td>129.03</td>
<td>277.80</td>
<td>30.94</td>
<td>57.25</td>
<td>89.41</td>
<td>143.51</td>
<td>333.88</td>
<td>14.32</td>
<td>40.79</td>
</tr>
<tr>
<td>CEX.IO</td>
<td>120.84</td>
<td>286.39</td>
<td>29.46</td>
<td>57.25</td>
<td>89.41</td>
<td>143.51</td>
<td>333.88</td>
<td>14.32</td>
<td>40.79</td>
</tr>
<tr>
<td>Gate</td>
<td>101.50</td>
<td>277.20</td>
<td>24.12</td>
<td>43.81</td>
<td>68.78</td>
<td>117.27</td>
<td>260.03</td>
<td>14.04</td>
<td>41.48</td>
</tr>
<tr>
<td>Gatecoin</td>
<td>196.89</td>
<td>219.90</td>
<td>2.62</td>
<td>46.70</td>
<td>81.69</td>
<td>136.05</td>
<td>305.50</td>
<td>24.44</td>
<td>40.60</td>
</tr>
<tr>
<td>Coinbase Pro</td>
<td>114.84</td>
<td>305.25</td>
<td>17.89</td>
<td>40.75</td>
<td>71.77</td>
<td>132.79</td>
<td>318.48</td>
<td>14.04</td>
<td>41.48</td>
</tr>
<tr>
<td>Gemini</td>
<td>115.36</td>
<td>343.30</td>
<td>21.07</td>
<td>43.27</td>
<td>72.42</td>
<td>130.54</td>
<td>309.53</td>
<td>24.44</td>
<td>40.77</td>
</tr>
<tr>
<td>HitBTC</td>
<td>101.22</td>
<td>287.97</td>
<td>19.10</td>
<td>37.64</td>
<td>62.72</td>
<td>112.79</td>
<td>273.14</td>
<td>14.14</td>
<td>41.36</td>
</tr>
<tr>
<td>Kraken</td>
<td>135.07</td>
<td>271.66</td>
<td>25.37</td>
<td>54.09</td>
<td>91.53</td>
<td>164.15</td>
<td>357.11</td>
<td>41.86</td>
<td>40.50</td>
</tr>
<tr>
<td>Liqui</td>
<td>90.79</td>
<td>60.20</td>
<td>23.51</td>
<td>49.96</td>
<td>77.40</td>
<td>115.62</td>
<td>201.88</td>
<td>28.97</td>
<td>39.98</td>
</tr>
<tr>
<td>Lykke</td>
<td>133.43</td>
<td>379.31</td>
<td>18.58</td>
<td>44.51</td>
<td>80.57</td>
<td>150.73</td>
<td>381.17</td>
<td>25.21</td>
<td>40.61</td>
</tr>
<tr>
<td>Poloniex</td>
<td>94.69</td>
<td>264.09</td>
<td>18.49</td>
<td>33.32</td>
<td>55.53</td>
<td>104.34</td>
<td>260.68</td>
<td>0.00</td>
<td>45.13</td>
</tr>
<tr>
<td>xBTCe</td>
<td>106.16</td>
<td>246.56</td>
<td>19.90</td>
<td>40.74</td>
<td>70.58</td>
<td>131.44</td>
<td>281.96</td>
<td>24.15</td>
<td>40.78</td>
</tr>
</tbody>
</table>

Notes: This table provides descriptive statistics of estimated arbitrage bounds for each sell-side market. We compute arbitrage bounds for a CRRA utility function with risk aversion parameter $\gamma = 2$. We estimate the bounds using the spot volatility estimator of Kristensen (2010) and out-of-sample predictions of the conditional moments of the latency based on a Gamma duration model. We report all values in basis points (except otherwise noted). Security gives the (percentage) contribution of the required number of confirmations to the median arbitrage boundary. Uncertainty corresponds to the (percentage) contribution of the uncertainty in latency to the median arbitrage boundary.

which in turn materializes into 7 bp of no-arbitrage regions per additional block.

Moreover, our theoretical framework allows us to directly analyze the relevance of the latency uncertainty. As the uncertainty of the arbitrageurs’ returns increases with the variance of the settlement latency, we can compare the estimated arbitrage bounds to the (hypothetical) case of a deterministic latency. The last column in Table 4 reports the percentage increase in arbitrage bounds when adjusting for the randomness in latency. The values correspond to the percentage difference between the median arbitrage bound and bounds based on the assumption $\mathbb{V}(\tau) = \mathbb{V}(\tau_B) = 0$. We find that the impact of the randomness in latency is substantial and accounts on average for 41% of the arbitrage bounds.

5.4 Price differences and latency-related arbitrage costs

To quantify to which extent observed cross-market price differences exceed the estimated arbitrage bounds and thus constitute potential arbitrage possibilities, we define the price
Figure 5: Price Differences in Excess of Arbitrage Bounds over Time

Notes: This figure shows daily average minute-level returns in excess of the estimated arbitrage bounds across all exchange pairs from January 1, 2018, to October 31, 2019. The solid red line corresponds to price differences based on the best bid and best ask of the individual exchange pairs, $\hat{E}_t$. The dashed blue line displays the corresponding excess price differences after adjusting for transaction costs, $\tilde{E}_t$.

Differences adjusted for transaction costs in excess of arbitrage bounds as

$$\tilde{E}_t := \Delta_t - \begin{pmatrix} \hat{a}_t^1 \\ \vdots \\ \hat{a}_t^N \end{pmatrix} \odot \Psi_t,$$

(29)

where the $(i, j)$-th element of $\Psi_t$ is defined as $\Psi_{t, i, j} = 1 \left\{ \frac{\tilde{b}_t^j(q_{t, i}^j)}{\hat{a}_t^j(q_{t, i}^j)} > \hat{b}_t^j \right\}$, $1 \{ \cdot \}$ is the indicator function, and $\odot$ corresponds to the element-wise multiplication operator.

Figure 5 plots the time series of cross-sectional daily average price differences in excess of the arbitrage bounds. The red solid line corresponds to price differences at the best bid and best ask not adjusted for transaction costs. The blue dashed line shows the corresponding excess price differences after adjusting for transaction costs. Taking transaction costs into account lowers the returns in excess of arbitrage bounds on average by 25%.

In our sample, we find that for a coefficient of relative risk aversion equal to 2, about 84% of observed price differences fall within the estimated bounds. After adjusting for
Figure 6: Excess Price Differences between Exchanges

<table>
<thead>
<tr>
<th>Sell-Side</th>
<th>Buy-Side</th>
</tr>
</thead>
<tbody>
<tr>
<td>Binance</td>
<td>Binance</td>
</tr>
<tr>
<td>Bitfinex</td>
<td>Bitfinex</td>
</tr>
<tr>
<td>Bitstamp</td>
<td>Bitstamp</td>
</tr>
<tr>
<td>Bittrex</td>
<td>Bittrex</td>
</tr>
<tr>
<td>CEX.IO</td>
<td>CEX.IO</td>
</tr>
<tr>
<td>Coinbase Pro</td>
<td>Coinbase Pro</td>
</tr>
<tr>
<td>Gate</td>
<td>Gate</td>
</tr>
<tr>
<td>Gatecoin</td>
<td>Gatecoin</td>
</tr>
<tr>
<td>Kraken</td>
<td>Kraken</td>
</tr>
<tr>
<td>HitBTC</td>
<td>HitBTC</td>
</tr>
<tr>
<td>Gemini</td>
<td>Gemini</td>
</tr>
<tr>
<td>Liqui</td>
<td>Liqui</td>
</tr>
<tr>
<td>Lykke</td>
<td>Lykke</td>
</tr>
<tr>
<td>Poloniex</td>
<td>Poloniex</td>
</tr>
<tr>
<td>bitFlyer</td>
<td>bitFlyer</td>
</tr>
<tr>
<td>Gatecoin</td>
<td>Gatecoin</td>
</tr>
<tr>
<td>xBTCe</td>
<td>xBTCe</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Mean (Excess) Price Differences (in bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
</tr>
<tr>
<td>20</td>
</tr>
<tr>
<td>40</td>
</tr>
<tr>
<td>60</td>
</tr>
</tbody>
</table>

