MARKET MICROSTRUCTURE INVARIANCE, BID-ASK SPREADS AND IMPACT COSTS IN THE SWEDISH STOCK MARKET

A TRANSACTION COST ANALYSIS FOR INTRADAY TRADING IN SWEDISH STOCKS

JIM DOMEIJ

OSCAR KRIEG

Bachelor Thesis

Stockholm School of Economics

2021

Market Microstructure Invariance, Bid-Ask Spreads and Impact Costs in the Swedish Stock Market : A Transaction Cost Analysis for Intraday Trading in Swedish Stocks

Abstract:

By studying high-frequency trading data for the Swedish stock market, as proxied by the OMXS30 index, we find that there exists an invariant relationship between transaction cost components and illiquidity. Specifically, we apply the notions of market microstructure and intraday trading invariance to confirm the existence of a proportional relationship between the relative bid-ask spread and an illiquidity measure comprised of observable financial market variables, such as trade volume, price and volatility. In addition, we conduct a case study for the Coronavirus Crash of 2020, through which we gain empirical support for the superiority of invariance-implied market impact models over conventional wisdom in predicting large price declines.

Keywords:

Market microstructure, invariance, bid-ask spread, liquidity, Swedish stock market, market impact, high-frequency trading, crashes

Authors:

Jim Domeij (24262)
Oscar Krieg (24552)

Tutors:

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

Introduction

Hasbrouck (2017) describes how markets have changed. At the end of the last century, financial markets started to become computerized and trading that had previously taken place on physical trading floors on exchanges moved over to electronic trading. Today the era of floor trading is over, and most markets are electronic. O'Hara (2015) calls this the new "high frequency world", where he defines High-Frequency Trading (HFT) as computer- and strategy-based trading performed at *very* high execution speeds. There is an ongoing arms race in the HFT industry. Firms invest billions just to gain an edge and beat competitors by just a few milliseconds according to Budish et al. (2015).

Due to its inherent demand for speed, HFT must employ trading algorithms, which are essentially computer programs provided with predetermined schedules for order execution. According to Aldridge (2013), the best HFT systems perform extensive analysis and simulation of different transaction-relevant characteristics, such as trading costs, based on high-frequency, tick-by-tick data. Intuitively, accurate quantitative models of transaction costs are of great importance to finance professionals.

O'Hara (1995) defines market microstructure as the analysis of how specific trading mechanisms affect the price formation process, while Harris (2002) explains it as: *"the branch of financial economics that investigates trading and the organization of markets."* Centralized regulation, digitalized clearing systems and stock exchange requirements can affect these trading mechanisms and can thus constitute explanatory factors in the study of market microstructure.

Within market microstructure, a more recent theory called market microstructure invariance has been postulated by Albert S. Kyle and Anna A. Obizhaeva. The definition of market microstructure invariance is based on the intuition that financial markets transfer risk in business time, and Kyle and Obizhaeva (2016) define it as: *"the hypotheses that the distribution of risk transfers ('bets') and transaction costs are constant across assets when measured per unit of business time"*. Andersen et al. (2018) and Kyle et al. (2016) define bets as risk transfers and meta-orders, executed through trades (alternatively a series of trades), and as part of a more long-run strategy for profiting on market development. The major benefit of this new invariance theory is that it can render accurate predictions of different quantities in market microstructure by using empirically observable variables in financial markets, such as returns variance, price and trade volume. However, for the market microstructure invariance theory to hold, the invariance relation(s) must hold empirically across all assets and time periods.

As a manifestation of its usefulness, several of Kyle and Obizhaeva's postulated invariance assumptions and hypotheses have already been proven empirically. For instance, Kyle and Obizhaeva (2017) and Kyle and Obizhaeva (2019) introduce scaling laws with important implications for two transaction cost components. Kyle and Obizhaeva (2017) confirm for Russian and US stocks that there exists a proportional relationship between relative bid-ask spread costs and an illiquidity measure, where the latter is proportional to the cube root of return variance to dollar volume. Secondly, Kyle and Obizhaeva (2019) derive a market impact model and find it superior to conventional wisdom in predicting price declines.

Given this promising new research, we want to test whether the proposed relationship between the relative bid-ask spread and illiquidity, as well as the market impact model holds for the Swedish stock market. In light of the increased demand for precise transaction cost models in HFT, we are particularly interested to investigate if the bid-ask spread model holds in intraday and high-frequency settings. Our first research question is:

Is the relative size of the bid-ask spread proportional to illiquidity, the latter defined as the cube root of the ratio of return variance to SEK volume, for Swedish stocks on a high frequency?

We answer this question by using HFT data spanning the time period between 2010[1] and 2020, gathered on a millisecond frequency and then aggregated on a five-minute level. Since we obtain tick-by-tick data,[2] this involves processing roughly 375 million rows of data. As a proxy for the Swedish stock market, we use OMXS30 index constituents.

Interestingly, we find that the proportional relationship between the relative bid-ask spread and illiquidity statistically holds for the full time period, although it only holds for one year on a standalone basis. This result provides us with the insight that the transaction cost model holds for several stock markets with different market structures, as well for extremely high frequencies. Prima facie, this should facilitate valuable benchmarks for HFT models, since this concerns a quantifiable market microstructure relationship based on empirically observable variables. The quantitative nature of this relationship makes it suitable for processing by a computerized algorithm.

Our second research question concerns whether an invariance-implied market impact formula renders more accurate predictions of price declines during the Great Coronavirus Crash of 2020[3] than conventional wisdom does. Here we compare the predictive power of the invariance model to conventional wisdom, as outlined by Kyle

[1] 2010-02-08.

[2] *Tick-by-tick* data shows every trade that occurs.

[3] Coy (2020).

and Obizhaeva (2019), on a more recent stock market crash. In this regard we specifically answer the research question:

Is invariance theory better to predict price declines during the Great Coronavirus Crash of 2020 than conventional wisdom?

We obtain millisecond trading data for OMXS30 index futures and stocks for 2019 and the spring of 2020. Since we only have one crash, we, similar to Kyle and Obizhaeva (2019), do not achieve statistically valid results in regard to this question, but this was not our initial intention either. However, we manage to prove that the invariance-implied market impact model generates predictions much closer in magnitude to actual price declines during the Great Coronavirus Crash of 2020, as compared to conventional wisdom.

The remaining paper is structured as follows. Section 1 defines fundamental concepts in market microstructure invariance to help the reader better grasp the content of this thesis. Section 2 touches upon previous research most adjacent to our research questions and illustrates our sources of inspiration as well as our contribution. Section 3 provides the theoretical market microstructure invariance framework which our research is based on. Section 4 formally postulates our empirical hypotheses, research design and the epistemological and theoretical motivation behind these. Section 5 describes our data and empirical variables; section 6 presents our empirical findings and lastly section 7 gives our concluding remarks.

1. Fundamental Concepts

Due to the inherent complexity and quantitative nature of the research area of market microstructure invariance, we here attach a brief explanation of a few central concepts. Understanding of these concepts are vital in the further reading of this thesis, and the explanation of these therefore deserves its own section.

Market Microstructure Invariance

Market microstructure invariance is the name of a theory or a set of theories within the research field market microstructure. Market microstructure is concerned with analysis of dynamics and mechanisms of the price formation process, see Zovko (2008). In general, *market microstructure invariance* is a theory hypothesizing that different relationships between these trading mechanisms and components of the price formation process are invariant across assets and time when certain transformations are made, see Kyle et al. (2016). These transformations are further elaborated below, but they essentially imply that one must adjust for security-specific risk transfer intensity by transforming calendar time to *business time*.

Bets

Kyle et al. (2016) defines a bet as a transaction of assets "*intended to produce an idiosyncratic gain*" in accordance with an institutional trader's view on market risk development. Market microstructure invariance, as proposed by Kyle and Obizhaeva (2016), implies viewing institutional trading in financial markets as trading games. Executed trades are essentially risk transfers, through which taking on and unloading risk are part of a more general and long-run profit-maximizing strategy. Due to the more long-run and overall character of bets, a single bet is often executed through a series of trades. The strategies underpinning bets are privately held information, and therefore hard to observe empirically. Trades are therefore usually used as a proxy for bets in the study of market microstructure invariance.

Business Time

Business time is another important concept within the study of market microstructure invariance. Kyle et al. (2016) defines business time as the expected calendar time between the arrival of bets in the market, with calendar time measured in conventional time units such as years, days, minutes etc. Stock-specific velocity is a related measure and is defined as the expected arrival rate of bets, and thus the inverse of business time. These definitions facilitate the economically relevant and illustrative interpretation that more liquid stocks, i.e. stocks trading at a higher frequency, have a higher velocity and consequently shorter business time than less liquid stocks. Kyle and Obizhaeva (2019) theorize that the length of business time is the only difference between different financial markets, and Kyle and Obizhaeva (2016) postulates that, in order for market microstructure invariance to hold, one must account for this difference by transforming calendar time into business time.

Bid-Ask Spread

The bid-ask spread is one measure of transaction costs. It originates from the fact that buyers and sellers in financial markets place orders at prices consistent with their view on market development and risk. In order for a trade to occur, traders in financial markets must meet the counterparty's price. The difference between the best (highest) bid price and the best (lowest) ask price is called the bid-ask spread.

Liquidity and Illiquidity

Liquidity of a financial asset is a measurement for how easily the asset can be converted to cash or cash equivalents. Illiquidity is simply the inverse of the chosen liquidity measure. Kyle et. al (2016), however, use a slightly different definition for liquidity and illiquidity. They define illiquidity as the average dollar cost of executing an average bet divided by the dollar value of the average bet. The illiquidity is a volume-weighted transaction cost and Kyle et. al (2016) express this as a percentage of the value of the average bet.

Tick Size

As described by Harris (2002), tick size refers to the smallest allowed price movement in a financial market. The bid-ask spread, measured in absolute terms, can never be smaller than the tick size. Tick sizes are set by regulators and exchanges. In. January 2018, the European Securities and Markets Authority (ESMA) enacted the second Markets in Financial Instruments Directive (MiFID II),[4] containing harmonizing regulation regarding securities markets in the EU. This regulation concerns, among other things, tick sizes.

[4] Directive 2014/65/EU of The European Parliament and of The Council of 15 May 2014 On Markets In Financial Instruments and Amending Directive 2002/92/EC and Directive 2011/61/EU.

2. Literature Review

This thesis revolves around three main articles, focusing on the derivation of key variables and analysis methodology within market microstructure invariance, intraday trading invariance, and transaction costs. The exhibited literature lays the theoretical foundation for the thesis and provides testable quantitative predictions for bid-ask spreads and market impact costs.

2.1. Dimensional Analysis, Leverage Neutrality, and Market Microstructure Invariance

Kyle and Obizhaeva (2017), applies dimensional analysis – an analytical method originating from physics, which simplifies scientific inferences by restricting the analysis to a set of assumed explanatory fundamental variables. Their analysis is supplemented by the financial equivalent of physics conservation laws, specifically the no-arbitrage principle of leverage neutrality. They ultimately pose invariance assumptions implying that some quantities are constant across assets and time. Kyle and Obizhaeva (2017) test their predictions empirically for US and Russian stocks. They find that scaling laws for bid-ask spreads and linear market impact costs, both comprised of their defined illiquidity measure, is consistent with financial data.

This article provides us with the necessary derivations, methodology and quantitative models. Our empirical analysis is performed for a different financial market, for a different time period and for high-frequency data. Our thesis provides additional empirical evidence for market microstructure invariance, in accordance with the central notion that invariance should hold across all assets and across time.

2.2. Intraday Trading Invariance in the S&P 500 Futures Market

Andersen et al. (2018) use tick-by-tick data on E-mini S&P 500 futures contracts to study how different trading activity variables are related on an intraday level. They propose a new theory called Intraday Trading Invariance (ITI) as a high-frequency extension of market microstructure invariance. This is new since market microstructure invariance theory usually involves bets over longer execution horizons (e.g. months). Andersen et. al (2018) test several invariance relationships by averaging tick-by-tick data, aggregated on a one-minute level, across all trading days in their sample. Their results confirm that the ITI hypothesis holds for the E-mini S&P 500 futures market.

We are most interested in the methodology from Andersen et al. (2018) for measuring variables, aggregating them and testing the above-mentioned ITI hypothesis. By applying the ITI analysis methodology on our own sample, we confirm that the specific scaling laws derived in Kyle and Obizhaeva (2017) hold for high-frequency data.

