PREDICTING LIQUIDITY IN THE CRYPTOCURRENCY MARKET

TESTING THE INVARIANCE THEORY ON A NEW MARKET STRUCTURE AND ASSET CLASS

CHARLES GYLLHAMN
JACOB WINBERG

Bachelor Thesis

Stockholm School of Economics

2022

Predicting Liquidity In The Cryptocurrency Market: Testing The Invariance Theory On A New Market Structure And Asset Class

Abstract:

By integrating dimensional analysis and principles of market microstructure invariance, this study documents a nearly invariant relationship between relative bid-ask spreads and illiquidity for the cryptocurrency market. The relationship is found by studying cryptocurrency trading data in two dimensions; Along a time series dimension, where data is aggregated on a daily level, and along an intraday dimension, where variables are aggregated at five-minute intervals across all trading days. The examined illiquidity measure is composed of directly observable asset characteristics such as price, trading volume, and volatility, and the findings provide support for the use of this universal transaction cost model for bid-ask spreads in the cryptocurrency market. Moreover, the results also demonstrate the presence of market frictions in this supposedly more efficient market structure.

Keywords:

Market microstructure, invariance, bid-ask spread, liquidity, Cryptocurrencies, Bitcoin, Cryptocurrency exchanges

Authors:

Charles Gyllhamn (24680)

Jacob Winberg (24774)

Tutor:

Olga Obizhaeva, Assistant Professor, Department of Finance

Examiner:

Adrien d'Avernas, Assistant Professor, Department of Finance

Bachelor Thesis
Bachelor Program in Business & Economics
Stockholm School of Economics

I. Introduction

Market microstructure is the branch of finance that studies how markets operate. The central area of inquiry concerns the structure of exchanges and trading venues, intraday trading behavior, the price formation process, and transaction costs (Kissell, 2014). A recent theory called Market Microstructure Invariance (MMI), postulated by Albert S. Kyle and Anna A. Obizhaeva (2016), has cemented itself in microstructure literature. It is based on the intuitive understanding that financial markets transfer risks in business time and can, in contrast to existing literature, be used to render accurate predictions of how different quantities in market microstructure vary across assets with different levels of trading activity by only employing empirically observable variables in financial markets such as returns variance, price, and trade volume.

After its publication, MMI theory has been supplemented and strengthened by additional papers and is today providing several empirical invariance hypotheses. One of them being derived from a paper published by Kyle and Obizhaeva (2017) where dimensional analysis, leverage neutrality, and a principle of market microstructure invariance is combined to ultimately introduce scaling laws, relevant for the creation of transaction cost functions. More specifically, Kyle and Obizhaeva (2017) derive an illiquidity measure, defined as the cube root of return variance to dollar volume, and predict that it has a proportional relationship to the relative bid-ask spread. This theory has received instant theoretical and practical acknowledgement in finance literature, and so far, it has been tested and confirmed empirically for various markets, e.g. Russian and US stock exchanges. To stretch and challenge the theory further, the predictions ought to be tested on new and unconventional types of assets and markets where there are reasons to believe that microstructure characteristics are unique.

An emerging asset class, cryptocurrencies, is built on novel technology that enables completely new market characteristics to take form. This divergent structure has never been seen in public markets before, and many believe it is paving the way for the development of financial markets. Because this market employs continuous trading, peer-to-peer transactions, and nearly instantaneous settlement, liquidity formation may differ from those of traditional asset markets (Bruneis et al., 2021). The nationless character of cryptocurrencies implies that they are completely decentralized from any organization, leading to autonomous pricing across cryptocurrency exchanges (Makarov and Schoar, 2020). A recent debate concerning the regulation of these exchanges has accelerated as they are serving a similar function as established stock exchanges, but at the same time, lack the appropriate regulations (CNBC, 2022).

The cryptocurrency market, which has previously been dominated by retail investors, has grown in terms of global importance, echoed by a pattern of increased engagement from institutional investors (Forbes, 2021). Cryptocurrencies are also becoming a larger element of many investment portfolios, many of which are being managed under rigorous trading philosophies, often based on High-Frequency Trading (HFT) data (Chu et al., 2020). Naturally, an increasing emphasis has been placed on finding an accurate and effective measurement for estimating liquidity. In recent research, Bruneis et. al (2021) investigate the efficacy of low-frequency liquidity measures to describe actual liquidity of two cryptocurrencies, Bitcoin and Ethereum. The Kyle and Obizhaeva (2016) estimator is employed in their study, and the authors claim that it outperforms other low-frequency measures.

The central contribution of this paper is twofold. First, prior studies have covered the market microstructure invariance theory, however, to the best of our knowledge, the theory has only been confirmed on equities, futures contracts, and portfolio transitions. We extend the invariance and microstructure literature by testing if our postulated invariance hypothesis holds for this unconventional, yet increasingly important, asset class and market. Second, by stretching Bruneis et al. (2021) work, we contribute to the field of research that focuses on liquidity measurement in cryptocurrency markets. By empirically testing if the invariant relationship underpinning the used estimator actually holds, we potentially confirm the performance of this generalizable transaction cost measure for a new asset class and market structure. With this in consideration, the aim of this paper is to answer to the following question:

Does the relative size of the bid-ask spread have a proportional relationship to the asset-specific illiquidity measure, defined as the cube root of the ratio of return variance to USD volume, for cryptocurrencies trading on the Kraken Exchange between 2019-01-01 and 2020-01-01?

We intend to answer our research question in two distinct dimensions; Along a time series dimension, where data is aggregated on a daily level, and along an intraday dimension, where variables are aggregated at five-minute intervals across all trading days. For both dimensions, we follow a similar theoretical framework to Kyle and Obizhaeva (2017) where scaling laws and transaction cost models are derived pursuant to the invariance assumptions. For the latter dimension, a method for variable aggregation described by Anderson et. al (2018) is used, allowing us to isolate and test the time-of-day effect.

We obtain tick-by-tick data recorded on a millisecond resolution, spanning from 2019-01-01 to 2020-01-01, for 16 different cryptocurrencies trading on the Kraken exchange. Our main aim is to test the hypothesized relationship on a cross-sectional level, but we also conduct individual tests for all cryptocurrencies to supplement our analysis. All tests are performed using Ordinary Least Square (OLS) regressions.

Along the time series dimension, we find that the relationship between relative bid-ask spread and illiquidity is broadly consistent with invariance. Although we statistically reject our hypothesis, the result is economically close to our predicted slope of one. On the level of individual cryptocurrency, the invariance relationship holds statistically for six out of the 16 cryptocurrencies in our sample. For the cross-sectional regression along an intraday dimension, we find that the relationship between relative bid-ask spread and illiquidity is broadly consistent with invariance, with a result economically close to the predicted slope of one.

In further investigation of the intraday plot, however, we observe a puzzling result. A practically horizontal pattern for most cryptocurrencies is found, suggesting that the bid-ask spread may be mechanically constrained. To investigate this seemingly systematic issue, we inspect the regression of all individual cryptocurrencies separately and find that the invariant relationship holds statistically for zero out of the 16 cryptocurrencies in our sample. The slopes of the best fitted lines range from -0.03 to 0.46, with a mean of 0.08. The R^2 ranges from 0.00 to 0.18, with a mean of 0.05. Kyle and Obizhaeva (2016) state that the empirical invariance hypothesis is not expected to hold exactly across all assets and time - if market frictions are high. If market makers are competitive, tick sizes are small, transaction fees and taxes are low, the invariance predictions may hold particularly well.

Our empirical results yield two distinct insights. First, that the examined relationship between relative bid-ask spread and illiquidity broadly holds for this nascent asset class and unique market structure, which supplements the findings of Bruneis et. al (2021) and provide support for using the Kyle and Obizhaeva (2017) transaction cost model to estimate liquidity in the cryptocurrency market. Second, along the intraday dimension and on the level of individual cryptocurrency, we overwhelmingly reject the hypothesized invariant relationship, indicating the existence of substantial market frictions. We therefore leave a recommendation to potential investors that trading should be exercised with caution in this market.

II. Fundamental Concepts

Cryptocurrencies

The whitepaper "Bitcoin: A Peer-to-Peer Electronic Cash System", published by Nakamoto (2008), describes a digital currency system in which transactions are recorded on a chain of connected blocks, hence "*blockchain*", and electronically validated by a decentralized network of users. The blockchain keeps track of all previous transactions since its inception, and all information on the blockchain is publicly available. Each block of transactions is linked to the one before it, ensuring that no one can alter data without it being visible to the other participants.

These novel characteristics allow Bitcoin to be fully decentralized, escaping the need of centralized banks and clearing houses to control government-backed currencies. The code upon which this is built is open source, which means that it is free to modify and redistribute. As a result, multiple other cryptocurrencies, known as altcoins, have been created, each hosted on its own blockchain and with unique features and characteristics. One of them is Ethereum, the world's first programmable blockchain, which builds on the same fundamental technology as Bitcoin, but with the addition of *smart contracts* (Buterin, 2014). Smart contracts allow for, among other things, the creation of *tokens*. Tokens are similar to other cryptocurrencies in the sense that they are fungible blockchain-based assets that can be sent and received. The main difference is that tokens are issued on another blockchain network, instead of running their own (Cong et al. 2019). In our analysis, we will test our postulated invariance hypothesis on both coins and tokens, which collectively are referred to as *cryptocurrencies*.

Cryptocurrency Exchanges

While cryptocurrency exchanges serve a similar function as a traditional stock exchange, i.e. to facilitate trades between accounts, the distinct technology underpinning this asset class have resulted in cryptocurrency exchanges being structured in a different way. Traditional exchanges execute trades through an order matching engine and they have no custody over the assets. Cryptocurrency exchanges, on the other hand, must custody their customers assets, match buyers and sellers, verify accounts, and finally process the trades (Forbes, 2019). The SEC has however expressed concerns that cryptocurrency exchanges are failing to set up sufficient barriers between different components of their businesses, such as custody, market-making, and providing a trading venue (Bloomberg, 2022).

A recent debate concerning the regulation of cryptocurrency exchanges has accelerated. In May 2022, the SEC declared initiatives to increase investor protection in this market (CNBC, 2022). Many of the major cryptocurrency exchanges are currently registered as Money Services Businesses (MSB), which requires them to register with the Financial Crimes Enforcement Network (FinCEN) in the US. It does however not imply that their trading activities are regulated and in reality, cryptocurrency exchanges lack a high level of transparency and operational resilience, among other requirements, when compared to other Alternative Trading Systems (ATS) which are used in equity and fixed income markets (Walker, 2021).

Bets

When portfolio managers trade financial assets, risks are being transferred. Kyle and Obizhaeva (2016) describes a bet as a portfolio decision which implements a trading idea (a risk transfer). Practically speaking, it can be described as a decision to acquire or divest a long-term position of a specific size independent of other such decisions. A bet is seldom easily observable for researchers as it theoretically can be placed in a series of trades executed over several days due to its long-run and strategic characteristics, and therefore, individual orders are often treated as a proxy for larger bets.

Business time

Kyle et al. (2016) defines business time as the expected *calendar time* between the arrival of bets in the market. The business time concept can be described through a simple analogy where bets can be thought of as strategic moves in a game of chess. Then business time can then be viewed as the time in between new moves being executed. For actively playing parties, or as in our case for actively traded assets, moves are being made fast and business time passes quickly.

Liquidity and Illiquidity

In conventional terms, *liquidity* is described as the efficiency in which an asset or security can be converted into cash or cash equivalence without influencing its market price (Investopedia, 2022). Illiquidity is simply the inverse of this measurement. Kyle el al. (2016) introduces a new definition of illiquidity that distinguishes from the former definition. Their measure of illiquidity is volume weighted and is defined as the percentage dollar cost of executing an average bet, divided by the average dollar value of the bet.

Market Microstructure Invariance

Market microstructure studies how different trading mechanisms, for instance, tick size and clearing systems, influence the price formation process in financial markets. Market microstructure characteristics, such as bet size, bid-ask spread, and market impact cost vary across assets and time. The theory of *market microstructure invariance* hypothesizes that these characteristics become constant (or microstructure invariant), when converting calendar time into *business time.* Kyle and Obizhaeva (2016) formulate two invariance principles as empirical hypotheses, assumed to apply for all assets and across time, which may be used to render accurate predictions on price formation processes by utilizing easy-to-access data from exchanges.