Notes: This heatmap shows the average price differences adjusted for transaction cost, \( \delta_{b,s} \), in excess of the arbitrage bounds \( d_t^{ac} \) across time for each exchange pair in our sample. Price differences are based on minute-level transaction cost adjusted bid and ask quotes for each exchange according to Equation (20). We account for exchange-specific taker fees according to Table 1 and compute the quantity which maximizes the return for each exchange pair using a grid search algorithm. The darker the color, the larger the average price difference through our sample period in the specific exchange pair. White or very light colors indicate that there are on average no or few price differences for a specific exchange pair.

transaction costs, on average, 91% of all observed price differences are in the no-trade region. Therefore, the vast majority of cross-exchange price differences on the Bitcoin market do not constitute arbitrage opportunities, but are still too small to be traded away by rational arbitrageurs taking into account transaction costs and the risks due to settlement latency.

Observations outside of the arbitrage bounds might arise due to additional market frictions, which are not captured by our theory, for instance, exchange-specific risks or capital controls. As exchanges themselves are vulnerable to different sources of risk (e.g., default or hacking risk), the disutility of being exposed to a specific exchange may prevent arbitrageurs from exploiting certain cross-market trades, unless these trades provide additional compensation outside of our current framework. Furthermore, local regulation or access restrictions may deter cross-market arbitrage activity altogether (e.g., Choi et al. (2018)). For instance, US-based exchanges typically do not allow European citizens to open an account. To adjust for such exchange-specific effects, we decompose
Figure 7: Implied Coefficient of Relative Risk Aversion over Time

Notes: This figure shows the daily average implied risk aversion parameter, $\hat{\gamma}_t$, from January 1, 2018, to October 31, 2019. We compute $\hat{\gamma}_t$ as the smallest relative risk aversion such that all observed price differences adjusted for transaction costs fall within the implied limits to arbitrage.

Price differences in excess of arbitrage bounds on the exchange-pair level. Figure 6 shows a heatmap of average excess price differences. In analogy to Figure 2, some exchanges in our sample seem to persistently quote lower prices than others (e.g. Gatecoin), even after adjusting for transaction costs and latency-implied price risk.

Observations outside of arbitrage bounds may not only occur due to additional market frictions but are also consistent with higher risk aversion. As estimates of $\gamma$ in the asset pricing literature range from as little as 0.35 to as much as 9.0 (see, e.g., Hansen and Singleton, 1982; Chetty, 2006), plausible levels of relative risk aversion are hard to pin down. Therefore, in Figure 7, we display the implied relative risk aversion, $\hat{\gamma}_t$, corresponding to the lowest value of relative risk aversion which prevents all traders from exploiting the observed cross-exchange price differences.\textsuperscript{17} We observe that the implied minimum risk aversion exhibits substantial variation over time with an average of around 17. This relatively high level indicates that risk aversion cannot be the only reason for observing large cross-exchange price differences, but suggests the presence of further market frictions.

\textsuperscript{17}See Appendix E for more details on the construction of $\hat{\gamma}_t$. 

31
6 Arbitrage Bounds and Cross-Exchange Activity

The preceding analysis demonstrates that the derived arbitrage bounds can reconcile a large fraction of the observed cross-exchange price differences in our sample. These results, however, do not reveal to which extent the time variation in these differences can be attributed to variations in arbitrage bounds or other explanatory factors. Recall that the arbitrage bounds computed in Section 5 are derived under the assumption that there is no possibility to bypass the settlement latency. In this sense, they constitute an hypothetical benchmark of economic frictions induced by a pure decentralized settlement system.

In practice, however, modern cryptocurrency exchanges offer services which in principle allow traders to bypass settlement latency. On the one hand, margin trading instruments offered by the exchanges enable short-selling. More importantly, however, modern cryptocurrency exchanges effectively bypass the premise of trustless intermediation by operating internal settlement procedures. As a result, leading cryptocurrency exchanges do not rely on settlement via distributed ledgers but instead establish themselves as intermediaries. The benefit of this approach is that it enables fast trading and execution within the exchange orderbook. The drawback relative to a purely decentralized system is that every trader that interacts with the exchange has to accept the counterparty risk of the exchange and vice versa. As a consequence, these intermediation services require collateralization from the perspective of cryptocurrency exchange providers and trust into these centralized intermediaries from investors.

To analyze whether latency-induced arbitrage bounds are significant drivers of time variations in cross-exchange price differences and how this relationship is affected by the presence of centralized intermediation services, we compute the cross-sectional hourly average of differences between the price level on each sell-side exchange and prices on all other exchanges.

To quantify investors’ inventory holdings, we extract the number of Bitcoins under the control of wallets that the data provider glassnode associated with the sell-side exchange. These holdings serve as a proxy for the implied costs of trusting the exchange. If investors perceive the risks associated with exchanges as low, they should be willing to store more of their holdings directly under the custody of exchanges which allows fast execution within the exchange. The time series of inventories at exchanges exhibits an annualized aggregate growth of 13.4% which indicates increasing trust in cryptocurrency exchanges. At the end of our sample, 12.4 Billion USD worth of Bitcoin were under the custody
of cryptocurrency exchanges. Whereas some exchanges increased their inventory by large amounts (e.g., 400% at Coinbase, 141% at Binance, and 32% at Bitstamp), some of the competing exchanges face net inventory outflows in our sample period. For cross-market trading that involves the latter exchanges, settlement latency thus became more relevant due to higher implied costs of inventory holdings.

Table 5 gives the estimation results of linear regressions of hourly averages of cross-exchange price differences of sell-side exchanges on exchange-specific fixed effects and various regressors. In columns (1) and (2), we include the average estimated exchange-specific arbitrage bound or, alternatively, its individual components, i.e., the average hourly sell-side spot volatility, the hourly median and the variance of realized waiting times of transactions entering the mempool until being included in a block for the first time (where we rescale the variance to have a mean of zero and a standard deviation of one). Consistent with our theoretical framework, we find a statistically significant positive relation between price differences and arbitrage bounds. The marginal effect of arbitrage bounds is statistically and economically significant: a 1 bp increase in arbitrage bounds is on average associated with a 0.3 bp increase of price differences. Substituting the (pre-estimated) arbitrage bounds with their components confirms that large price differences are consistent with periods of high price risk due to settlement latency.

In columns (3) and (4), we interact the arbitrage bounds with sell-side exchange-specific dummy variables indicating whether the exchange offers margin trading instruments (Margin) and access for institutional traders (Business Accounts). We find that exchanges with margin trading are less sensitive to arbitrage bounds, but still exhibit a significant relation between price differences and arbitrage bounds. The costs of margin trading for investors thus seem to exceed the risk-adjusted latency-implied price risk, presumably due to substantial margin requirements by the exchanges as the absence of central clearing requires costly insurance against counterparty risk. Similarly, exchanges which feature access for institutional traders are less sensitive to arbitrage bounds, consistent with the notion that large institutions are more likely to exhibit a lower risk aversion than individual arbitrageurs.

