Market Microstructure and Algorithmic Execution

A post-trade analysis on global futures markets

S.T. Morseld

ANR: 982376

s.t.morseld@tilburguniversity.edu

A thesis submitted in partial fulfillment of the requirements for the degree of MSc Quantitative Finance & Actuarial Science

Tilburg School of Economics and Management

Tilburg University

Supervised by:

dr. Nikolaus Schweizer

January 28, 2022

Abstract

Combining a unique data set of futures order book data and trade execution data from a hedge fund across 50 global futures markets over a 9-months period with a total transaction value of 2.3 billion dollars, we measure and benchmark the actual transaction costs incurred by a large trader. Building on Perold (1988), we define total costs as the sum of trading cost and delay cost in our research and derive a futures-specific categorization of transaction costs. Our results on trading costs are an order of magnitude smaller than previous studies on execution in equities suggest and we observe limited evidence of a buy-sell asymmetry. Finally, we present unique insights into how trading costs vary across asset classes globally and apply the quantile regression approach to estimate the impact of several trade- and market-specific characteristics on trading costs. This model outperforms the homoskedastic OLS model often suggested in the literature and we observe that trade duration and market-specific volatility are the most important variables in explaining variation in trading costs.

Contents

1 Introduction 1

2 Futures 3

2.1 Specification of a Futures Contract . . . 3

2.2 Futures Pricing . . . 5

3 Market Microstructure 8

3.1 Basic Concepts of Market Microstructure Theory . . . 8

3.2 Mechanics/Dynamics of a Modern (Futures) Market . . . 8

3.3 Liquidity, Price Discovery and Volume . . . 10

3.4 Futures Markets Specifics . . . 11

4 Transaction Costs 12

4.1 Categorization of Transaction Costs . . . 12

4.2 Transaction Cost Analysis . . . 17

5 Algorithmic Execution 21

5.1 Objective Function . . . 21

5.2 Solving The Problem . . . 22

5.3 Execution Algorithms . . . 22

6 Data Description 24

6.1 Portfolio Construction and Trade Execution . . . 24

6.2 Data . . . 24

6.3 Sample Properties of Varick's Trades . . . 26

7 Quantifying/Measuring The Implicit Transaction Costs 29

7.1 Implicit Cost . . . 29

7.2 Delay Cost . . . 29

7.3 Trading Cost . . . 31

7.4 Benchmarking . . . 38

8 Quantile Regression 41

8.1 Introduction . . . 41

8.2 Estimation Results . . . 42

8.3 Formal Testing . . . 43

8.4 Goodness-of-fit . . . 44

8.5 Interpretation of Results . . . 44

8.6 Relative importance of variables . . . 47

9 Conclusion & Discussion 49

9.1 Conclusions . . . 49

9.2 Discussion & Future Research . . . 50

10 References 52

A Appendix A: Definitions **56**
A.1 Definition order-size classes 56

B Appendix B: Quantile Regression **57**
B.1 Introduction 57
B.2 Formal Testing 59
B.3 Goodness-of-fit 60
B.4 Relative Importance of Variables 60

1 Introduction

Nowadays, a wide range of assets can be traded electronically. Stocks, bonds and a variety of derivatives like futures are actively traded on many exchanges throughout the world (Hull, 2014) and can all be bought and sold at the push of a button. With the proliferation of computer technology, investors are demanding faster, cheaper, more reliable, and smarter access to financial markets. As a consequence, algorithmic execution has become an important part of modern financial markets. Banks and hedge funds are taking advantage of this trend and have begun an arms race to find the best electronic trading systems and execution algorithms.

The theoretical basis for algorithmic execution is found mainly in the fields of financial econometrics and market microstructure. Following Easley and O'Hara (1995), "market microstructure research focuses on the interaction between the mechanics of the trading process and its outcomes, with the specific goal of understanding how actual markets and market intermediaries behave". Due to the rapid development of algorithms and electronic trading, market microstructure analysis is still one of the fastest growing fields in financial research.

Within market microstructure research, transaction cost analysis is one of the key areas. According to Pedersen (2018), transaction costs are an unavoidable and crucial part of implementing any investment strategy, including "passive" ones like index strategies. In practice, a theoretically profitable trading strategy may not be profitable because transaction costs outweigh expected profits. Empirical evidence indicates that transaction costs significantly impact investment performance and suggests these costs should be carefully managed.

To date, extensive research has been devoted to analyzing the magnitude and determinants of transaction costs. While recent studies[1] examine proprietary trade data from institutions, most studies have relied solely on publicly available data.[2] In these publicly available datasets, individual trades are not reliably classified as buyer- or seller-initiated and, more importantly, each trade typically exists in isolation; there is no information on sequences of trades that form part of a large order (Almgren et al., 2005). Overall, most studies use limited data in terms of breadth, typically only covering trades in the United States and a small cross-section of stocks.

Furthermore, much of the existing work analyzes transaction costs in equity markets, whereas comparable research in futures markets is scarce. There are no studies that we are aware of that have available proprietary execution data and corresponding order book data to measure transaction costs in futures markets. Although some research based on clearing house settlement data is available, these studies generally cover a small cross-section of stock index or bond futures.[3] The scarcity of research based on futures markets is surprising, given the importance of futures as a mechanism for obtaining market exposure to underlying assets. In addition, there are several distinctions between equity and futures markets, raising questions about the applicability of equity market findings to futures markets. For instance, futures markets are generally more liquid *in nature* than equity markets, have a lower probability of private information[4] and are not constrained by short-selling restrictions.

[1] Bikker et al. (2007); Fraenkle et al. (2011); Frazzini et al. (2018)
[2] Rydberg and Shephard (2003) have, for instance, worked with the trade and quote (TAQ) tick record from the NYSE.
[3] Frino & Oetomo (2005); Frino et al. (2007)
[4] Gorton & Pennacchi (1991); Subrahmanyam (1991)

To summarize, due to a lack of access to both order book and execution data, measuring and analysing trading costs thoroughly has been an empirical challenge in academia, especially for futures markets. We intend to fill this gap by utilizing proprietary execution data from a hedge fund across 50 global futures markets over a 9-months period with a total transaction value of 2.3 billion dollars to quantify a large money manager's actual transaction costs. The data includes a significant number of trade- and market-specific characteristics for each order, providing a unique opportunity to analyze algorithmic execution in futures markets.

Building on the Implementation Shortfall definitions proposed by Perold (1988), we first redefine the total transaction costs incurred in futures trading. Moreover, since our data set allows us to quantify both the delay cost and trading cost, it allows for a richer measurement than is common in the literature. Furthermore, we present unique insights into how trading costs vary across asset classes. Even though the mean trading cost for the different asset classes is shown to be similar, we observe a substantial difference in terms of outliers. To the best of our knowledge, we obtain the most comprehensive examination to date on the transaction costs incurred by trading a real portfolio of futures contracts.

In addition, we have access to underlying order book data from the broker for each executed trade. This provides us variables like the average spread during the trading horizon, enabling us to measure costs in terms of spreads and compare their values across different contracts. Moreover, it allows us to evaluate the performance of the specific execution algorithm used, by comparing the execution prices against two (intraday) benchmarks. As a result, we obtain a more comprehensive picture of the order execution itself as well.

Finally, previous research has already defined some variables related to trade execution that could explain trading costs. In this study, we assess to which extent the variables examined in previous research explain the variation of trading costs in our sample and whether these variables affect each part of the cost distribution in a similar way. While most studies generally apply the homoskedastic OLS to analyze the average impact of several characteristics on trading costs, we apply the quantile regression approach to estimate the impact of trade- and market-specific characteristics on orders with the $100\theta\%$ highest trading cost, where θ can take any value in the range $(0, 1)$. Overall, we find the quantile regression to better capture the relationship between trading costs and various factors than the homoskedastic OLS. In conclusion, our results show that trade duration and market-specific volatility are the most important variables in explaining variation in trading costs. Furthermore, we find momentum to have an amplifying effect on the likelihood of incurring significant trading costs.

The remainder of this thesis is structured as follows. Section 2 starts by presenting general theory on futures contracts. After that, we provide a literature review of the important studies that have paved the way for algorithmic execution. Specifically, Section 3 describes the theory behind market microstructure, Section 4 addresses transaction cost analysis (TCA) and Section 5 discusses issues related to cost optimization. Subsequently, Section 6 presents the data used in this research and provides descriptive statistics. Section 7 outlines the methodology for measuring implicit costs from the execution data and compares the measured costs to those found in previous studies. Furthermore, the performance of the execution algorithm is analyzed by applying several benchmarks. Section 8 describes the quantile regression approach and discusses the estimation results. Finally, Section 9 summarizes and concludes.

2 Futures

Over the last 40 years, derivatives such as futures have become increasingly important in finance (Hull, 2014). A derivative is a financial instrument whose value is determined by the values of other underlying variables or is derived from them. Futures are listed derivatives, which are standardized contracts traded on an exchange. The following subsection reports the most important specifications of such a contract and closely follows the definitions in Hull (2014).

2.1 Specification of a Futures Contract

A futures contract is a legally binding agreement between two parties to buy or sell a particular asset at a certain time in the future at a predetermined price (Hull, 2014). A futures exchange, such as the CME (Chicago Mercantile Exchange), facilitates this transaction. Depending on the underlying asset, different types of futures contracts are available for trading. Stock index futures, for example, are futures contracts that track stock market indexes. The quality, quantity, (physical) delivery time, and location are all specified in an exchange-traded futures contract, making the contract's specifications the same for all participants. Using these standardized features of the futures contract, contract ownership can be easily transferred to another party.

2.1.1 Contract Size & Value

As an example of a contract specification, the *contract size* indicates the quantity of the asset that has to be delivered for a single contract. Contract sizes like 'Treasury bonds with a face value of $50,000', '100 ounces of Gold' or '5,000 bushels of Soybeans' are all defined in the futures contract specification. Related to this, the contract notional value, also known as *contract value*, is the financial expression of the contract size and the current futures price, which is calculated by multiplying the contract size by the price at time t of the underlying future i:

$$\text{Contract value}_{it} = \text{Contract Size}_i \cdot \text{Futures Price}_{it}$$

Hence, assuming March 2022 E-mini S&P500 futures are trading at $4778.50,[5] whose contract size is 50, the contract value of one ESH2 contract (at that specific time) is equal to $50 \cdot \$4778.50 = \$238,925$.

2.1.2 Tick Size

A *tick* is the smallest price fluctuation, which is also part of a futures contract's specifications. The exchange determines tick sizes, which vary based on the underlying contract. Tick sizes are chosen to ensure maximum liquidity and narrow bid-ask spreads.

2.1.3 Futures Trading Hours

Each type of futures contract (e.g., interest rate or stock index) has its own trading hours, which are generally different from the spot market hours of the underlying asset. While the stock market in the United States, for instance, is most active from 9:30 am to 4:00 pm Eastern time, stock index futures trade nearly 24 hours a day, six days a week. The E-mini S&P500 Futures on the CME exchange, for instance, start trading at 5 pm Eastern time and close at 4 pm Eastern, both on weekdays and Sundays.

[5] Last quoted price at the CME Exchange on 3 January 2022 10:00 AM UTC +1.