III. Literature Review

The theory of measuring and estimating liquidity in markets is nothing new. Various ways have been introduced and tested over many years of research. But it was not until 2016, when Kyle and Obizhaeva introduced their market microstructure invariance hypothesis, that transaction costs could be measured through a generic model. They propose two invariance principles, *Invariance of Bets* and *Invariance of Transaction Costs*, conjectured to hold for all assets and across time. They perform empirical tests on portfolio transition orders, viewed as natural experiments for measuring transaction costs. The empirical tests support the postulated invariance hypothesis, which lay the foundation for further research in this exciting new field of financial study.

One year later, Kyle and Obizhaeva (2017) follow up their previous work on market microstructure invariance by applying *dimensional analysis* to introduce scaling laws for different financial variables such as bid-ask spread, bet sizes, and transaction cost. Dimensional analysis is an analytical framework borrowed from physics, where the number of explanatory variables are restricted to simplify scientific inferences. The principle of leverage neutrality is also introduced as a financial analog to traditional conservation laws in physics. In isolation, dimensional analysis does not provide functional market microstructure predictions, as some of the quantities are not directly observable. To generate empirically useful predictions, assumptions implying that some quantities are invariant across assets and time are introduced. Kyle and Obizhaeva (2017) use data from the US and Russian stock market to test their hypothesis empirically. The introduced scaling laws were found to be consistent with the financial data. The derivations, methods, and quantitative models which will be of essential use in our thesis are all provided in this work.

A paper published by Anderson et al. (2018) propose a new theory called Intraday Trading Invariance (ITI). In their study, they focus on how quantitative relationships between distinct market activity variables can be found on an intraday level, based on the invariance intuition of turning calendar time into business time. Anderson et al. measure and aggregate variables at one-minute frequency using tick-by-tick data, and we apply their proposed methodology when measuring and aggregating our variables.

Although the microstructure invariance hypotheses should theoretically hold across all assets and time, the theory has this far only been tested on a few asset classes and markets in actuality. Bruneis et al. (2021) study the efficiency of liquidity measures on the two largest cryptocurrencies, Bitcoin and Ethereum. They compare high-frequency measures of liquidity with easy to compute low-frequency measures, and one of their key findings is that the Kyle and Obizhaeva (2016) estimator outperform. It is however important to note that the market microstructure invariance theory, which this estimator is based on, has only been confirmed empirically for equities, futures, and portfolio transitions, but not for cryptocurrencies.

IV. Theoretical Background

The aim of our thesis is to test if there exists a proportional relationship between bid-ask spread and illiquidity. In order to test this relationship empirically, we must first derive appropriate measures of these hypothesized microstructure characteristics. The dimensional analysis process used by Kyle and Obizhaeva (2017) will serve as the structural backbone for this section. Additionally, we introduce the Intraday Trading Hypothesis (ITI) by Andersen et al. (2018) as we intend to test our relationship on an intraday level as opposed to the daily averages examined by Kyle and Obizhaeva (2017).

Dimensional analysis

Dimensional analysis is a physics-based practice that studies the links between physical properties. The analysis begins by defining the base quantities (e.g. length, mass, and time) and units of measurement (e.g. kilometers and kilograms), which are then tracked as you carry out calculations or comparisons between these dimensions (Berenblatt, 1996). Dimensional analysis has been utilized for a variety of objectives in the past. From inferring the size and number of molecules in a mole of gas to determining the magnitude of explosive energy in an atomic burst (Kyle et al., 2017). In this thesis, we employ a process established by Berenblatt (1996) that is similar to how dimensional analysis is applied to problems in physics, but for analyzing economic problems instead. In finance, the base dimensions are value (measured in units of currency), asset quantity (measured in units of shares), and time (measured in units of years, months, days, hours, minutes, seconds, milliseconds, or microseconds).

When creating our variables, we use the subscripts *jt* to refer to cryptocurrency *j* at time *t*. Let G_{jt} denote the expected price impact cost of executing a bet of Q_{jt} coins[1]. Let P_{jt} denote the cryptocurrency price, V_{jt} its volume in coins, σ^2_{jt} its return variance, and Q_{jt} as the number of coins traded in a bet. Kyle and Obizhaeva (2017) also introduce C as the unconditional expected dollar cost of executing a bet, which can be expressed as,

$$\mathrm{C} := \mathrm{E}\left\{ G_{jt} P_{jt} |Q_{jt}| \right\}. \quad (1)$$

The market microstructure hypothesis, which is explored in further detail below, allows C to be expressed without the subscript *jt*. Using dimensional analysis properly includes selecting the appropriate collection of variables to describe and explain the variables of interest, G_{jt} in our case. Kyle and Obizhaeva (2017) assume that the five aforementioned variables can be combined to a function explaining the market impact cost G_{jt} of executing a bet of Q_{jt} coins,

[1] One unit of cryptocurrency is defined as a coin throughout this thesis, being analog to a share in traditional finance.

$$G_{jt} := g\left(Q_{jt},\ P_{jt},\ V_{jt}, \sigma^2_{jt},\ C\right). \tag{2}$$

Further, we provide a table below to clarify in which dimensions our main variables are measured in as dimensional analysis pays careful attention to maintaining consistency of dimensions and units of measurement. The variables are divided into two distinct groups, those measured in a base dimensional quantity, and those measured in dimensionless quantities.

$$\left[G_{jt}\right] := 1, \qquad \left[V_{jt}\right] := \text{coins/day},$$

$$\left[Q_{jt}\right] := \text{coins}, \qquad \left[\sigma^2_{jt}\right] := 1/\text{day},$$

$$\left[P_{jt}\right] := \text{currency/coins}, \quad \left[C\right] := \text{currency}.$$

The expected price impact cost, G_{jt}, is measured as a fraction of the value traded, and can therefore be considered to be dimensionless. Using dimensional analysis in the finance field, we observe the three basic dimensional quantities that are time, quantity, and value. We can see that $P_{jt},\ Q_{jt},\ and\ \sigma_{jt}^2$ span these three independent dimensions since Q_{jt} has units of quantity, $Q_{jt}\ P_{jt}$ has units of currency, and $1/\sigma_{jt}^2$ has units of days. In addition, the dimensions of these variables are complete in the sense that they can be combined to express the dimensions of the remaining two variables, C and V_{jt}. This allows us to construct two new dimensionless quantities, denoted L_{jt} and Z_{jt},

$$L_{jt} := \left(\frac{m^2 P_{jt} V_{jt}}{\sigma^2_{jt} C}\right)^{\frac{1}{3}}, \tag{3}$$

$$Z_{jt} := \frac{P_{jt} Q_{jt}}{L_{jt} C}\ . \tag{4}$$

A dimensionless scaling constant m^2 is added when constructing these variables. The dimensionless nature of this scaling constant will be motivated in the market microstructure invariance section below. The exponent of one-third in the formulation of L_{jt} is selected purposefully for key reasons linked to leverage neutrality, which will be motivated in that section. Equation (2) can now be written as,

$$G_{jt} := g\left(Q_{jt}, P_{jt}, \sigma^2_{jt}, L_{jt}, Z_{jt}\right). \tag{5}$$

In accordance with dimensional analysis, we can simplify this model further by dropping Q_{jt}, P_{jt}, σ_{jt}^2, motivated by the fact that the function $g(Q_{jt}, P_{jt}, \sigma_{jt}^2, L_{jt}, Z_{jt})$ is dimensionless in itself, implying that it cannot depend on any of the dimensional quantities included in the model. According to Kyle and Obizhaeva (2017), the logic for this reduction is that physical laws do not depend on the units used to measure variables. We are now left with the following model,

$$G_{jt} := g\big(L_{jt}, Z_{jt}\big). \tag{6}$$

A more intuitive rationale to why we end up with this greatly simplified dimensionless model is that informed investors are not misled by different units of measure when calculating transaction costs. Then, the transaction costs should stay constant regardless of the currency in which the coin trades in, how it is divided, or what timescale (business time or calendar time) is used for measurement.

Leverage neutrality

Kyle and Obizhaeva (2017) continues by introducing leverage neutrality which is treated as a financial proxy for traditional conservation laws in physics. They make three different assumptions that support the use of leverage neutrality to simplify the transaction cost model further. *The first assumption* is that trading cash or cash-equivalent assets incur no cost, and hence, trading a bundle of cash and risky securities will always result in the same economic cost regardless of the amount of cash included. For a numerical example, suppose that cash worth $P_{jt}(A\text{-}1)$ is bundled with each share of stock for some number A, which will result in a new share price of $P_{jt}A$. The economic risk transferred in a bet of Q_{jt} remains unchanged, and hence, the number of shares Q_{jt} and the trading volume V_{jt} remains unchanged as well. The dollar cost of executing a bet C is also intact since the economic risk is unaltered and trading cash incurs no costs. The dollar risk of each share $P_{jt}\sigma_{jt}$ remains, but the return variance σ_{jt}^2 and standard deviation σ_{jt} changes to σ_{jt}^2/A^2 and σ_{jt}/A, respectively. By incorporating the exponent of one third into the definition of L_{jt}, it can scale proportionally with A when leverage changes, just like price P_{jt}. As a result, L_{jt} is dimensionless but not leverage neutral, whereas Z_{jt} is dimensionless and leverage neutral.

The magnitude of the exponent has a straightforward notion that stems from the heart of invariance. The invariance hypothesis' implications may be explained using the concept of *trading activity*, which is defined as the product of dollar volume and return volatility. Kyle and Obizhaeva (2016) proves that the number of bets per calendar day is proportional to two-thirds power of the trading activity, while the bet size is proportional to one-thirds power. This relationship implies that the arrival rate of bets must rise twice as rapidly as their size in order to maintain an invariant distribution. This composition may be shown with an example: assume that the arrival rate of bets increases by a factor of four but volatility in calendar time remains constant. The volatility per business time is then reduced by a factor of two, necessitating the trade size to grow by a factor of two in order for the distribution to stay invariant. The resulting increase in volume by a factor of eight is stemming from an increase of bets by a factor of four ($8^{2/3}$) and an increase in bet size of a factor of 2 ($8^{1/3}$). Thus, if not including the cube root in our illiquidity measure, the ratio of return variance to dollar volume would increase by a factor of eight instead of doubling the percentage transaction cost as required.

The percentage cost of executing a bet G_{jt} transforms to $G_{jt}A^{-1}$ due to the fact that the dollar cost of executing a bundled bet is invariant while the dollar value of the of the bundled bet scales proportionally with price. The transformations are summarized below,

$$Q_{jt} \rightarrow Q_{jt}, \qquad L_{jt} \rightarrow L_{jt}A,$$

$$V_{jt} \rightarrow V_{jt} \qquad Z_{jt} \rightarrow Z_{jt},$$

$$P_{jt} \rightarrow P_{jt}A, \qquad C \rightarrow C,$$

$$\sigma^2_{jt} \rightarrow \sigma^2_{jt}A^{-2}, \qquad G_{jt} \rightarrow G_{jt}A^{-1}.$$

The second assumption incorporates Modigliani Miller’s equivalence of capital structure to market microstructure in the sense that changing the firm's leverage ratio has no influence on the economic cost of trading the firm's securities, assuming that the firm's debt is riskless.

The third assumption relates to repo-haircuts and margin requirements. Margin requirements state what percentage of cash or cash-equivalent assets is needed as collateral in order to trade on margin. A repo (repurchase agreement) is the sale of a security for cash in conjunction with the agreement to repurchase the security at a later date. If the seller defaults, the buyer assumes the collateral as repayment. The more volatile the collateral is, the more risk is assumed by the buyer. To compensate for this risk, the buyer applies a haircut to the valuation of the collateral. According to Kyle and Obizhaeva (2017), changes in repo-haircuts and/or margin requirements should not alter the economic cost of trading risky securities.