2.3. Large Bets and Stock Market Crashes

Kyle and Obizhaeva (2019) apply an invariance-implied model to predict market impacts during stock market crashes, characterized by intense selling pressure and large price declines. They compare their predictions with conventional wisdom, which assumes that price elasticity in financial markets is such that selling 1% of market capitalization has a market impact of less than 1%. The main difference between conventional wisdom and the invariance-implied model is that the latter accounts for business time. The consideration of business time implies larger price impacts in more liquid markets. Kyle and Obizhaeva (2019) contrast their model with conventional wisdom to compare the relative accuracy of these models on five large crash events between 1929 and 2010 and find their model superior.

This article provides us with a market microstructure invariance model for predicting price movements induced by large sell orders. We also test the relative accuracy of the invariance model for the more recent Coronavirus Crash of 2020.

3. Theoretical Background

As one purpose of this thesis is to investigate the potentially proportional relationship between the relative size of bid-ask spreads and illiquidity, we must formally derive a consistent and logical measure of these market microstructure phenomena. The computation of the bid-ask spread is generally accepted, and even though we transform this to a relative measure by relating it to the mid-price, this part of the theory is relatively straight-forward.

Market impact costs are incurred for traders when prices move against them as a result of their own trading activities. The result of buy and sell orders is that the trader will face new adverse prices if he or she continues to trade in the same direction. The difference between the previous execution price and the new one the trader would face, is the market impact cost. Bid-ask spread is another transaction cost and can be interpreted as the cost that the trader would incur if he or she would simultaneously buy and sell the same security at the best prevailing ask and bid price, respectively. The bid-ask spread is defined as the difference between the best bid and ask prices.

The other purpose of this thesis is to test a market microstructure invariance model for predicting price developments induced by intense selling pressure during stock market crashes. The core premise and hypothesis presented in Kyle and Obizhaeva (2019) is that market microstructure invariance can predict price declines in stock market crashes more accurately than conventional wisdom. In this section we briefly derive the invariance-implied market impact formula from Kyle and Obizhaeva (2019).

Section 3.1 presents the relevant theoretical background, relevant derivations and necessary assumptions underpinning our used illiquidity measure and its proposed relationship to the relative size of the bid-ask spread. This largely mimics Kyle and Obizhaeva (2017). Section 3.2 includes a brief constructive presentation of the invariance-implied market impact formula and the intuitive rationale behind its superiority to conventional wisdom.

3.1. Dimensional Analysis, Leverage Neutrality, and Market Microstructure Invariance

Kyle and Obizhaeva (2017) use dimensional analysis, an analytical method which originates from physics which facilitates inferences regarding complex relationships between variables. The method simplifies these complicated relationships by imposing certain restrictions, which reduces the number of explanatory variables to a restricted set of fundamental variables.

An illustrative example of dimensional analysis in physics is the case of the British physicist G. I. Taylor's estimation regarding the atomic bomb developed within the Manhattan project (Trinity Test, 1945, New Mexico). More specifically, Taylor was asked to estimate the developed bomb's energy output, see Deakin (2011). The details of the Manhattan Project were classified and thereby not available to Taylor, he had only the press images from the test explosion to rely on. Taylor applied the fundamental units of physics, such as length, mass, time and energy, along with assumed scaling arguments, to formulate a proposed relationship between the radius of the sphere, elapsed time since the explosion and the energy of the explosion. Once the project details were later exposed, Taylor was proven to have arrived at an estimate which was very close to reality.

Analogously to G.I. Taylor, Kyle and Obizhaeva (2017) apply dimensional analysis in finance to derive scaling laws for bid-ask spread and market impact costs. The financial application of dimensional analysis initially requires deciding on (assuming) the right set of explanatory variables, and identifying the relevant base dimensional quantities in which these variables are measured. The financial equivalents of physics' base dimensions are by Kyle and Obizhaeva (2017) identified as value, measured in currency (e.g. USD), asset quantity, measured in number of contracts or securities, and time, measured in conventional time units such as years, days, minutes, milliseconds and so forth.

3.1.1. Dimensional Analysis Application

When institutional traders execute bets through order placement in financial markets, they move market prices. This occurs due to increased demand as well as information asymmetries and adverse selection.[5] Buy orders of institutional sizes signal to the market that a stock might be undervalued, while a sell order conversely signals that a stock might be overvalued. Buy bets puts upward pressure on prices while sell bets work in the opposite direction, and we denote this market impact cost of a bet G. Specifically, G denotes the expected price impact cost of executing a bet of a certain size, expressed as a fraction of the unsigned value traded through the bet. This is, however, a general notion of transaction costs and there are at least three special cases, which G takes on in the form of a more specific transaction cost. For now we just mention them here as bid-ask spread, linear market impact cost and square-root market impact cost, but below we show the specific derivation of the general transaction cost formula, as outlined by Kyle and Obizhaeva (2017).

Market microstructure is concerned with trading mechanisms at work in financial markets. The beauty of financial markets, and a fact that simplifies quantitative analysis and derivation of relationships, is that there is a limited amount of observable, quantitative variables at work. A lot of trading-relevant variables also consist of

[5] Ong et al. (2020) illustrates some of these effects by proving that institutional ownership has effects of IPO pricing.

combinations of more fundamental variables whose existence can be taken as given. By deducing the financial equivalents of the physics base dimensions, the isolation of fundamental variables becomes even more intuitive.

Going forward, stock- and time-specific variables are denoted with subscripts *j* and *t* to indicate stock *j* at time *t*. Kyle and Obizhaeva (2017) assume that the market impact cost G_{jt} is a function of the fundamental financial variables asset quantity Q_{jt}, asset price P_{jt}, volume V_{jt}, return variance σ_{jt}^2. We initially assume that the market impact cost G_{jt} is any function of these variables. For clarity, the difference between Q_{jt} and V_{jt} is that Q_{jt} represents asset quantity relating to an individual bet, and V_{jt} represents trading activity in the market by measuring the amount of trading in a specific stock over a specific time interval. Kyle and Obizhaeva (2017) also introduce the expected bet cost as an explanatory variable for the market impact cost G_{jt}. They claim that the assumption underpinning this inclusion is less intuitive than the inclusion of the other chosen variables. The formula for the unconditional expected bet cost is:

$$C \coloneqq E\{G_{jt}P_{jt}|Q_{jt}|\} \tag{1}$$

As can be seen in equation (1), C is the expected product of the stock- and time-specific market impact cost G_{jt} and price P_{jt} for executing a bet of the unsigned bet size Q_{jt}. The reason for the absence of subscripts with regards to C, is that it is assumed to be constant across all stocks and time. This constitutes an additional market microstructure invariance assumption. In summary, Kyle and Obizhaeva (2017), assume that these five variables are explanatory with regards to G_{jt}, and thus constitute arguments in the function of G_{jt}, rendering us the assumed formal relationship:

$$G_{jt} \coloneqq g\left(Q_{jt,}\, P_{jt}, V_{jt}, \sigma_{jt}^2, C\right) \tag{2}$$

At this stage, we have yet to assume or conclude the exact composition of function *g* in equation (2). We initially restrict our analysis to this set of five assumed explanatory variables. This is a key step in dimensional analysis. Kyle and Obizhaeva (2017) mention that due to the inherent procedure of dimensional analysis, some explanatory variables might have been omitted. Since we also apply dimensional analysis, we might also suffer from omitted variables and misspecifications. This is, however, mitigated by the fact that our intention is to essentially test their proposed market microstructure invariance relationship. Therefore, despite our derivation of these relationships and inherent variables, we take this specification as exogenously given, provided by existing theory in the field of market microstructure invariance.

The assumed explanatory variables are divided into (1) variables measured in base dimensional quantities, and (2) those measured in dimensionless quantities. Dimensionless quantities are quantities other than the base dimensional quantities, and also includes constants. For clarification purposes, we here summarize the basic dimensional quantities in which our variables are measured:

$[G_{jt}] = 1;$ $[Q_{jt}] = shares;$

$[P_{jt}] = currency/shares;$ $[V_{jt}] = shares/day;$

$[\sigma_{jt}^2] = 1/day;$ $[C] = currency.$

The brackets indicate that what is stated on the right side of the equation is the basic dimensional quantity in which the stated variables are measured in. The quantity of 1 mentioned in relation to G_{jt} indicates that this quantity is dimensionless, since it is a fraction not consisting of basic dimensional quantities.

All variables that are not measured in the financial, basic dimensional quantities (i.e., currency, asset quantity or time), or combinations hereof, are considered dimensionless. As can be seen in the notation above, in this setup, G_{jt} is the only dimensionless variable. Through this exposition one key benefit of choosing these explanatory variables begins to emerge. Most of these variables are accessible from financial markets data and measured in easily comprehensible and generally accepted basic dimensional quantities.

The three dimensional variables Q_{jt}, P_{jt} and σ_{jt}^2 are dimensionally independent and span the three basic dimensions. Further, V_{jt} and C can be expressed by combining these three variables. The fact that the basic dimensions in finance are three – currency, quantity and time – and we have five variables, allows us to replace V_{jt} and C by introducing two newly constructed dimensionless variables. The new dimensionless variables L_{jt} and Z_{jt} are constructed by adding a dimensionless scaling constant m^2:

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

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

For the general methodology (the Buckingham theorem), see Kyle and Obizhaeva (2017). Following the subscript notation above, L_{jt} and Z_{jt} represent the stock- and time-specific liquidity measure and scaled bet size, respectively. More specifically, Z_{jt} is the size of a bet expressed as a multiple of the mean of the unsigned bet size. Since the market impact cost G_{jt} is dimensionless, and for consistency in units on both sides of equation (2), the dimensional quantities Q_{jt}, P_{jt} and σ_{jt}^2 on the right hand-side must be dropped as well.

$$G_{jt} := g\left(Q_{jt,} P_{jt}, V_{jt}, \sigma_{jt}^2, C\right) = g\left(L_{jt}, Z_{jt}\right) \tag{5}$$

In equation (5) we note that G_{jt} is expressed using only two dimensionless quantities and thus independent of the units of measurement. This implication is based on the assumption that rational investors are not confused by units of measurement. Kyle and Obizhaeva (2017) claim that rational investors do not suffer from money illusion, and thereby do not change their behavior when shares are split. Nor do rational investors confuse calendar time with business time. Conclusively, it is irrelevant if transaction costs are measured in different currencies, since these currencies are interchangeable.

3.1.2. Leverage Neutrality

Kyle and Obizhaeva (2017) supplement their derivation with a universal, financial equivalent of a physics conservation law, leverage neutrality. This principle is a generally accepted sub-principle of the Modigliani and Miller principle of no-arbitrage. In a transactional setting, this principle implies that adding a riskless asset, i.e. cash, to a transaction in risky securities, will not alter the cost of such a transaction, since cash is a riskless asset. Leverage neutrality also involves the capital structure irrelevance theory as outlined by Modigliani and Miller (1958), which states that a firm's undertaking of riskless debt does not alter the cost of trading the firm's securities.

Kyle and Obizhaeva (2017) makes a supplementing interpretation of leverage neutrality, from a macro perspective. This interpretation implies that margin requirements and repo haircuts are irrelevant for the economics of trading, i.e. irrelevant for transaction costs. To bring clarity to the terminology, repo haircuts and margin requirements are distinct from each other but both concern collateralization. Repo haircuts, or repurchase agreements haircuts, are the conservative valuation discount applied to collateral by, for instance, a securities dealer. Margin requirements concern maintaining the value of posted collateral, and if this drops too low in, for instance, a margin account, the potential debtor is contractually forced to post additional collateral, see Dang et al. (2013).

By invoking these implications of leverage neutrality, Kyle and Obizhaeva (2017) show that, since liquidity L_{jt} scales proportionally with the cash included in a transaction and is thereby not leverage neutral, impact cost G_{jt} scales inversely to liquidity L_{jt} in equation (6). Z_{jt} is leverage neutral since it represents the scaled bet size:

$$G_{jt} \coloneqq g\left(Q_{jt}, P_{jt}, V_{jt}, \sigma_{jt}^2, C\right) = g\left(L_{jt}, Z_{jt}\right) = \frac{1}{L_{jt}} f\left(Z_{jt}\right) \tag{6}$$

3.1.3. Market Microstructure Invariance Assumptions

The third step in Kyle and Obizhaeva (2017) is formulating different invariance hypotheses. They initially assume that the function f in the general transaction cost formula in equation (6) takes the form of a power function with exponent w:

$$f\left(Z_{jt}\right) = \bar{\lambda}\left|Z_{jt}\right|^{w} \tag{7}$$

By assuming different values for this exponent in equation (7), they make predictions about bid-ask spread costs, linear and square-root market impact costs. For our thesis, bid-ask spreads and linear market impact costs are most interesting. Therefore, we show them below.