2.1.4 Clearing Houses

Whenever two traders have reached an agreement on a trade, the clearing house validates and finalizes (*clears*) it. In futures transactions, a clearing house acts as an intermediary, standing between the two traders and managing the risks. Hull (2014) provides a clear description of how a trade or transaction is typically implemented. Say, for instance, that a specific trader (trader 1) has reached an agreement to buy 200 ounces of gold from another trader (trader 2) at some date in the future for $1,800 per ounce. As a consequence, trader 1 will actually have a contract to buy 200 ounces of gold from the clearing house at $1,800 per ounce, while trader 2 will have a contract to sell 200 ounces of gold to the clearing house at the same price. The clearing house then manages credit risk by forcing each trader to deposit funds (known as *margin*) with the clearing house in order to guarantee that both parties meet their obligations.

Overall, it is the clearing house's job to settle accounts and clear trades; collect and maintain margin funds; regulate delivery; and report trading data. By ensuring that all parties comply with the system and procedures, transactions can proceed smoothly, which increases the confidence of market participants and thus the liquidity of the market.

2.1.5 Lifespan of a Futures Contract

The lifespan of a futures contract is limited, which will affect trading results and the exit strategy. The expiration date of the contract is the final day the contract can be traded. In fact, every futures contract is identified by its delivery month, which is represented by *contract display codes* or *expiry date codes*. These are usually one- to three-letter codes that identify the product, followed by characters that indicate the expiration month and year. The format of an expiry date code varies per asset class and trading platform. As an example, the CME Globex expiry date code for the E-mini S&P500 (ES) futures contract expiring in March 2022 is *ESH2*.

Specifically, the exchange specifies when a contract's delivery will take place within the month, as well as when it will start and end trading. As a result, prior to expiration, traders typically have three options:

Offsetting or closing out a position
Due to the fact that most traders prefer to close out their positions before the expiry date stated in the contract, a predominant part of futures contracts does not result in delivery (settlement). *Closing out* or *offsetting* a position refers to entering into a trade that is the exact opposite of the original one, which is also known as *neutralizing* the trade. When a trader offsets a position, he or she is able to realize all of the profits or losses associated with the closed position.

Rollover
When a trader transfers his position from the current month's contract to a futures contract that is further in the future, this is called a *rollover* or *roll*. In order to roll forward, the trader closes out his current position while additionally establishing a new position in another contract month further in the future. A trader who is long two S&P500 futures contracts that expire in March 2022, for example, will sell two ESH2 contracts and purchase two ESM2 (expiring in June 2022) or further distant ES contracts at the same time.

Settlement/Delivery
The contract will expire if a trader has not closed out or rolled his or her position before the contract expiry date, and the trader will be obliged to *settle*. The trader with a short position is now required to deliver the underlying asset in line with the original contract's terms. This delivery may take the form of physical delivery of the underlying asset for some contracts (mostly commodity futures). However, due to the fact that only a tiny fraction of all (commodity) futures contracts are physically delivered, the majority of deliveries will result in cash settlements. In the case of a cash-settled contract, settlement occurs in the form of a credit or debit based on the contract value at the expiration date of the contract.

2.1.6 First & Last Notice Day

Lastly, for a contract, the first notice day, the last notice day, and the last trading day are the three most important days. The *first notice day* is the first day on which notices of intent to deliver are authorized, whereas the *last notice day* is the last day to submit such a notice. Generally, the *last trading day* is a few days before the last notice day, but investors with long positions who do not wish to be at risk of taking delivery should close out their positions before the first notice day.

2.2 Futures Pricing

Since the above-mentioned contract specifications are standardized, *price* is the only contract variable. As previously stated, relatively few futures contracts result in the delivery of the underlying asset; the majority of them are offset early. Nonetheless, the price of futures is determined by the possibility of delivery at expiration. Moreover, we know that the value of a futures instrument is derived from the price of its underlying asset and moves in synchronization with it. That is, if the underlying asset's price drops, so will the futures price, and vice versa. In fact, in a well-functioning futures market, the futures price at expiration equals the price of the underlying asset, i.e, the spot price, which is due to the *spot – future parity* (Hull, 2014). However, before expiration, the spot price and the futures price may differ. We typically distinguish between two situations: *normal backwardation* and *contango*. Normal backwardation is the situation where the futures price is below the expected future spot price, while contango is when the futures price is above the expected future spot price. This is shown in Figure 1.

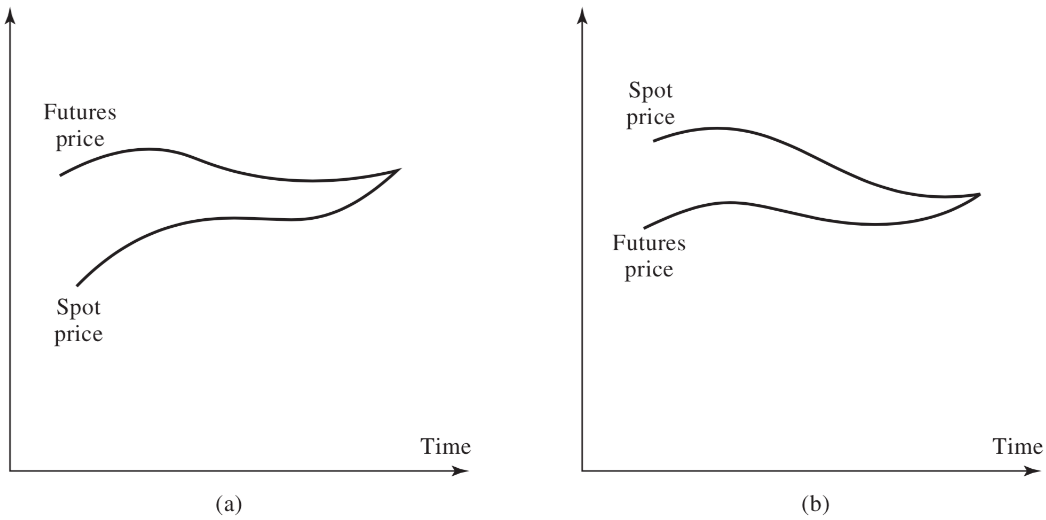

Figure 1: Relationship between futures price and spot price as the time to expiry decreases: (a) *contango*; (b) *normal backwardation. Retrieved from Hull (2014)*

The difference between spot and futures prices is driven by variables like interest rates, dividends, and time to expiry. In general, there is a mathematical equation that equates the spot price of the underlying asset and its futures price. This equation is often referred to as the *futures pricing formula*, which is presented for several asset classes below,[6] using the definitions in Hull (2014) and Baltas (2017). Throughout this subsection, we will assume that r_t is the annual, continuously compounded (zero-coupon) risk-free interest rate, for an investment maturing at the delivery date, that is, in $T - t$ years.

Stock Index Futures

A stock index can be thought of as an investment that pays dividends based on the underlying stock portfolio of the index. The dividends paid by the investment asset are the dividends that would usually be paid to the holder of the underlying portfolio of stocks. However, rather than providing a known cash income, the dividends are assumed to provide a known *dividend yield*. Hence, the futures price F_t of a stock index is given by

$$F_t = S_t e^{(r_t - q_t)(T-t)}$$

Here, S_t is the spot price of the underlying index and q_t is the dividend yield. In practice, the dividend yield q_t on the underlying portfolio of an index fluctuates over time.

Currency Futures

The underlying asset here is the foreign currency. Assuming the local currency is USD, we can define S_t as the spot price of one unit of the foreign currency (in USD) at time t and F_t as the futures price of one unit of the foreign currency (in USD). The futures pricing formula for currency futures then looks as follows:

$$F_t = S_t e^{(r_t - r_t^*)(T-t)}.$$

The idea behind this formula is that the holder of the foreign currency can earn interest at the risk-free interest rate r_t^* prevailing in the foreign country. Moreover, r_t is simply the risk-free rate when investing money in local currencies for the time period $T - t$. Hence, the futures price increases with the time to maturity $T - t$ of the futures contract at a rate of the difference in interest rates between the foreign and local country (i.e., $r_t - r_t^*$).

Commodity Futures

In the absence of storage costs and income, the futures price of a commodity is simply given by

$$F_t = S_t e^{r_t(T-t)},$$

where S_t is the spot price of the commodity (in USD) at time t, r_t is the risk-free interest rate at time t and $T - t$ is the time to maturity. However, some commodities incur significant storage costs, which can be treated as negative income. This leads us to two additional ways of pricing commodity futures.

Firstly, the futures price may be expressed as

$$F_t = (S_t + U)e^{r_t(T-t)},$$

where U is the present value of all storage costs (net of income) during the lifespan of the contract.

[6] More asset classes for futures exist in general, but the ones mentioned here correspond to the actual futures contracts traded by Varick Capital. See Table 14 in the Appendix for a complete overview.

Secondly, storage costs are often assumed to be proportional to the price of the commodity, that is,

$$F_t = S_t e^{(r_t+u_t)(T-t)},$$

which implies that storage costs are basically treated as "negative yield". Consumption commodities, such as crude oil, are examples of commodities that typically generate no income but might incur considerable storage costs.

In general, however, owning a physical asset might enable a specific company to keep their production process operating and maybe even profit from (transitory) local shortages. Owning a futures contract, on the other hand, does not allow for this and, for this reason, the benefits from holding the physical asset are often included in a variable known as the commodity's *convenience yield*. Assuming the storage costs per unit are a constant proportion u, we may define the futures price as

$$F_t = S_t e^{(r_t+u_t-y_t)(T-t)}.$$

Here, y_t is the convenience yield, which reflects the market's expectations concerning the future availability of the commodity: the greater the likelihood of shortages, the higher the convenience yield.

Bond Futures

Generally, a bond futures contract is a contract where the asset to be delivered is a (government or) Treasury bond. Following Hull (2014), we can state that such contracts can be interpreted as contracts on a traded asset (the bond) that provide the holder with known income.[7] Hence, the relationship between the futures price F_t and the spot price S_t is defined as

$$F_t = (S_t - I)e^{r_t(T-t)},$$

where I is the present value of the coupons during the lifespan of the contract, $T - t$ is the time until expiration (maturity), and r_t is the risk-free interest rate for the time period $T - t$.

Price Discovery

Using the futures pricing formula above, the *fair value* of futures can be calculated. However, the actual market price is discovered by *quoting*, that is, bidding and asking until a match (trade) is found. If we, for instance, assume that on January 1st, the February futures price of gold is quoted at $1,800, this implies that traders can agree to buy or sell gold for February delivery at this specific price.[8] It is worth emphasizing that this price is simply determined by the laws of supply and demand.

[7]This assumes that the cheapest-to-deliver bond and delivery date are both known.
[8]This price is exclusive of commissions.

3 Market Microstructure

Generally, asset-pricing theory tends to separate itself from the underlying mechanics of trading and concentrates purely on fundamental values (Johnson, 2010). Despite the elegance and simplicity of the Modern Portfolio Theory (MPT) and Capital Asset Pricing Model (CAPM), the theory on which they are based - the Efficient Market Theory - is too simplistic and idealistic in comparison to real-world market conditions. The following section will expand on this by introducing the concepts behind market microstructure, a competing field of study. Following that, some key aspects of a modern electronic market are covered, before delving into the specifics of futures markets.

3.1 Basic Concepts of Market Microstructure Theory

As expressed in Labadie and Lehalle (2010), one of the hypotheses of Efficient Market Theory is the existence of a single market price, which reflects the fundamental value of the asset. However, as mentioned in their paper, the concept of price itself is ambiguous, as there are typically numerous different prices in any market, such as the ask price, bid price, mid price, and last traded price. Furthermore, Labadie and Lehalle (2010) argue that assuming the existence of a single market price ignores the price formation process, which is dependent on the nuances of each market and explains why prices differ between marketplaces. Furthermore, one of the main assumptions of the famous CAPM is that markets are perfect, implying that all assets are perfectly liquid and there are no transaction costs, which clearly does not hold in real markets, as we will see later on. The field of study that seeks to comprehend these effects is known as market microstructure.