Leverage neutrality imposes an additional restriction to the general transaction's cost formula, which implies that for any A, the function g should fulfill the following condition,

$$g\left(AL_{jt},\ Z_{jt}\right)=A^{-1}g\left(L_{jt},\ Z_{jt}\right). \tag{7}$$

If we set $A = L_{jt}^{-1}$, the market impact cost function g can be written as,

$$g(L_{jt},\ Z_{jt}) = A^{-1} g(1,\ Z_{jt}). \tag{8}$$

By defining the univariate function f as $f(Z_{jt}) := g(1,\ Z_{jt})$, a more simple market impact cost model now emerge,

$$G_{jt} = L_{jt}^{-1} f(Z_{jt}). \tag{9}$$

This illustrates that the market impact cost of executing a bet, G_{jt} that scales inversely with our illiquidity measure L_{jt} as a consequence of the proportional relationship with the cash included in the transaction. Moreover, Z_{jt} is leverage neutral, which in combination with above results in the following simplified function,

$$G_{jt} := g\left(Q_{jt},\ P_{jt},\ V_{jt},\ \sigma_{jt},\ C\right) = g(L_{jt},\ Z_{jt}) = \frac{1}{L_{jt}} f(Z_{jt}). \tag{10}$$

Market Microstructure Invariance

Dimensional analysis in itself will not suffice for testing our hypothesis empirically. To ultimately generate useful microstructure predictions, it is vital to understand how to measure the appropriate quantities. Kyle and Obizhaeva (2017) point out three of the quantities, return volatility σ_{jt}, trading volume V_{jt}, and asset price P_{jt}, as asset characteristics that are easy to observe from public data. It is less apparent, however, how to measure the cost of a bet C and the scaling parameter m^2, and how these quantities vary across different assets. Kyle and Obizhaeva (2017) make the invariance assumption that C and m^2 are constant across all assets and time. This assumption on invariance does not follow from dimensional analysis, nor leverage neutrality, it is solely motivated by Ockham's razor: that it is the simplest possible empirical hypothesis.

For simplicity in the exposition, the cost of a bet C and the dimensionless scaling parameter m^2, is dropped. Further, it is assumed that function f in equation (10) takes the form of a power function with exponent w, $f(Z_{jt}) = \lambda|Z_{jt}|^{w}$. By assuming that market microstructure invariance hold, Kyle and Obizhaeva (2017) derive a distinct transaction cost model used for the proportional bid-ask spread cost model by setting the exponent w to a value of zero,

$$G_{jt} = \text{const} \cdot \frac{1}{L_{jt}}. \tag{11}$$

The model suggests that the transaction cost G_{jt} is a constant fraction of the asset value and thus not dependent on the size of the bet Q_{jt}. Kyle and Obizhaeva (2017) predict the bid-ask spreads to be inversely proportional to the illiquidity measure, and that the proportionality constant is invariant across all assets. As a last step before testing the model empirically, Kyle and Obizhaeva (2017) sets the market impact cost G_{jt} to the relative bid-ask spread S_{jt}/P_{jt}. For clarification purposes, the relative bid-ask spread is the size of the bid-ask spread divided by mid-price, the latter being defined as the average of the current best bid and ask price. This is the relationship we will test empirically to answer our research question.

Intraday Trading Invariance

Andersen et al. (2018) develop an alternative hypothesis called Intraday Trading Invariance (ITI), which examines whether the same invariance relationships that apply to bets also apply to individual transactions conducted over short horizons. Many existing tests for market microstructure invariance are conducted over longer time horizons, typically monthly. Thus, the emphasis has naturally been on the variation of these monthly aggregates on distinct assets, and tests have relied heavily on cross-stock comparisons. Andersen et al. (2018), on the other hand, only test their hypothesis on one asset, the E-mini S&P 500 Futures, but analyze the variation in the distinct intraday trading cycle to determine if invariant relationships still remain. This is conceivable given today's nearly 24-hour-a-day global markets, where considerable variation in trading volatility and volume can be examined between the European, North American, and Asian trading hours. Conveniently, Andersen et al. (2018) confirm that the scaling laws introduced through dimensional analysis by Kyle and Obizhaeva (2017) are consistent with the intraday trading invariance relationships investigated.

The intraday trading invariance is obtained by stipulating that the invariant relationships of transactions conducted over short horizons are similar to those discovered when the market microstructure invariance theory was first developed, when invariance principles were applied on large speculative bets over longer time horizons.

V. Hypotheses and Research Design

Hypotheses

In this section, we present our hypothesis concerning the invariant relationship between bid-ask spread and our illiquidity measure.

Does the relative size of the bid-ask spread have a proportional relationship to the asset-specific illiquidity measure, defined as the cube root of the ratio of return variance to USD volume, for cryptocurrencies trading on the Kraken Exchange between 2019-01-01 and 2020-01-01?

$$\ln\left(\frac{\mathrm{S_{jt}}}{\mathrm{P_{jt}}}\right) = \mathrm{const} + 1\cdot \ln\left(\frac{1}{\mathrm{L_{jt}}}\right) \tag{12}$$

Where

$$\left(\frac{1}{\mathrm{L_{jt}}}\right) = \mathrm{const}\cdot \left[\frac{\mathrm{P_{jt}V_{jt}}}{\sigma_{\mathrm{jt}}^2}\right]^{-\frac{1}{3}} \tag{13}$$

And

L = Liquidity
S = Absolute bid-ask spread
P = Mid-price
V = Share Volume
σ = Volatility

The relationship in equation (13) is statistically tested by assigning the log relative bid-ask spread as the dependent variable and the log illiquidity measure as the independent variable. The log illiquidity measure's coefficient is denoted by $\boldsymbol{\beta}$. The alternative hypothesis and null hypothesis can now be developed.

$\mathrm{H_0}$: $\boldsymbol{\beta} = 1$
$\mathrm{H_1}$: $\boldsymbol{\beta} \neq 1$

This hypothesis is postulated in accordance with how Kyle and Obizhaeva (2017) postulated their hypothesis. Under $\mathrm{H_0}$, the slope coefficient is equal to one, confirming the invariant relationship between relative bid-ask spread and illiquidity. Hence, failure to reject $\mathrm{H_0}$ implies that our postulated market microstructure invariance hypothesis holds for cryptocurrencies.

Research Design

Our research question entails looking into and empirically testing a purported proportional relation between the defined illiquidity measure and the percentage bid-ask spread. Kyle and Obizhaeva (2017), like us, do not investigate causality between the two variables. As a result, the findings of this thesis cannot be used to determine whether the bid-ask spread causes illiquidity or vice versa.

We collect data for 16 cryptocurrencies that were trading on Kraken from the beginning to the end of our chosen time interval (2019-01-01 to 2020-01-01). We perform both an individual statistical analysis for each cryptocurrency, and a cross sectional analysis as in Kyle and Obizhaeva (2017). Cross sectional analysis involves obtaining data from multiple participants at one point in time Allen (2017), the participants being different cryptocurrencies in this case. The cryptocurrencies trading on the Kraken exchange are not cherry-picked based on the size of their market capitalization, implying that our aggregation could potentially include smaller and less representative cryptocurrencies. However, the theory says that supposedly invariant relationships should hold across all assets and time, meaning that our dataset still can be used as a proxy to draw conclusions if market microstructure invariance relationships hold for cryptocurrencies.

Our analysis will be performed along the time series dimension with daily averages over the course of one year. We will then supplement our empirical testing with an analysis performed along the intraday dimension, with variables aggregated at five-minute intervals. When testing the invariance theory across the intraday pattern we are more likely to capture short-lived intraday fluctuations, and thus allow for a more granular check whether the invariance between the relative size of the bid-ask spread and illiquidity holds. When aggregating our empirical variables, we will apply the proven methodology that was used by Andersen et al. (2018) in which they aggregated their data for E-mini S&P 500 futures at one-minute intervals.

It is worth emphasizing that a short time interval can result in several setbacks, e.g. statistically insufficient number of observations in each of our five-minute intervals, making our data potentially subject to higher standard errors of arithmetic means. Considering this, we perform an additional robustness check as part of the empirical testing of our hypothesis, as well as constructing a confidence interval centered around the regression slope, allowing us to confirm our postulated hypothesis if the predicted slope of one is contained in the confidence interval. As part of the robustness check, we also aggregate the data to a one-hour frequency. If the results remain quantitatively similar, we can present our intraday analysis with a higher degree of certainty, thus allowing for an as-granular-as-possible analysis without being too optimistic that there will be a sufficient number of observations or any other microstructure limitations interrupting our test.

The hypothesis' correlational nature makes regression analysis the obvious choice for establishing if there is a statistical link between the variables. The premise of proportionality when multiplying the illiquidity measure with a constant provides further support to use the ordinary least square (OLS) regression model as the connection should provide a correlational, linear relationship regardless of the intercept and hence the value of this constant. This regression model creates a best-fit line for each time period based on observations, which in our instance are the relative magnitude of the bid-ask spread and the values of our calculated illiquidity measure.

VI. Data

Data source

We use cryptocurrency trading data from the Kraken exchange, which has been purchased from CryptoTick, a well-reputed market data provider of cryptocurrency data. The cost of obtaining data from CryptoTick is relatively high, which is compensated for by backtested, bias-free, and processed data. Furthermore, CryptoTick gives us access to the Kraken exchange, which is one of the most reputable and established cryptocurrency exchanges, accounting for more than 10% of global cryptocurrency trading volume in 2019[2]. According to a widely referenced article by Hougan et al. (2019), up to 95% of exchange-reported Bitcoin trading activity may not represent economically relevant transactions or may even be blatantly fake. The authors put 83 cryptocurrency exchanges through a series of tests. Ten exchanges passed and could be called "real-volume exchanges", one of them being the Kraken exchange.

CryptoTick provides order book and trades data on intraday resolution. The order book data is a snapshot representing the 50 best levels of bid price, ask price, bid volume, and ask volume in one-minute intervals. The relevant trades data for this thesis is the trade volume and trade price, which is provided on a millisecond resolution. We use data from 2019-01-01 (BOD) to 2020-01-01 (EOD).

Data cleaning process

The exchange provides crossed books due to an internal matching engine that does not match orders immediately by design. This causes negative bid-ask spreads when large buy orders are placed and is purely a behavior that stems from how the matching engine is built. In an ideal world, matching orders should be executed immediately and not placed in the book. Therefore, we rigidly remove the order book updates where the bid-ask spread is non-positive and thus avoid this undesired behavior. Furthermore, the minimum tick size appears to be binding, with more than 10% of the bid-ask spread equaling the minimum tick size. As a result, as part of the data cleaning process, we filter out these observations. We filtered out approximately 10 million negative bid-ask spread observations and nearly 30 million observations where the bid-ask spread matched the minimum tick-size, meaning that we in total left out 16% of the order book data as it did not display meaningful activity (see Appendix C for details).

Several referenced articles discuss the remarkably high volatility profile of cryptocurrencies and even the largest cryptocurrency, Bitcoin, can even during regular market conditions be said to have over 2 times higher volatility than the average stock. For smaller cryptocurrencies, this number is even higher and it is not uncommon that the annualized volatility exceeds 100 percent. Even with these high oscillations in mind, we can still identify datasets with a few short-lived observations that deviate

[2] Average of 10 trading days in April, 2019.

significantly from the rest. We suspect them originating from individuals behaving in unpredictable and irrational ways, trading at low volumes, or flaws in the dataset caused by the exchange. As we test for invariance only when the market displays meaningful activity, we also remove what is below and above the 0.1th and 99.9th percentiles as a general outlier treatment. This filter is applied to all variables as a process of removing extraneous microstructure noise while retaining a high number of economically meaningful observations in our variables.

Methodology for measuring High-Frequency Trading data

When working with a hypothesis that requires a market microstructure scope, the measurement procedure will involve variables that are subject to high-frequency expectations or real-time interactions among rapidly fluctuating variables. Therefore, we must employ robust and powerful frameworks to explore the regressions.

First, we present the measurement procedure we use for testing the invariance hypothesis on an intraday level. The systematic diurnal variation is captured by averaging the observations for each intraday interval across all trading days which formally can be described through the following scheme,

$$y_t = \frac{1}{D} \cdot \sum_{d=1}^{D} \tilde{y}_{dt} \approx c + \frac{1}{D} \cdot \sum_{d=1}^{D} y_{dt}, \qquad \text{for } t=1,\ldots,\ T,$$

The intraday dimension consists of five-minute intervals, implying that t is ranging from 1 to 288. The constructed five-minute intervals will serve as a foundation when testing the invariance hypothesis on a daily level. The time series variation is obtained by aggregating five-minute intervals each day, thus generating one data point per trading day. The daily time series is formally obtained by performing the following aggregation scheme,

$$y_d = \frac{1}{T} \cdot \sum_{t=1}^{T} \tilde{y}_{dt} \approx c + \frac{1}{T} \cdot \sum_{t=1}^{T} y_{dt}, \qquad \text{for } d=1,\ldots,\ D,$$

Data processing

To obtain a consistent dataset, we exclude 16 of 32 cryptocurrencies trading on the Kraken exchange as they either were introduced beyond the initiation of our time period, ceased trading, or were trading against a currency other than the US dollar. The mix between trading periods and trading conditions would merely confuse and result in a disproportionate comparison in the analysis of empirical results.