In the last two columns of Table 5, we control for inventory holdings. As expected, we find that an increase of inventory holdings on an exchange reduces cross-market price differences as investors can quickly react to arbitrage opportunities without facing settlement latency. Most importantly, however, arbitrage bounds still remain statistically and economically significant drivers of cross-market price differences even when we control for inventory holdings on the exchanges. This result illustrates that the frictions
of a decentralized system cannot easily be resolved by reintroducing intermediation services. A possible reason is the traders’ lack of trust to store funds under the custody of an exchange. Our findings indicate also that this lack of trust is still too high or that intermediation services are still too underdeveloped in order to remove economically relevant deviations from the law of one price. Traders are obviously reluctant to accept the costs and risks implied by exchange-specific intermediation services in order to effectively bypass the market frictions implied by distributed ledger technologies.

Finally, in line with Roll et al. (2007), we find that trading costs, measured by the magnitude of bid-ask spreads, are an additional significant market friction that increase cross-market price differences.

In the last step of our analysis, we exploit the pre-estimated arbitrage bounds in order to shed some light on the relation between cross-market price differences and transfers of assets between exchanges. In fact, during periods of large price differences, i.e., in periods, where price differences likely exceed the arbitrage bounds, transfers of funds between exchanges should increase, consistent with the notion that arbitrageurs have to wait until the transfer of assets between wallets associated with the cryptocurrency exchanges is recorded on the blockchain. This should be particularly relevant if the premises of a blockchain-based market without any trusted intermediation services are fully internalized such that, e.g., inventory holdings on several exchanges are not an feasible option.

We hence extend our data by cross-exchange asset flows. Since exchanges are reluctant to provide the identity of their customers, it is virtually impossible to identify actual transactions by arbitrageurs. However, we take the overall transfer of assets between two different exchanges as a measure for the trading activity of cross-market arbitrageurs. For each exchange, we thus collect a list of addresses that are likely under the control of the exchanges in our sample.\textsuperscript{18} Bitcoin transactions are pseudonymous in the sense that each transaction publicly reveals all addresses associated with the transaction, but it is hard to map these addresses to their respective physical or legal owners. Exchanges typically control a large number of addresses to keep track of individual users’ assets. However, algorithms are available which link addresses to certain exchanges (e.g., Meiklejohn et al., 2013; Foley et al., 2019). Usually, the matching procedure is based on either having observed an address being advertised to belong to an exchange or by actively sending small amounts of Bitcoin to exchanges. We gather 62.6 million unique exchange addresses which allow us to identify 3.9 million cross-exchange transactions with an average daily

\textsuperscript{18}We thank Sergey Ivliev for his tremendous support on this front.
### Table 5: Price Differences and Sources of Price Risk

<table>
<thead>
<tr>
<th>Dependent Variable:</th>
<th>Price Differences</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
</tr>
<tr>
<td>Arbitrage Bound (in %)</td>
<td>0.307***</td>
</tr>
<tr>
<td></td>
<td>(15.98)</td>
</tr>
<tr>
<td>Spot Volatility (in %)</td>
<td>5.416***</td>
</tr>
<tr>
<td></td>
<td>(16.99)</td>
</tr>
<tr>
<td>Latency Median (in Min)</td>
<td>0.003***</td>
</tr>
<tr>
<td></td>
<td>(3.92)</td>
</tr>
<tr>
<td>Latency Variance</td>
<td>0.078***</td>
</tr>
<tr>
<td></td>
<td>(3.53)</td>
</tr>
<tr>
<td>Arbitrage Bound × Margin</td>
<td>-0.258***</td>
</tr>
<tr>
<td></td>
<td>(-7.07)</td>
</tr>
<tr>
<td>Arbitrage Bound × Business</td>
<td>-0.220***</td>
</tr>
<tr>
<td></td>
<td>(-5.38)</td>
</tr>
<tr>
<td>Inventory</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Spread (in %)</td>
<td>0.111***</td>
</tr>
<tr>
<td></td>
<td>(2.91)</td>
</tr>
<tr>
<td>Exchange Fixed Effects</td>
<td>Yes</td>
</tr>
<tr>
<td>Adjusted $R^2$</td>
<td>0.162</td>
</tr>
<tr>
<td>Exchange-Hour Observations</td>
<td>213,984</td>
</tr>
</tbody>
</table>

**Notes:** This table provides OLS estimates based on a regression of hourly average sell-side exchange-specific price differences and the main components of price risk due to stochastic settlement latency. *Price Differences* is the sell-side exchange-specific average hourly price difference from all other exchanges (in percent). *Spot Volatility* is the average hourly sell-side spot volatility estimate based on one-sided Gaussian kernel estimates (Kristensen, 2010). *Latency* denotes the hourly median (variance) of the waiting time of transactions entering the Bitcoin mempool, where we rescale the variance to have a mean of zero and a standard deviation of one. *Arbitrage Bound* corresponds to the average hourly sell-side exchange calibrated arbitrage bound. *Margin* is a dummy variable that indicates the availability of margin trading instruments and *Business Accounts* indicates whether exchanges offer access for institutional investors. *Inventory* is the number of Bitcoins controlled by all wallets associated with the sell-side exchange at hour $t - 1$. We compute *Spread* as the hourly sell-side exchange-specific average percentage spread. We report $t$-statistics based on heteroskedasticity-robust standard errors in parentheses. ***, **, and * indicate statistical significance on the 1%, 5% and 10% levels (two-tailed), respectively.

Volume of 72 million USD in our sample period.$^{19}$

Table 6 gives the estimates of a two-stage least squares regression of hourly cross-exchange flows on hourly averaged cross-market price differences as well as exchange-$^{19}$We compute the average daily volume by extracting the hourly sum of net flows (inflows to an exchange minus the outflows in BTC) and multiplying it by the hourly average midquote across all exchanges.
Table 6: Cross-Exchange Flows and Arbitrage Opportunities

<table>
<thead>
<tr>
<th>Dependent Variable:</th>
<th>Exchange Inflows (in 100k USD)</th>
<th>Log(Exchange Inflows)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>Price Differences (in %)</td>
<td>2.407*** (17.03)</td>
<td>2.525*** (15.09)</td>
</tr>
<tr>
<td>Spread (in %)</td>
<td>-0.355*** (-3.74)</td>
<td>-0.376*** (-3.73)</td>
</tr>
<tr>
<td>Exchange Fixed Effects</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Exchange-Hour Observations</td>
<td>213,984</td>
<td>213,984</td>
</tr>
</tbody>
</table>

Notes: This table provides the estimated marginal effects based on a two-stage least square regression of cross-exchange asset flows on price differences and bid-ask spreads. Inflows are the average hourly inflows (in BTC) to market $s$ from all other markets in our sample. Price Differences denote the price differences on sell-side market $s$ and are the fitted values of the regression outlined in Table 5 and denote price differences on sell-side market $s$. In columns (1) and (3), we instrument price differences with all components of arbitrage bounds. Columns (2) and (4) correspond to the estimation results where we directly use the estimated arbitrage bounds as an instrument. We compute Spread as the hourly sell-side exchange-specific average percentage spread. We report $t$-statistics based on heteroskedasticity-robust standard errors in parentheses. ***, **, and * indicate statistical significance at the 1%, 5% and 10% levels (two-tailed), respectively.