In contrast to the vast majority of macro-based theory, market microstructure literature concentrates on the actual trading process and examines how specific trading mechanisms impact both observed prices and traded volumes (Johnson, 2010). Market microstructure is defined by Easley and O'Hara (1995) as "the study of the process and outcomes of exchanging assets under explicit trading rules". It helps understand why asset prices may differ from their fundamental values. According to micro-based models, trading is not an auxiliary market activity that can be ignored when examining price behavior. Rather, trading is a critical element of the price formation process.

Following Johnson (2010), market microstructure theory can generally be broken down into three key areas: i) market structure and design, ii) trading mechanism research and iii) transaction cost measurement and analysis. The subsections that follow provide an introduction to market (micro)structure and design, while Section 4 discusses transaction costs.

3.2 Mechanics/Dynamics of a Modern (Futures) Market

We follow Johnson (2010) by first discussing some of the fundamental elements of trading and markets before delving deeper into the major principles of market microstructure. First of all, markets generally exist to facilitate trade. Its fundamental purpose is to bring together buyers and sellers. Although there are several approaches to establishing an electronic market, financial markets all boil down to allowing participants to indicate their intent to trade and having a matching mechanism connect potential buyers and potential sellers. In order to provide a basic understanding of the organization of a market, we now first survey the different types of markets and orders.

Order types

Orders play an important role in market structure. Following the definitions in Hull (2014), an electronic market, in its basic configuration, has two types of orders: *market* orders and *limit* orders. Market orders are generally considered aggressive orders that seek to execute a trade immediately. With a market order, a trader expresses its intent to buy or sell a certain amount of an asset at the best available price, which (usually) results in an immediate trade. On the other hand, limit orders are considered passive orders, which specify a particular limit price. As a result, the order will only be executed at this price or at a price that is either lower (for buy orders) or higher (for sell orders). For example, for an investor who wants to buy at a limit price of $50, the order will only be filled at a price of $50 or less. However, since the limit price may never be attained, there is no certainty that the order will be executed at all.

Using specific conditions and incorporating additional special behaviours has enabled venues to offer a wide range of order types. For example, hybrid orders such as market-to-limit orders actually have some of the properties of both market orders and limit orders. In this research, we shall not cover these though.

The Limit Order Book

According to market microstructure theory, any market's price formation processes and trade dynamics are determined by that market's unique organization. In modern electronic marketplaces, the limit order book (LOB) is the primary market mechanism. Buy and sell orders are listed in the limit order book at varying prices and for varying quantity levels.

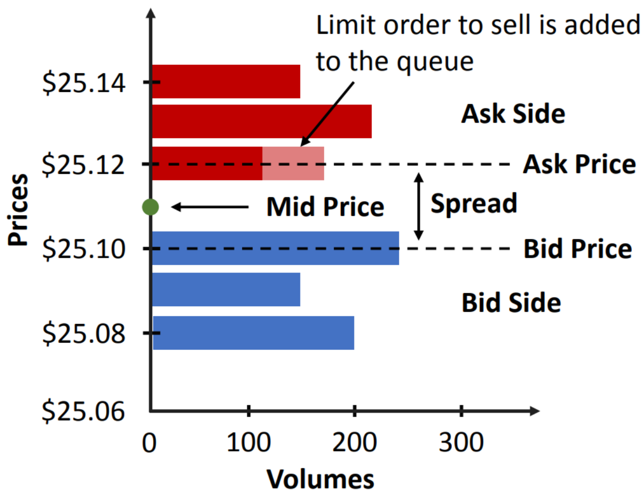

Figure 2: Visualization of the LOB structure. *Retrieved from Vyetrenko and Xu (2019)*

Figure 2 visualizes an example of a LOB structure. In this specific case, buying 200 contracts at $25.08 or less, or selling 150 contracts at $25.14 or higher, would be examples of limit orders. The best *bid* (buy order) and the best *ask* (sell order) at any particular time can be determined by combining all of these orders into a single system, such as an exchange.

Spread

The *mid price* is defined as the average of the best current bid and ask prices being quoted, whereas the difference between the bid and the ask price is referred to as the *full spread*.

Following Hedayati et al. (2018), another common spread measure is the difference between the execution price and the prevailing mid price prior to execution, which is often referred to as the *effective spread*. When a trader is willing to cross the spread, that is, to purchase at someone else's ask price or sell at someone else's bid price, a trade occurs. The market is typically anonymous, and trade can only take place based on price and quantity. This is referred to as a *continuous auction*.

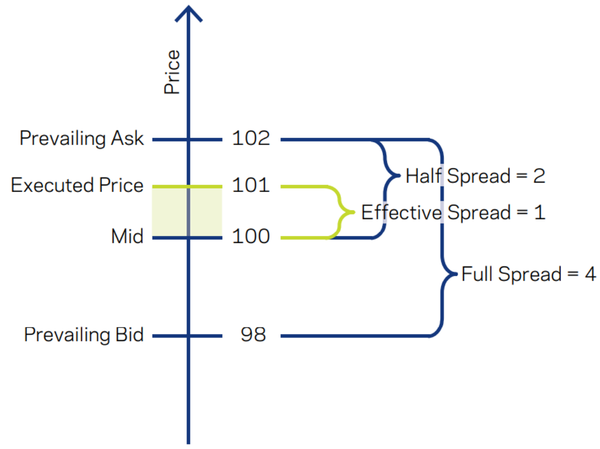

Figure 3: Prevailing market prices and spreads. *Retrieved from Hedayati et al. (2018)*

Figure 3 illustrates the prevailing market prices and spreads for a buyer that has executed a trade at an average price of 101. The difference between the execution price of 101 and the prevailing mid price of 100, is the effective spread and is equal to 1. Evidently, the mid price itself is halfway between the bid and ask prices of 98 and 102, respectively, so that the prevailing full spread is equal to 4.

3.3 Liquidity, Price Discovery and Volume

The liquidity or effectiveness of a market can be described by determining the number of active market participants and the quantity of a particular asset that these participants are ready to trade at any given time. Defining and quantifying liquidity is one of the main aspects of market microstructure research and an important step in minimising transaction costs. Overall, a well-functioning market has many participants who trade significant amounts of an asset with one another on a continuous basis. At any given point in time, the price of the asset being quoted will thus reflect what many participants believe it should be. Otherwise, some of the orders would be filled, and the market would either move up or down in the order book. This is known as *price discovery*. As a result, the spread, or the difference between the bid and ask price, is a general indicator of liquidity.

Another indicator that is frequently used to determine how liquid a trading environment will be is *volume*. However, liquidity and volume are not interchangeable. Liquidity refers to how easily assets can be traded, while volume represents how much trading occurs for a given instrument, or set of instruments, over a given time period. Hence, volume is a related but indirect proxy for liquidity. In general, the higher the volume, the lower the relative size of the investor's trade and the larger the probability of executing a trade quickly and with minimal trade cost. However, despite the fact that volume is crucial to liquidity, there are cases where high volume is accompanied by low liquidity.

3.4 Futures Markets Specifics

We conclude this section by stating a number of differences between futures markets and, for instance, equity markets, which bring into question the applicability of equity market findings to futures markets.

Firstly, futures markets are generally more liquid *in nature* than equity markets, since futures are standardized contracts traded on regulated exchanges. Note the emphasis on 'in nature', since it is not necessarily true that, for instance, commodity futures markets are more liquid than equity markets. However, there has been some research indicating a large difference in absolute spreads between, for instance, the nearby contract of stock index futures and its constituent stocks. As an example, Fleming et al. (1996) find that the average absolute bid/ask spread for the nearby contract of S&P500 futures is \$0.0558, while the average absolute spread for the constituent S&P500 stocks is \$0.2185.

Secondly, there is typically a low probability of private information in most futures markets. The fact that futures are standardized contracts and traded on regulated exchanges makes them highly transparent. Moreover, information asymmetry due to stock-specific private information is typically diversified away in stock index futures, resulting in lower adverse selection costs for trades in stock index futures than in underlying equity markets (Gorton & Pennacchi, 1991; Subrahmanyam, 1991). Similarly, it is less likely that trades in bond futures contain private information. For instance, Ederington and Lee (1993) document price reaction on Treasury bond futures at and following, but not prior to, macroeconomic announcement releases by the U.S. Treasury.

Thirdly, some equity markets can place restrictions on the short positions that investors may take. Futures markets, by contrast, are not constrained by such short-selling restrictions. Hence, unlike equity markets, futures markets are as likely to facilitate purchases as sales.

4 Transaction Costs

As reported by Pedersen (2018), transaction costs are an unavoidable and crucial part of implementing any investment strategy, including "passive" ones like index strategies. Each time a certain asset is traded, transaction costs are incurred and, for this reason, transaction costs are a key area for market microstructure research. By carefully measuring and analysing transaction costs, one can try to minimise these. Recently, as a result of the proliferation of algorithmic execution, transaction cost analysis (TCA) has regained a newfound interest in the financial community (Kissell, 2013). However, performing such a TCA requires complete knowledge and understanding of the underlying elements.

This section, therefore, starts by providing an overview of the transaction cost components. Based on these components, TCA is then classified into three categories: i) cost measurement, ii) cost estimation, and iii) trade cost optimization. In this section, cost measurement is covered by discussing studies that have quantified transaction costs and examined a relationship between the observed costs and some explanatory factor(s) ex-post. Furthermore, this section briefly covers cost estimation by providing a review of the model proposed by Almgren and Chriss (1997), which serves as the foundation for all regression models that attempt to estimate costs ex-ante. Section 5 covers trade cost optimization.

4.1 Categorization of Transaction Costs

Prior research has already thoroughly investigated transaction costs. Perold (1988), for instance, introduces Implementation Shortfall (IS) as the total cost of executing an investment idea. As pointed out in Perold's paper, the Implementation Shortfall is generally calculated as the difference between a portfolio's paper return, which assumes all contracts are traded at a benchmark price, and the portfolio's real return, which takes into account actual execution prices and the number of contracts traded. That is,

$$\text{Implementation Shortfall} = \text{Paper Return} - \text{Actual Return}$$

Perold (1988) breaks down this total cost into trading cost, opportunity cost and fixed fees. The trading cost component represents the cost that is incurred in the market by trading, which he refers to as *price impact*. Wagner and Edwards (1993) expand on Perold's Implementation Shortfall methodology by splitting out the trading costs into specific components (timing, delay, impact and opportunity costs), in order to more accurately identify and classify total trading costs based on where, when, and how they occur within the investment cycle (Johnson, 2010). These components can be even further expanded by breaking the total transaction costs down into nine distinct components, following Johnson (2010). Moreover, we find that there are several ways of classifying these constituents. By slightly altering Johnson's (2010) classification to design a futures specific categorization of transaction costs, we obtain Table 1. Each of the components is presented in detail below, following the definitions in Kissell (2006).

		Explicit	Implicit	Fixed	Variable
Investment	Taxes	✓			✓
	Delay Cost		✓		✓
Trading	Taxes	✓		✓	
	Commission	✓		✓	
	Fees	✓		✓	
	Spreads		✓		✓
	Market Impact		✓		✓
	Price Trend		✓		✓
	Timing Risk		✓		✓
	Opportunity Cost		✓		✓

Table 1: Categorization of transaction costs (futures specific)

Distinguishing between *investment* and *trading related* costs is helpful in determining at which phase these costs can be controlled. All costs that occur before t_0 (the time the order starts executing) may be classified as investment related costs, while the remainder is classified as trading related costs. As expressed in Kissell (2006), these components can be further divided into fixed and variable components as well as explicit and implicit costs.