As we aggregate our variables to five-minute intervals, it is critical to create consistency in the periods when constructing our main empirical variables. Therefore, we construct a dummy dataset that ticks one time every five minutes over the course of our chosen time-interval. This is then merged with our original datasets, which leaves us with distinct five-minute trading periods in our data. The additional dummy rows are filled with duplicated data from the previous trade. We are aware that this manipulated dataset contains fictitious trades, and therefore, we solely use it in the construction of the empirical variables that are time-weighted inside five-minute intervals. For variables that are summed in five-minute periods and not weighted, we use the original dataset which contains no dummy data.

Data limitations

We collect data over a one year time span, similar to Kyle and Obizhaeva (2017). However, in addition to testing the invariance hypothesis on a time series dimension, we also perform tests on the intraday dimension with variables aggregated in five-minute intervals. Other studies that test the hypothesized intraday relationship often gather data over considerably longer time frames; for example, Anderson et al. (2018) collect data over a four-year period.

Our choice of time period has been restricted by three factors. First, considering the research from Haugen et al. (2020) regarding fake trading volume discussed above, our set of exchanges to retrieve data from is considerably reduced. Second, most of the remaining exchanges all have very strict processes and requirements in respect to which cryptocurrencies are listed on the exchange, and hence, the further back we go in time, the fewer coins are listed. For instance, Coinbase, which is one of the most reputable exchanges in the industry, only supported trading of five cryptocurrencies in September 2018. Third, due to the highly unusual trading activity caused by Covid-19, we deem the period from 2020 to 2022 as unsatisfactory and unrepresentative of normal market conditions. This leaves us with a period ranging from 2019-01-01 to 2020-01-01.

Our conviction is that a shorter time horizon is worth as a trade off in order to get trading under normal market conditions. We have to reserve, however, that previous studies on the intraday dimension have had varying results when testing each year on a standalone basis, with many of the regressions statistically rejecting invariance. Nevertheless, it should have no effect on the empirical tests performed on the time series dimension.

Construction of main empirical variables

We rebase our trading data to five-minute intervals, leaving us with 288 trading periods per day. Our main empirical variables are order size Q_{jt}, USD volume $P_{jt}V_{jt}$, realized volatility σ_{jt}, main illiquidity index $1/L_{jt}$, and the percentage bid-ask spread S_{jt}/P_{jt^*}, where subscript j denotes cryptocurrency j, ranging from 1 to 16, and subscript t denotes each five-minute interval per day, ranging from 1 to 288.[3]

[3] Asterix implies that the variable is calculated using order book data. Otherwise, it is calculated using trades data.

In the construction of our empirical variables, the time-matching of recorded quantities is critical. Our order book data is time stamped at irregular intervals which has no connection or similarity to the timing of trades in the other data set, making a direct matching procedure impossible. We are therefore left with two options. Either to average the order book data set per five-minute intervals, to then duplicate the average with the number of individual trades and timing from the other data set. This alternative gives a direct comparison between the two data sets, but it relies on a significant assumption when averaging the data. Instead, we construct the percentage bid-ask spread variable from the order book data, and the illiquidity measure from the trades data. The limitation of doing this is that we assume that the prevailing price is similar to the average of the best bid and the best ask, which intuitively makes a better estimate than averaging the data on five-minute periods. In this thesis, we use the best bid and offer when calculating the bid-ask spread. Further, it is worth emphasizing that we use quoted order book data instead of realized. It is not exactly similar, but it makes a prediction about bid-ask spreads that should still hold according to Kyle and Obizhaeva (2017).

When creating our main empirical variables, we are first calculating $V_{jt,}$ σ_{jt}, P_{jt}, S_{jt}, and P_{jt^*} for each individual row in our datasets. The first two variables are calculated from the original trades data set, the third is created from the trades data set, including dummy trades, and the last two are calculated from the order book data set which also includes dummy data. Before we construct our main illiquidity index, $1/L_{jt}$, and the percentage bid-ask spread, S_{jt}/P_{jt^*}, we aggregate the aforementioned variables into consecutive five-minute periods for all trading days. By doing this, we receive the same number of rows in all our data sets and can then construct the rest of our empirical variables.

Robustness test

We perform a robustness test where the intraday observations are aggregated at hourly intervals. By allowing for less granularity, we obtain periods consisting of more data points which in turn reduces microstructure noise. The constructed hourly intervals derive from the previously constructed five-minute periods, following the method Bruneis et al. (2021) use when they conduct similar liquidity tests for the cryptocurrency market. When forming their hourly time intervals, they require 80% of the subintervals to have available data, implying that if more than two five-minute periods inside one hour contain no data, the corresponding hour should be omitted from the analysis. We rigidly mimic Bruneis et al. (2021) in the aggregation of trades data, causing 67.2% of the hourly periods to be omitted from our one-hour robustness test (see appendix B for details).

We discard the 80% rule when performing aggregations of the order book data as the original order book data is updated at frequent time stamps, leaving no sub-periods empty of data. Therefore, the rule would be meaningless. Instead, we form a new rule that omits an hourly interval if trading is absent in at least one of the hour's five-minute intervals. The attentive reader may suggest that this would leave no effect as all sub-intervals should be filled with order book data, and therefore, we underline that we use an outlier cleaning technique removing 0.2% of the most extreme observations based on their bid-ask spread relationship. This implies that all data inside a five-minute interval possibly could be removed if it is within an extremely divergent trading period. By applying this filter, we remove 1,469 of 140,544 hourly periods.

The fact that the two data sets have an unmatching number of remaining rows is not an issue as this test is run in an intraday dimension, with data for each cryptocurrency being averaged across 24 one-hour intervals after the date stamp is removed.

We find that the intraday robustness check produces similar results to our main intraday test. Even though the slope of the cross sectional plot is broadly similar to our main result, we observe that the individual plots of each cryptocurrency is displaying a somewhat more vertical pattern, hinting on the insight that depending on which frequency our plots are presented on, mechanically constraining issues are more or less apparent. We will expand on this result in section VII. The specific results of the regression on the intraday robustness check can be observed in further detail in Appendix B.

Summary statistics – Time series dimension

The daily values for the activity variables are shown in figure 1 and summarized in table 1, which were obtained by averaging the data throughout each trading day in our sample. Trading volume (dollar volume) and transaction intensity show a considerable degree of covariation, as anticipated. Return volatility is high, although it is decreasing on average throughout the year. Despite the fact that the volatility scheme does not follow a similar pattern as trading, we find that persistent volatility swings are paralleled by essentially comparable movements in trading, even if the latter series appears more erratic.

TABLE 1 – Summary statistics for daily variables

	Mean	Median	Min	Max	Ratio
Q	361	621	244	964	3.9
Dollar volume	18,802	14,412	2,984	116,198	38.9
Volume	7,304	6,301	2,172	30,035	13.8
N	14	12	5	55	10.4
VAR	177%	165%	84%	398%	4.7
1/L	3.36	3.19	2.25	7.13	3.2
S/P	3.67	3.44	1.82	9.06	5.0

TABLE 1. THIS SUMMARY STATISTICS ARE REPORTED FOR THE DAILY VARIABLES, FOLLOWING AVERAGES OF THE FOLLOWING FIVE-MINUTE OBSERVATIONS FOR EACH TRADING DAY. THE VOLATILITY MEASURE IS PRESENTED ON ANNUAL BASIS AND IS REPRESENTING THE REALIZED VOLATILITY. THE PERCENTAGE BID-ASK SPREAD AS WELL AS OUR ILLIQUIDITY MEASURE ARE PRESENTED IN 10^{-4}.

FIGURE 1 – Trading patterns along the time-series dimension

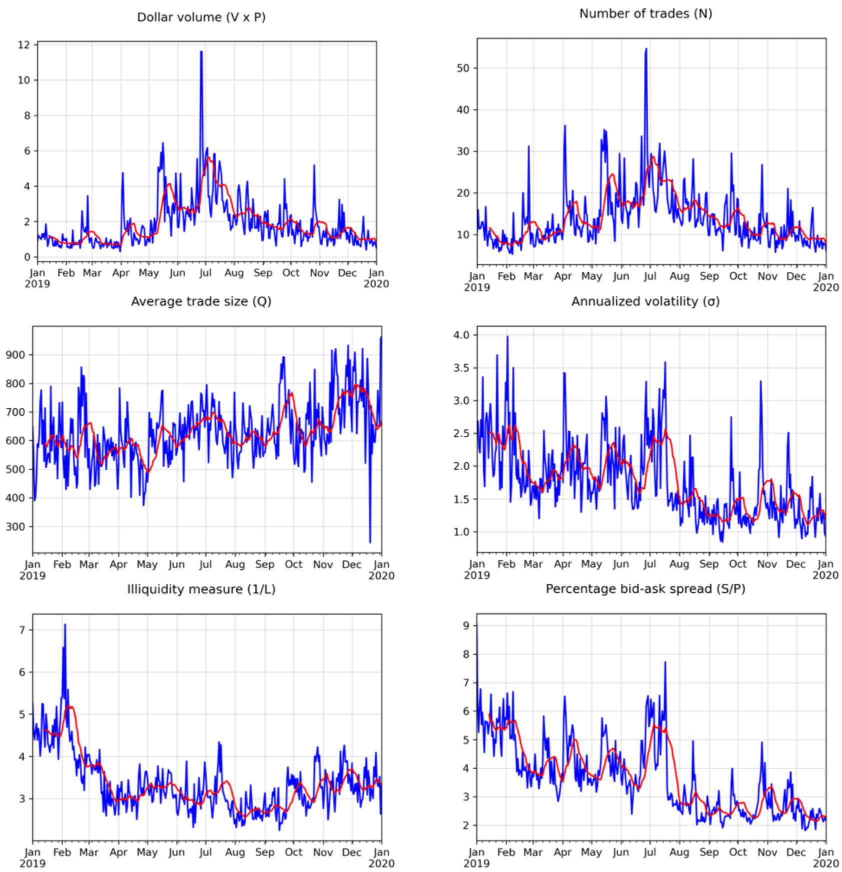

FIGURE 1. THESE FIGURES PLOT THE TIME SERIES AVERAGES OF THE FOLLOWING FIVE-MINUTE OBSERVATIONS FOR EACH TRADING DAY FOR DOLLAR VOLUME V_dP_d; IN 10^4, VOLATILITY σ_d (ANNUALIZED), NUMBER OF TRANSACTIONS N_d, TRADE SIZE Q_d, OUR ILLIQUIDITY MEASURE 1/L; IN 10^{-3}, AND PERCENTAGE BID-ASK SPREAD S/P; IN 10^{-3}. THE RED LINE INDICATES THE 2-WEEK MOVING AVERAGE. THE SAMPLE PERIOD RANGES FROM 2019-01-01 TO 2020-01-01.

Summary statistics - Intraday dimension

Our activity variables' intraday trends are shown in figure 2 and summarized in table 2, which were obtained by averaging the data for each five-minute intraday period for all trading days in the sample. The noticeable shadowing defines three regional trading zones. The green regime includes observations from 00:00 to 07:00, which corresponds to Asian trading hours; the blue regime includes observations from 07:00 to 16:00, which corresponds to European trading hours; and the red regime includes observations from 13:30 to 20:00, which corresponds to American trading hours.

Unlike Andersen et al. (2018), we will not devote as much attention to the different trading regimes and how these reflect in the intraday variation. This is motivated by three reasons. First, as cryptocurrency markets never open or close, the typical intraday patterns documented in stock markets will be difficult to observe. Second, we believe that the trading activity will be heavily skewed to the US trading regime because we gathered data from a US-based crypto exchange and only on USD trading pairs. Third, it has been shown by Dyhrberg et al. (2018) that most cryptocurrency trades are still made by retail investors, implying that working hours should have less of an influence on intraday trading patterns.