Bid-ask spread cost ($w = 0$):

$$G_{jt} = constant \cdot \frac{1}{L_{jt}} \tag{8}$$

Linear market impact ($w = 1$):

$$G_{jt} = constant \cdot \frac{P_{jt}\left|Q_{jt}\right|}{C L_{jt}^2} \tag{9}$$

Bid-ask spreads are not dependent on bet size Q_{jt} and the proportionality constant between bid-ask spreads and illiquidity in equation (8) should be constant and dimensionless across all assets. On the other hand, we see that linear market impact depends on bet size in equation (9). Finally, Kyle and Obizhaeva (2017) proceed to set the market impact cost G_{jt} in equation (8) to the relative bid-ask spread S_{jt}/P_{jt}, i.e. the size of the bid-ask spread as a fraction of the mid-price. The mid-price is the price located in the middle of the prevailing best bid and ask prices. Like G_{jt}, the relative bid-ask spread S_{jt}/P_{jt} is a dimensionless fraction. This relationship provides the theoretical foundation for our first hypothesis.

3.2. Invariance-Implied Market Impact in Stock Market Crashes

Kyle and Obizhaeva (2019) build on the core premise that the only difference between different financial markets is that business time passes faster in more liquid markets than less liquid markets. They use the concepts of transformation of calendar time to business time and the inversely correlated measure velocity to predict price impacts observed in financial markets during times of intense selling pressure, and contrasts this to predictions according to so-called conventional wisdom.

3.2.1. Predicting Market Impact Through Conventional Wisdom

The rationale underpinning the conventional wisdom model for predicting market impact, is by Kyle and Obizhaeva (2019) said to be the Capital Asset Pricing Model (CAPM), as well as competitive and efficient financial markets. Since the CAPM model implies that all stocks included in the market portfolio contributes to the level of systematic risk, trading a certain quantity of a single stock for a limited time carries significantly less rewardable risk than the entire market. According to PwC (2021), the market risk premium historically has varied around 5-7%. In light of this, conventional wisdom states that price elasticity in financial markets is such that a sell order corresponding to 1% of market capitalization generates a price impact of less than 1%. The conventional wisdom is by Kyle and Obizhaeva (2019) interpreted mathematically such that the expected unsigned market impact $\Delta P/P$, approximated logarithmically as $\Delta \ln P$, can be expressed as the fraction of average daily volume (ADV) traded, wherein Kyle and Obizhaeva (2019) assume that a stock has an asset turnover of a 100% over the course of a trading year (250 days):

$$\Delta \ln P \approx \frac{\Delta P}{P} = \frac{Q}{250 \cdot V} \tag{10}$$

Where $V$is the share volume over a certain interval (normally a calendar day). In application, this leads to Kyle and Obizhaeva (2019) taking the currency volume PV over the total market cap of the relevant instrument(s).

3.2.2. Predicting Market Impact Through Market Microstructure Invariance

Kyle and Obizhaeva (2019) emphasize the importance of accounting for the asset-specific business time, which they mean is the only difference between different financial markets. Henceforth we return to the conventional notation wherein we let P denote price, V denote daily share volume, and σ returns volatility. The derivation of the invariance-implied market impact formula in Kyle and Obizhaeva (2019) rests upon the core premise within market microstructure invariance - the transformation of calendar time to business time.

In our review of Kyle and Obizhaeva (2017) we derive the market microstructure conjecture that expected transaction costs of executing bets are constant across markets, contingent on that the bets transfer the same currency risk $(P \cdot \sigma)$ per unit of business time. Kyle and Obizhaeva (2019) pair this transaction cost invariance conjecture with the invariance of returns volatility, given the transformation of calendar time to business time. Kyle and Obizhaeva (2019) and Kyle et. al (2016) define trading activity W as a measure of the rate at which a market transfers risk:

$$W = P \cdot V \cdot \sigma \tag{11}$$

Trading activity W is measured in dollars/day$^{3/2}$ due to the fact that it is comprised of currency volume PV, measured in dollars/day, and volatility σ measured in units per day$^{1/2}$. Kyle and Obizhaeva (2019) further introduces a benchmark stock with business time corresponding to one calendar day, and is said to approximately correspond to a stock at the bottom of the S&P 500 index. This stock is introduced to standardize units in the step-by-step formula derivation, and is kept throughout the derivation of the linear price impact model. This stock has a price $P^* =$ \$40/share, expected trading volume $V^* =$ 1,000,000 shares/day and daily volatility $\sigma =$ 2%/day$^{1/2}$.

Kyle and Obizhaeva (2019) derive the relationship between trading activity W and variables observable in financial markets, and invoke the deduced proposition that invariance holds for the probability distribution of random bet sizes, contingent upon scaling by business time over one business day. They denote the used scaling constant Z^*. The introduction of this scaling constant "forces" Q to become invariant. This is necessary when deriving an invariance-based price impact model, since this allows for model application across markets. Once Kyle and Obizhaeva (2019) derive Q as invariant, the equated bets are said to render equivalent price impacts, denoted $\Delta P/P$, contingent of them being expressed as a fraction of returns volatility over one business day. In light of this, it becomes possible to estimate market impacts across markets with different characteristics, by accounting for differences in business time.

Using a log-linear impact model $\Delta P/P \approx \Delta lnP$, the linear price impact formula, along with the involvement of the benchmark stock constants, renders an invariance-implied market impact formula for empirical testing:

$$\Delta \ln P = \frac{\bar{\lambda}}{10^4} \cdot \left(\frac{PV}{40 \cdot 10^6}\right)^{\frac{1}{3}} \cdot \left(\frac{\sigma}{0.02}\right)^{\frac{4}{3}} \cdot \frac{Q}{(0.01)V} \tag{12}$$

Kyle and Obizhaeva (2019) estimate the term $\bar{\lambda}/10^4$ to be five basis points (0.05%), and is for the purpose of our application of this model, exogenously given. This function takes unsigned bet size Q, dollar volume PV, and return volatility σ as arguments. The result is the predicted invariance-implied market impact.

In summary, through their step-by-step derivation, Kyle and Obizhaeva (2019) derive an invariance-implied market impact formula to be compared to conventional wisdom in predicting price developments in stock market crashes. They hypothesize and confirm that differences in business time constitutes an explanatory consideration necessary for a higher level of accuracy in predicting price movements caused by large sell orders preceding stock market crashes.

4. Hypotheses and Research Design

This section includes our explicitly formulated hypotheses and the theoretical motivation behind them. Further, this section presents our research design, which includes a description of our practical approach, coupled with the theoretical motivations behind it.

4.1. Hypotheses

We formulate two market microstructure invariance hypotheses about transaction costs on the Swedish stock market. First, we formulate our invariance hypothesis regarding the bid-ask spread on the Swedish stock market:

The relative size of the bid-ask spread has a proportional relationship to the asset-specific illiquidity measure, defined as the cube root of the ratio of return variance to SEK volume, for Swedish stocks observed on a millisecond frequency (tick-by-tick data) for the period of 2010-02-08 to 2020-12-31.

$$\ln\left(\frac{S_{jt}}{P_{jt}}\right) = constant + 1 \cdot \ln\left(\frac{1}{L_{jt}}\right) \tag{13}$$

Where

$$\frac{1}{L_{jt}} = constant \cdot \left[\frac{P_{jt}V_{jt}}{\sigma_{jt}^2}\right]^{-\frac{1}{3}} \tag{14}$$

And

L = Liquidity

S = Absolute bid-ask spread

P = Mid-price

V = Share volume

σ = Volatility

In order to statistically test the hypothesis in equation (13), we choose the log relative bid-ask spread as the dependent variable and the log illiquidity measure as the independent variable. Let β denote the coefficient of the log illiquidity measure. We then formulate our null and alternative hypothesis as follows:

$H_0: \beta = 1$

$H_1: \beta \neq 1$

Under the null hypothesis, the invariant relationship between the bid-ask spread and illiquidity measure holds for the Swedish stock market, i.e. the slope is equal to 1. This is similar to the manner in which Kyle and Obizhaeva (2017) postulate their hypothesis. In this setting, failure to reject the null hypothesis would confirm our market microstructure invariance hypothesis. This hypothesis corresponds to the first special case of the general transaction cost formula in equation (8). The hypothesis above is also coherent with the general market microstructure notion argued by Kyle and Obizhaeva (2016), that for invariance to truly hold, the proposed relationship must hold for all assets and across time. A successful outcome with regards to this hypothesis would add yet another market and another time period to the area of applicability of this transaction cost model. Furthermore, this would also extend the applicability of this transaction cost to an intraday setting, which would be particularly valuable due to the importance of precise and granular transaction cost models in HFT.

Our second hypothesis concerns the invariance-implied linear impact cost in the Swedish stock market, which is another type of transaction cost. It is also the second special case of the general transaction cost formula in equation (9). We hypothesize the following:

The invariance-implied market impact model comes closer in magnitude to predicting price declines in the Swedish stock market during the Coronavirus Crash of 2020, than conventional wisdom.

Here our null hypothesis is that the linear market impact model is able to predict the magnitude of price declines during large stock market crashes more accurately than conventional wisdom. The alternative hypothesis is that the magnitude of the price fall suggested by conventional wisdom comes closer to the actual price impact. We are not able to statistically test this second set of hypotheses for two reasons. First, these are the same hypotheses as formulated in Kyle and Obizhaeva (2019) and they do not provide an appropriate test statistic. Second, we have only looked at one stock market crash and therefore, we do not have a sufficient amount of data to run statistical tests. However, being able to quantitatively show the relative superiority of this impact model would provide future stock market crash stakeholders with a complementary model, which at least historically has generated more accurate predictions than conventional wisdom. Since risk models generally make use of sensitivity and scenario analyses, this could provide yet another simulation pattern and dimension to such models.

4.2. Research Design

In the step-by-step dimensional analysis in section 3, we pose necessary assumptions and show derivations of some formulas in a condensed manner to promote readability. The attentive reader will notice that this step-by-step methodology in relevant parts closely mimics Kyle and Obizhaeva (2017), whose methodology in turn largely originates from Barenblatt (1996) for non-financial applications. One reason for the conformity between our thesis and Kyle and Obizhaeva (2017), is that Albert S. Kyle and Anna A. Obizhaeva are the main proponents of invariance theories within market microstructure. The second and dominant reason is that we wish for a potentially successful empirical result of our research to add to the accumulated empirical findings in this area, by adding another market and longer time period.

Cross-Sectional Analysis

A cross-sectional research method implies collecting data from different participants at one point in time, see Allen (2017). The participants in our case are the different stocks that are, at any point in time, included in the Swedish stock index OMXS30. We gather stock-specific data on certain frequencies in order to construct a proxy for the Swedish stock market (the reason for us not to study the index directly is that it is not securitized and data on the index lacks certain relevant trading variables). Since market microstructure invariance theories say that supposedly invariant relationships should hold across all assets and time, we perform statistical analysis for both individual stocks and on an aggregate level across all stocks as in Kyle and Obizhaeva (2017). We also choose to look at the aggregate sample of stocks in order to test if invariance holds for our proxy for the Swedish stock market.

Intraday Analysis

Bets are inherently part of more general, long-run strategies, and in Kyle and Obizhaeva (2017) the analysis is performed on a daily frequency and over the course of one year. To test the durability of the invariance theory even more, we wish to see whether invariance between the relative size of the bid-ask spread and illiquidity holds for an intraday frequency as well. We do this by applying the adjusted invariance analysis method of intraday trading invariance, a term coined by Andersen et al. (2018). We aggregate our empirical variables by certain time intervals, e.g. five-minute intervals, to study intraday trading patterns for Swedish stocks similarly as Andersen et. al (2018) do for E-mini S&P 500 futures. With this approach we can capture the time-of-day effect for intraday trading. The considerations underpinning our chosen level of frequency (length of time interval), entails balancing between a high enough level of granularity and sufficient number of observations (trades) in each time interval. Longer intervals gives a greater number of observations, but are less precise and less likely to capture short-lived intraday patterns/fluctuations. On the other hand, shorter intervals may only contain a few observations, leading to larger standard errors of arithmetic averages.