Fixed cost components reflect all costs that remain constant regardless of the execution strategy. These costs cannot be controlled or minimized during execution. *Variable* cost components, on the other hand, may vary depending on the asset, the order, market conditions and the execution strategy.

Furthermore, transaction costs are usually divided into explicit and implicit cost components. *Explicit* costs are those for which the cost or fee structure is known beforehand. Explicit costs are clearly identified and quantifiable and may, for instance, be specified as a percentage of the value traded. *Implicit* transaction costs, on the other hand, are costs for which the cost or fee structure is not known beforehand with any certainty. These costs are generally less readily visible than explicit costs and hence more difficult to quantify. For instance, until the order is initiated, it is not known exactly what the market charges for executing large orders (e.g., market impact costs).

An illustration of an in-depth transaction cost breakdown is provided by Johnson (2010), as shown in Figure 4. The following subsections provide more detailed descriptions for each of the individual cost components and, as indicated before, closely follow the definitions in Kissell (2006).

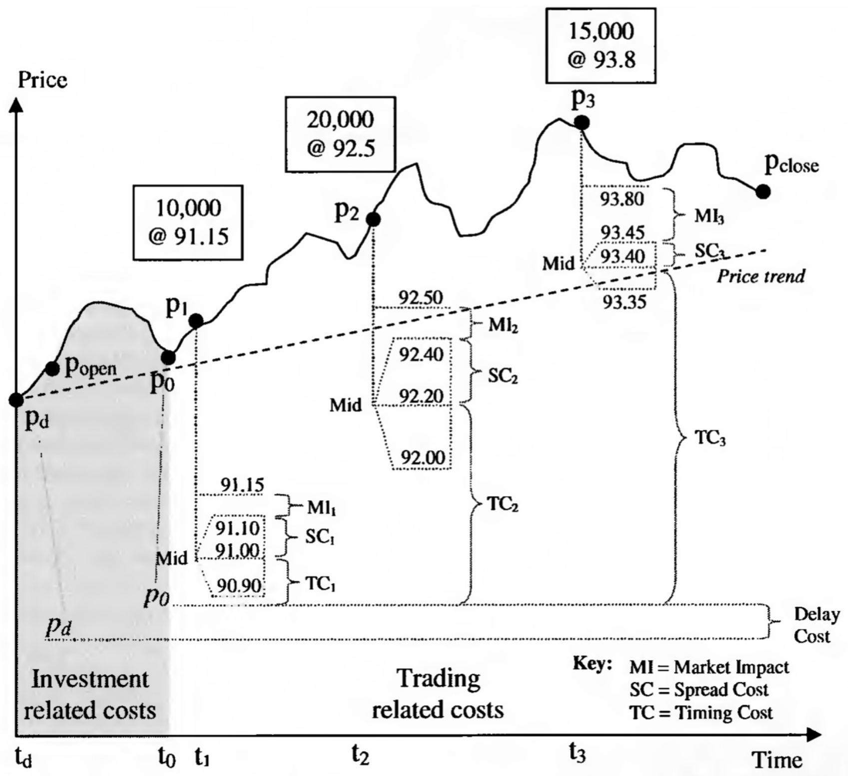

Figure 4: Detailed transaction cost breakdown for an example buy order. *Retrieved from Johnson (2010)*

4.1.1 Investment Related Costs

Following Kissell et al. (2004), investment related costs are "implicit byproducts of investment decisions", which arise between the time the decision to invest is made (t_d) and the time the order execution is initiated (t_0). These costs, which include delays and taxes, can account for a significant portion of the overall transaction costs.

Delay Cost

As expressed by Labadie and Lehalle (2010), *delay cost* is caused by any price change between the initial decision to invest (t_d) and the time the order is released to the market (t_0). This is depicted as the shaded region in Figure 4. Clearly, delay cost arises during the investment decision phase prior to commencement of trading. Delay cost is an implicit variable transaction cost component and is in a way a penalty associated with waiting to trade.

Taxes

Investment related *taxes* are charges placed on funds based on their realized profits. Tax rates differ depending on the investment and type of return. For example, capital gains, long-term profits, dividends, and short-term earnings typically have varying tax rates. Since tax rates are typically known in advance, but the cost quantity is determined by the execution price, investment related taxes are an explicit and variable cost component.

4.1.2 Trading Related Costs

Commissions and fees are the explicit trading related costs. Often, these will be quoted in advance of trading as percentages of the traded value. The most significant costs, however, are the implicit trading costs, which include spread, market impact and timing risk, as well as price trend, and opportunity cost.

These costs make up the majority of transaction costs, and while they cannot be totally eliminated, they can be minimized by choosing the right execution strategy. As a result, traders (or algorithms) must be especially aware of these components. They can make a great investment opportunity become marginally profitable and a rather profitable opportunity to turn out poorly if not properly quantified and controlled (Kissell, 2006). Figure 4 shows the breakdown of these costs for an example buy-order.[9]

Taxes

For U.S. futures, there is an additional duty on contracts traded. For every contract traded, an additional 2¢ of taxes is charged by the CFTC/NFA. Hence, for completeness, taxes should be included as a fixed and explicit trading related component of transaction costs when considering U.S. futures.

Commissions

Commissions are the explicit fees charged by the broker for handling orders and executing the trades. These costs are usually calculated in cents per contract. Hence, commissions are categorized as a fixed and explicit component of transaction costs. It is worth noting that, while commission costs are fixed and known in advance, their value might fluctuate from broker to broker.

Fees

Fees represent the actual charges during execution, charged by brokers for clearing and settlement costs and are typically incorporated into total commissions by the broker. Fees are a fixed and explicit component of the transaction costs.

Spread Cost

Spread cost represents the difference between the best offer (ask) and best bid price at any particular time. Spread cost is an implicit and variable transaction cost component; spreads can be easily measured and observed, but must be recorded each time an order is split in order to determine the spread costs. In addition, spreads vary considerably across markets, assets, and even throughout the day. For some assets, spreads, for example, tend to be wider around the open and close.

Market Impact

Market impact is defined as the price movement caused by a particular trade or order. In general, it has an adverse effect, for instance, driving up prices when placing a buy order. The exact market impact cost of a trade or order is the difference between the price trajectory of the stock with the order and the hypothetical price trajectory that would have occurred if the order had not been released to the market (Kissell, 2006). Unfortunately, one cannot observe both price trajectories at the same time; one can either observe the price evolution with the order or the price evolution without the order. As a result, market impact has often been described as the *Heisenberg uncertainty principle of finance.*

Market impact is usually broken down into two components: temporary and permanent impact. According to Kissell (2006), *temporary* impact reflects the cost of demanding liquidity, while *permanent* or *persistent* impact corresponds to the long-term effect of the order, representing the information content that it has exposed to the market. This is generally caused by the market reevaluating the true intrinsic value of the asset based on the newly arrived information.

[9] Note that for convenience, the time related costs are often grouped together as an overall timing cost: Timing Cost = Price Trend + Timing Risk.

Price Trend

Price trend represents the natural price movement of an asset. Asset prices sometimes exhibit broadly consistent trends, which are also referred to as *drift* or *momentum.* Typically, an upward trend implies that costs will increase when buying an asset, whilst savings will be made if selling. Conversely, the opposite is true for a downward price trend. In Figure 4, the trend line is shown as the dashed line. Price trend is an implicit, variable transaction cost component.

Market or Timing Risk

Kissell and Glantz (2003) use *timing risk* to represent the uncertainty of the transaction cost estimate, where the two main sources of uncertainty are volatility of the asset price and traded volume. However, according to Kissell (2006), timing risk is also dependent upon intra-day trading patterns, cumulative market impact cost caused by other participants, and the underlying execution strategy. Timing risk is an implicit variable cost component.

Opportunity Cost

Opportunity cost represents the cost of failing to (completely) execute the investment decision. The reason this cost arises is usually due to insufficient liquidity or prices moving away. Either way, it represents a missed opportunity, since after the trading horizon prices *may* move even further away. Note that unlike the other cost components, opportunity cost represents a 'virtual' loss rather than an actual one. The loss is only realised when a new order makes up the remainder at a less favourable price. Opportunity cost is an implicit, variable transaction cost component.

Conclusion

The classification scheme described above is essential for understanding transaction costs. Clearly, transaction costs encompass much more than just fixed costs. In fact, as Table 1 shows, the majority of transaction costs are categorized as implicit and variable. Hence, our discussion for measuring and estimating transaction costs and cost optimization will focus on the implicit variable costs, since these are the costs that can actually be attempted to be minimized. It is worth noting, however, that since these cost structures are implicit, they usually have to be estimated using statistical inference or some other estimation technique. The next subsection will elaborate on this.

Moreover, following the expanded IS methodology devised by Wagner and Edwards (1993), it follows that for futures trading, the total transaction costs for order i consist of

$$\mathrm{IS}_i = \underbrace{\underbrace{\left(\sum_j x_j\right)(p_0 - p_d)}_{\text{Delay Cost}} + \underbrace{\sum_j x_j p_j - \sum_j x_j p_0}_{\text{Trading Cost}}}_{\text{Execution Cost}} + \underbrace{\left(X - \sum_j x_j\right)(p_N - p_0)}_{\text{Opportunity Cost}} + \text{Explicit Costs}$$

Here, p_0 is the futures price when the order was released to the market, p_d is the manager's decision price, p_N is the futures price at the end of execution and p_j is the price of the jth trade. Moreover, X is the total number of contracts, multiplied by the contract size, with $X > 0$ for a buy, and $X < 0$ for a sell. Lastly, x_j is the number of contracts executed in the jth trade, multiplied by the contract size. Hence, $\sum_j x_j$ and $X - \sum_j x_j$ are the total number of *executed* and *unexecuted* contracts, respectively, multiplied by the contract size.

4.2 Transaction Cost Analysis

Transaction cost analysis (TCA) is a tool used by investors to strive to get the best possible execution and primarily comprises pre-trade and post-trade analysis. Following the definitions in Kissell (2013), pre-trade analysis occurs prior to the commencement of trading and primarily consists of *cost estimation.* Post-trade analysis, on the other hand, does not include any form of trading decision, either pre-trade or intraday and consists of two parts: *measuring costs* and *evaluating performance.*

Measurement/Estimation

Before we can continue our transaction cost analysis, it is first important to distinguish between measuring and estimating transaction costs. Costs are *measured* after execution (ex-post) to determine the portfolio slippage due to trading (Kissell et al., 2004). On an ex-post basis, we simply measure the aggregate of the trading cost quantities, since there is no way to distinguish between these components. On the other hand, costs are *estimated* before execution (ex-ante) and are used to develop a suitable execution strategy (Kissell et al., 2004). Hence, with regards to cost estimation, it is necessary to estimate each of the components. For instance, there need to be separate estimates for market impact, price trend and volatility. This is due to the fact that the actual order only affects the market impact cost, price trend will occur with or without the order, and volatility is the corresponding uncertainty surrounding actual market conditions and executions.

4.2.1 Cost Estimation

The following part will present an example of a cost estimation model. Even though cost estimation models are generally referred to as market impact models, they often incorporate other components such as price trend and price volatility, whilst they were mentioned as separate components before. Therefore, we can state that these models actually try to estimate the total expected implicit trading costs.