TABLE 2 – Summary statistics for intraday variables

	Mean	Median	Min	Max	Ratio
Q	679	680	534	836	1.6
Dollar volume	32,414	31,978	21,659	56,448	2.6
Volume	9,306	9,194	6,537	16,050	2.5
N	20	20	15	33	2.2
VAR	110%	109%	93%	158%	1.7
1/L	2.24	2.23	2.01	2.65	1.3
S/P	3.66	3.66	3.51	3.86	1.1

TABLE 2. THIS SUMMARY STATISTICS ARE REPORTED FOR THE INTRADAY VARIABLES, FOLLOWING AVERAGES OF THE FOLLOWING FIVE-MINUTE OBSERVATIONS ACROSS ALL TRADING DAYS. THE VOLATILITY MEASURE IS PRESENTED ON ANNUAL BASIS AND IS REPRESENTING THE REALIZED VOLATILITY. THE PERCENTAGE BID-ASK SPREAD AS WELL AS OUR ILLIQUIDITY MEASURE ARE PRESENTED IN 10^{-4}.

FIGURE 2 – Trading patterns along the intraday dimension

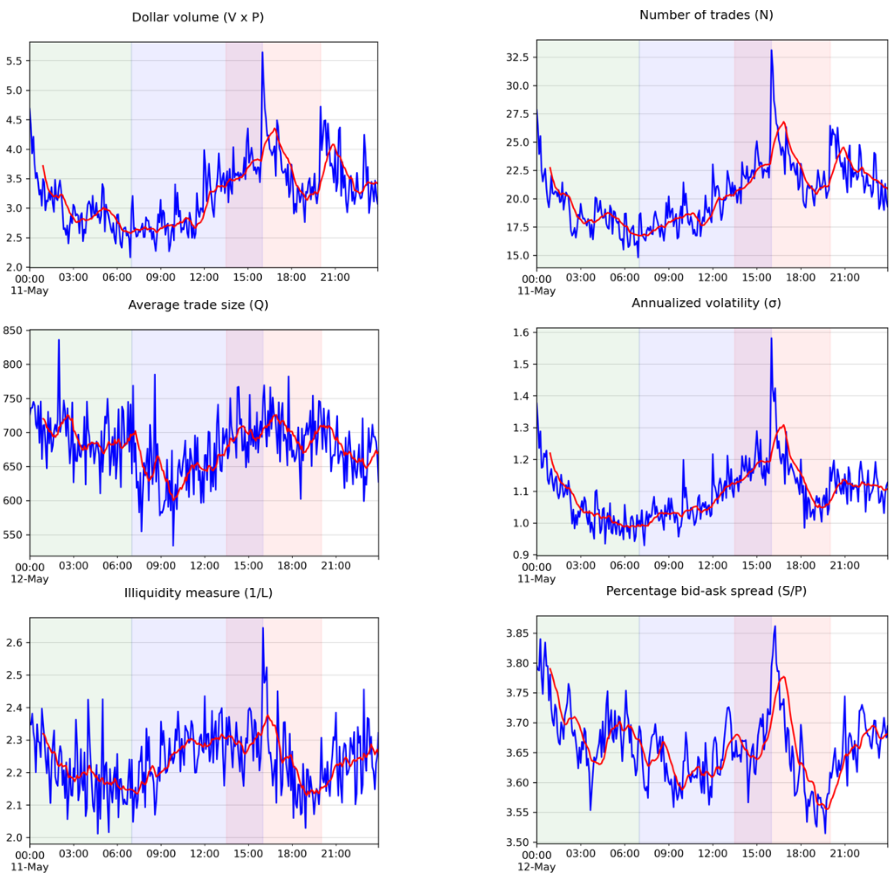

FIGURE 2. THESE FIGURES PLOT THE AVERAGES OF THE FIVE-MINUTE INTERVALS ACROSS ALL TRADING DAYS FOR DOLLAR VOLUME V_dP_d; IN 10^4, VOLATILITY σ_d (ANNUALIZED), NUMBER OF TRANSACTIONS N_d, TRADE SIZE Q_d, OUR ILLIQUIDITY MEASURE 1/L; IN 10^{-3}, AND PERCENTAGE BID-ASK SPREAD S/P; IN 10^{-3}. THE RED LINE INDICATES THE 2-WEEK MOVING AVERAGE. THE GREEN ZONE HIGHLIGHTS ASIAN TRADING HOURS (00:00 TO 07:00), THE BLUE ZONE HIGHLIGHTS EUROPEAN TRADING HOURS (07:00 TO 16:00), AND THE RED ZONE

HIGHLIGHTS AMERICAN TRADING HOURS (13:30 TO 20:00). THE SAMPLE PERIOD RANGES FROM 2019-01-01 TO 2020-01-01.

Considering that cryptocurrency markets lack opening and closing times, the typical u-shaped intraday pattern documented in stock market volume by Jain and Joh (1988), and McInish and Wood (1985), should in theory not emerge. However, as depicted above, we still exhibit an interesting intraday pattern in which the number of trades and dollar volume is at its lowest around 08:00, corresponding to the opening times of the European markets. The trading activity then increases rapidly and reaches its peak around 16:00, two and a half hours after the US market opened and approximately the same time European markets close. This intraday pattern is similar to what Wang et al. (2020) found for Bitcoin, where volume and volatility are noticeably higher for periods that dovetail with the European and US exchange trading hours. The trading volume and number of trades decrease promptly following the 16:00 spike. A second, albeit smaller, spike appears around 20:00, corresponding to the closing of US markets. Interestingly, another spike in the trading activity appears around 00:00, as the Asian markets open. This despite Kraken being an US based exchange and all currency pairs being in USD.

The intraday pattern that emerges across our main empirical variables can to some extent be compared to conventional intraday patterns, even though it cannot be drawn from similar arguments. Despite the fact that cryptocurrency exchanges do not have regular opening and closing hours, trading activity appears to occur during the same hours as traditional stock exchanges, which might be due to the fact that the average trader is most active when his equivalent stock market is open. This might lead to a correlation between cryptocurrencies' intraday variables and major international stock exchanges' opening hours. According to Wang et al. (2020), this relationship might indicate that cryptocurrencies are being used as an alternative investment by market participants. When trading conventional equities, investors may coordinate their trading operations and change their cryptocurrency positions in line with current market conditions.

VII. Empirical findings

This section will present our empirical findings along with descriptive statistics. We will first present the results for the daily cross sectional regression, followed by the intraday cross-sectional regression with respective analysis.

Daily invariance

TABLE 3 - Daily OLS regressions

No. obs.	c	β	Se(c)	Se(β)	R^2	CI Limits (β) 95% LL	CI Limits (β) 95% UL
5,853	1.3161	1.1755	0.033	0.005	0.900	1.165	1.186

TABLE 3. THIS TABLE DISPLAYS THE OLS REGRESSION OF OUR CROSS SECTIONAL TEST ALONG THE TIME SERIES DIMENSION. 16 CRYPTOCURRENCIES ARE INCLUDED IN THE REGRESSION, CORRESPONDING TO 5,853 OBSERVATIONS WHICH ARE FOUND INSIDE OUR TIME PERIOD BETWEEN 2019-01-01 AND 2020-01-01. THE CONSTANT, C, AND SLOPE, β, IS FOLLOWED BY THE STANDARD ERROR OF EACH. R^2 AND A 95% CONFIDENCE INTERVAL ARE REPORTED AS AN INDICATION OF ROBUSTNESS OF THE EMPIRICAL RESULTS.

Table 3 presents results from the OLS regression for the cross sectional analysis along the time series dimension. Figure 3 plots the log percentage bid-ask spread, $\ln(S_{jt}/P_{jt})$, against log illiquidity, $\ln(1/L_{jt})$, for all cryptocurrencies. Each point represents one daily observation, with different colors representing different cryptocurrencies. The predicted and empirical slopes are displayed in the figure by a dashed and solid line, respectively. The empirical slope is broadly consistent with our predicted slope of one, displayed by our fitted line which is represented by the function; $\ln(S_{jt}/P_{jt}) = 1.316 + 1.176 \cdot \ln(1/L_{jt})$. The confidence interval spans from 1.166 to 1.186, on a 95% confidence level, and the R^2 is 0.900. As a result, we statistically reject the null hypothesis, that the slope is equal to one. The slope of 1.176 is however economically close to the one predicted.

FIGURE 3 - Cross sectional plot along the time series dimension

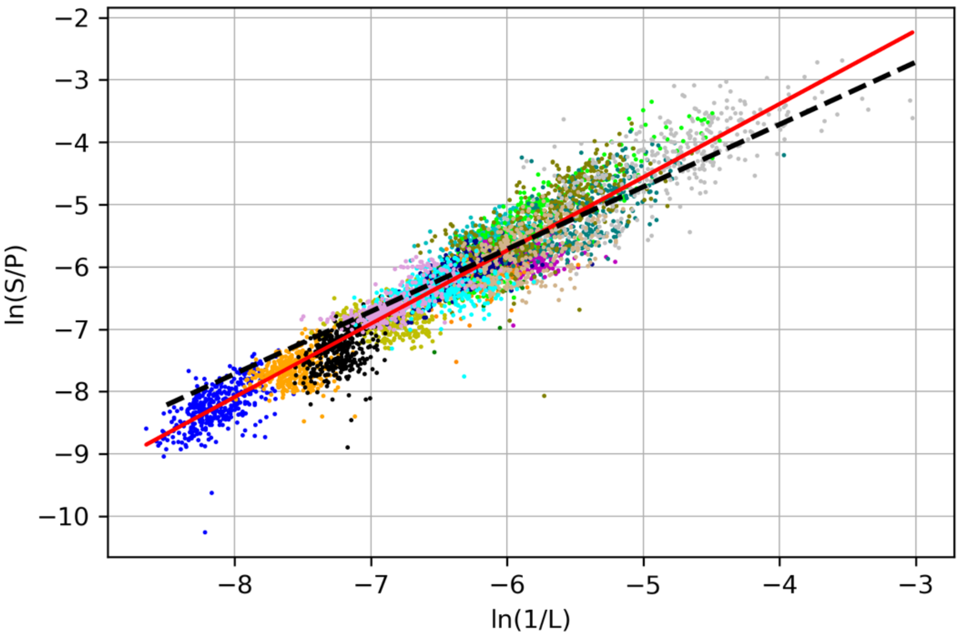

FIGURE 3. THIS FIGURE PLOTS THE LOG PERCENTAGE BID-ASK SPREAD AGAINST THE LOG ILLIQUIDITY OF THE FOLLOWING FIVE-MINUTE OBSERVATION FOR EACH DAY. ALL OF THE 16 CRYPTOCURRENCIES IS REPRESENTED BY AN INDIVIDUAL COLOR, AND EACH POINT REPRESENTS ONE DAY BETWEEN 2019-01-01 AND 2020-01-01. THE DASHED LINE IS ADDED AS A BENCHMARK TO REPRESENT THE PREDICTED REGRESSION SLOPE OF ONE ACCORDING TO OUR HYPOTHESIS. THE RED LINE IS THE BEST FITTED LINE FOR OUR SAMPLE AND IS GIVEN BY THE FOLLOWING EQUATION: LN(S/P) = 1.316 + 1.176 · LN(1/L).

As depicted above, the observations cluster around the benchmark line. In terms of R^2, we observe a higher number compared to those found by Kyle and Obizhaeva (2017), of 0.450 and 0.876 for US and Russian stocks, respectively, indicating that, even though the invariant relationship was statistically confirmed for both US and Russian stocks, the almost invariant relationship can be better explained in our sample.

The fitted lines of individual cryptocurrencies have varying slopes, ranging from 0.1 to 1.4, with a mean of 0.73. We statistically confirm the invariant relationship for six out of the 16 cryptocurrencies, on a 95% confidence interval. Moreover, we obtain slopes that are economically close to one for two more cryptocurrencies. For summary statistics on all individual cryptocurrency regressions, see Appendix A.