Specific fixed effects and the bid-ask spread as a proxy for trading costs. The dependent variable is the sum of cross-market flows into a given exchange per hour, which we use in both absolute numbers, as well as in logarithmic terms to reduce the impact of outliers.20

We have to take into account that cross-market asset flows and price differences are jointly determined, giving rise to a simultaneity problem. On the one hand, arbitrage activity is expected to increase with higher price differences (in excess of arbitrage bounds). On the other hand, price differences should decrease in response to arbitrage trades as arbitrageurs enforce adjustments towards the law of one price. We therefore instrument the price differences by the estimated arbitrage bounds (columns (2) and (4)) and, alternatively, by their respective components, i.e., the spot volatility, median settlement latency and variance of realized latencies (columns (1) and (3)). These variables satisfy the two necessary conditions for the validity as an instrument. First, we find a positive correlation between price differences and arbitrage bounds after controlling for other exogenous variables (see Table 5). Second, the only role arbitrage bounds play in influencing cross-market flows is through their effect on the endogenous price differences.

20Note that for any given hour and exchange, arbitrage opportunities involving a particular exchange might arise using the exchange as a sell-side market in some trades and as a buy-side market in some other trades. Therefore, to distinctly quantify the direction of flows, we compute the aggregate sum of cross-market flows into a given exchange.
Throughout all specifications, we find a significant positive relation between cross-exchange flows into an exchange and (instrumented) price differences: a one percentage point increase in price differences is on average associated with a 0.5% increase in asset flows into an exchange in a given hour. These results are robust when we control for bid-ask spreads, which are negatively related to inflows coming from other exchanges. The negative marginal effect of the bid-ask spread is consistent with the notion that higher transaction costs deter arbitrageurs’ activity. Hence, the regression results indicate that cross-exchange flows increase in response to larger price differences triggered by larger arbitrage bounds. This provides evidence for arbitrageurs chasing profitable arbitrage opportunities by actively transferring assets across markets.

7 Conclusions

Many market participants believe that distributed ledger technologies have the potential to radically transform the transfer of assets. Replacing trusted intermediaries and central clearing parties with decentralized consensus protocols may increase efficiency and security and lower transaction costs. However, a new friction emerges as the potential merits come at the cost of latency in the settlement process, which severely slows down the speed of trading. The inability to trade quickly implies limits to arbitrage as market participants cannot react sufficiently fast to exploit potential cross-exchange price differences. The fact that in pure blockchain-based markets arbitrageurs have to wait for the validation of a transaction before they can dispose of their position renders arbitrage trades inherently risky. This price risk increases with an exchange’s security standard, requiring not only one but several verifications through time-consuming consensus protocols.

We show that settlement latency – and the associated inability to trade quickly – implies limits to arbitrage as it is not worthwhile for risk-averse arbitrageurs to exploit cross-exchange price differences in periods with high volatility and long validation times. We formally derive the resulting no-trade price bounds for arbitrary concave utility functions and a general class of latency distributions.

Using data from the Bitcoin market in 2018 and 2019, we quantify arbitrage bounds of on average 121 bp. We interpret these bounds as the economic friction that arises in a distributed ledger system without any intermediation services that would allow traders to bypass settlement latency. In fact, we show that on average, 91% of observed Bitcoin cross-market price differences, adjusted for transaction costs, are within the corresponding arbitrage bounds. We document that these arbitrage bounds are significant drivers
of the time variation of cross-exchange price differences, which indicates that market participants perceive these restrictions. Deviations from the law of one price are hence particularly large during times of high latency-implied price risk. These results are supported by an analysis of cross-exchange transfers of assets. We collect a novel data set of exchange wallets to examine inventories and cross-exchange Bitcoin flows, and find that asset transfers between exchanges respond to price differences. We thus provide evidence that price differences are only actively explored by risk averse arbitrageurs if potential gains offset latency-implied price risk.

These results shed some new light on the inherent trade-off between costs and benefits of central clearing (via central counterparties) versus decentralized validation (via distributed ledger systems). While central clearing counterparties take on counterparty risk to guarantee instantaneous trading on non-settled positions, distributed ledger systems provide fast settlement and trustless intermediation. The degree of trustworthiness in blockchain-based markets, however, depends on the complexity in the validation process, which ultimately causes settlement latency and a slowdown of trading. We document that the economic costs of the latency-related trading frictions of decentralized settlement (in terms of arbitrage bounds) are substantial.

Our findings explain recent developments in blockchain-based markets where market operators exert considerable effort to circumvent these frictions by (re-)introducing intermediation services. In particular, trading venues act as custodians of traders’ funds and operate fast inter-exchange settlement networks. Both features aim at providing fast trading environments without the need to validate every single transaction on the blockchain. Exchanges hence take on counterparty risks and compete for the trust of market participants.

However, to put trading funds under the custody of an exchange and to rely on an exchange’s protection against counterparty risk by providing collateral requires trustworthiness of the exchange. Although we observe an increase in trust in exchanges, measured by the increase in funds under the custody of exchanges, our results indicate that intermediation services are still insufficiently utilized to exploit cross-exchange price differences. In fact, we demonstrate that the estimated arbitrage bounds remain statistically and economically significant driving forces of time-varying price differences, even when we control for exchange-specific inventory holdings and margin trading possibilities. These results indicate that circumventing distributed settlement via alternative strategies is not sufficiently pervasive to completely offset the impact of arbitrage bounds. A possible reason is a lack of trust in the capabilities of cryptocurrency exchanges to serve as central
counterparties.

This paper thus contributes to an ongoing debate on the organization of clearing on financial markets and the role of third-party intermediation for reliable settlement systems. Our analysis demonstrates that a decentralized system cannot easily replace central clearing. Removing the frictions (and costs) induced by third-party intermediation causes novel trading frictions with non-trivial implications for pricing. First, limits to arbitrage implied by settlement latency may harm price efficiency, as the lower activity of arbitrageurs reduces the information flow across markets. Second, deviations from the law of one price affect the pricing of securities, as risk neutral probabilities are not uniquely defined. Third, the implied costs of settlement latency depend on the design of the distributed ledgers and should influence the decision whether to migrate to a decentralized settlement system.
References


Appendix

A Proofs

Proof of Lemma 1. The proof of the lemma is an application of Equation (2.2) in Barndorff-Nielsen et al. (1982).

Proof of Theorem 1. First, note that the characteristic function in Lemma 1 yields the first moment \( \mu_r \) of the returns as

\[
E_t(r_{(t:t+\tau)}^{b,s}) = (-i) \frac{\partial}{\partial u} \varphi_r^{b,s}(u) \bigg|_{u=0} \]

\[
= \delta_t^{b,s} e^{iu\delta_t^{b,s} \tau} \left( \frac{iu\mu_t^s - \frac{1}{2}u^2(\sigma_t^s)^2}{u} \right) + e^{iu\delta_t^{b,s} \tau} \left( \frac{iu\mu_t^s - \frac{1}{2}u^2(\sigma_t^s)^2}{u} \right) \bigg|_{u=0}
\]

\[
= \delta_t^{b,s} + E_t(\tau)\mu_t^s,
\]

(A1)

since \( m_\tau(0) = 1 \) and \( m_\tau'(0) = E_t(\tau) \) by definition of the moment generating function.