In 1997, Almgren and Chriss provided the financial sciences with a seminal paper to estimate trade related transaction costs (Kissell et al., 2004). Their work is based on a random walk model, which incorporates both permanent and temporary impact as well as price drift and serves as the foundation for many transaction cost models. Cost is computed as the difference between the actual transaction value and the transaction value that would have occurred had all the trades been executed at the arrival price. The market impact model proposed by Almgren and Chriss implements four transaction cost components and looks as follows:

$$MI_{\$}(x_k) = \sum_t x_t P_t - S \cdot P_0,$$

with

$$P_t = P_0 + \underbrace{\sum_{j=1}^{t} \Delta p_j}_{\text{Price Trend}} + \underbrace{\sum_{j=1}^{t} k(x_j) e^{-(t-j)c}}_{\text{Temporary Market Impact}} + \underbrace{\sum_{j=1}^{t} f(x_j)}_{\text{Permanent Market Impact}} + \underbrace{\sum_{j=1}^{t} \epsilon_j}_{\text{Price Volatility}}$$

Here, P_t is the price of the nth transaction, P_0 is the price at commencement of trading, Δp_j is the expected price change between the $(j-1)$th and jth trade. In addition, $k(x_j)$ is the temporary impact function, $f(x_j)$ is the permanent impact function, $e^{-(t-j)c}$ is the temporary impact dissipation function, $c > 0$ is the rate of temporary impact decay and ϵ_j is a random noise $\overset{i.i.d.}{\sim} N(0, \sigma_\epsilon^2)$.

This model poses a problem, because it requires accurate estimation of temporary and permanent impact, and temporary dissipation functions at the trade level (Kissell et al., 2004). As already stated earlier in this section, numerous researchers have documented the difficulty of determining an approximate market impact function at the trade level, because, as stated by the Heisenberg Uncertainty Principle of Trading, it is impossible to simultaneously observe price evolution with and without the order. Nevertheless, since 1997, many cost estimation models have been proposed in the literature, of which most focus on determining an approximate market impact function and attempting to estimate market impact cost by utilizing order attributes.

4.2.2 Cost Measurement

As stated before, a cost measure is an *ex-post* or *after the fact* measure (Kissell, 2013). The aforementioned Wagner's Expanded Implementation Shortfall, for example, is a measure that represents the total cost of executing the investment idea:

$$\text{IS}_i = \underbrace{\underbrace{\left(\sum_j x_j\right)(p_0 - p_d)}_{\text{Delay Cost}} + \underbrace{\sum_j x_j p_j - \sum_j x_j p_0}_{\text{Trading Cost}}}_{\text{Execution Cost}} + \underbrace{\left(X - \sum_j x_j\right)(p_N - p_0)}_{\text{Opportunity Cost}} + \text{Explicit Costs}$$

As reported by Kissell (2013), the trading cost component is generally measured as the difference between the average execution price and the price of the asset at the time the order was entered into the market, p_0 (arrival price). However, in most cost measurement studies, trading costs are typically reported relative to p_0, either in percentages or basis points. We will elaborate on this way of measuring trading costs later.

Cost Measurement Studies

Quantifying trading costs has been the subject of a large number of studies, the majority of which focus on equity trading. For instance, Chan and Lakonishok (1997) and Keim and Madhavan (1997) examine the trading costs of packages of trades that are executed by identifiable institutional investors. Furthermore, Chiyachantana et al. (2004) investigate the features of institutional trading in international stocks from 37 countries between 1997 and 2001.

While most of these studies use publicly available data, Bikker et al. (2007) examine the equity trading costs incurred by ABP, one of the world's largest pension funds, using a proprietary execution data set. More recently, Frazzini et al. (2018), for example, utilize trade execution data from AQR Capital - one of the largest institutional wealth managers in the world - across 21 developed stock markets over a 19-year period.

Frino and Oetomo (2005) provide empirical evidence of the trading costs of institutional trades in futures markets. By examining various stock index and bond futures contracts traded on the Sydney Futures Exchange, their research shows that for these specific contracts, (i) the trading costs are significantly smaller than previously documented for equity markets, and (ii) there is no evidence of an asymmetry between the trading costs of buy and sell orders. Berkman et al. (2005) find similar results for the FTSE100 stock index futures traded on the ICE exchange.

Relevant Order Attributes

In addition to measuring trading costs, cost measurement studies typically examine the relationship between trading costs and some explanatory factor(s) ex-post. These factors may then be used for determining an approximate market impact function and attempting to estimate costs *ex-ante*.

Trading costs have already been shown to be affected by a variety of trade and market-specific variables. Based on a review of some relevant cost estimation studies,[10] we can conclude that the order attributes which are most commonly linked to trading costs are: i) relative order size, ii) volatility, iii) trading style/intensity, and iv) liquidity-related variables.[11] Based on empirical findings and following Kissell (2013), we explicitly find the following main properties for trading costs:

> **P1.** *Trading costs increase with size.* Larger orders will incur a higher cost than smaller orders in the same asset and with the same strategy.
>
> **P2.** *Trading costs increase with volatility.* Volatile assets incur higher costs for the same number of shares (contracts) than less volatile assets.
>
> **P3.** *Trading cost and its risk depend on the trading strategy.* Trading at a fast rate will incur higher expected cost but less market risk. Trading at a slower rate will incur less expected cost but more market risk. This is known as the trader's dilemma, which we will elaborate on later.
>
> **P4.** *Trading costs are inversely related to liquidity.* Assets with wider bid-ask spreads, which typically indicates lower liquidity, have higher trading costs than assets with narrow spreads (all other factors held constant). Moreover, large cap assets or assets with high volume generally have lower trading costs than small cap or low volume assets (holding all other factors constant).

Another variable that is less often incorporated in cost estimation models, but may be important as well, especially for trend-following strategies, is *momentum*. Generally, positive momentum indicates a buying trend, whereas negative momentum indicates a selling trend. Following the reasoning by Bikker et al. (2009), one would expect that as momentum rises, so will the liquidity costs of buy orders (in order to persuade more futures contracts owners to sell their contracts). In a similar way, as momentum declines, liquidity costs are expected to decrease, making it more appealing for market participants to purchase futures contracts. Furthermore, Bikker et al. (2009) conclude that an upward (downward) market trend may signal the presence of positive (negative) news, thus an increase (decrease) in momentum usually translates to an increase in the information content of a buy (sell) order as well. Hence, aside from the properties stated above, we add the following additional property:

> **P5.** *Trading costs increase with the magnitude of momentum.* Buy orders for assets with a strong upward trend are expected to incur higher costs than buy orders for assets with a weaker upward trend, holding other factors constant. On the other hand, sell orders for assets with a strong downward trend are expected to incur higher costs than sell orders for assets with a weaker downward trend, holding other factors constant.

There are several examples of existing studies that investigate the relationship between momentum and trading costs. Korajczyk and Sadka (2004), for instance, assess whether momentum-based strategies that have been previously demonstrated to yield significant returns are profitable after accounting for trading costs. They find that accounting for trading costs results in a significant decrease in the apparent profitability of several previously examined momentum-based strategies, particularly equally weighted ones.

[10] An overview of these studies can be found in Table 12.

[11] Market capitalization, volume, and bid-ask spreads are examples of liquidity-related variables.

4.2.3 Benchmarking

Finally, performance analysis is a key tool for comparing trader/algorithm performances on an ex-post basis, with benchmarking being the most widely used method. This boils down to selecting an appropriate benchmark and then comparing the average execution price with it. Choosing an appropriate benchmark is critical since it serves as the measuring stick for determining whether an execution strategy is profitable or not. As stated by Labadie and Lehalle (2010), a good benchmark "should be easy to track, verifiable, and provide an accurate performance measurement".

Generally, we distinguish between three types of benchmarks: *pre-trade*, *intraday* and *post-trade*. Post-trade benchmarks will not be known until after the trade has been fully executed. Pre-trade benchmarks, on the other hand, are known before the trading even starts. Lastly, intraday benchmarks need to be recalculated as the day progresses. Table 2 shows the various benchmarks grouped in terms of when they may be determined.

Pre-trade	**Intraday**	**Post-trade**
Previous Close	OHLC	Close
Opening Price	TWAP	Future Close
Decision Price	VWAP	
Arrival Price		

Table 2: Three types of benchmarks

Pre-trade benchmarks are readily observable prices, which may be used to directly measure performance. *Post-trade* benchmarks are generally based on closing prices, either for the same trading day as the specific trades or some time in the near future.

Intraday benchmarks, on the other hand, use average prices to reflect the intraday market conditions. As an example, the OHLC (Open High Low Close) average has often been used as a proxy for the mean market price. However, as an average of only four data points, extreme values can easily distort it.

The *Time Weighted Average Price* (TWAP) benchmark is a time-weighted average of observed transaction prices. Since the TWAP incorporates a new price to the existing ones for each new trade and updates the average, it is referred to as a dynamic benchmark. Given a certain time period $t_i \in [t_1, t_2, \ldots, t_T]$, with a specific price p_i (e.g., the mid price), the time-weighted average price is simply defined as

$$\text{TWAP}_T = \frac{\sum_i p_i}{T}$$

Because all TWAP values are weighted in the same way, extreme prices can have a significant impact on the average.

The *Volume Weighted Average Price* (VWAP) benchmark is arguably the most accurate indicator of market price movement over time. Given N trades in a certain period, each with a specific price p_i and traded volume v_i, the volume-weighted average price is defined as

$$\text{VWAP}_N = \frac{\sum_i p_i v_i}{\sum_i v_i}$$

Small trades at extreme prices are smoothed out, while the largest trades dominate the average. As a result, large trades have a greater impact on the benchmark price than small trades.

5 Algorithmic Execution

The automated programs that implement strategies to execute (large) orders in markets with electronic access are widely referred to as execution algorithms (EAs). As the trading environment has become more competitive, investors have turned to such algorithms (Kissell, 2006). Moreover, under the current macro-economic environment, with COVID-19 still present, a higher adoption rate of algorithmic execution is observed, in order to assist traders in achieving best execution. A report by the IMARC Group (2021) states that in March 2020, more than 60% of trades for ticket sizes over $10 million were executed via an algorithm, while this number was less than 50% in 2019. Moreover, a survey by The TRADE (2021) reveals that hedge funds in Europe and the United States are increasingly relying on execution algorithms; over half of the hedge funds surveyed use algorithms to execute the majority of their total value traded.

5.1 Objective Function

The optimal execution strategy is generally obtained via an optimization process. Using the knowledge obtained in Section 4 and Lillo (2016), we are now able to define the optimization problem. An investor wants to trade (buy or sell) a given number of contracts and wants to minimize costs by trading incrementally. Hence, suppose for order i an investor has X futures contracts to trade in T time periods, where all the X contracts have the same trade direction. Next, let v_t ($t = 1, \ldots, T$) be the (signed) number of contracts to be traded in interval t, p_t be the price at which the investor trades at interval t and p_0 be the price before the start of the execution. A very often used objective function for order i is then defined as

$$C_i(v_t) \equiv \sum_{t=1}^{T} v_t p_t - X p_0,$$

that is, the difference between the actual cost and the cost in an infinitely liquid market, which is very similar to how we defined trading cost before. Generally, this cost is a stochastic variable, so one typically wants to minimize $E\left[C_i(v_t)\right]$. Note that this assumes a risk-neutral profile.