Correlational Analysis

In general, our first and main research question implies investigating and empirically testing a proposed relationship or correlation between two variables. Kyle and Obizhaeva (2017) do not investigate causality between the two variables, and neither do we. The results of this thesis can thus not facilitate inferences regarding whether the bid-ask spread causes illiquidity or vice versa. However, it can provide normative benchmarks and predictive transaction cost models.

Ordinary Least Squares Regression

The correlational character inherent in the bid-ask spread hypothesis renders regression analysis as the evident choice in determining whether there exists a statistical relationship between them. The assumption of proportionality, also incorporated in our research question and hypothesis, further facilitates the choice of methodology. The proportionality, wherein the illiquidity measure is multiplied with a constant, renders that this relationship regardless of intercept and thus the value of this constant, should render a correlational, linear relationship. This motivates us to apply the ordinary least squares (OLS) regression model. This regression model generates a best fitted line based on made observations, which in our case are the values of the relative size of the bid-ask spread and the values of our computed illiquidity measure respectively, for each time interval.

Confidence Intervals and Robustness Checks

As part of the empirical statistical testing of our hypothesis we also construct a confidence interval, centered around our regression slope. This allows us to validate the theory if the confidence interval spans a range which includes our desired slope of 1. This confidence interval could also, upon failure to validate our postulated market microstructure invariance hypothesis, provide a quantitative basis for qualitative statements regarding *economic closeness* to the slope 1.

We also perform robustness checks wherein we test the sensitivity of our obtained results by making certain alterations to assumptions and inputs, without changing the general methodology. These alterations are thus not made to the assumptions underpinning the theoretical framework or the derivations of our variables, but rather to our testing procedure. We analyze different time periods, both in regard to our hypothesis test but also in regard to our construction of confidence intervals, where we aggregate our data on different time periods.

Non-Statistical Test of Invariance-Implied Market Impact Model

For our second hypothesis, concerning the relative accuracy of the invariance-implied market impact model we do not have sufficient data for performing a statistically valid hypothesis test. Such a test is however not undertaken by Kyle and Obizhaeva (2019) either, and this is probably due to the lack of a large enough data sample. We suffer from the same problem. We can thus make quantitatively based statements regarding the relative accuracy, or closeness in magnitude, of the invariance-implied market impact model. We perform this test in the same way as Kyle and Obizhaeva (2019).

We compare our predictions according to the conventional wisdom model to the predictions generated by the invariance-implied market impact model from Kyle and Obizhaeva (2019). The relative accuracy of these two methods are illustrated by comparing them to the actual price decline. By accounting for which method renders the prediction closest to the actual price decline, we are able to make quantitative, but not statistically supported, statements regarding the relative accuracy of these predictive models.

As a substitute for statistical robustness checks, and since we only have one stock market crash in our sample, we will divide the period for this crash (later defined as the time period between the all-time high listing and the bottom listing of 2020) into several shorter sub-periods. We will also investigate how sensitive the invariance-implied market impact model is to input estimates, which is why we gradually alter the computation ("calibration") period for the average daily SEK volume and volatility used as inputs in the model.

5. Data

This section presents our data, outlined with regards to each of our respective hypotheses. The disposition is motivated by the use of different data for different hypotheses, and our first and main hypothesis require a lot more data processing. In each subsection, we first present the data in general, then our cleaning, processing and aggregating of this data. Lastly, we present identified and potential data issues as well as summary statistics regarding each dataset.

The subsection regarding futures data follows, despite the absence of subheadings, the sequence outlined in this introduction. The absence of subheadings is motivated by the simpler data processing undertaken in regard to the futures data.

5.1. Stock Data

We use high-frequency trading data for the Swedish stock market to test the market microstructure invariance between percentage bid-ask spread and the illiquidity measure in Kyle and Obizhaeva (2017). We use data from NASDAQ OMX Stockholm (formerly known as the Stockholm Stock Exchange) from February 2010 to December 2020, which was obtained through the Swedish House of Finance National Research Data Center. The reason for us to choose this period, and not include older periods, is that NASDAQ changed their trading system from SAXXESS to the more advanced Inet, and data from SAXXESS is not freely available.

For this thesis we use a reconstructed limit order book for each stock included in the OMXS30 index. A reconstructed limit order book includes the first and second level bid price, bid volume, ask price, ask volume, volume traded, and trade price timestamped to the millisecond. For this thesis, we only look at the first level of the limit order book, i.e. the top bid price, bid volume, ask price and ask volume. The top bid and top ask prices are used to calculate the bid-ask spread and the trade price and trade volume are used to compute the dollar volume (turnover).

The data covers all the historical constituents in the OMX Stockholm 30 Index (OMXS30) between 2010-02-08 to 2020-12-31. The index consists of the 30 most traded stocks on NASDAQ OMX Stockholm and tracks the Swedish stock market. Its composition is reviewed semi-annually, and it happens that incumbent constituents are replaced by new stocks with a higher turnover in the market. We have adjusted the data for all exclusions and inclusions that occurred during our time period of interest, to include these new constituents. It deserves mentioning that the constituent companies were largely the same for the entire time period, and in total we have reviewed 35 companies.

We only consider trades between 09:00 and 17:25 (CET),[6] which corresponds to the continuous trading hours for equities on NASDAQ OMX Stockholm.[7] We also exclude 51 trading days from our sample because of erroneous start and end times inconsistent with official trading hours.

Since Sweden is a member of the EU, NASDAQ OMX Stockholm must follow certain trading rules which apply to all stock exchanges in the EU. Since January 2018, the tick size of all stocks listed on EU exchanges is regulated by MiFID II. The tick-size, the minimum increment in bid or ask price, of each stock traded on NASDAQ OMX Stockholm depends on both the price of the stock and its "liquidity band", according to MiFID II. The average daily number of trades (ADNT) in a stock on "the relevant market in terms of liquidity" determines which liquidity band it belongs to. The European Securities and Markets Authority (ESMA) decides what the relevant market is and calculates the average daily number of trades for each stock on an annual basis. This information is then released to the public.

5.1.1. Data Limitations

We do not have data on realized spread costs and use quoted bid-ask spreads instead. However, according to Kyle and Obizhaeva (2017) the predictions about bid-ask spreads from dimensional analysis, leverage neutrality and market microstructure invariance should still hold for quoted bid-ask spreads as well.

Our choice of time period and frequency differs from Kyle and Obizhaeva (2017), which use minute-data from January to December 2015 aggregated on a daily basis. We study a longer period of time, more than 10 years of financial data on a millisecond level, which we then aggregate to five-minute intervals. In addition, since we use a longer time period our dataset is posed to contain more stocks split events, which in turn change the price of our selected stocks. Fortunately, neither stock splits nor different frequencies of data have an impact on these securities-specific measures, such as liquidity, according to Kyle and Obizhaeva (2017).

Our granularity-induced purpose of intraday data would have been promoted by the use of futures on the Swedish stock market as opposed to stocks, with relevant necessary adjustments to our hypothesis and research question.

---

[6] http://www.nasdaqomxnordic.com/tradinghours

[7] NASDAQ OMX Stockholm is open between 09:00-17:30 CET, but there is no order-matching during the last five minutes (17:25-17:30).

Index futures are more liquid than stocks,[8] hence they would be beneficial to study, especially with regards to the high level of granularity induced by our high-frequency data. A possible motivation behind the usage of E-mini futures in Andersen et. al (2018) is that such financial instruments are highly liquid and trades intensively, which adds value when gathering tick-by-tick data.

Futures on OMXS30 exist, but there are two key reasons why we do not include these in our sample. Futures can be divided into regular futures and mini-futures. Mini-futures, like the ones studied in Andersen et. al (2018), is more easily accessible to the large masses of retail investors due to the lower quantity of underlying,[9] and thereby lower trading thresholds, contained in each contract. Mini-futures on the OMXS30 index were first issued in the summer of 2020, and to study these we would have had to settle for a short time period. We deemed this time period as insufficient, and macroeconomically significant events during this particular time period, including the occurrence of both a pandemic as well as a turbulent presidential election in the US, would likely not render such a sample representative of normal market conditions.

Regular futures on the OMXS30 index do, to our knowledge, exist for our time period of interest. However, the reason for not including these in our sample instead of the shares constituting the index itself, is due to limited access to a longer time series of data. There are, however, benefits of studying individual stocks instead. By studying individual companies on the Swedish stock market, we can gain insights about how the regression slope varies between the different companies in the index.

Despite the fact that we constructed this index-like proxy ourselves in order to obtain the necessary trading-related data, a large index is what generally is referred to as "the market". The OMXS30-included stocks make up approximately 60.5% of the market capitalization of the entire Stockholm Stock Exchange.[10] Despite the fact that the sample most coherent with the terms "Swedish stocks" or "Swedish stock market" would be a sample of all publicly traded Swedish stocks, our chosen sample strikes, in our opinion, a good balance between feasibility, representation, data quality and precision, respectively.

---

[8] Nasdaq OMXS30 statistics.

[9] Kurov and Lasser (2004) highlights the main differences between regular futures and mini futures.

[10] Data obtained through the FinBas database, accessed through Swedish House of Finance. Market capitalization EOD per 2020-12-30 for OMXS30 stocks still included per 2021-05-12, and EOD data per 2020-12-30 for Stockholm Stock Market.

5.1.2. Construction of Main Empirical Variables

In this thesis we mainly use variables similar to Kyle and Obizhaeva (2017). This is motivated by the fact that financial market microstructure variables usually have a conventional and established composition, but also due to the fact that reproducing certain assumptions are necessary in order to test whether the bid-ask spread has a proportional relationship to illiquidity for Swedish stocks on an intraday level.

Let the subscript j denote stock j, and subscript t denote each five-minute intraday interval during a full trading day (only continuous trading), that starts at 09:00 and closes at 17:25. As our entire sample contains 35 Swedish stocks, subscript j ranges from 1 to 35, and subscript t ranges from 1 to 101. Our main empirical variables are order size Q_{jt}, SEK 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}.

We use tick-by-tick data with timestamp to the millisecond. For each stock we aggregate trades recorded on a millisecond frequency for each consecutive five-minute intraday interval for all trading days. More specifically, we calculate time-weighted averages for the mid-price P_{jt} and the percentage bid-ask spread S_{jt}/P_{jt} within each five-minute intraday interval. However, time-weighting is not possible for trades that are recorded at the exact same timestamp in our data sample. This happens when a marketable order is executed against several standing limit orders. Our remedy for this is to use a volume-weighting similar to the approach by Pohl et. al (2020), before the time-weighting. When multiple trades are recorded at the exact same timestamp on the millisecond t_k, the observed mid-prices P_{jt_k} and percentage bid-ask spreads $S_{jt_k}/P_{jt_k}$are volume-weighted to give us one observation per variable for that timestamp and stock. The time-weighting follows what is done by McInish and Wood (1992). The procedure for calculating these volume- and time-weighted averages is explained in more detail in Appendix C.

5.1.3. Summary Statistics

Our empirical variables of interest are the formula input variables for the measurements whose relationships are tested. Apart from the directly computable mid-price P_{jt}, we focus on volume V_{jt} and volatility σ_{jt} as these are our main inputs for our illiquidity measure $1/L_{jt}$, whose composition is further derived and elaborated as one of our empirical variables in Appendix C. For these variables, we divide each trading day in 101 five-minute intervals, assign them a chronological number based on their occurrence during the trading day, so that the time interval 09:00-09:05 (CET) receives the number 1 since this is the first five-minute interval of a trading day, 09:05-09:10 receives the number 2 since this is the second five-minute interval of a trading day, and so on. Once this is done we group all variables by each five-minute interval and average them across our entire time period (February 2010-December 2020), in order to obtain average variable measurements for each five-minute interval.