In spirit of Arditti (1967) and Scott and Horvath (1980), we express the expected utility of the arbitrageur by a Taylor expansion which results in a function of the higher-order moments of the return distribution. A Taylor expansion of a general utility function \( U_\gamma(r) \) around the mean \( \mu_r \) yields

\[
U_\gamma(r_{(t:t+\tau)}^{b,s}) = \sum_{k=0}^{\infty} \frac{U_\gamma^{(k)}(\mu_r)}{k!} \left( r_{(t:t+\tau)}^{b,s} - \mu_r \right)^k,
\]

(A2)

where \( U_\gamma^{(k)}(\mu_r) := \frac{\partial^k}{\partial \mu_r^k} U_\gamma(\mu_r) \). Then, taking expectations yields

\[
E_t \left( U_\gamma \left( r_{(t:t+\tau)}^{b,s} \right) \right) = U_\gamma(\mu_r) + \sum_{k=2}^{\infty} \frac{U_\gamma^{(k)}(\mu_r)}{k!} E_t \left( \left( r_{(t:t+\tau)}^{b,s} - \mu_r \right)^k \right).
\]

(A3)

Following Markowitz (1952), we next consider a first-order Taylor expansion for the CE. We thus implicitly assume that the risk premium, \( \mu_r - CE \), is small and that higher-order moments vanish:

\[
E_t \left( U_\gamma \left( r_{(t:t+\tau)}^{b,s} \right) \right) = U_\gamma(CE) = U_\gamma(\mu_r) + U_\gamma'(\mu_r)(CE - \mu_r).
\]

(A4)
Moreover, the first-order Taylor expansion provides a convenient closed-form approximation of the certainty equivalent which is linear in the moments of the return distribution. We obtain the equation in the theorem by equating (A3) and (A4), plugging in (A1), and solving for $CE$.

**Proof of Lemma 2.** The proof follows directly from applying Theorem 1 together with the derivatives of the isoelastic utility function which yields

$$d_t^s - \frac{1}{2} \gamma \frac{\sigma_t^s}{d_t^s} E_t(\tau) \left( - \frac{1}{8} \frac{\gamma(\gamma + 1)(\gamma + 2)}{(d_t^s)^3} (\sigma_t^s)^4 E_t(\tau^2) \right) = 0. \quad (A5)$$

Details regarding the moment generating function of the returns are provided in Appendix D. Then, by Descartes’ rule of signs there is exactly one positive real root to the polynomial

$$(d_t^s)^4 - \frac{1}{2} \gamma (\sigma_t^s)^2 E_t(\tau) (d_t^s)^2 - \frac{1}{8} \gamma(\gamma + 1)(\gamma + 2)(\sigma_t^s)^4 E_t(\tau^2) = 0. \quad (A6)$$

All four solutions of the quartic polynomial are given by

$$d_t^s = \pm \frac{1}{\sqrt{2}} \sqrt{\frac{\gamma}{2} (\sigma_t^s)^2 E_t(\tau) \pm \sqrt{\frac{\gamma^2}{4} (\sigma_t^s)^4 E_t(\tau)^2 + \frac{\gamma(\gamma + 1)(\gamma + 2)}{2} (\sigma_t^s)^4 E_t(\tau^2)}}. \quad (A7)$$

However, since

$$\frac{\gamma}{2} (\sigma_t^s)^2 E_t(\tau) < \sqrt{\frac{\gamma^2}{4} (\sigma_t^s)^4 E_t(\tau)^2 + \frac{\gamma(\gamma + 1)(\gamma + 2)}{2} (\sigma_t^s)^4 E_t(\tau^2)} \quad (A8)$$

holds for all $\gamma > 0$, $\sigma_t^s > 0$ and $E_t(\tau^2) > 0$, the unique positive real root is given by the expression in the lemma.

**Proof of Lemma 3.** The Taylor representation of $U_\gamma(\tilde{r})$ yields for $\rho^* := \log \left( \frac{1 + \rho^{b,A}(q)}{1 - \rho^s(q)} \right)$:

$$E_t(U_\gamma(\tilde{r})) = \delta_t^{b,s} + E_t(\tau) \mu_t^s - \rho^*$$

$$+ \sum_{k=2}^{\infty} U_\gamma^{(k)} \left( \delta_t^{b,s} + E_t(\tau) \mu_t^s - \rho^* \right) \frac{1}{k!} E_t \left( \left( r_{(t:t+\tau)}^{b,s} - \rho^* - \delta_t^{b,s} - E_t(\tau) \mu_t^s \right)^k \right). \quad (A9)$$

Let $d_t^s$ be the arbitrage boundary (in absence of transaction costs) as defined in Equa-
tion (8). Then, $d_t^* + \ln \left( \frac{1+r_{i,t}^{b,A}(q)}{1-r_{i,t}^{s,B}(q)} \right)$ is a root of the function

$$
\tilde{F}(d) := d + \mathbb{E}_t(\tau)\mu_t^* - \rho^* + \sum_{k=2}^{\infty} \frac{U_t^{(k)}(d + \mathbb{E}_t(\tau)\mu_t^* - \rho^*)}{k!U_t^{(k)}(d + \mathbb{E}_t(\tau)\mu_t^* - \rho^*)} \mathbb{E}_t \left( \left( \tau^{b,s}_{(t+\tau)} - \rho^* - d - \mathbb{E}_t(\tau)\mu_t^* \right)^k \right).
$$

(A10)

Therefore, $\mathbb{E}_t(U_\gamma(\tilde{r}))$ is positive if and only if

$$
\delta_{t}^{b,s} > d_t^* + \ln \left( \frac{1+r_{i,t}^{b,A}(q)}{1-r_{i,t}^{s,B}(q)} \right).
$$

(A11)

Proof of Lemma 4. The proof directly follows from Lemma 3 and Theorem 1.

Proof of Lemma 5. We cast the arbitrageur’s optimization problem in terms of the Lagrangian

$$
\mathcal{L}(q,f;\xi) = B_t^*(1 - \rho^{s,B}(q))q + A_t^b(1 + \rho^{b,A}(q+f))(q+f) - \xi \left( d_t^*(f) - \delta_t^{b,s} + \log (1 + \rho^{b,A}(q)) - \log (1 - \rho^{s,B}(q)) \right)
$$

(A12)

and observe that the corresponding Karush-Kuhn-Tucker (KKT) conditions imply

$$
q = 0 \lor B_t^* \left( (1 - \rho^{s,B}(q)) - \rho^{s,B}(q)q \right)
$$

$$
- A_t^b \left( (1 + \rho^{b,A}(q+f)) + \rho^{b,A}(q+f)(q+f) \right)
$$

$$
- \xi \left( \frac{\rho^{b,A}(q+f) + \rho^{b,A}(q+f)}{1 + \rho^{b,A}(q+f)} - \frac{\rho^{s,B}(q)}{1 + \rho^{s,B}(q)} \right) = 0
$$

(A13)

$$
f = 0 \lor - A_t^b \left( (1 + \rho^{b,A}(q+f)) + \rho^{b,A}(q+f)(q+f) \right)
$$

$$
- \xi \left( \frac{d}{df}d_t^*(f) + \frac{\rho^{b,A}(q+f)}{1 + \rho^{b,A}(q+f)} \right) = 0
$$

(A14)

$$
\xi = 0 \lor d_t^*(f) - \delta_t^{b,s}
$$

$$
+ \log (1 + \rho^{b,A}(q+f)) - \log (1 - \rho^{s,B}(q)) = 0,
$$

(A15)
We first consider the case of $\xi = 0$. Conditions (A13) and (A14) now become
\[ q = 0 \lor B^*_i \left( (1 - \rho^{s,B}(q)) - \rho^{s,B'}(q)q \right) \]
\[ - A^*_b \left( (1 + \rho^{b,A}(q + f)) + \rho^{b,A'}(q + f)(q + f) \right) = 0 \]  
(A16)
\[ f = 0 \lor - A^*_b \left( (1 + \rho^{b,A}(q + f)) + \rho^{b,A'}(q + f)(q + f) \right) = 0 \]  
(A17)
which only holds if
\[ 1 + \rho^{b,A}(q + f) = -\rho^{b,A'}(q + f)(q + f). \]  
(A18)
Since $\rho^{b,A'}(q + f) > 0$ by Assumption 4, this cannot be the case for any $q > 0$ or $f > 0$. Also note that $\xi = q = f = 0$ implies a contradiction. Therefore, the constraint (16) cannot be slack at the optimum and there does not exist a candidate solution for $\xi = 0$.