Trader's Dilemma

However, traders are generally not risk-neutral and, therefore, need to balance the trade-off between cost and risk. According to Kissell and Malamut (2005), a trader generally faces a dilemma of trading quickly (aggressively) and trading slowly (passively). Mandes (2016) expands on this, explaining that, at one extreme, a market order can be used to immediately execute an order, resulting in a high expected trading cost (i.e., a lower execution price due to market impact). Moreover, Mandes (2016) states that, on the other hand, the order can be equally split and set to execute at a consistent rate throughout the execution period. This strategy is often referred to as a TWAP strategy. It typically has the lowest impact but an inherent price risk, that is, the difference between the effective execution price and the arrival price benchmark as a result of random price changes (shortfall). The optimal execution rate generally lies somewhere in-between this range, which is bounded on one side by the least variance strategy and on the other by the minimum impact strategy.

Problem Formulation

Almgren and Chriss (2001) provide a mathematical formulation of this problem. As expressed in their paper, the objective is to compute a trajectory function $x(t)$, representing the remaining number of contracts to be executed at time t.

Here, the initial target is $x(0) = X_0$, that is, the number of contracts at trade initiation, and the completely executed order at the end of the execution period is referred to as $x(T) = 0$. Next, we define $v(t) = x(t-1) - x(t)$ as the corresponding execution rate (i.e., the number of contracts to be executed in one time interval). For a specified level of risk aversion λ, an optimal trading strategy may then be determined via the following cost minimization:

$$\min_x \left(E[C_i(x)] + \lambda \cdot Var[C_i(x)]\right),$$

where $C_i(x)$ is the trading cost for order i and $Var[C_i(x)]$ is the trading cost variance as a proxy for risk.

5.2 Solving The Problem

Solving the optimization problem described above for various levels of risk will result in numerous optimal trading strategies. Each strategy has the lowest cost for a specific level of risk as well as the lowest risk for the specified cost. The solution of the cost minimization for all values of λ constitutes the Efficient Trading Frontier (ETF), which was first introduced by Almgren and Chriss (2001).

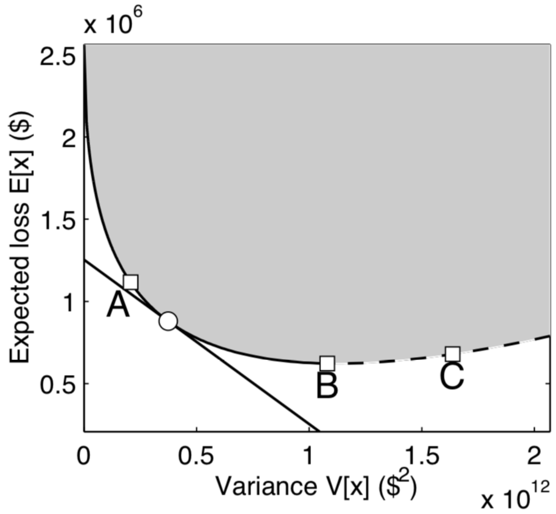

Figure 5: Efficient Trading Frontier as introduced by *Almgren and Chriss (2001)*

Figure 5 shows an example of the ETF, produced by Algmren and Chriss (2001). Each point on the frontier represents a unique strategy for executing the same order in an optimal way. The tangent line represents the best solution for a specific risk parameter λ. Strategy A is chosen by a risk-averse trader wishing to buy or sell quickly to reduce exposure to market risk, despite incurring transaction costs in doing so. Strategy B is called the "naive" strategy, since it represents the optimal strategy corresponding to simply minimizing expected trading costs without regard to variance. Strategy C would be chosen only by a trader who prefers risk. The trader postpones execution, thus incurring both higher expected trading costs and higher variance during the extended period in which the order is executed.

5.3 Execution Algorithms

As the trading environment has become more complex and competitive, investors have turned to "efficient" algorithms for order execution, in order to solve the optimization problem formalized above (Kissell, 2006). These algorithmic execution strategies are mainly driven by a style of trading and their objective is to minimize either absolute or risk-adjusted costs relative to a benchmark.

The investors specifically need to define input parameters, such as the start time of the order and the urgency (participation rate). The execution algorithms are then generally based on a stochastic market impact model, which represents both the volatile price evolution of assets and the market impact during execution (Gatheral & Schied, 2013). The cost criterion has to incorporate both the liquidity costs arising from market impact and the price/timing risk resulting from late execution. We have already seen such a framework in Section 4.2.1. Optimal trading trajectories, which are the basis for execution algorithms, are then obtained as minimizers of the cost criterion among all trading strategies that liquidate a given asset position within a given time frame, as was discussed in Section 5.2.

It is worth emphasizing that the main purpose of execution algorithms is, without a doubt, to achieve optimal execution. This concept, however, has numerous conflicting aspects and will be defined differently for various users based on their specific objective functions. While different types of execution algorithms can accommodate several options for execution, no single execution algorithm can optimize all aspects of concurrency.

6 Data Description

The underlying data sample contains all algorithmic trades originated from the trading activity of Varick Capital at several exchanges via their broker, combined with tick data for each specific trade. The following section describes the data, provides a definition of the variables and presents descriptive statistics.

6.1 Portfolio Construction and Trade Execution

With respect to trading at Varick Capital, we distinguish three types of potential trading orders. The first category relates to the so-called rebalancing activities of the hedge fund. Rebalancing is defined as adjustment of the portfolio weights by selling and buying futures. The second category of trading orders are referred to as *signal* orders, which originate from an underlying investment system. The third and last category are *rolls*, which are orders to roll a position by closing the current contract and (simultaneously) establishing the same position in another contract month further in the future. Roll orders are excluded from the analysis though to avoid any bias in measuring trade costs caused by transactions associated with rolling over a position from near to deferred contracts.

The implied (trading) orders are uploaded into an execution management system, where each futures market has an associated execution algorithm. The objective of the execution algorithms is to minimize trading costs and not to make any explicit portfolio decisions. Hence, the portfolio construction process is separate from the execution process (Frazzini et al., 2018). The execution algorithm uses electronic exchanges to directly obtain market liquidity and dynamically breaks up (parent) orders into smaller (child) orders to minimize trading costs. Given a start time and urgency (participation rate) parameter, the goal ultimately is to solve the trader's dilemma at the specified level of urgency (as a proxy for risk aversion λ), starting at a predefined time. Specifically, the algorithm used by Varick Capital attempts to strategically attract passive fills while participating with volume.

6.2 Data

The execution data is drawn from the post-trade analysis data maintained by Varick Capital and covers all orders executed algorithmically, except for the *rolls*. The remaining data set consists of 681 worldwide futures orders, executed between January and September 2021, with a total transaction value of \$2.26 billion. The fund's futures contracts universe consists of 50 worldwide futures on stock indices, bonds, commodities and currencies. For all of these futures contracts, the data set provides specifications like tick size and contract size, which were taken from Bloomberg. Table 14 provides a complete overview of the contracts traded and their specifications.

Moreover, what distinguishes our data set from others is that, in addition to execution data, we have access to tick data collected by the broker. Tick data is the most granular intraday data and is the sequence of each executed trade or bid/ask quote. This specifically includes transaction prices, time stamps, quantities, as well as the best bids and best asks derived from the order book and allows us to extract intraday data on spreads, volumes, and prices for all of the orders. As a result, we are able to extract the mid price at the exact start time of the order, calculate several benchmark prices and measure all costs and performances in terms of spreads.

6.2.1 Data Constituents and Definitions

The data set *unambiguously* classifies each order as either buyer- or seller-initiated, so that proxies for trade direction are not required. In addition, the data set contains fields that document the date and time at which the order was executed, the expiry date code, the exchange on which the order was executed, the number of contracts traded, and the asset class of the traded futures contract. Lastly, the data provides several relevant prices and detailed information on a variety of trade- and market-specific characteristics, which are discussed below. A complete overview of the variables in the data set, as well as their definitions, can be found in Table 13.

Prices

For each order, the data provides the execution price, the arrival price and the decision price. The *execution price* is defined as the volume-weighted average price of all child fills of the particular order throughout the execution period. The *arrival price* is defined as the futures mid price at the time the order was initiated. The *decision price* is the futures price when the decision to buy or sell the asset was made. Together, these prices allow us to capture the costs incurred at different phases of the investment cycle. Cost measured against the decision price, for example, captures all price changes from the moment the decision is made, while cost measured against the arrival price captures price movements from the start of the execution (Hedayati et al., 2018). Lastly, the data sample provides two benchmark prices: the interval TWAP and interval VWAP, which will be explained in more detail in Section 7.4.

Characteristics

Moreover, in Section 4.2.2, we concluded that the order attributes that are expected to have the most influence on trading costs are i) relative order size, ii) volatility, iii) trading style/intensity, iv) liquidity-related variables and v) momentum. The variables utilized in this analysis are based on these order attributes and are described below.

Market-Specific Characteristics

According to market microstructure theory, the price formation process is dependent on the distinctive characteristics of each futures market. Hence, regarding market-specific characteristics, we distinguish the daily trading volume, momentum, volatility and spread. The *daily trading volume* is defined as the total volume (in contracts) of the relevant market on a specific date, taking into account the expiry date code. The annualized *volatility* is calculated based on the closing prices (in local currencies) of the last thirty trading days prior to execution[12] and is expressed in % per year. Similar to the definition used by Bikker et al. (2007), *momentum* is calculated as the average daily return of a specific futures contract over the last five trading days prior to execution and is again expressed in %. Like for volatility, these returns are based on the closing prices of the specific trading days. Typically, momentum indicates whether there is an upward or a downward trend for a particular futures contract. The quantities above are all calculated using data from Bloomberg. Lastly, using the best bid and ask prices from the tick data, the volume-weighted average bid-ask *spread* during the life of the order is calculated. The spread is recorded for each child fill of the specific order and is subsequently weighted by the volume of that specific fill to obtain a volume-weighted average.

[12]Following the methodology used by Bikker et al. (2007), we choose a period of thirty days to ensure that recent price fluctuations are incorporated in the measure of volatility.

Trade-Specific Characteristics

Regarding trade-specific characteristics, we first distinguish the relative size of the order. Since there is no definition of what a large order in futures markets constitutes, we construct 2 order-size classes based on their (relative) executed size. See Appendix A.1 for a precise definition of these relative order-size classes. The idea behind these size classes is that orders in size class 1 are generally not expected to move the market even if executed as a market order. Orders in size class 2, on the other hand, are expected to cause significant impact even if executed gradually. Moreover, the data provides information on the *duration* of the trade. Similar to Bikker et al. (2007), we define duration as the time elapsed between the moment the order was initiated and the moment it was fully executed. Trade duration may be regarded as a proxy for liquidity, since trades that are executed quickly *generally* indicate that the algorithm had little difficulty in finding liquidity without having to accept a significant price change. Lastly, one way for the execution algorithm to minimize trading costs is to split each *parent* order into smaller *child* orders and submit them as separate trades. Therefore, the number of separate trades executed is directly represented by the variable *child fills*.

6.3 Sample Properties of Varick's Trades

Market-Specific

The 681 orders in total consist of 100 stock index orders, 118 currency orders, 242 commodity orders and 221 bond orders. Table 3 provides information on the variety of futures traded by Varick Capital in terms of their characteristics by presenting the aggregate results for the entire data sample.