TABLE 1. Summary Statistics for OMXS30 Constituents

	Mean	Median	Min	Max
SEK Volume	2 763 481	1 687 939	42	912 627 641
Volume	24 155	11 179	2	6 856 204
# Trades	26	18	2	4 292
Trade Size	929	643	1	93 548
Relative Bid-Ask Spread	8.15	7.59	1.11	340.07
Illiquidity	8.85	8.16	0.00	176.94
Volatility	0.23	0.15	0.00	20.39

Table 1 – The summary statistics are reported for all OMXS30 stocks and the entire continuous trading session. Relative bid-ask spread is the ratio between bid-ask spread and the quoted mid-price, and it is measured in 10^{-4}. Illiquidity is defined as the cube root of the ratio of return variance to SEK volume, and it is measured in 10^{-5}. Volatility is *realized* volatility and is calculated by summing squared returns between each quoted mid-price within every five-minute interval. Realized volatility is then averaged across all five-minute intervals and stocks, and finally annualized by multiplying with $101 \cdot 252$. All other variables are reported on a five-minute level.

This thesis focuses on share volume V_{jt}, volatility σ_{jt}, trade size Q_{jt}, bid-ask spreads S_{jt}/P_{jt} and illiquidity $1/L_{jt}$. We plot these variables together with the number of trades $N_{jt} = V_{jt}/Q_{jt}$ in figure 1. All six trading variables show interesting intraday patterns. Perhaps surprisingly, bid-ask spreads in our sample do not follow the U-shaped pattern reported for NYSE stocks by Brock and Kleidon (1992) and Lee et. al (1993), nor the reversed J-shaped pattern for NYSE stocks documented by McInish and Wood (1992). In figure 1 bid-ask spreads for OMXS30 constituents drastically drop during the first minutes after market open, slowly decrease through the rest of the trading day (except a small jump around 15:30 when US markets open), and then begin to slightly fall around 16:00. This more closely matches the (weakly) reversed J-shape for bid-ask spreads in DAX30 constituents that Hussain (2011) finds, with the main difference being that our spreads decrease rather than increase near the close. Chan et. al (1995) also see intraday bid-ask spreads for NASDAQ securities narrow near the close, in contrast to classical U-shape observed for NYSE securities. Their explanation for these different patterns is structural differences between the two markets (NASDAQ is a dealer market, while NYSE is a specialist market). Traders' inventory control of securities in dealer markets is said to be consistent with narrow spreads preceding the close, while specialists' market power is thought to contribute to wider spreads at the end of the day. However, in figure 1 we can see a U-shaped pattern for share volume, similar to that of Brock and Kleidon (1992) and Chan et. al (1995), as well as a U-shape for the number of trades, matching the intraday curve in Chan et. al (1995). In our sample, realized volatility has a reverse J-shaped pattern: it falls significantly directly after the market opens and this decreasing pattern continues until later in the afternoon when volatility suddenly jumps up at 15:30. It stays close to this new level for the rest of the day. This is in line with results found by Hussain (2011) and Chan et. al (1995).

FIGURE 1. Intraday Trading Patterns for OMXS30 Constituents

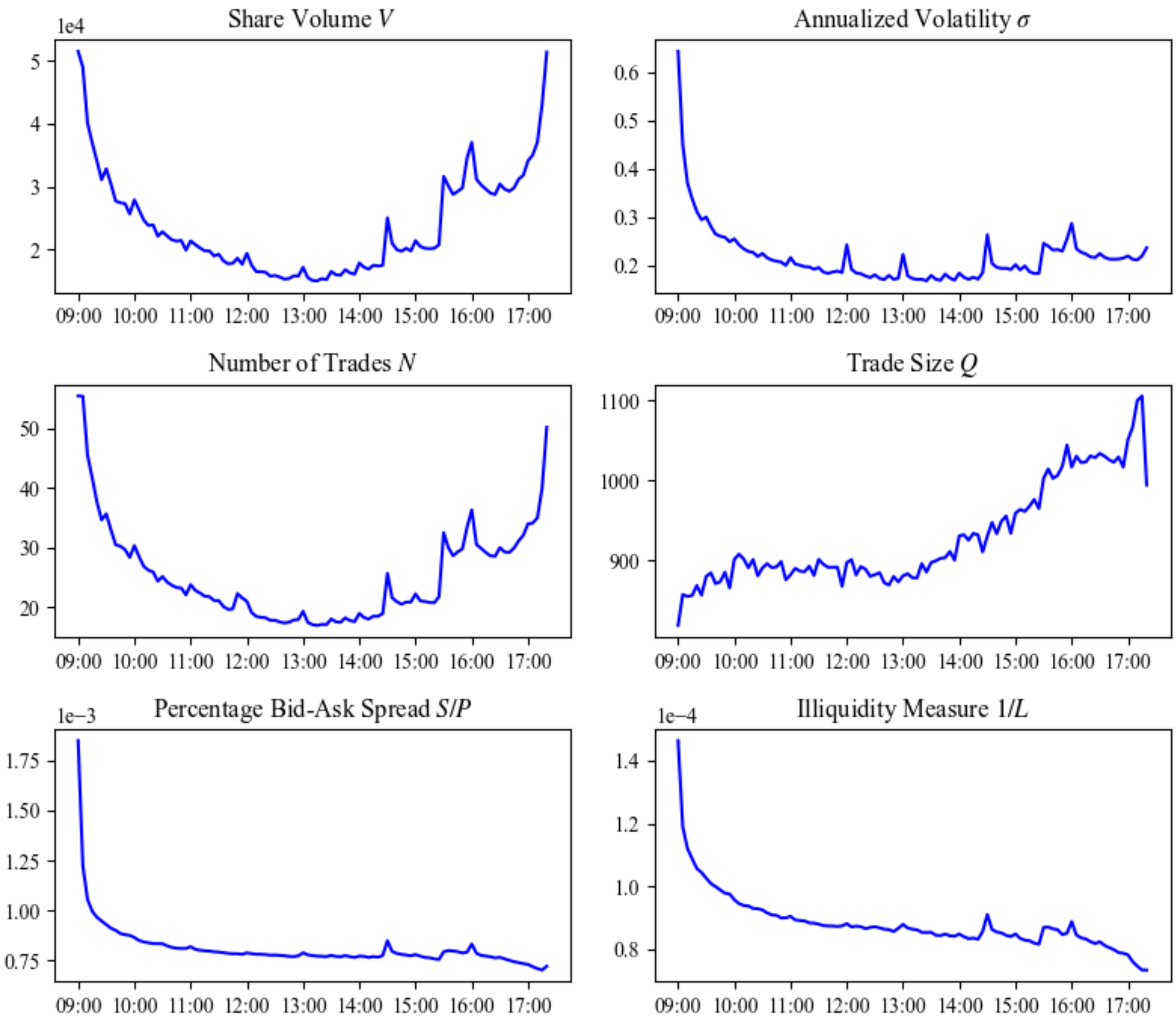

Figure 1 – The figure shows averages (per five minutes) across all stocks and days for volatility, share volume, trade size, number of trades, illiquidity and percentage bid-ask spread. Illiquidity is defined as the cube root of the ratio of return variance and SEK volume. Percentage bid-ask spread is defined as the dollar bid-ask spread divided by the mid-point price between the best bid and ask price. The figure shows how these averages develop over the average trading day for the average stock. All times are in CET.

5.2. Futures Data

To test the invariance-implied market impact formula in equation (12) we use three datasets. The first dataset consists of tick-by-tick (millisecond) trade data on OMXS30 derivatives (futures and options) from January to March 2020 and is provided by NASDAQ's office in Vilnius. This sample covers seven different futures contracts with maturities in January, February, March, April, May, June and September 2020, and includes 62 trading days. The second dataset contains a time series of the daily closing value of the OMXS30 index between January and March 2020, and is provided by NASDAQ. The third dataset is a time series of the continuous OMXS30 futures contract and is retrieved directly from Refinitiv Eikon. This particular continuous contract always gives the nearest month futures contract.

Data cleaning procedures include what follows. The data contains trades in both futures and options contracts. Since we are only interested in the futures market in order to explain price change on the stock market, we drop all option trades (191,432 trades) and are left with 9,361,855 futures trades. We only have futures trades data for the first quarter of 2020.

The main empirical variables for our analysis of the 2020 Coronavirus Crash are average daily volume (ADV), daily volatility σ of futures prices, futures gross sales in SEK, stock gross sales in SEK and the actual price impact of the OMXS30 index. For the alternative conventional wisdom model we use the same gross sales as for the invariance-implied model, but for market capitalization we use the market capitalization of the entire Stockholm Stock Exchange.

We construct average daily volume (ADV) by first summing the total SEK volume of futures each day and then taking the mean of daily SEK volumes (for x days preceding 18 February 2020). To calculate daily volatility σ, we compute daily returns from the price of the continuous nearest month OMXS30 futures contract (for x days preceding 17 February 2020). Then we take the standard deviation of the returns to get daily volatility. We calculate futures (gross) sales for a specified time period as the total sales in SEK during the same period. We calculate actual price impact for the same period as for gross sales, by computing the percentage change between end-of-day (EOD) index values of the OMXS30.

TABLE 2. Summary Statistics for OMXS30 Futures

Exp. Date	Daily Volatility	Daily # Trades	Daily Volume	Total Volume	Total # Trades	Trade Size
2020-01-17	0.80%	53 830	197 732	2 175 047	592 134	4
2020-02-21	0.83%	51 101	178 877	6 439 565	1 839 619	4
2020-03-20	2.31%	101 643	287 181	14 646 236	5 183 812	3
2020-04-17	5.27%	51 331	122 725	4 172 643	1 745 251	2
2020-05-15	17.05%	62	231	2 076	562	4
2020-06-18	8.35%	29	452	7 227	468	15
2020-09-18	14.85%	3	3	8	8	1

Table 2 – This table presents daily volatility, average daily number of trades, average daily volume, total volume, total number of trades and average trade size for futures with the expiration dates stated in the left column. Volatility is calculated by taking the standard deviation of EOD futures returns. Total (daily) volume is calculated for each futures series (based on expiration date) by summing (averaging) the number of traded contracts for all trades within the entire time period.

6. Empirical Findings

This section is devoted to illustrating the empirical results that we obtain through the analysis of our sample data. The first subsection presents our results with regards to our first hypothesis, proposing a proportional relationship between the relative size of the bid-ask spread and illiquidity in the Swedish stock market. We combine descriptive statistics with illustrating plots and a written review of our results. This is concluded by a brief statement concerning whether or not the hypothesis is statistically confirmed.

The second subsection concerns our second hypothesis, whether or not the invariance-implied market impact model is more accurate than conventional wisdom in predicting sell order-induced price impacts in the Swedish stock and futures market during the Great Coronavirus Crash of 2020. As aforementioned, we do not have large enough data samples to confirm or reject this hypothesis statistically, hence we merely validate the results from Kyle and Obizhaeva (2019) for another dataset, obtained from a more recent crash.

6.1. The Relative Bid-Ask Spread vs. Illiquidity

It deserves mentioning that the results presented here are the aggregated results for the entire time series. We present results for individual years in Appendix B as an indication of level of robustness. The reason behind this disposition is that our hypothesis concerns the Swedish stock market taken over the entire time period. The sampling of data for individual stocks is also a necessary step in our compilation of the stocks contained in the OMXS30 index, but not particularly relevant to the hypothesis test. Therefore results for individual stocks are placed in Appendix A. As a reminder, the investigated and supposedly proportional relationship is $\ln(S_{jt}/P_{jt}) = constant + 1 \cdot \ln(1/L_{jt})$. Illiquidity $1/L_{jt}$ is in turn defined as the cube root of the ratio of return variance σ_{jt}^2 to SEK volume $P_{jt}V_{jt}$.

Table 3 presents the results from the OLS regression for $\ln(S_{jt}/P_{jt})$ against $\ln(1/L_{jt})$ on the aggregate sample, i.e. the entire time period (2010-02-08 to 2020-12-31) and includes 35 different stocks. The number of stocks covered in this sample is greater than 30, because it includes all stocks that at some point during the time period of interest have been included in the OMXS30 index. Each trading day contains 101 five-minute intervals, which renders a total of $35 \cdot 101 = 3535$ data points.

TABLE 3. OLS Regression

Start Date	End Date	# Obs	c	β	Se(c)	Se(β)	R^2	CI Limits 95% LL	95% UL
2010-02-08	2020-12-31	3535	2.0908	0.9881	0.0827	0.0089	0.7786	0.9707	1.0054

Table 3 – The table presents the OLS regression of $\ln\left(S_{jt}/P_{jt}\right) = c + \beta \cdot \ln\left(1/L_{jt}\right)$. For each of the 35 stocks contained in the OMXS30 index over the full time period, 101 aggregated five-minute intervals are used in the regression, corresponding to 3535 points. Coefficients c and β, coefficient of determination R^2 and 95 % confidence interval for the aggregate slope β are estimated for the entire period (2010-2020).