Next, we turn to the analysis of $\xi > 0$. The simple case of $q = 0$ does not deliver any positive returns and it does not make sense for the arbitrageur to pay any fee $f > 0$. If anything, the arbitrageur would prefer not to trade at all, i.e., $q = f = 0$. We are left with the two interesting cases of $q > 0$.

For $f = 0$, the KKT conditions give the candidate solution $\{q_1, f_1, \xi_1\}$ as solutions to the system of equations
\[ B^*_i \left( (1 - \rho^{s,B}(q_1)) - \rho^{s,B'}(q_1)q_1 \right) - A^*_b \left( (1 + \rho^{b,A}(q_1)) + \rho^{b,A'}(q_1)(q_1) \right) \]
\[ -\xi_1 \left( \frac{\rho^{b,A'}(q_1)}{1 + \rho^{b,A}(q_1)} - \frac{\rho^{s,B'}(q_1)}{1 + \rho^{s,B}(q_1)} \right) = 0 \]  
(A19)
\[ d^*_i(f_1) - \delta^b_{t,s} + \log \left( 1 + \rho^{b,A}(q_1) \right) - \log \left( 1 - \rho^{s,B}(q_1) \right) = 0 \]  
(A20)
\[ f_1 = 0. \]  
(A21)
For $f > 0$, we can get the candidate solution $\{q_2, f_2, \xi_2\}$ as solutions to
\[
B^s_t \left( (1 - \rho_t^{s,B}(q_2)) - \rho_t^{s,B'}(q_2)q_2 \right) \\
-A^b_t \left( (1 + \rho_t^{b,A}(q_2 + f_2)) + \rho_t^{b,A'}(q_2 + f)(q_2 + f_2) \right) \\
-\xi \left( \frac{\rho_t^{b,A'}(q_2 + f_2)}{1 + \rho_t^{b,A}(q_2 + f_2)} - \frac{\rho_t^{s,B'}(q_2)}{1 + \rho_t^{s,B}(q_2)} \right) = 0 \quad (A22) \\
-A^b_t \left( (1 + \rho_t^{b,A}(q_2 + f_2)) + \rho_t^{b,A'}(q_2 + f)(q_2 + f_2) \right) \\
-\xi \left( \frac{d^2 q_2}{df^2} + \frac{\rho_t^{b,A'}(q_2 + f_2)}{1 + \rho_t^{b,A}(q_2 + f_2)} \right) = 0 \quad (A23) \\
d^2 d_1(f_2) - \delta^s_t + \log (1 + \rho_t^{b,A}(q_2 + f_2)) - \log (1 - \rho_t^{s,B}(q_2)) = 0. \quad (A24)
\]

However, combining (A22) and (A22) shows that the solutions are only admissible if
\[
\xi = \frac{B^s_t \left( (1 - \rho_t^{s,B}(q_2)) - \rho_t^{s,B'}(q_2)q_2 \right)}{\frac{1}{d^2 q_2} - \frac{\rho_t^{s,B'}(q_2)}{1 + \rho_t^{s,B}(q_2)}} > 0. \quad (A25)
\]

Equation (A25) now provides us with necessary conditions for a solution to the problem that entails a strictly positive settlement fee. Namely, $q_2 > 0$, $f_2 > 0$, $\xi_2 > 0$ can only be solution if one of the following two conditions holds
\[
(i) \quad -\frac{d^2 q_2}{df^2} > \frac{\rho_t^{s,B'}(q_2)}{1 + \rho_t^{s,B}(q_2)} \quad \text{and} \quad 1 - \rho_t^{s,B}(q_2) > \rho_t^{s,B'}(q_2)q_2
\]
\[
(ii) \quad -\frac{d^2 q_2}{df^2} < \frac{\rho_t^{s,B'}(q_2)}{1 + \rho_t^{s,B}(q_2)} \quad \text{and} \quad 1 - \rho_t^{s,B}(q_2) < \rho_t^{s,B'}(q_2)q_2.
\]

However, condition (ii) cannot hold at the maximum since $1 - \rho_t^{s,B}(q_2) < \rho_t^{s,B'}(q_2)q_2$ means that the trading quantity is such that the marginal price impact exceeds the average price impact. In this case, the arbitrageur would reduce the trading quantity to raise her total return. Consequently, (i) remains as the necessary condition for a candidate solution with a positive settlement fee which completes the proof. 

\[\square\]

**B Latency Distribution under Stochastic Volatility**

We can relax the assumption that $\sigma_t^s$ is constant over the interval $[t, t+\tau]$ by allowing $\sigma_t^s$ to vary over time. More specifically, let $\sigma_t^s : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ with $\theta(t) := \int_t^{t+\tau} (\sigma_k^s)^2 \, dk < \infty \quad \forall \tau$, i.e., the volatility of the sell-side market follows a (deterministic) path with bounded integrated variance. Assuming $\mu_t^s = 0$, we can then rewrite the log returns of the arbitrageur
for given latency \( \tau \) as
\[
  r_{(t:t+\tau)}^{b,s} = \delta_t^{b,s} + \int_t^{t+\tau} \sigma_s^k dW_k^s. \tag{B1}
\]

The integral above corresponds to a Gaussian process with independent increments. More specifically, we get
\[
  \mathbb{E}_t \left( \left( r_{(t:t+\tau)}^{b,s} - \delta_t^{b,s} \right)^2 \right) = \theta(\tau) - \theta(0) = \mathbb{E}_t \left( W_s^{\theta(\tau)} - W_s^{\theta(0)} \right). \tag{B2}
\]

In other words, the time-changed Brownian motion \( W_s^{\theta(\tau)} \) has the same distribution as the log returns given in Equation (B1) (e.g., Durrett, 1984; Barndorff-Nielsen et al., 2002). We can thus rewrite the return process as
\[
  r_{(t:t+\tau)}^{b,s} = \delta_t^{b,s} + \int_t^{t+\tau} \theta(\tau) dW_k^s. \tag{B3}
\]

The implications of Lemma 1 still hold, but we need to compute the moment generating function of the transformed latency \( m_{\theta(\tau)}(u) \), which depends on the latency distribution and the dynamics of the volatility process. First, note that, as \( \theta(\tau) \) is strictly increasing, the probability integral transformation yields the distribution of \( \tau(\theta) \),
\[
  \mathbb{P}_t \left( \theta(\tau) = y \right) = \mathbb{P}_t \left( \tau = \theta^{-1}(y) \right) \quad \forall y > 0. \tag{B4}
\]

Finally, the distribution of \( \theta(\tau) \) is fully described via its characteristic function which is of the form
\[
  \varphi_{\theta(\tau)}(u) = \mathbb{E}_t \left( e^{i\theta(\tau)u} \right) = \frac{1}{2\pi} \int_0^\infty \int_{-\infty}^\infty \varphi_\tau(s) e^{-is\tau} ds e^{i\theta(\tau)u} d\tau. \tag{B5}
\]