	Volatility (%)	**Daily volume (10^5)**	**Contract value (\$100,000)**	**Spread (bps)**
Mean	13.59	2.24	1.16	2.02
St. dev.	11.78	3.49	0.71	1.82
Median	10.92	0.98	1.18	1.22
5% Quantile	0.37	0.14	0.25	0.35
95% Quantile	33.9	8.69	2.21	5.83

Table 3: Descriptive statistics of futures contracts traded

As we would expect, we observe a wide variation in characteristics between the markets in our sample. We will use this variation to estimate the coefficients in the quantile regression of Section 8. Even though there exist differences in characteristics between markets belonging to the same asset classes as well, the following part will focus on the differences across the asset classes. Table 15 in the Appendix illustrates where these discrepancies may come from by presenting the distribution of characteristics across the asset classes.

From Table 15, we can, for instance, see that while the bond futures have an average annualized volatility of 2.91%, the commodity futures have an average annualized volatility of 25.89%. This could point to significant differences in the price formation process. Moreover, daily volumes vary widely, which could indicate large differences in liquidity. While the average daily volumes for the commodity and currency orders are approximately 86,000 and 100,000 contracts, respectively, the average daily volume for the bond orders is nearly 5 times higher.

Looking within markets, the average daily volume (in contracts) for the S&P500 futures contracts in our data sample is approximately 1.4 million, which is equivalent to approximately $287 billion in traded value.[13] As a comparison, the average daily volume in Apple stocks - a constituent of the S&P500 and one of the most actively traded stocks in the US stock market - over the past year was about $11 billion.[14]

Moreover, the volume-weighted average spread during the life of the order, relative to the arrival price, is reported in basis points. Again, we observe a clear variation in relative spreads, confirming the hypothesis that there are substantial differences across the markets in terms of liquidity. While the average spread for the bond futures contracts is 0.75 bps, the average spread for the commodity futures contracts is 3.69 bps. Moreover, Table 15 shows that the average relative spread for the stock indices contracts in our data sample is 1.76 bps. Comparing this to other research, Frazzini et al. (2018) report an average bid-ask spread at arrival of 21.33 bps for their sample of trades from 21 developed equity markets and conclude that the average spread found for their sample of stocks is consistent with spreads quoted in other studies from TAQ data. Altogether, this confirms the conjecture that stock indices futures markets are generally more liquid than equity markets (Fleming et al., 1996).

Trade-Specific

Table 4 reports sample statistics of several trade characteristics such as trade duration and trade size. The *trade* or *transaction value* of an order is calculated by multiplying the arrival price of each order by the number of contracts executed and the contract size of the specific contract.[15] Among the 681 orders, 415 are buy orders, with a transaction value of $1.24 billion, and 266 are sell orders, with a transaction value of $1.02 billion.

	Duration (mins)	**No. of contracts**	**No. of child fills**	**Trade value ($ million)**	**Rel. trade size (bps)**
Buy Orders					
Mean	1.91	24.22	6.12	3.00	3.16
St. dev.	5.36	41.94	7.53	6.35	7.56
Median	0.05	10.00	3.00	0.79	1.07
5% Quantile	0.00	1.00	1.00	0.13	0.05
95% Quantile	7.34	87.30	22.00	11.04	12.14
Sell Orders					
Mean	1.85	29.39	5.79	3.84	3.14
St. dev.	3.89	45.42	6.34	7.87	4.99
Median	0.23	11.5	3.00	0.86	1.26
5% Quantile	0.00	1.00	1.00	0.08	0.07
95% Quantile	7.34	113.00	17.75	11.91	10.17

Table 4: Trade-specific characteristics of buy and sell orders

[13] Here, we used the volume-weighted average arrival price of all the S&P500 futures contracts in our sample and multiplied by the contract size to calculate the traded value in USD.

[14] Obtained from Bloomberg.

[15] Since we want to obtain the USD value of the trade in this research, the arrival price is first converted from a foreign currency to USD. For some of the traded futures contracts, prices are denominated in a foreign currency, as can be seen from Table 14.

From Table 4, we observe that, on average, trade durations are similar for sell and buy orders; the average trade takes about 2 minutes to be completed. The average trade size for buy (sell) orders is 24.22 (29.39) contracts, with an average trade value of $3.00 ($3.84) million. Expressed as a fraction of daily trading volume, average trade size equals 3.16 bps for buy orders and 3.13 bps for sell orders.[16]

Lastly, we conclude that the execution algorithm has split the total 681 parent orders into 4081 smaller child orders in order to minimise trading costs. As a result, Table 4 reports that the average order was broken up into approximately 6 separate trades, or stated differently, it on average took 6 *child fills* in order to fully execute the order.

[16] We used a slightly smaller sample size to compute the relative size sample statistics than we did for the other trade-specific sample statistics. This is due to the fact that, for some orders, we did not have any data on the daily volume. Eventually, we calculated the statistics based on 672 (parent) orders.

7 Quantifying/Measuring The Implicit Transaction Costs

This section digs deeper into the data presented in Section 6 in order to quantify the cost of Varick Capital's trades. As stated in the conclusion of Section 4.1, we will specifically quantify the implicit variable costs, since we are mainly interested in (relative) price changes, and not fixed costs such as commissions. In this section, we describe our methodology for measuring costs. Moreover, we present our results in detail and compare our findings with previous research. The last part of this section covers benchmarking.

7.1 Implicit Cost

In Section 4.2.2, we defined the total transaction costs (Implementation Shortfall) for order i as

$$\mathrm{IS}_i = \underbrace{\underbrace{\left(\sum_j x_j\right)(p_0 - p_d)}_{\text{Delay Cost}} + \underbrace{\sum_j x_j p_j - \sum_j x_j p_0}_{\text{Trading Cost}}}_{\text{Execution Cost}} + \underbrace{\left(X - \sum_j x_j\right)(p_N - p_0)}_{\text{Opportunity Cost}} + \text{Explicit Costs}$$

Hence, total transaction cost (IS) is equal to the sum of all implicit and explicit costs. Since all trades in our sample were fully executed, opportunity costs are negligible. This is consistent with the findings of Keim and Madhavan (1997), who discovered a 95% completion rate in institutional trades. As a result, the total transaction cost (IS) in this research is equal to the sum of execution cost and explicit cost. As defined in Kissell (2013), execution cost represents the costs incurred as a result of executing contracts at a less favorable price than the original decision price. It can be further broken down into delay cost and trading cost, which are the components of interest in the upcoming part; Subsection 7.2 will focus on delay costs, while Subsection 7.3 will concentrate on trading costs.

7.2 Delay Cost

Following the definitions in Kissell (2006), the (implicit) delay cost component of order i is measured as the difference in price between the initial decision to invest (t_d) and the time the order is initiated (t_0):

$$\text{Delay Cost}_i = \left(\sum_j x_j\right)(p_0 - p_d),$$

where p_0 is the arrival price and p_d is the decision price. In this analysis, the investment decision is based on the closing price of the previous trading day. This decision is subsequently implemented the next trading day at a time determined by the start time parameter that is provided to the execution algorithm. Hence, the delay cost is defined as the price change from the previous close (t_d) to the start time of the trade (t_0) and the decision price p_d is the closing price of the last trading day before t_0. By setting the start time parameter, the investor cannot participate in this price change, resulting in a sunk cost in the case of an adverse price change or a savings in the case of a favorable price change (Kissell, 2006). In this analysis, we measure the *relative delay cost* for order i in the following way:

$$\text{Relative Delay Cost}_i \text{ (bps)} = d_i \cdot \frac{p_i^0 - p_i^d}{p_i^d} \cdot 10^4,$$

where p_0 and p_d are the arrival price and decision price for order i, respectively, and the direction d_i is 1 for all buy orders and -1 for all sell orders. By using this definition, a positive number implies that a trade has been initiated at a worse price than the prevailing price at the moment the investment decision was made, both for buy and sell orders.

7.2.1 Results

Table 5 reports sample means, standard deviations, and quantiles of delay costs for both buy and sell orders. Moreover, it reports the volume-weighted average cost to account for the volume effect and minimize data noise, which is obtained by weighting each observation by the volume of the specific contract at the specific date.

	Relative Delay Cost (bps)
Buy Orders	
Volume-weighted average	32.05
Mean	33.23
Standard deviation	87.00
Proportion > 0	0.67
Median	16.35
5% Quantile	-76.37
95% Quantile	187.13
Sell Orders	
Volume-weighted average	-17.59
Mean	-27.44
Standard deviation	115.55
Proportion > 0	0.45
Median	-0.60
5% Quantile	-250.36
95% Quantile	103.36

Table 5: Sample statistics of relative delay cost (in bps)

For the buy orders, the volume-weighted average relative cost of investment delay is approximately 32 basis points. This implies that for the buy orders, due to the time gap between the investment decision and initiation of the order, there has been an average *adverse* (upward) price change of approximately 32 basis points, relative to the decision price p_d. This number reflects the average penalty or sunk cost associated with waiting to trade. Interestingly, for the sell orders, we observe something different. The negative sign implies that, on average, prices have actually increased between the investment decision and initiation of the sell orders. Hence, due to the gap between the investment decision and trade initiation, on average, there has been a *favorable* relative price change of approximately 18 basis points. Consequently, for the sell orders, waiting to trade has on average resulted in a savings relative to the decision price p_d.

Although the average delay cost is positive for buy orders and negative for sell orders, there is a significant disparity in delay costs between orders in the left and right tails of the cost distributions. In particular, 33% of the buy orders have actually incurred a negative delay cost, whereas 45% of the sell orders have incurred a positive delay cost. Hence, further study may be required to determine which asset classes or specific markets contribute the most to these averages.

7.3 Trading Cost

Referring back to Section 4.1's Implementation Shortfall methodology, the (implicit) trading cost component for order i is generally measured as

$$\text{Trading Cost}_i = \sum_j x_j p_j - \sum_j x_j p_0,$$

where p_j is the executed price of the jth child fill and p_0 is the arrival price. Using this equation, the (absolute) dollar value of costs incurred during the execution of a specific order might be measured. However, analogous to the previous subsection, we will use a slightly different definition to measure the relative trading cost and its relation to market impact, as we will see in the next subsections. Defining the trade cost in this way is mainly to be able to compare our results with previous research.

7.3.1 Relative Trading Cost

We specifically follow Fraenkle et al. (2011) by defining the *relative trading cost* for parent order i as the relative price change (in bps) between the average execution price p^{exe} and the arrival price p^0:

$$\text{Relative Trading Cost}_i = d_i \cdot \frac{p_i^{exe} - p_i^0}{p_i^0} \cdot 10^4,$$

where p_{exe} is the average execution price for order i, p_0 is the arrival price and the direction of the order d_i is 1 for all buy orders and -1 for all sell orders. For each order executed, our data set provides both of these prices. By using this definition, for both buy and sell orders, positive trading cost implies that an order has been executed against a worse price than the arrival price.

To acquire some further insights, we also measure the relative trading costs in terms of spreads and ticks. As indicated in Section 4.1, spreads may vary considerably across markets, assets, time and even throughout the day. In fact, Table 15 showed a substantial difference in spreads across the markets in our sample. Hence, if we are averaging costs (performance) across different futures contracts we should allow for differences in (absolute) spread by measuring in terms of spreads instead of bps. This allows us to distinguish between cost due to spread and cost due to other components. Moreover, following Frino and Oetemo (2005), the cost of executing an order may be measured in price ticks, which is the convention used in futures markets. In conclusion, we write

$$\text{Relative Trading Cost}_i \text{ (spreads)} = \mathrm{d}_i \cdot \frac{\mathrm{p}_i^{exe} - \mathrm{p}_i^0}{\mathrm{S}_i}$$
$$\text{Relative Trading Cost}_i \text{ (ticks)} = \mathrm{d}_i \cdot \frac{\mathrm{p}_i^{exe} - \mathrm{p}_i^0}{\mathrm{MT}_j},$$

where S_i represents the volume-weighted average bid-ask spread during the life of order i and MT_j represents the minimum tick of the underlying futures contract j.