Intraday Invariance

TABLE 4 - Intraday OLS regressions

						CI Limits (β)	
No. obs.	c	β	Se(c)	Se(β)	R^2	95% LL	95% UL
4,608	0.7196	1.1216	0.040	0.007	0.865	1.109	1.134

TABLE 4. THIS TABLE DISPLAYS THE OLS REGRESSION OF OUR CROSS SECTIONAL TEST ALONG THE INTRADAY DIMENSION. 16 CRYPTOCURRENCIES ARE INCLUDED IN THE REGRESSION, CORRESPONDING TO 4,608 OBSERVATIONS WHICH ARE FOUND INSIDE OUR TIME PERIOD BETWEEN 2019-01-01 AND 2020-01-01. THE CONSTANT, C, AND SLOPE, β, IS FOLLOWED BY THE STANDARD ERROR OF EACH. R^2 AND A 95% CONFIDENCE INTERVAL ARE REPORTED AS AN INDICATION OF ROBUSTNESS OF THE EMPIRICAL RESULTS.

Table 4 presents results from the OLS regression for the cross sectional analysis along the intraday dimension. Figure 4 plots the log percentage bid-ask spread, $\ln(S_{jt}/P_{jt})$, against log illiquidity, $\ln(1/L_{jt})$, for all cryptocurrencies. Each point represents one five-minute interval, with different colors representing different cryptocurrencies. The predicted and empirical slopes are displayed in the figure by a dashed and solid line, respectively. Similar to the daily aggregation, the empirical slope is also broadly consistent with our predicted slope of one, displayed by our fitted line which is represented by the function; $\ln(S_{jt}/P_{jt}) = 0.720 + 1.122 \cdot \ln(1/L_{jt})$. The confidence interval spans from 1.109 to 1.134, on a 95% confidence level, and the R^2 is 0.865. As a result, the null hypothesis that the slope is one is statistically rejected. However, the slope of 1.122 is economically close to the one predicted.

FIGURE 4 - Cross sectional plot along the intraday dimension

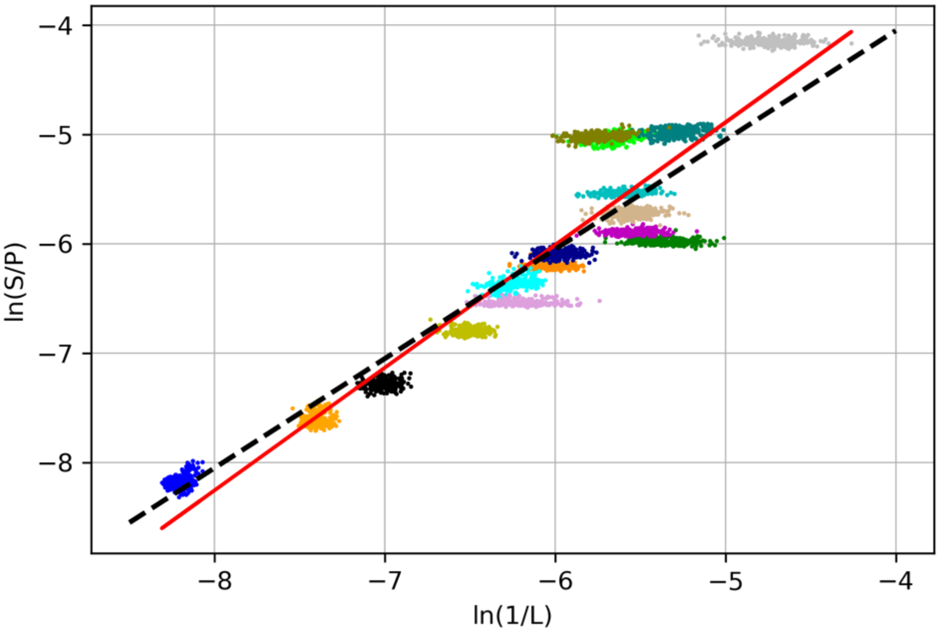

FIGURE 4. THIS FIGURE PLOTS THE LOG PERCENTAGE BID-ASK SPREAD AGAINST THE LOG ILLIQUIDITY OF THE AVERAGES OF FIVE-MINUTE OBSERVATIONS ACROSS ALL TRADING DAYS. ALL OF THE 16 CRYPTOCURRENCIES IS REPRESENTED BY AN INDIVIDUAL COLOR, AND 288 POINTS FOR ALL FIVE-MINUTE INTERVALS ACROSS THE DAY ARE PLOTTED. THE DASHED LINE IS ADDED AS A BENCHMARK TO REPRESENT THE PREDICTED REGRESSION SLOPE OF ONE ACCORDING TO OUR HYPOTHESIS. THE RED LINE IS THE BEST FITTED LINE FOR OUR SAMPLE AND IS GIVEN BY THE FOLLOWING EQUATION: $\mathrm{LN}(S/P) = 0.720 + 1.122 \cdot \mathrm{LN}(1/L)$. THE SAMPLE PERIOD RANGES BETWEEN 2019-01-01 AND 2020-01-01.

Interestingly, even though the cross sectional intraday sample lines up well in accordance with invariance, we observe puzzling results at a level of individual cryptocurrency. The slopes of the fitted lines are significantly lower compared to the cross sectional sample, ranging between -0.03 to 0.46. There seems to be a systematic issue binding the bid-ask spread in a mechanical way, especially pronounced for more illiquid cryptocurrencies. The R^2 for the individual regressions is zero for five of the 16 cryptocurrencies, with the mean being 0.05.

Discussion of empirical findings

If market frictions are large, empirical invariance is unlikely to hold across all assets and time, according to Kyle and Obizhaeva (2016). There are several factors that can influence how well invariance holds, including tick size, fees, taxes, market maker competitiveness, regulation, and clearing systems. Table 5 shows that tick-sizes expressed in index points and notional values are reasonably normal for our sample.

TABLE 5 - Extensive tick-size statistics

	Minimum tick-size	Notional tick-size
ADA	0.01	0.18
BCH	1,000	3.85
BTC	1,000	0.14
DASH	10	0.10
ETC	10	1.78
ETH	100	0.55
EOS	1	0.25
GNO	100	6.34
LTC	100	1.45
QTUM	0.1	0.04
REP	10	0.75
XLM	0.01	0.11
XMR	100	1.51
XRP	0.1	0.33
XTZ	1	1.00
ZEC	100	1.78

TABLE 5. THIS TABLE DISPLAYS EXTENSIVE TICK-SIZE STATISTICS IN FORM OF MINIMUM AND NOTIONAL TICK-SIZES ACROSS ALL CURRENCIES. THE NOTIONAL TICK-SIZE IS CALCULATED BY DIVIDING THE MINIMUM TICK-SIZE WITH THE CURRENCY'S AVERAGE TRADING PRICE ACROSS THE YEAR AND IS EXPRESSED IN BASIS POINTS. MINIMUM TICK IS OBTAINED DIRECTLY FROM THE EXCHANGE AND IS PRESENTED IN 10^{-4}.

Anderson et al. (2018) find that the average bid-ask spread in their sample barely exceeds the minimum value, indicating that their tick size is binding. Despite this, they obtain almost perfect slopes in accordance with invariance, and therefore, we do not believe our striking results are attributable to the tick size. Another factor that could link to our puzzling results are the fees on the exchange. Kraken uses a tiered fee schedule, ranging from 0.26% to 0%, which results in no noteworthy difference to traditional stock exchanges. From a tax perspective, there are no dissimilarities to stocks either, as the IRS treats cryptocurrencies as property. Further, there are no reasons to believe that the competitiveness of market makers should be lower compared to traditional markets.

This leaves two plausible factors to explain our surprising results, clearing mechanisms and regulation. As previously mentioned, we notice a large amount of negative bid-ask spread observations. This is caused by Krakens internal matching engine, which causes the book to be crossed, meaning that the orders are not instantaneously matched. This is far from ideal, and it leads us to believe that Kraken's flawed matching engine is affecting our intraday plots in a highly mechanical manner. Additionally, the debate of regulation has heated up recently. Traditional stock exchanges being regulated as Alternative Trading Systems (ATS) requires them to meet stringent transparency, operational resilience, and trading dependability standards, while cryptocurrency exchanges, such as Kraken, are evade these requirements. As a matter of fact, Kraken's poor matching engine might stem from these insufficient regulations. But analyzing this further takes us beyond this thesis scope.

VIII. Concluding remarks

In this paper we test if there exists a proportional relationship between the relative bid-ask spread and a specific illiquidity measure, predicated on assumptions of market microstructure invariance. This relationship is tested on tick-by-tick data from 16 cryptocurrencies trading on the Kraken exchange during the period from 2019-01-01 to 2020-01-01. We do this in large part by replicating the Kyle and Obizhaeva (2017) process for applying dimensional analysis to finance, with the addition of testing the relationship along an intraday dimension as well, inspired by Andersen et al. (2018).

We discover that the proportional relationship between relative bid-ask spread and illiquidity for cryptocurrencies is economically close to the predicted value along both the time series and intraday dimensions. We validate the invariant relationship for six out of 16 cryptocurrencies along the time series dimension at the level of individual cryptocurrency. Along the intraday dimension for individual cryptocurrencies, we strongly reject the invariant relationship. This puzzling result is moderately investigated and discussed, with the most probable explanation being Krakens flawed order matching engine, which in turn might be linked to careless regulation of cryptocurrency exchanges. We leave for further research to examine if this is an isolated event for the Kraken exchange over this specific timespan, or a more systemic phenomenon present in the broader cryptocurrency market.

The illiquidity measure we test is composed of variables that are directly observed or readily estimated from public data on securities transactions. Its supposedly universal applicability to all assets and time would intuitively be of high value to risk managers and investors, who otherwise have to employ different transaction cost models depending on which market is being traded. Despite being relatively nascent, the cryptocurrency market has already reached a combined market capitalization of two trillion dollars, which corresponds to roughly 2% of the global equities market. Our paper contributes not only to the invariance literature by testing the hypothesized relationship on an unconventional market structure, but also, provides clarity for risk managers and traders seeking reliable transaction cost models applicable for this specific, albeit increasingly influential, asset class. Furthermore, the discovered problem with the bid-ask spread data can be of useful insight for financial authorities and provide arguments for imposing more stringent regulations for cryptocurrency exchanges.

References

Literature

ANDERSEN, T.G., BONDARENKO, O., KYLE, A.S. and OBIZHAEVA, A.A., 2018. Intraday Trading Invariance in the E-Mini S&P 500 Futures Market. Rochester, NY.

BARENBLATT, G. I., 1996. Scaling, Self-Similarity, and Intermediate Asymptotics. Cambridge University Press.

BRAUNEIS, A., MESTEL, R., RIORDAN, R., THEISSEN, E. 2021. How to measure the liquidity of cryptocurrency markets?, *Journal of Banking & Finance*

CHU, J., CHAN, S., ZHANG, Y. 2020, High frequency momentum trading with cryptocurrencies, *Research in International Business and Finance*

CONG, L.W and HE, Z., 2019. Blockchain Disruption and Smart Contracts, *The Review of Financial Studies*, pp. 1754–1797

DYHRBERG, A.H., FOLEY, S., SVEC, J., 2018. How investible is Bitcoin? Analyzing the liquidity and transaction costs of Bitcoin markets, *Economics Letters*, pp. 140-143

JAIN, P. C., & JOH, G.-H, 1988. The Dependence between Hourly Prices and Trading Volume. *The Journal of Financial and Quantitative Analysis*, 23(3), 269–283.

KYLE, A.S. and OBIZHAEVA, A.A., 2017. Dimensional Analysis, Leverage Neutrality, and Market Microstructure Invariance". Rochester, NY.

KYLE, A.S. and OBIZHAEVA, A.A., 2016. Market Microstructure Invariance: Empirical Hypotheses. Econometrica, 84(4), pp. 1345-1404.

MAKAROV, I., SCHOAR, A., 2020. Trading and arbitrage in cryptocurrency markets, *Journal of Financial Economics*, Volume 135, Issue 2, 2020, pp. 293-319

MODIGLIANI, F. and MILLER, M.H., 1958. The Cost of Capital, Corporation Finance and the Theory of Investment. The American Economic Review, 48(3), pp. 261-297.