Lévy’s characterization allows to extend these ideas to more general non-deterministic integrands and to stochastic time-changes. Although Equation (B5) allows to derive theoretical arbitrage bounds based on Theorem 1 for every continuous local martingale, we restrict our analysis to analytically more tractable and intuitive dynamics of the price process and the associated settlement latency.
C Return Distribution for Exponentially-Distributed Latency

To provide an illustrative example, we parameterize the probability distribution of the settlement latency as an exponential distribution with locally-constant scale parameter $\lambda_t := \lambda (I_t)$. The probability density function of the latency is then given by

$$\pi_t (\tau) = \lambda_t e^{-\lambda_t \tau}, \quad (C1)$$

with conditional mean $E_t (\tau) = \lambda_t^{-1}$ and conditional variance $V_t (\tau) = \lambda_t^{-2}$. The moment generating function of the exponential distribution is

$$m_{\tau} (u) = \frac{1}{1 - \lambda_t u} \quad \text{(e.g., Kotz et al., 2012)}.$$ 

Without a drift ($\mu_s t = 0$), the distribution collapses to a symmetric Laplace distribution with location parameter $\delta^{b,s}_t$, scale parameter $\sigma_t \sqrt{2 \lambda_t}$, and corresponding probability density function

$$\pi_t \left( r^{b,s}_{(t:t+\tau)} \right) = \frac{\sqrt{2 \lambda_t}}{2 \sigma_t} \exp \left( -\frac{\sqrt{2 \lambda_t}}{\sigma_t} \left| r^{b,s}_{(t:t+\tau)} - \delta^{b,s}_t \right| \right), \quad (C3)$$

with $E_t \left( r^{b,s}_{(t:t+\tau)} \right) = \delta^{b,s}_t$ and $V_t \left( r^{b,s}_{(t:t+\tau)} \right) = (\sigma_t)^2 E_t (\tau)$. Hence, not surprisingly, in the absence of a drift in the underlying Brownian motion, the (conditionally) expected return implied by the arbitrage strategy is equal to the instantaneous return $b^s_t - a^b_t$. The (conditional) variance equals the (locally constant) spot variance on market $s$, $(\sigma_t^s)^2$, scaled by the (conditional) expected waiting time until the settlement of the transaction, $\lambda_t^{-1}$. Hence, the higher the volatility on the sell-side market or the longer the expected waiting time until the transfer is settled, the higher is the risk of extreme adverse price movements.

Figure C1 provides a graphical illustration of the resulting distribution. The plot shows simulated draws from a Brownian motion stopped at randomly sampled waiting times. The marginal distribution at the top of the figure illustrates the exponential distribution of the waiting times. The marginal distribution on the right-hand side shows the resulting sampling distribution of the price process which converges in the limit to a Laplace distribution. The figure shows the resulting asymmetric Laplace distribution for a price process with a negative drift, whereas a price process without drift would yield a
Figure C1: Return Distribution under Exponentially-Distributed Latency.

Notes: This figure illustrates the impact of stochastic latency (horizontal axis) on the distribution of returns (vertical axis) if log prices follow a Brownian motion and if latencies are exponentially distributed. The individual paths correspond to sample draws of the price process and the dots correspond to the terminal value of the stopped Wiener process. The marginal distribution on the top corresponds to the sampled latencies. The marginal distribution on the right-hand side corresponds to the sampled distribution of returns which converges in the limit to a Laplace distribution. The figure shows the resulting distributions for a price process with negative drift $\mu_s^* < 0$.

D Arbitrage Bound under Constant Absolute Risk Aversion

We provide a further application of our main result to the case of the commonly-used utility function with constant absolute risk aversion (CARA). Again, we ignore the impact of higher order moments above the fourth degree of the Taylor representation in Equation (7). These assumptions yield an analytically tractable formulation of the arbitrage bound.

Lemma 6. If, in addition to Assumptions 1 and 2, the arbitrageur has an exponential utility function $U_\gamma(r) := \frac{1-e^{-\gamma(1+r)}}{\gamma}$ with risk aversion $\gamma > 0$, then the arbitrage boundary
Thus, we have

\[
\mu_t^* = -\mathbb{E}_t (\tau) \mu \mu_t^* + \frac{\gamma}{2} \left( \mathbb{V}_t (\tau) (\mu_t^*)^2 + (\sigma_t^*)^2 \mathbb{E}_t (\tau) \right)
- \frac{\gamma^2}{6} \left( 3 \mu_t^* (\sigma_t^*)^2 \mathbb{V}_t (\tau) + (\mu_t^*)^3 \mathbb{E}_t ((\tau - \mathbb{E}_t (\tau))^3) \right)
+ \frac{\gamma^3}{24} \left( (\mu_t^*)^4 \mathbb{E}_t ((\tau - \mathbb{E}_t (\tau))^4) + 6 (\sigma_t^*)^2 (\mu_t^*)^2 (\mathbb{E}_t (\tau)^3 + \mathbb{E}_t (\tau)^3) - 2 \mathbb{E}_t (\tau)^2) \right)
+ \frac{\gamma^3}{8} \mathbb{E}_t (\tau^2) (\sigma_t^*)^4.
\]  

(D1)

Proof. For the exponential utility, we have

\[
U^{(k)} (r) / U' (r) = (-\gamma)^{k-1} \text{ for } k \geq 1.
\]

Therefore, from Theorem 1 we have

\[
CE = \delta_t^{b,s} + \mathbb{E}_t (\tau) \mu_t^* - \frac{\gamma}{2} \mu_{t,t+r}^{b,s} (2)
+ \frac{\gamma^2}{6} \mu_{t,t+r}^{b,s} (3) - \frac{\gamma^3}{24} \mu_{t,t+r}^{b,s} (4) + O (r),
\]

(D2)

where \( \mu_{t,t+r}^{b,s} (k) := \mathbb{E}_t \left( \left( r_{(t+t+r)}^{b,s} - \delta_t^{b,s} - \mathbb{E}_t (\tau) \mu_t^* \right)^k \right) \) is the \( k \)-th order central moment of the returns and \( O (r) \) corresponds to the Taylor approximation error which we neglect subsequently. Recognizing that by definition \( m_{r_{(t+t+r)}}^{b,s} (iu) = \varphi_{r_{(t+t+r)}}^{b,s} (u) \), we can derive the moment generating function of the returns given by

\[
m_{r_{(t+t+r)}}^{b,s} (u) = e^{u \delta_t^{b,s}} \mu_t^* \left( u \mu_t^* + \frac{1}{2} u^2 (\sigma_t^*)^2 \right).
\]

(D3)

The central moment generating function is defined as

\[
C_{r_{(t+t+r)}}^{b,s} (u) = \mathbb{E}_t \left( \exp \left( u \left( r_{(t+t+r)}^{b,s} - \mathbb{E}_t (r_{(t+t+r)}^{b,s}) \right) \right) \right)
= \exp \left( -u \mathbb{E}_t (r_{(t+t+r)}^{b,s}) \right) m_{r_{(t+t+r)}}^{b,s} (u).
\]

(D4)

Thus, we have

\[
\mu_{r_{(t+t+r)}}^{b,s} (k) = \frac{\partial^k}{\partial u^k} C_{r_{(t+t+r)}}^{b,s} (u) \bigg|_{u=0}
= \frac{\partial^k}{\partial u^k} \exp (- \mathbb{E}_t (\tau) \mu_t^* u) m_{\tau} \left( u \mu_t^* + \frac{1}{2} u^2 (\sigma_t^*)^2 \right) \bigg|_{u=0}.
\]