Figure 2 plots log bid-ask spread $\ln(S_{jt}/P_{jt})$ against log illiquidity $\ln\left(1/L_{jt}\right)$. Each point represents each (non-overlapping) five-minute interval between 09:00 and 17:25 (CET), aggregated from our millisecond data, for each stock. Each color of the points represents a different stock. A fitted line for each stock is also plotted in the same color as the corresponding points. This graph is supposed to capture the time-of-day effect for each stock on its average trading day. Hence, the number of trading days are aggregated into five-minute intervals and do not contribute to the number of points plotted in the graph. The blue fitted line for the aggregate data, which we obtain through our OLS regression, is $\ln\left(S_{jt}/P_{jt}\right) = 2.0908 + 0.9881 \cdot \ln\left(1/L_{jt}\right)$. Our confidence interval ranges from 0.9707 to 1.0054, while the coefficient of determination R^2 is 0.7786. As a benchmark, we add a red line with a slope of 1, with its exact equation being $\ln\left(S_{jt}/P_{jt}\right) = 2.2022 + 1 \cdot \ln\left(1/L_{jt}\right)$ as predicted by the market microstructure invariance hypothesis.

FIGURE 2. Aggregate OLS Regression

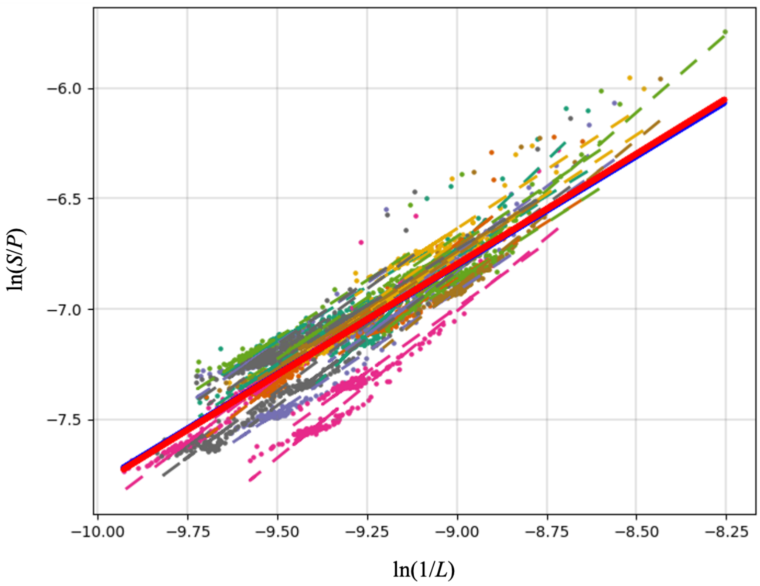

Figure 2 – This figure plots log percentage bid-ask spread against log illiquidity for each of the 35 stocks in the OMXS30 index from 2010-02-08 to 2020-12-31. Different colors represent different stocks. For each stock, 101 points for all the five-minute intervals between 09:00 and 17:25 are plotted. Each point

represents a five-minute bid-ask spread and illiquidity average of a particular stock. All stocks have been averaged across all days. Dashed lines represent OLS regressions for individual stocks. The red line $\ln(S_{jt}/P_{jt}) = 2.2022 + 1 \cdot \ln(1/L_{jt})$ is added as a benchmark and represents the predicted regression slope of one according to our hypothesis. The slope is fixed at exactly one and only its intercept is estimated for the aggregate sample. The blue fitted line for the aggregate sample is $\ln(S_{jt}/P_{jt}) = 2.0908 + 0.9881 \cdot \ln(1/L_{jt})$. The blue and red line almost overlap because our estimated blue line is very close to the hypothetical red line.

As we can see, both by the aggregate slope coefficient β stated in Table 3 as well as by the spread of the observations in Figure 2, the estimated slope is not exactly one. However, if we instead turn our attention to our statistically constructed 95 % confidence interval, which ranges from 0.9707 to 1.0054, we see that the value one is contained in this interval. On a 95 % confidence level the invariance prediction that the slope is one (null hypothesis) is not statistically rejected. Thus, we can confirm, with a relatively high level of certainty, that invariance holds for our entire sample, i.e. the Swedish stock market over the period 2010-2020,[11] when we aggregate trades on a five-minute frequency. Kyle and Obizhaeva (2017) use daily averages for US and Russian stocks in their log-linear regression. For the Russian stocks, their fitted line is $\ln(S_{jt}/P_{jt}) = 2.093 + 0.998 \cdot \ln(1/L_{jt})$. For the US sample, their fitted line is $\ln(S_{jt}/P_{jt}) = 1.011 + 0.961 \cdot \ln(1/L_{jt})$. Hence, our results are similar to the results reported by Kyle and Obizhaeva (2017), in particular their fitted line for the Russian stocks.

For the 35 dashed lines in figure 2, which are fitted based on data on each of the 35 individual stocks, we can see that the slope estimates vary around and above one. In table 4 we see that the slopes of individual stocks vary between 0.9867 and 1.5776, with a mean slope of 1.1454. See Appendix A for the slopes of all stocks. These estimates are higher than implied by invariance. For Russian stocks, Kyle and Obizhaeva (2017) get individual slopes which are instead lower than one, and for both US and Russian stocks they report larger spread in individual slopes.

TABLE 4. OMXS30 Constituents Regression Statistics

			Individual slopes		
Start date	**End date**	**# Obs**	Max	Min	Mean
2010-02-08	2020-12-31	3535	1.5776	0.9867	1.1454

Table 4 – The table presents the maximum, minimum and mean individual slopes across all OMXS30 constituents and days for the period 2010-02-08 to 2020-12-31.

As a robustness check over subperiods, we perform yearly aggregate regressions across all stocks. The results from these regressions can be found in Appendix B. We

[11] 2010-02-08 to 2020-12-31.

statistically reject the hypothesis that the slope is one for all individual years 2010-2020, except for 2017. However, for 2010, 2014, 2016 and 2018 the slope is economically close to one. There are many explanations for these results since market microstructure relationships are affected by a number of factors, such as regulation, clearing mechanisms, digitized trading, tick sizes and so on.

This simple model does not account for these market frictions. Based on the regression statistics presented in Appendix B, we notice, for instance, that a shift occurs 2017/2018 and the yearly slope seems to increase thereafter. This shift interestingly coincides with the EU's introduction of MiFID II in January 2018, which harmonized tick sizes on stock exchanges throughout the union. This causes us to suspect that regulation concerning minimum tick sizes could affect bid-ask spreads.

Kyle and Obizhaeva (2017) also derive a more general bid-ask spread formula that accounts for the minimum tick size K_{jt}^{min} and lot size Q_{jt}^{min}, which may be theoretically relevant here:

$$\frac{S_{jt}}{P_{jt}} = \frac{1}{L_{jt}} s\left(\frac{K_{jt}^{min} L_{jt}}{P_{jt}}, \frac{Q_{jt}^{min} \sigma_{jt}^2 L_{jt}^2}{V_{jt}}\right) \tag{15}$$

However, they do not test this extended formula empirically. We test this new specification for the bid-ask spread by taking a naive approach, assuming that $s(x, y) = xy$, and taking the natural logarithm of both sides in equation (15). We get a log-linear relationship with three explanatory variables: log illiquidity, log ratio between actual and optimal tick size, and log ratio between actual and optimal lot size. Optimal tick and lot size are derived from invariance theory. We have been able to show that these two new variables improve R^2 of our regression from 0.7786 to 0.8110.

Lastly, it deserves mentioning that although we manage to confirm our postulated market microstructure invariance hypothesis regarding the relationship between the relative size of the bid-ask spread and our illiquidity measure, we are not able to make any causal inferences. Similar to Kyle and Obizhaeva (2017) we are only able to state that market microstructure invariance holds for the relationship between the relative size of the bid-ask spread and illiquidity, when empirically tested for the Swedish stock market between the years 2010[12] and 2020. Our research design does not allow for us to make any inferences regarding whether the relative size of the bid-ask spread *causes* illiquidity, or vice versa.

[12] 2010-02-08.

6.2. Invariance Predictions for the Coronavirus Crash of 2020

Our second hypothesis concerns the relative accuracy of the invariance-implied market impact model developed by Kyle and Obizhaeva (2019). They confirm their model's superiority to conventional wisdom in predicting price declines in previous stock market crashes, by testing it empirically for five different market crashes, with the latest being in 2008.

During the Great Coronavirus Crash of 2020 the OMXS30 index fell from 1 900.28 (2020-02-19) to 1 292.27 (2020-03-23) and thus experienced a price decline of 32%. The 23rd of March represented the lowest mark of 2020 and by studying the subsequent development, it can be said that the recovery started after this date.

FIGURE 3. OMXS30 2020-02-10 to 2020-04-30

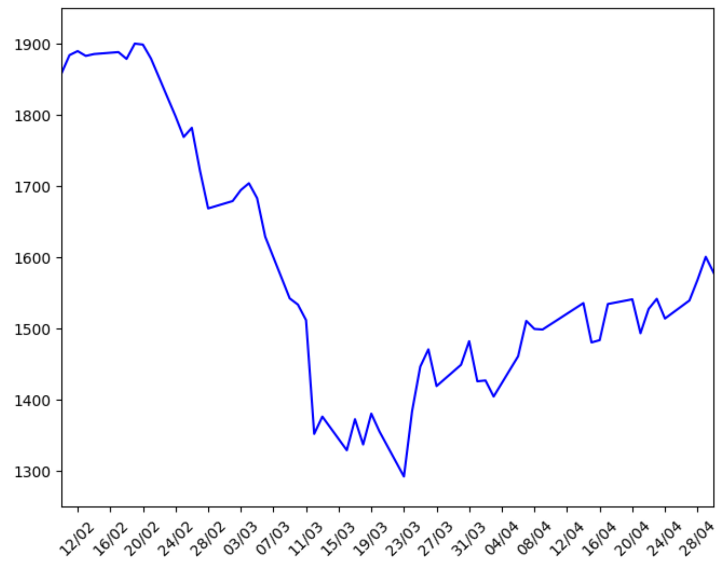

Figure 3 – The figure plots the crash of the OMXS30 index during the outbreak of Covid-19 and the beginning of its subsequent recovery. The graph includes dates between 2020-02-10 and 2020-04-30, with the highest and lowest point being 1 900.28 (2020-02-19) and 1 292.27 (2020-03-23), respectively. Specific dates are plotted along the x-axis, with corresponding index values along the y-axis.

Emergency Information from Swedish Authorities (2020) reported Sweden's first registered case of Covid-19, the disease caused by the Coronavirus, at the end of January 2020. The OMXS30 index, however, continued to rise until the top listing on the 19th of February. The 11th of March the World Health Organization (WHO) declared Covid-19 a pandemic, and consequently, on the 12th of March the OMXS30 index experienced a decline of 10.57%. Table 5 illustrates the development of the index during the period between 2020-02-19 and 2020-03-23:

TABLE 5. OMXS30 2020-02-19 to 2020-03-23

	19/2	21/2	28/2	6/3	13/3	20/3	23/3
OMXS30 Value (EOD)	1 900.28	1 879.10	1 668.84	1 629.00	1 376.46	1 356.54	1 292.27
Change	-	-1.11%	-11.19%	-2.39%	-15.50%	-1.52%	4.67%

Table 5 – The table presents EOD values of the OMXS30 index for selected dates during the 2020 Coronavirus Crash. The first date is the all-time-high listing on the 19th of February, and subsequent values are obtained for the following Fridays, with the last value being the bottom listing on Monday the 23rd of March.