ROBERT KISSELL, Chapter 2 - Market Microstructure, The Science of Algorithmic Trading and Portfolio Management, Academic Press, 2014, Pages 47-85

WANG, J.N., LIU, H.C., HSU, YT., 2020. Time-of-day periodicities of trading volume and volatility in Bitcoin exchange: Does the stock market matter?, *Finance Research Letters*, WOOD, R.A., McINISH, T.H. and ORD, J.K. 1985, An Investigation of Transactions Data for NYSE Stocks. *The Journal of Finance*, 40: 723-73

Electronic

ANKNEY, LESLIE. 2019, What's So Different About Cryptocurrency Exchanges?, Forbes, Jan 4 Available:https://www.forbes.com/sites/leslieankney/2019/01/04/whats-so-different-about-cryptocurrency-exchanges/ [Retrieved: March 16th 2022].

BLOOMBERG, 2022, SEC's Gensler Says Crypto Exchanges Trading Against Clients, May 13. Available: https://www.bloomberg.com/news/articles/2022-05-10/sec-chief-questions-whether-crypto-exchanges-bet-against-clients [Retrieved: May 13th 2022].

BUTERIN., V. 2014. Ethereum: A Next-Generation Smart Contract and Decentralized Application Platform. https://ethereum.org/669c9e2e2027310b6b3cdce6e1c52962/Ethereum_Whitepaper_-_Buterin_2014.pdf

CNBC, 2022, SEC Chair Gensler says agency is planning greater oversight of crypto markets to protect investors, April 4, Available: https://www.cnbc.com/2022/04/04/sec-chairman-proposes-expanded-protections-for-crypto-investors.html [Retrieved: April 10th 2022].

NAKAMOTO, S. 2008. Bitcoin: A Peer-to-Peer Electronic Cash System. https://bitcoin.org/bitcoin.pdf

WALKER, C. W, MARTIN, 2021, Regulated cryptocurrency exchanges: sign of a maturing market or oxymoron?, LSE Business review, April 13, Available: https://blogs.lse.ac.uk/businessreview/2021/04/13/regulated-cryptocurrency-exchanges-sign-of-a-maturing-market-or-oxymoron/ [Retrieved: March 28th 2022].

WINTERMEYER, LAWRENCE, 2021 Institutional Money Is Pouring Into The Crypto Market And Its Only Going To Grow, August 12, Available: https://www.forbes.com/sites/lawrencewintermeyer/2021/08/12/institutional-money-is-pouring-into-the-crypto-market-and-its-only-going-to-grow/?sh=7cb4785c1459 [Retrieved: March 10th 2022].

Appendix A

FIGURE A1 - Daily plots for individual cryptocurrencies

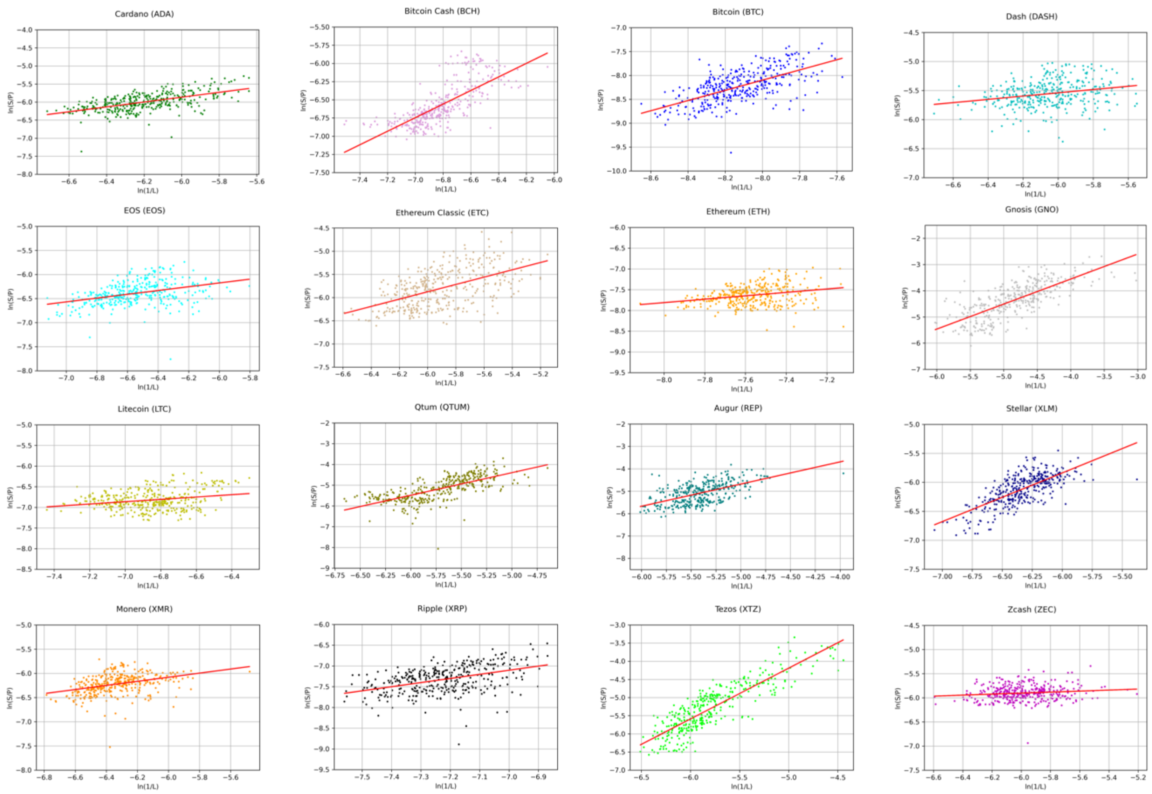

FIGURE A1. DEPICTED ABOVE IS THE LOG PERCENTAGE BID-ASK SPREAD AGAINST THE LOG ILLIQUIDITY FOR EACH DAY AND CRYPTOCURRENCY. EACH POINT IS REPRESENTING A SINGLE DAY, RANGING BETWEEN 2019-01-01 AND 2020-01-01. THE RED LINES REPRESENTS THE BEST FITTED SLOPE OF EACH REGRESSION.

	TABLE A1 - Individual OLS regressions on a daily level						CI Limits (β)	
	No. obs.	c	β	Se(c)	Se(β)	R^2	95% LL	95% UL
ADA	365	-1.83	0.67	0.30	0.05	0.34	0.58	0.77
BCH	366	-0.22	0.93	0.31	0.05	0.54	0.85	1.02
BTC	366	0.45	1.10	0.60	0.07	0.37	0.93	1.21
DASH	366	-3.85	0.28	0.34	0.06	0.06	0.17	0.39
ETC	366	-1.14	0.79	0.38	0.07	0.29	0.66	0.92
ETH	366	-4.53	0.41	0.58	0.10	0.07	0.26	0.56
EOS	366	-3.82	0.39	0.34	0.05	0.14	0.30	0.50
GNO	366	0.27	0.96	0.22	0.05	0.56	0.87	1.04
LTC	366	-4.88	0.28	0.38	0.06	0.07	0.17	0.39
QTUM	366	1.04	1.09	0.34	0.06	0.49	0.97	1.20
REP	366	0.27	0.99	0.41	0.08	0.33	0.84	1.14
XLM	366	-0.80	0.84	0.27	0.04	0.51	0.76	0.93
XMR	366	-3.50	0.42	0.35	0.06	0.14	0.31	0.53
XRP	365	-0.15	0.99	0.80	0.10	0.20	0.79	1.20
XTZ	366	2.80	1.40	0.24	0.04	0.77	1.32	1.48
ZEC	366	-5.28	0.10	0.24	0.04	0.02	0.02	0.19

TABLE A1. THIS TABLE DISPLAY A STATISTICAL OVERVIEW OVER THE INDIVIDUAL REGRESSIONS ALONG THE TIME SERIES DIMENSION. THE CONSTANT, C, AND SLOPE, β, IS FOLLOWED BY THE STANDARD ERROR OF EACH. R^2 AND A 95% CONFIDENCE INTERVAL ARE REPORTED AS AN INDICATION OF ROBUSTNESS OF THE EMPIRICAL RESULTS.

FIGURE A2 - Intraday plots for individual cryptocurrencies

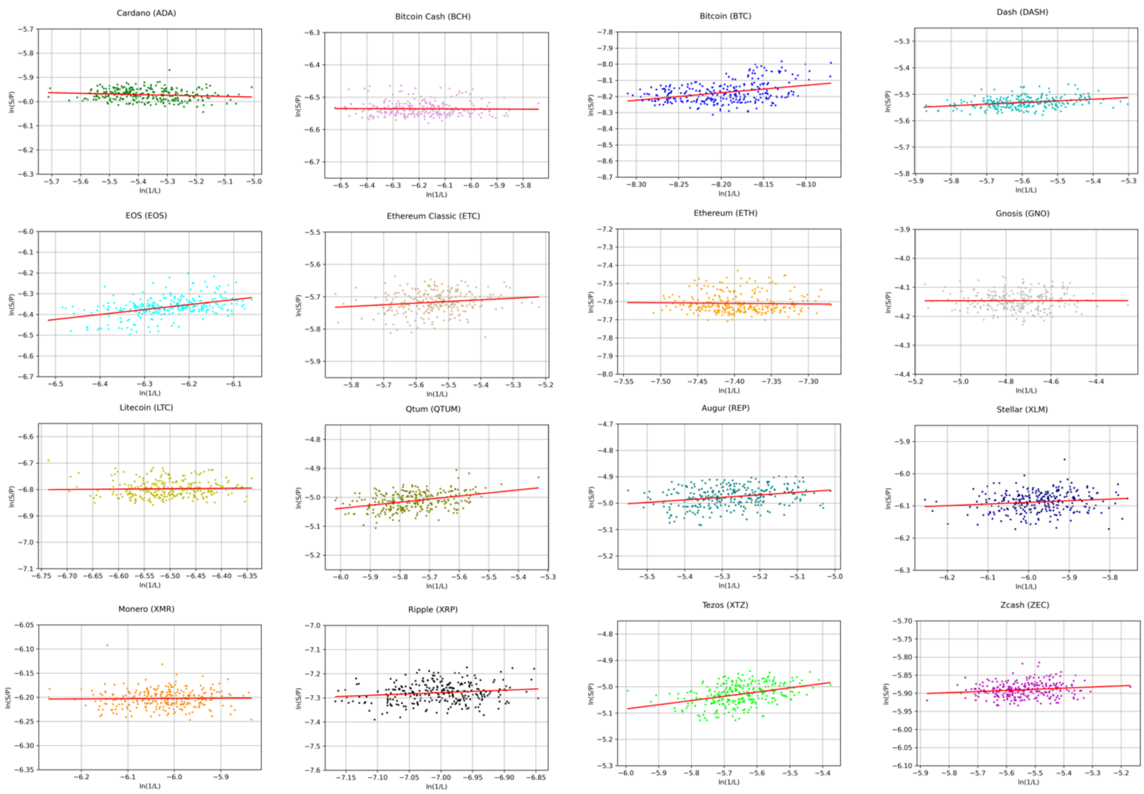

FIGURE A2. DEPICTED ABOVE IS THE LOG PERCENTAGE BID-ASK SPREAD AGAINST THE LOG ILLIQUIDITY FOR EACH CRYPTOCURRENCY WHICH IS OBTAINED BY AVERAGING THE FIVE-MINUTE OBSERVATIONS ACROSS ALL TRADING DAYS BETWEEN 2019-01-01 AND 2020-01-01 ON AN INDIVIDUAL LEVEL. 288 POINTS FOR ALL FIVE-MINUTE INTERVALS ACROSS THE DAY ARE PLOTTED. THE RED LINES REPRESENTS THE BEST FITTED SLOPE OF EACH REGRESSION.