(D5)
Basic calculus then yields

$$\mu_{t,(t+\tau)}^{b,s} (2) = \nabla_t (\tau) (\mu_t^s)^2 + (\sigma_t^s)^2 \mathbb{E}_t (\tau)$$  \hspace{1cm} (D6)

$$\mu_{t,(t+\tau)}^{b,s} (3) = 3 \mu_t^s (\sigma_t^s)^2 \nabla_t (\tau) + (\mu_t^s)^3 \mathbb{E}_t ((\tau - \mathbb{E}_t (\tau))^3)$$  \hspace{1cm} (D7)

$$\mu_{t,(t+\tau)}^{b,s} (4) = (\mu_t^s)^4 \mathbb{E}_t (((\tau - \mathbb{E}_t (\tau))^4) + 3 \mathbb{E}_t (\tau^2) (\sigma_t^s)^4$$

$$+ 6 (\sigma_t^s)^2 (\mu_t^s)^2 (\mathbb{E}_t (\tau)^3 + \mathbb{E}_t (\tau^3) - 2 \mathbb{E}_t (\tau) \mathbb{E}_t (\tau^2)).$$  \hspace{1cm} (D8)

Then, we plug in equations (D6)-(D8) into (D2). Finally, recognizing that the arbitrageur exploits price differences if and only if $CE > 0$, we can solve for the minimum instantaneous price differences $\delta_t^{b,s}$ which completes the proof.

In the absence of a drift ($\mu_t^s = 0$), the arbitrage boundary of Lemma 6 further simplifies to

$$d_t^s = \frac{\gamma}{2} (\sigma_t^s)^2 \mathbb{E}_t (\tau) + \frac{\gamma^3}{8} (\sigma_t^s)^4 \left( \nabla_t (\tau) + \mathbb{E}_t (\tau)^2 \right).$$  \hspace{1cm} (D9)

Just like in the case of CRRA, the arbitrage boundary $d_t^s$ positively depends on (i) the arbitrageur’s risk aversion, (ii) the local volatility on the sell-side market $s$, (iii) the expected waiting time until settlement, and (iv) the variance of the waiting time, $\nabla_t (\tau)$.

**E  No-Arbitrage Implied Relative Risk Aversion**

We compute the implied relative risk aversion $\hat{\gamma}_{t}^{b,s}$ such that all observed price differences of exchange pair $\{b,s\}$ at time $t$ are located within the implied limits to arbitrage. The interpretation of $\hat{\gamma}_{t}^{b,s}$ is straightforward: if the risk aversion of an arbitrageur is below $\hat{\gamma}_{t}^{b,s}$, it would be rational to trade. We compute $\hat{\gamma}_{t}^{b,s}$ according to the following lemma.

**Lemma 7.** Define $\hat{\gamma}_{t}^{b,s}$ as the root of the cubic polynomial

$$\left( \tilde{\delta}_t^{b,s} \right)^4 - \frac{1}{8} (\tilde{\sigma}_t^s)^4 c_2 \left( \hat{\gamma}_{t}^{b,s} \right)^3 - \frac{3}{8} (\tilde{\sigma}_t^s)^4 c_2 \left( \hat{\gamma}_{t}^{b,s} \right)^2$$

$$- \frac{1}{2} (\tilde{\sigma}_t^s)^2 \left( c_1 \left( \delta_t^{b,s} \right)^2 + \frac{1}{2} (\tilde{\sigma}_t^s)^2 c_2 \right) \hat{\gamma}_{t}^{b,s} = 0,$$  \hspace{1cm} (E1)

where, analogously to Equations (27) and (28), $c_1 = \hat{\nabla}_t (\tau) + \hat{\mathbb{E}} (\tau_B) \cdot (B^s - 1)$ and $c_2 = \hat{\nabla}_t (\tau) + \hat{\mathbb{E}} (\tau_B) \cdot (B^s - 1)^2 + \left( \hat{\mathbb{E}} (\tau_B) \cdot (B^s - 1) + \hat{\mathbb{E}} (\tau) \right)^2$. Then, price differences (adjusted for transaction costs) $\tilde{\delta}_t^{b,s}$ constitute a (statistical) arbitrage opportunity for an arbitrageur with risk aversion $\gamma$ only if $\gamma < \hat{\gamma}_{t}^{b,s}$.
Proof. The proof follows directly from applying Theorem 1 together with the derivatives of the utility function which yields

\[ d^*_t - \frac{1}{2} \frac{\gamma}{d^*_t} (\sigma^*_t)^2 \mathbb{E}_t(\tau) - \frac{1}{8} \frac{\gamma(\gamma + 1)(\gamma + 2)}{(d^*_t)^3} (\sigma^*_t)^4 \mathbb{E}_t(\tau^2) = 0. \]  
(E2)

Then, by Descartes’ rule of signs there is exactly one positive real root to the polynomial

\[ (d^*_t)^4 - \frac{1}{2} \gamma (\sigma^*_t)^2 \mathbb{E}_t(\tau) (d^*_t)^2 - \frac{1}{8} \gamma(\gamma + 1)(\gamma + 2) (\sigma^*_t)^4 \mathbb{E}_t(\tau^2) = 0. \]  
(E3)

By definition, \( d^*_t \) corresponds to the arbitrage boundary for a given risk aversion \( \gamma \). The arbitrageur prefers to trade if observed price differences \( \tilde{\delta}_b,s^*_t \) exceed the boundary. Therefore, rewriting Equation (E3) in terms of \( \gamma \) and replacing \( d^*_t \) with \( \tilde{\delta}_b,s^*_t \) yields a cubic polynomial in \( \gamma \):

\[ \left( \tilde{\delta}_b,s^*_t \right)^4 - \frac{1}{8} (\hat{\sigma}^*_t)^4 \mathbb{E}_t(\tau^2) \left( \tilde{\gamma}_b,s^*_t \right)^3 - \frac{3}{8} (\hat{\sigma}^*_t)^4 \mathbb{E}_t(\tau^2) \left( \tilde{\gamma}_b,s^*_t \right)^2 \]
\[ - \frac{1}{2} (\hat{\sigma}^*_t)^2 \left( \mathbb{E}_t(\tau) \left( \tilde{\delta}_b,s^*_t \right)^2 + \frac{1}{2} (\hat{\sigma}^*_t)^2 \mathbb{E}_t(\tau^2) \right) \tilde{\gamma}_b,s^*_t = 0 \]  
(E4)

Replacing the (conditional) expected latencies with the values given by Equations (27) and (28) completes the proof. \( \square \)

The exchange-pair specific implied risk aversion \( \hat{\gamma}_b,s^*_t \) is defined in a way such that the observed price differences \( \tilde{\delta}_b,s^*_t \), adjusted for transaction costs, coincide with the arbitrage bounds for an isoelastic utility function with risk aversion parameter \( \hat{\gamma}_b,s^*_t \). As the arbitrage bounds monotonically increase with risk aversion, any value of \( \gamma < \hat{\gamma}_b,s^*_t \) constitutes a trading opportunity for the arbitrageur. Conversely, \( \gamma > \hat{\gamma}_b,s^*_t \) reflects that the observed price differences do not justify (unconstrained) trading because an arbitrageur with a higher risk aversion obtains higher (expected) utility by trading less or not at all. As the asset is traded on \( N \) markets, we define \( \hat{\gamma}_b,s^*_t \) as the smallest risk aversion parameter for which all observed price differences fall within the implied arbitrage bounds, i.e.,

\[ \hat{\gamma}_b,s^*_t := \max_{i,j \in \{1,...,N\}} \hat{\gamma}^{ij}_t. \]  
(E5)