We apply the predictive models to three separate time periods during the Coronavirus Crash of 2020. First, we compare the models to the actual price decline for the entire time period, from top to bottom listing, i.e. from 2020-02-19 to 2020-03-23. This is motivated by a desire to capture the entire downward movement. Due to the construction of the invariance-implied market impact model, wherein a longer measurement period renders higher predicted price impacts (unless ADV is calculated based on a rolling window), we also choose to divide this period into two sub-periods as well. The first sub-period begins on 2020-03-09 and ends on 2020-03-23. This period contains a major price decline, as well as the individual days with the steepest value decline (12th and 23rd of March). The second sub-period consists of a single day, 12th of March, motivated by the fact that it was the largest daily decline (-10.57%) during the Coronavirus Crash of 2020 for the Swedish market. Table 6 contains a summary of the actual and predicted price declines for our selected time and sub-periods respectively:

TABLE 6. Comparison of Predictive Market Impact Models

		Price Decline		
Period	**%ADV**	Actual	Invariance Prediction	Conventional Prediction
19/2-23/3	111.56%	32.00%	17.97%	0.75%
9/3-23/3	63.47%	18.83%	10.66%	0.42%
12/3	12.95%	10.57%	0.98%	0.04%

Table 6 – The table summarizes actual and predicted price declines during the Coronavirus Crash of 2020. The table compares actual price declines of the OMXS30 index during this period with predicted price declines by market microstructure invariance theory and conventional wisdom, respectively. The invariance predictions are based on the first model calibration period 2020-02-03 to 2020-02-18.

Below we present the invariance-implied market impact predictions for all three time periods in tables 7-9. Each table includes the input variables we use for the invariance-implied market impact formula. We make predictions for stocks and futures separately as well as combined, with the combined impact as our main result. All currency values are expressed in 2020 SEK, thus without the use of a GDP deflator. The model inputs are calibrated for three different time periods, the first one covering the period from the beginning of February 2020 up until the day before the top listing on the 19th of February. The second calibration period is intended to correspond to trading days in 2020 preceding the crash, and is therefore measured from the first trading day of 2020, i.e. the 2nd of January. The third and last calibration period contains an entire year, starting at the 18th of February 2019.

TABLE 7. Invariance-Implied Market Impact 2020-02-19 to 2020-03-23

	Model Calibration Period Start Date		
	2020-02-03	2020-01-02	2019-02-18
OMXS30 Futures ADV (in 2020 SEK)	51 952 694 128	42 702 353 675	25 772 486 352
Stocks ADV (in 2020 SEK)	283 246 622	266 112 126	234 661 810
Daily Volatility	0.8609%	1.0112%	0.9105%
19/2 - 23/3 OMXS30 Futures Sales (in 2020 SEK)	52 372 466 565	52 372 466 565	52 372 466 565
19/2-23/3 OMXS30 Stock Sales (in 2020 SEK)	5 900 397 707	5 900 397 707	5 900 397 707
19/2 - 23/3 Sales (of ADV)	111.56%	135.62%	224.06%
Price Impact of Sales Combined	**17.97%**	**24.40%**	**28.82%**
Price Impact of OMXS30 Future Sales	16.37%	22.31%	26.46%
Price Impact of OMXS30-included Stock Sales	47.80%	56.82%	54.80%

Table 7 – The table shows the invariance-predicted price impact of futures sales of SEK 52 372 466 565 and stock sales of SEK 5 900 397 707, combined as well as separately, between 2020-02-19 and 2020-03-23 given average daily SEK volume and volatility between 2020-02-18 and 2020-02-03, 2020-01-02, and 2019-02-18 respectively. Conventional wisdom predicts a price impact of 0.75% for the same time period. The actual price decline over this period was 32%. Note that all price impacts are unsigned values, and all price impacts stated represent price declines.

TABLE 8. Invariance-Implied Market Impact 2020-03-09 to 2020-03-23

	Model Calibration Period Start Date		
	2020-02-03	2020-01-02	2019-02-18
OMXS30 Futures ADV (in 2020 SEK)	51 952 694 128	42 702 353 675	25 772 486 352
Stocks ADV (in 2020 SEK)	283 246 622	266 112 126	234 661 810
Daily Volatility	0.8609%	1.0112%	0.9105%
9/3 - 23/3 OMXS30 Futures Sales (in 2020 SEK)	29 811 566 530	29 811 566 530	29 811 566 530
9/3-23/3 OMXS30 Stock Sales (in 2020 SEK)	3 343 432 828	3 343 432 828	3 343 432 828
9/3 - 23/3 Sales (of ADV)	63.47%	77.16%	127.48%
Price Impact of Sales Combined	**10.66%**	**14.71%**	**17.58%**
Price Impact of OMXS30 Futures Sales	9.67%	13.39%	16.05%
Price Impact of OMXS30-included Stock Sales	30.81%	37.87%	36.23%

Table 8 – The table shows the invariance-predicted price impact of futures sales of SEK 29 811 566 530 and stock sales of SEK 3 343 432 828, combined as well as separately, between 2020-03-09 and 2020-03-23 given average daily SEK volume and volatility between 2020-02-18 and 2020-02-03, 2020-01-02, and 2019-02-18 respectively. Conventional wisdom predicts a price impact of 0.42% for the same time period. The actual price decline over this period was 18.83%. Note that all price impacts are unsigned values, and all price impacts stated represent price declines.

TABLE 9. Invariance-Implied Market Impact 2020-03-12

	Model Calibration Period Start Date		
	2020-02-03	2020-01-02	2019-02-18
OMXS30 Futures ADV (in 2020 SEK)	51 952 694 128	42 702 353 675	25 772 486 352
Stocks ADV (in 2020 SEK)	283 246 622	266 112 126	234 661 810
Daily Volatility	0.8609%	1.0112%	0.9105%
12/3 OMXS30 Futures Sales (in 2020 SEK)	2 451 416 950	2 451 416 950	2 451 416 950
12/3 OMXS30 Stock Sales (in 2020 SEK)	439 916 324	439 916 324	439 916 324
12/3 Sales (of ADV)	5.54%	6.73%	11.12%
Price Impact of Sales Combined	**0.98%**	**1.38%**	**1.67%**
Price Impact of OMXS30 Futures Sales	0.83%	1.17%	1.43%
Price Impact of OMXS30-included Stock Sales	4.73%	6.07%	5.75%

Table 9 – The table shows the invariance-predicted price impact of futures sales of SEK 2 451 416 950 and stock sales of SEK 439 916 324, combined as well as separately, on 2020-03-12 given average daily SEK volume and volatility between 2020-02-18 and 2020-02-03, 2020-01-02, and 2019-02-18 respectively. Conventional wisdom predicts a price impact of 0.04% for the same time period. The actual price decline over this period was 10.57%. Note that all price impacts are unsigned values, and all price impacts stated represent price declines.

As we can see from tables 6-9, the invariance-implied market impact model generates predictions that are closer in magnitude to actual price declines than the conventional wisdom model. This is in a general sense coherent with the findings of Kyle and Obizhaeva (2019).

7. Concluding Remarks

Market microstructure invariance theory provides several empirical hypotheses, many of which are presented by Kyle and Obizhaeva (2016), Kyle and Obizhaeva (2017) and Kyle and Obizhaeva (2019).

In this thesis we use high-frequency trading data for Swedish stocks and index futures to examine scaling laws for bid-ask spreads and linear market impact costs derived through dimensional analysis, leverage neutrality and from principles of market microstructure invariance. Kyle and Obizhaeva (2017) finds that these scaling laws hold for data on bid-ask spreads for US and Russian stocks, and Kyle and Obizhaeva (2019) apply scaling laws to explain price declines during stock market crashes.

We examine the proportional relationship between the bid-ask spread and an illiquidity measure, the latter composed of observable financial market variables, and find that it also holds for the Swedish stock market for more than a decade. In the spirit of invariance, we thus add yet another time period and another financial market to the discourse of market microstructure invariance.

We extend the daily analysis used by Kyle and Obizhaeva (2017) to analyze bid-ask spreads in a high-frequency setting, similar to the intraday analysis of S&P 500 E-mini futures performed by Andersen et al. (2018). We combine dimensional analysis with intraday trading invariance, the high-frequency version of market microstructure invariance, and the main finding of this thesis is that the proportional relationship between bid-ask spreads and illiquidity holds on an intraday level as well.

Furthermore, we show that the proportional bid-ask spread-illiquidity relationship still holds for OMXS30, a proxy for the Swedish stock market, despite a different tick size regime. Russian and US stock markets are affected by different institutional microstructure details or market frictions, such as minimum tick sizes, than the Swedish stock market. This thesis provides additional statistical evidence of the explanatory power of such details for bid-ask spreads to the market microstructure literature.

In addition to bid-ask spread costs we investigate linear market impact, which is another transaction cost component. Kyle and Obizhaeva (2019) construct an invariance-implied market impact model with the purpose of predicting the price impact of order pressure. This model accounts for security-specific differences in the length of business time, the inverse of bet frequency. When testing this model for the recent Coronavirus Crash of 2020, we find that it predicts price declines more accurately than conventional wisdom. However, we as Kyle and Obizhaeva (2019), are not able to statistically validate the general applicability of this model due to a small sample size.

This thesis uses regression analysis and in light of our findings, we are therefore not able to make statements regarding causality. We do however provide additional support to a robust and universal transaction cost model for bid-ask spreads that seems to hold across different markets and time periods. This is desirable, since otherwise a number of different models will be needed for different markets and time periods. Market microstructure invariance for bid-ask spread costs would effectively reduce the number of transaction cost models needed. This is of course of great interest to risk managers and traders, who according to Kyle and Obizhaeva (2017) want to measure and minimize transaction costs. Moreover, the relationship between tick sizes and bid-ask spreads found here can provide valuable insights to financial regulators and stock exchanges. Likewise, a model for understanding how order flows move prices is important from a financial stability perspective.

Conclusively, our findings can serve as a benchmark to which market concepts and theories in behavioral finance can be compared. We hope that our thesis alone, as well as paired with further studies, can facilitate normative implications of value to academics, finance professionals and society at large.

References

ALLEN, M., 2017. Cross-Sectional Design. The Sage Encyclopedia of Communication Research Methods (Vols. 1-4). SAGE Publications, Thousand Oaks, CA.

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. Camebridge University Press.

BROCK, W.A. and KLEIDON, A.W., 1992. Periodic Market Closure and Trading Volume: A model of Intraday Bids and Asks. Journal of Economic Dynamics & Control, 16(3), pp. 451-489.

CHAN, K.C., CHRISTIE, W.G. and SCHULTZ, P.H., 1995. Market Structure and the Intraday Pattern of Bid-Ask Spreads for NASDAQ Securities. The Journal of Business, 68(1), pp. 35-60.

COY, P., 2020-03-19, The Great Coronavirus Crash of 2020 Is Different. Available: https://www.bloomberg.com/news/articles/2020-03-19/the-great-coronavirus-crash-of-2020-is-different [Retrieved: May 16th 2021].

DANG, T., GORTON, G. and HOLMSTRÖM, B., 2013, Haircuts and Repo Chains*, Working Paper, Columbia University.

DEAKIN, M.A.B., 2011. G.I. Taylor and the Trinity test. International Journal of Mathematical Education in Science and Technology, 42(8), pp. 1069-1079.

HARRIS, LARRY, 2002, Trading and Exchanges: Market Microstructure for Practitioners (Draft, Forthcoming Oxford University Press, Oxford, UK).

HASBROUCK, J., 2017. Securities Principles and Procedures, NYU, pp. 19-27. http://people.stern.nyu.edu/jhasbrou/TeachingMaterials/STPPms12a.pdf [Retrieved: May 16th 2021].

HUSSAIN, S.M., 2011. The Intraday Behavior of Bid-Ask Spreads, Trading Volume and Return Volatility: Evidence from DAX30. International Journal of Economics and Finance, 3(1), pp. p23

KYLE, A.S., OBIZHAEVA, A.A. and KRITZMAN, M., 2016. A Practitioner's Guide to Market Microstructure Invariance. Journal of portfolio management, 43(1), pp. 43-53.

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., 2019. Large Bets and Stock Market Crashes. Working Paper. Rochester, NY.

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

LEE, C.M.C., MUCKLOW, B. and READY, M.J., 1993. Spreads, Depths, and the Impact of Earnings Information: An Intraday Analysis. The Review of Financial Studies, 6(2), pp. 345-374.

MCINISH, T.H. and WOOD, R.A., 1992. An Analysis of Intraday Patterns in Bid/Ask Spreads for NYSE Stocks. The Journal of Finance, 47(2), pp. 753-764.

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.

KUROV, A. and LASSER, D.J., 2004. Price Dynamics in the Regular and E-Mini Futures Markets. The Journal of Financial and Quantitative Analysis, 39(2), pp. 365-384.

POHL, M., RISTIG, A., SCHACHERMAYER, W. et al, 2020. Theoretical and Empirical Analysis of Trading Activity. Math. Program. 181, 405–434.

O'HARA, M., Market Microstructure Theory, 1995, Blackwell, Oxford, United Kingdom.