TABLE A2 - Individual OLS regressions on a intraday level

							CI Limits (β)	
	No. obs.	c	β	Se(c)	Se(β)	R^2	95% LL	95% UL
ADA	288	-6.11	-0.03	0.06	0.01	0.02	-0.05	0.00
BCH	288	-6.60	-0.00	0.06	0.01	0.00	-0.02	0.02
BTC	288	-4,38	0.46	0.57	0.07	0.13	0.33	0.60
DASH	288	-5.18	0.06	0.06	0.01	0.10	0.04	0.08
ETC	288	-5.43	0.05	0.11	0.02	0.02	0.01	0.09
ETH	288	-7.84	-0.03	0.48	0.07	0.00	-0.16	0.10
EOS	288	-4.87	0.24	0.19	0.03	0.18	0.18	0.30
GNO	288	-4.14	0.00	0.06	0.01	0.00	-0.02	0.03
LTC	288	-6.70	0.02	0.18	0.03	0.00	-0.04	0.07
QTUM	288	-4.40	0.11	0.09	0.02	0.14	0.08	0.14
REP	288	-4.47	0.10	0.11	0.02	0.07	0.06	0.14

XLM	288	-5.78	0.05	0.13	0.02	0.02	0.01	0.09
XMR	288	-6.17	0.01	0.09	0.01	0.00	-0.02	0.03
XRP	288	-6.57	0.10	0.27	0.04	0.02	0.03	0.18
XTZ	288	-4.10	0.16	0.13	0.02	0.14	0.12	0.21
ZEC	288	-5.72	0.03	0.07	0.01	0.02	0.01	0.06

TABLE A2. THIS TABLE DISPLAY A STATISTICAL OVERVIEW OVER THE INDIVIDUAL REGRESSIONS ALONG THE INTRADAY DIMENSION. THE CONSTANT, C, AND SLOPE, β, IS FOLLOWED BY THE STANDARD ERROR OF EACH. R^2 AND A 95% CONFIDENCE INTERVAL ARE REPORTED AS AN INDICATION OF ROBUSTNESS OF THE EMPIRICAL RESULTS.

Appendix B

TABLE B1 - Outliers for robustness check (trades data)

	Initial observations	Outliers detected and removed	% removed
ADA	8,784	7,449	84.8%
BCH	8,784	3,063	34.9%
BTC	8,784	59	0.7%
DASH	8,784	8,400	95.6%
ETC	8,783	7,862	89.5%
ETH	8,784	272	3.1%
EOS	8,784	7,536	85.8%
GNO	8,782	8,724	99.3%
LTC	8,784	4,133	47.1%
QTUM	8,784	5,467	62.2%
REP	8,783	8,577	97.7%
XLM	8,781	7,190	81.9%
XMR	8,784	7,939	90.4%
XRP	8,784	2,236	25.5%
XTZ	8,784	7,067	80.5%
ZEC	8,784	8,427	96.0%
Total	**140,537**	**94,401**	**67.2%**

TABLE B1. THIS TABLE REPORTS THE OUTLIERS REMOVED FROM THE TRADES DATA FOR THE ONE HOUR ROBUSTNESS CHECK. A ONE HOUR PERIOD IS OMITTED IF IT IS MISSING TWO OR MORE FIVE-MINUTE PERIODS OF TRADING ACTIVITY WITHIN IT. FOR THE TWO LARGEST CRYPTOCURRENCIES, BITCOIN AND ETHEREUM, TRADING IS MORE FREQUENT AND THEREFORE ONLY 0.7 AND 3.1% FIVE-MINUTE INTERVALS ARE OMITTED, RESPECTIVELY, AND FOR THE SMALLER CRYPTOCURRENCIES, WE DELETE OVER 90% OF THE OBSERVATIONS.

TABLE B2 - Outliers for robustness check (order book data)

	Initial observations	Outliers detected and removed	% removed
ADA	8,784	110	1.3%
BCH	8,784	102	1.2%
BTC	8,784	147	1.7%
DASH	8,784	91	1.0%
ETC	8,784	79	0.9%
ETH	8,784	102	1.2%

EOS	8,784	113	1.3%
GNO	8,784	94	1.1%
LTC	8,784	97	1.1%
QTUM	8,784	68	0.8%
REP	8,784	69	0.8%
XLM	8,784	69	0.8%
XMR	8,784	86	1.0%
XRP	8,784	79	0.9%
XTZ	8,784	76	0.9%
ZEC	8,784	87	1.0%
Total	**140,544**	**1,469**	**1.0%**

TABLE B2. THIS TABLE REPORTS THE OUTLIERS REMOVED FROM THE ORDER BOOK DATA FOR THE ONE HOUR ROBUSTNESS CHECK. A ONE HOUR PERIOD IS OMITTED IF IT IS MISSING ONE OR MORE FIVE-MINUTE PERIODS OF ORDER BOOK UPDATES WITHIN IT. AS WE HAVE REMOVED THE MOST EXTREME DAYS FROM OUR SAMPLE, THIS RULE IS REMOVING THE MOST EXTREME INTERVALS FROM OUR ROBUSTNESS TEST.

FIGURE B1 - Cross sectional plot for the robustness test

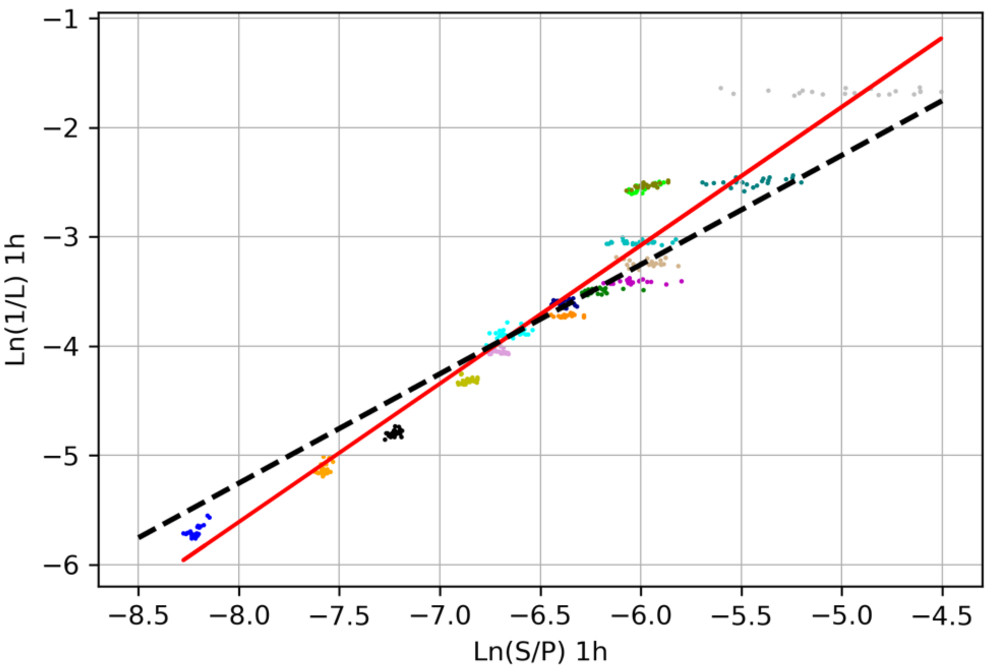

FIGURE B1. THIS FIGURE IS THE RESULT OF OUR ONE HOUR ROBUSTNESS TEST AND PLOTS THE LOG PERCENTAGE BID-ASK SPREAD AGAINST THE LOG ILLIQUIDITY OF THE AVERAGES OF ONE-HOUR OBSERVATIONS ACROSS ALL TRADING DAYS. ALL OF THE 16 CRYPTOCURRENCIES IS REPRESENTED BY AN INDIVIDUAL COLOR, AND 24 POINTS FOR ALL ONE HOUR INTERVALS ACROSS THE DAY ARE PLOTTED. THE DASHED LINE IS ADDED AS A BENCHMARK TO REPRESENT THE PREDICTED REGRESSION SLOPE OF ONE ACCORDING TO OUR HYPOTHESIS. THE RED LINE IS THE BEST FITTED LINE FOR OUR SAMPLE AND IS GIVEN BY THE FOLLOWING EQUATION: $LN(S/P) = 4.5188 + 1.2660 \cdot LN(1/L)$. THE OLS REGRESSION OF THIS SAMPLE GIVES AN R^2 OF 0.936. THE SAMPLE PERIOD RANGES BETWEEN 2019-01-01 AND 2020-01-01.

Appendix C

TABLE C1 - Outliers removal (based on tick-size)

	Negative	Equal to minimum tick-size	Total observations	% removed
ADA	643,652	193,650	14,799,986	6%
BCH	690,282	7,803,473	16,631,565	51%
BTC	715,553	6,721,055	20,052,584	37%
DASH	664,229	123,322	15,966,830	5%
ETC	622,077	436,885	14,555,249	7%
ETH	705,155	5,502,487	18,512,424	34%
EOS	588,878	360,534	15,243,666	6%
GNO	191,637	83,314	4,019,522	7%
LTC	626,413	2,733,812	15,874,329	21%
QTUM	544,717	68,544	12,828,176	5%
REP	495,109	207,512	12,262,771	6%
XLM	691,596	381,691	15,144,778	7%
XMR	536,128	566,950	12,545,375	9%
XRP	835,929	2,906,114	17,612,230	21%
XTZ	684,041	648,697	16,799,648	8%
ZEC	450,582	218,622	12,162,470	6%
Total	**9,685,978**	**28,956,662**	**235,011,603**	**16%**

TABLE C1. THIS TABLE REPORTS THE OBSERVATIONS REMOVED FROM THE ORDER BOOK DATA WHERE THE TICK-SIZE EITHER WAS NEGATIVE OR EQUAL TO THE MINIMUM OF RESPECTIVE CURRENCY. EVEN THOUGH IT IS NOT FORMALLY INVESTIGATED, WE OBSERVE A PATTERN OF % REMOVED AND TRADING ACTIVITY. THIS MIGHT BE EXPLAINED BY THE LIMITATIONS OF KRAKEN'S MATCHING ENGINE'S WHICH IS MORE APPARENT WHEN LARGE ORDERS ARE PLACED.

Appendix D

In this part of the appendix, we present the construction of our empirical variables more thoroughly. We ignore the subscript j below for simplification purposes. We distinguish the timing of observations by defining t_n as the timestamp of an individual trade, where n ranges from 1 to N_i inside the five-minute interval i.

Part I – Variables constructed using trades data

Let $\mathrm{Q_{t_n}}$ denote the number of coins traded at time t_n,

then, let $\mathrm{Q_i} = \frac{1}{\mathrm{N_i}}\sum_{\mathrm{n=1}}^{\mathrm{N_i}} \mathrm{Q_{t_n}}$ denote the average trade size in the five-minute interval i, and

$\mathrm{V_i} = \sum_{\mathrm{n=1}}^{\mathrm{N_i}} \mathrm{Q_{t_n}}$ denote the total trade volume inside the five-minute interval i.

Further, let $\mathrm{P_i} = \sum_{\mathrm{n=1}}^{\mathrm{N_i}} \frac{\mathrm{P_{t_n}} \cdot (\mathrm{t_{n+1}} - \mathrm{t_n})}{\mathrm{T'}-\mathrm{T}}$ denote the volume-weighted price for the five-minute interval i. where $t_0 = T$ is the first trade, and $\mathrm{T_{N_i}} = T'$ is the last trade, inside the interval i. $\mathrm{P_{t_n}}$ is given by the execution price of the corresponding trade at time t_n.

Let $\sigma_{\mathrm{i}}^2 = \sum_{\mathrm{n=2}}^{\mathrm{N_i}} \left(\ln(\mathrm{P_{t_n}}) - \ln(\mathrm{P_{t_{n-1}}})\right)^2$ denote realized variance over interval i.

Then let $\frac{1}{\mathrm{L_i}} = \left(\frac{\mathrm{P_i V_i}}{\sigma_{\mathrm{i}}^2}\right)^{\frac{-1}{3}}$ denote our defined illiquidity measure.

Part II – Variables constructed using order book data

Let $\mathrm{P_{t_n}} = \left(\frac{\mathrm{A_{t_n}} + \mathrm{B_{t_n}}}{2}\right)$ denote the quoted mid-price at time t_n,

where $\mathrm{A_{t_n}}$ and $\mathrm{B_{t_n}}$ is representing the ask and bid, respectively, and is given by the top-of-book at time t_n.

Let $S_{t_n} = \mathrm{A_{t_n}} - \mathrm{B_{t_n}}$ denote the best bid-ask spread at time t_n,

then, we can compute the percentage bid-ask variable, $\left(\frac{\mathrm{S_i}}{\mathrm{P_i}}\right) = \sum_{\mathrm{n=1}}^{\mathrm{N_i}} \frac{\left(\frac{\mathrm{S_{t_n}}}{\mathrm{P_{t_n}}}\right) \cdot (\mathrm{t_{n+1}} - \mathrm{t_n})}{\mathrm{T'}-\mathrm{T}}$

where $t_0 = T$ is the first trade, and $T_{N_i} = T'$ is the last trade, inside the interval i.