O’HARA, M., 2015. High Frequency Market Microstructure. Journal of Financial Economics, 116(2), pp. 257-270.

ONG, C.Z., MOHD-RASHID, R. and TAUFIL-MOHD, K.N., 2020. Do Institutional Investors Drive the IPO Valuation? Borsa Istanbul Review, 20(4), pp. 307-321.

PWC, 2021, Riskpremiestudien 2021. Available at: https://www.pwc.se/sv/pdf-reports/corporate-finance/riskpremiestudien-2021.pdf?fbclid=IwAR35A7bXn7Edlzj-X6-HCm4wc-IK0qcQ-dVevm9kQ0S0X_9Kh3xqSR8LOg8 [Retrieved: May 16th 2021].

SWEDISH CIVIL CONTINGENCIES AGENCY (MSB), 2020, Emergency Information from Swedish Authorities. Available at: https://www.krisinformation.se/en/news/2020/january/who-classes-the-outbreak-of-the-corona-virus-as-an-international-threat-to-human-life2 [Retrieved: May 16th 2021].

ZOVKO, I.I., 2008. Topics in Market Microstructure. Amsterdam, Netherlands, The Amsterdam University Press.

Appendix A – Cross-Sectional Plots

FIGURE A1. Individual OLS Regressions of Bid-Ask Spreads on Illiquidity

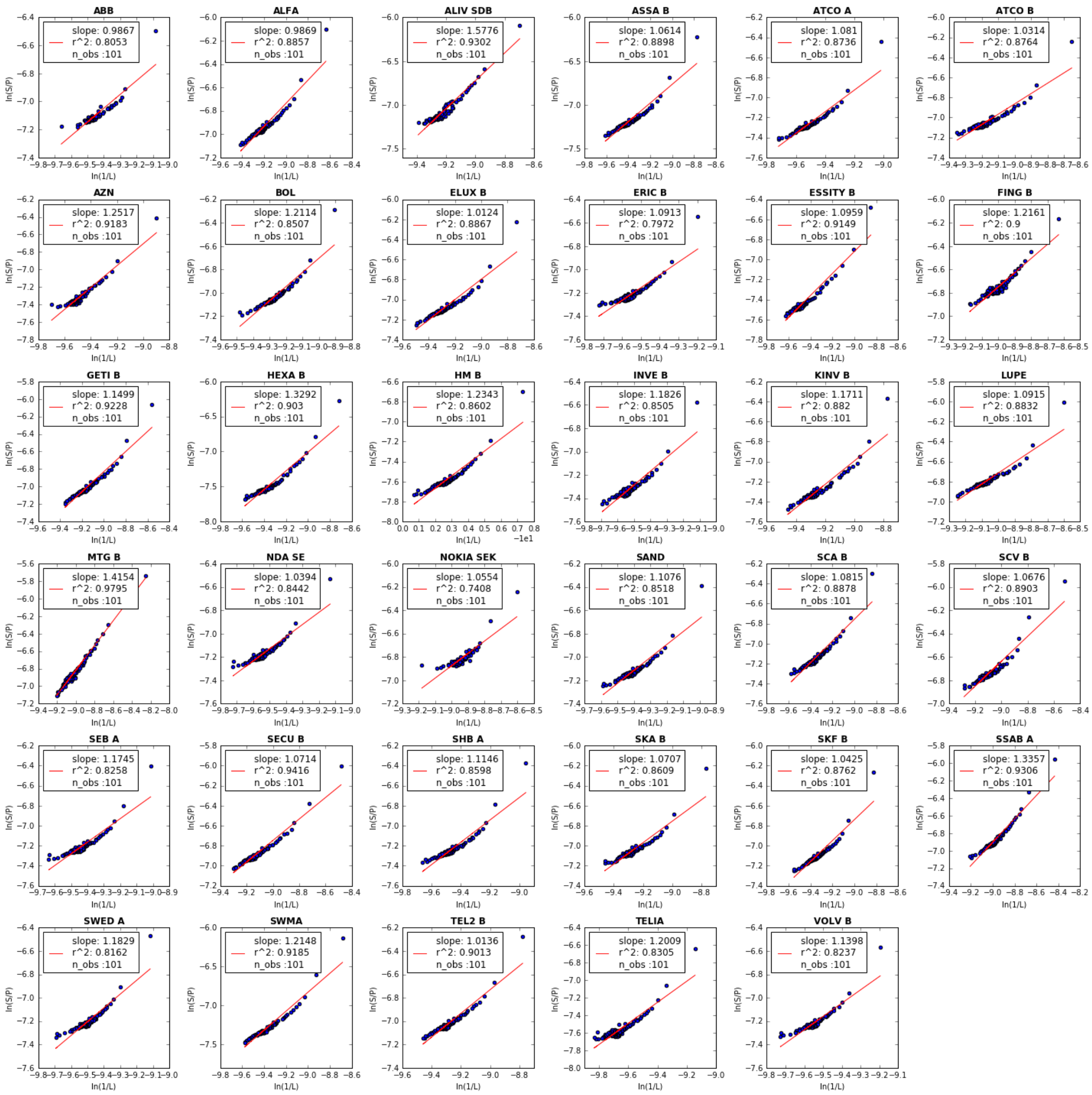

Figure A1 – plots individual OLS regressions for 35 OMXS30 constituents during the period 2010-02-08 to 2020-12-31. Each subplot contains 101 points corresponding to the number of five-minute intervals between 09:00-17:25 (CET).

Appendix B - Robustness Check

TABLE B1. Aggregate OLS Regressions for Individual Years 2010-2020

								CI Limits	
Start Date	**End Date**	**# Obs**	c	β	**Se(c)**	**Se(β)**	R^2	95% LL	95% UL
2010-02-08	2010-12-31	3030	2.1856	0.9687	0.0805	0.0086	0.8064	0.9518	0.9856
2011-01-01	2011-12-31	3030	0.2322	0.7797	0.0912	0.0099	0.6705	0.7602	0.7992
2012-01-01	2012-12-31	3030	1.3776	0.9012	0.1149	0.0124	0.6362	0.8769	0.9255
2013-01-01	2013-12-31	3030	1.4238	0.9065	0.1142	0.0121	0.6488	0.8827	0.9303
2014-01-01	2014-12-31	3131	1.8841	0.9520	0.0853	0.0090	0.7807	0.9344	0.9697
2015-01-01	2015-12-31	3030	1.0079	0.8584	0.1186	0.0127	0.6021	0.8336	0.8833
2016-01-01	2016-12-31	3030	2.7654	1.0557	0.1203	0.0129	0.6881	1.0304	1.0810
2017-01-01	2017-12-31	3131	2.3069	1.0116	0.1898	0.0200	0.4491	0.9724	1.0509
2018-01-01	2018-12-31	3131	2.5839	1.0744	0.0963	0.0102	0.7803	1.0544	1.0944
2019-01-01	2019-12-31	3030	3.4480	1.1661	0.1052	0.0111	0.7832	1.1442	1.1879
2020-01-01	2020-12-31	3030	4.0801	1.2350	0.0971	0.0104	0.8224	1.2145	1.2554

Table B1 – The table shows linear regression statistics for intercept c and slope β across all stocks, for each individual year between 2010 and 2020.

TABLE B2. Individual Stock Regression Statistics 2010-2020

			Individual Slopes		
Start Date	**End Date**	**# Obs**	Max	Min	Mean
2010-02-08	2010-12-31	3030	1.4044	0.6926	1.0885
2011-01-01	2011-12-31	3030	1.4405	0.7249	1.1018
2012-01-01	2012-12-31	3030	1.4308	0.7209	0.9849
2013-01-01	2013-12-31	3030	1.4491	0.6636	1.0118
2014-01-01	2014-12-31	3131	1.2774	0.6802	0.9761
2015-01-01	2015-12-31	3030	1.3510	0.7410	0.9777
2016-01-01	2016-12-31	3030	1.3234	0.7951	1.0301
2017-01-01	2017-12-31	3131	1.7660	0.4750	0.9373
2018-01-01	2018-12-31	3131	2.0931	0.8903	1.1955
2019-01-01	2019-12-31	3030	1.8743	0.9177	1.2220
2020-01-01	2020-12-31	3030	2.2254	1.1332	1.4774

Table B2 – The table shows a statistical overview of the individual slopes in each yearly linear regression between 2010 and 2020.

Appendix C – Construction of Empirical Variables for Stocks

Step 1: Volume-weighted aggregation

We drop subscript *j* for stock *j* for simplification of all aggregation formulas below. Furthermore, suppose that L trades occur simultaneously at time t_k , so that the timestamp is recorded L times ($t_{k_1}, t_{k_2}, \dots , t_{k_L}$). First, these L trades are counted as <u>one</u> trade.

First, let $Q_{t_k} = \sum_{\ell=1}^{L} Q_{t_{k_\ell}}$ denote the total number of shares traded at time t_k, where $Q_{t_{k_\ell}}$ is the trade size of limit order ℓ.

Let $P_{t_{k_\ell}} = \frac{A_{t_{k_\ell}} + B_{t_{k_\ell}}}{2}$ denote the quoted mid-price at time t_k after limit order ℓ, where $A_{t_{k_\ell}}$ and $B_{t_{k_\ell}}$ denote the best ask price and best bid price at time t_k, after limit order ℓ, respectively.

Finally let $S_{t_{k_\ell}} = A_{t_{k_\ell}} - B_{t_{k_\ell}}$ denote the bid-ask spread at time t_k, after limit order ℓ.

Then we can volume-weight our empirical variables as follows:

$P_{t_k} = \frac{1}{Q_{t_k}} \sum_{\ell=1}^{L} Q_{t_{k_\ell}} P_{t_{k_\ell}}$ denotes the volume-weighted mid-price at time t_k.

$\frac{S_{t_k}}{P_{t_k}} = \frac{1}{Q_{t_k}} \sum_{\ell=1}^{L} Q_{t_{k_\ell}} \frac{S_{t_{k_\ell}}}{P_{t_{k_\ell}}}$ denotes the volume-weighted relative bid-ask spread at time t_k.

Each variable now has exactly one observation per timestamp. The steps in our time-weighted aggregation are outlined below.

Step 2: Time-weighted aggregation

Temporarily let *i* denote intraday time interval *i* (replacing *t*). Now let trade k represent the volume-weighted trade at time t_k above. Suppose that N_i trades (marketable orders) occur at irregularly spaced times $t_1, t_2, \dots , t_{N_i}$ during each intraday interval *i* with a fixed length $T' - T$ = 5 minutes.

First,

$Q_i = \frac{1}{N_i} \sum_{k=1}^{N_i} Q_{t_k}$ denote the average trade size in five-minute interval *i,* and

$V_i = \sum_{k=1}^{N_i} Q_{t_k}$ denote the total share volume traded in five-minute interval *i.*

For the first intraday interval *i* each day (no preceding trades), the time-weighted mid-price and relative bid-ask spread are:

$$P_i = \sum_{k=1}^{N_i} \frac{P_{t_k}(t_{k+1} - t_k)}{T' - t_1}$$

$$\frac{S_i}{P_i} = \sum_{k=1}^{N_i} \frac{(S_{t_k}/P_{t_k})(t_{k+1} - t_k)}{T' - t_1}$$

For each subsequent intraday interval *i*, the time-weighted mid-price and relative bid-ask spread are:

$$P_i = \sum_{k=0}^{N_i} \frac{P_{t_k}(t_{k+1} - t_k)}{T' - T}$$

$$\frac{S_i}{P_i} = \sum_{k=0}^{N_i} \frac{(S_{t_k}/P_{t_k})(t_{k+1} - t_k)}{T' - T}$$

where $t_0 = T$ is the start time of the interval and $t_{N_i + 1} = T'$ is the end time of the interval. P_{t_0} and S_{t_0} are the outstanding mid-price and bid-ask spread at the beginning of the interval, respectively. The realized variance over interval *i* is calculated by summing the squared returns between mid-prices at each time of trade:

$$\sigma_i^2 = \sum_{k=2}^{N_i} \left(\ln\left(P_{t_k}\right) - \ln\left(P_{t_{k-1}}\right)\right)^2$$

Finally, the illiquidity index over interval *i* can be computed from our time-weighted mid-price, total volume and realized variance:

$$\frac{1}{L_i} = \left(\frac{P_i V_i}{\sigma_i^2}\right)^{-1/3}$$