7.3.2 Trading Cost and Market Impact

Given the objective of this research, trading costs are quantified in a manner that corresponds closely to market impact measures used in previous examinations in, for instance, equity markets. Market impact is generally defined as the interaction of the specific order with the market, that is, how large is the price change influenced by this order (Fraenkle et al., 2011). As a result, market impact is often defined as the price difference between a benchmark price, which should be influenced by the specific order as little as possible, and a price incorporating the full impact, which is basically what we do here as well.

What is neglected in many of these studies, however, is that the (relative) price change may also contain price movements which are not attributable to the market impact of the own order, which Fraenkle et al. (2011) refer to as externally triggered movements. As stated in Section 4.2, on an *ex-post* basis, we can only measure the aggregate of the implicit trading cost component, since there is no way to distinguish between its constituents.[17] One can, for instance, never be sure that the price change was caused by the market impact of the own order and was not due to the impact of other trades, a price trend, or other market movements that caused the price to change. For simplification, we may write the relative price change for order i as a sum of two components here:

$$\begin{aligned}\text{Relative Trading Cost}_i &= \text{Price Move}_i + \text{Impact}_i \\ &= \mathrm{r}_i^e + \mathrm{I}_i\end{aligned}$$

The component I is the market impact, while the other components are bundled together as one component here, which we refer to as the price move r_e. Note that this component is assumed to be induced by external influences.

Next, we define the market impact as the mean value of the distribution of the relative trading cost. The advantage of this definition by Fraenkle et al. (2011) is that the mean value of the r_e distribution is assumed to be zero. This implies that the mean value of the relative trading cost distribution is an unbiased estimator of the empirical market impact. Given the relatively short durations of the trades and the fact that external price movements might have both an adverse and favorable effect on prices, we expect that, on average, market movements[18] do not affect $\mathbb{E}[\mathrm{C}]$. To summarize, we can define the market impact of a trade as

$$\mathbb{E}[C] = \mathbb{E}[r_e] + \mathbb{E}[I] = \mathbb{E}[I].$$

This implies that, in expectation, trading costs are assumed to be equal to the actual market impact of the orders. However, although the external price movements do not contribute to the average impact $\mathbb{E}[I]$, these movements may significantly increase the variance of the trading costs (Fraenke et al., 2011). The consequences of this will be discussed in Section 7.3.3.

7.3.3 Results

Following previous studies, Table 6 reports sample means, standard deviations, medians and quantiles of trading costs for both buy and sell orders. Moreover, following Bikker et al. (2009), this table reports the principal-weighted average cost, which is obtained by weighting each observation by the USD value of the trade,[19] so that smaller orders contribute less to the average trading costs than larger ones. In this way, we account for the size/volume effect and try to minimize data noise. This approach is standard in the investment industry and allows for the evaluation of the overall dollar amount of the relative price change. Table 6 also includes standard deviations, to give an indication of the significance of the average relative trading costs estimates. However, it is worth noting that these standard deviations have been calculated under the assumption that observations are mutually uncorrelated (Bikker et al., 2007).

[17] As stated before, spread cost is the only component that can be measured separately, which will be shown later on.

[18] External market movements might, for instance, result from a daily upward drift in prices or market impact caused by other participants.

[19] The price used for calculating the USD value of the trade is the arrival price, after converting it to USD from the local currency.

	Relative Trading Cost (bps)	Relative Trading Cost (spreads)	Relative Trading Cost (ticks)
Buy Orders			
Principal-weighted	0.15	0.25	0.27
Mean	0.07	0.18	0.20
Standard deviation	4.13	1.44	4.03
Proportion > 0	0.51	0.51	0.51
Median	0.07	0.05	0.06
5% Quantile	-4.78	-1.50	-3.82
25% Quantile	-0.79	-0.50	-0.50
75% Quantile	0.79	0.50	0.50
95% Quantile	4.91	2.39	4.46
Sell Orders			
Principal-weighted	0.16	0.34	0.24
Mean	0.06	0.13	0.10
Standard deviation	4.34	1.58	3.76
Proportion > 0	0.50	0.50	0.50
Median	0.04	0.03	0.04
5% Quantile	-4.98	-1.95	-3.00
25% Quantile	-0.67	-0.50	-0.50
75% Quantile	0.47	0.50	0.50
95% Quantile	6.73	2.77	3.97

Table 6: Sample statistics of relative trading costs (in bps, spreads and ticks)

From Table 6, we find that the principal-weighted average trading costs are 0.15 basis points (bps) for buy orders and 0.16 bps for sell orders. These numbers tell us how far the execution price is from the arrival price. A positive number indicates an adverse price change, that is, prices moving against the trader. Hence, in this case, on average, execution prices have moved away 0.15 bps from the arrival price for the buy orders and 0.16 bps for the sell orders. As indicated in Section 7.3.2, this might be either due to market impact, market movement (the market as a whole is moving) or a combination of both, but we expect these values to on average reflect the real market impact of the orders.

We would generally expect trading costs to be positive since, all other things being equal, any trading tends to push the price in the direction traded by the investor. Table 6 shows that this is indeed the case, on average. However, by testing the deviation of the mean trading costs (in bps) from zero with the use of a t-test (adjusted for sample size), we conclude that the means for both buy and sell orders are indistinguishable from zero at any reasonable significance level.[20] Moreover, we see that approximately half of the orders executed result in profits rather than losses, both for buy and sell orders. The costs for the 5% and 25% quantiles are negative, whereas from the median these numbers become positive. As stated before, this might be due to idiosyncratic price movements and does not necessarily imply the orders had a negative impact on the prices. Looking at the quantiles in more detail, we notice a substantial difference in trading costs between orders in the left and right tails of the cost distribution. However, the trades that really matter in terms of trading costs, are the ones in the right tail of the distribution. These are crucial in cost management and will play a key role in the remainder of this research.

[20]The statistics of the t-test for buy and sell orders are 0.36 and 0.24, with a p-value of 0.72 and 0.81, respectively.

Furthermore, the median costs are similar to the mean costs, suggesting that trading costs are neither positively nor negatively skewed. In addition, in all of the measures used, the unweighted average price effects for buy and sell orders are of similar magnitude to the principal-weighted ones. This indicates that it is not necessarily the largest orders that lead to the highest trading costs,[21] which is an interesting observation to which we will return later.

Lastly, analyzing the results in terms of spreads, we observe that the principal-weighted average trading costs are 0.25 spreads for buy orders and 0.34 spreads for sell orders. Hence, following the same reasoning as before, on average, execution prices have moved away 0.25 spreads from the arrival price for the buy orders and 0.34 spreads for the sell orders. Clearly, the quantiles are much smaller in terms of spreads, suggesting that a substantial part of trading cost is attributable to spread cost. In Section 7.4, we will investigate these numbers more thoroughly and examine their implications for the execution algorithm's performance.

Comparison with Previous Research

It is interesting to compare our findings with previous research. As mentioned in Section 4.2.2, Bikker et al. (2007) examine the equity trading costs incurred by ABP, one of the world's largest pension funds. For an average trade value of approximately 1.5 million euros, the paper reports average trading costs of 20 bps for buy orders and 30 bps for sell orders. More recently, Frazzini et al. (2018) examine trading costs incurred using a proprietary execution algorithm across 21 developed stock markets over a 19-year period. For an average trade value of $607,200, value-weighted mean costs of 16 basis points are reported. Keep in mind that the average trade value in this research was shown to be over $3 million.

Even though we are considering a variety of futures from different asset classes and cannot compare the costs incurred in trading, for example, commodity futures with equity trading, we can conclude that the order of magnitude of the average trade costs in our study is significantly lower than the average costs reported above. Even the 95% quantile values in our results are significantly lower than the average costs reported in those studies. As stated in Section 3.4, there is typically a low probability of private information in futures markets. The low probability of private information translates to low information asymmetry and therefore, low adverse selection costs in futures markets. In turn, this implies that orders executed in futures markets are likely to have a relatively small effect on prices.

Indeed, studies like Frino and Oetomo (2005) and Frino et al. (2007) report substantially lower trading costs. More specifically, we can compare the costs incurred for the S&P500 index futures orders in our sample to the numbers reported in, for instance, Frino et al. (2007). The volume-weighted average cost for the orders in S&P500 contracts in our sample is 0.60 bps. In comparison, Frino et al. (2007) report a volume-weighted trade cost for buy orders of 0.46 bps[22] for trades initiated by institutional clients in S&P500 index futures. These findings back up previous research showing a significant difference in absolute spreads between the S&P500 futures contract and its constituent stocks (Fleming et al., 1996), which typically indicates a substantial difference in liquidity (and thus we expect costs to be lower).

[21]We also calculated principal-weighted standard deviations, medians, and quantiles to test this hypothesis. However, the weighted quantiles are actually lower in magnitude than the unweighted ones, confirming our initial thought: there is no evidence that a few large orders with high costs dominate the weighted average.

[22]There are no sell orders for the S&P500 in our sample, so we can only compare buy orders. Moreover, for the comparison, we assumed that orders are categorized as Group 3 based on their size (i.e., 11 up to 20 contracts), since the average number of contracts traded for the S&P500 orders in our sample is 11 contracts.

In addition, Frino and Oetomo (2005) examine orders executed on the Sydney Futures Exchange (SFE) and report their results in terms of ticks. In particular, for orders executed in the SPI200 futures markets with a size of 9-29 contracts, average trading costs of 0.64 ticks are reported for buy orders. In comparison, for the buy orders in SPI200 contracts in our sample, the volume-weighted average cost is 2.19 ticks, which is slightly higher. In addition, for 10-year Australian bond futures, Frino and Oetomo (2005) report trading costs of 0.76 ticks for buy packages and 0.73 ticks for sell packages. In comparison, the volume-weighted average trading cost for the 10-year Australian Bond orders in our sample is 0.24 ticks for buy orders and 0.32 ticks for sell orders. Hence, despite the fact that we employed a limited number of orders, our results appear to be consistent with earlier findings in futures markets in terms of cost magnitude.

Overall, it is clear that the relative trading costs found in this research are of a *similar magnitude* as the corresponding costs documented in previous studies of the S&P500 index futures and SFE markets, but significantly lower than the costs documented in previous studies for equity markets. These findings confirm conjectures in previous research that we expect a low probability of private information in futures markets. Moreover, futures markets are generally liquid in nature and trading costs are inversely related to market liquidity.

Thirdly, in contrast to the findings from equity markets, this research finds limited evidence of a buy-sell asymmetry in trading costs in futures markets. As stated in Section 3.4, futures markets are not constrained by short-selling restrictions. Hence futures markets typically are as likely to facilitate purchases as sales. This implies that the systematic difference in trading costs of buy and sell orders documented in equity markets is unlikely to occur in futures markets. As concluded by Chan and Lakonishok (1993), the buy-sell asymmetry found in equity studies "is due to the high cost of short selling and the general reluctance of traders to short sell on stock markets". Our results are consistent with the findings in other futures markets studies, for instance, those reported in Frino and Oetomo (2005) and Berkman et al. (2005).

7.3.4 Analysis of Orders with 1% Highest Cost

Summarizing, average trading costs are shown to be insignificantly different from zero, implying that the average impact of Varick Capital's trades on prices seems rather small. However, analyzing the quantiles, we observed a substantial difference in costs, which prompts us to investigate which orders lead to the highest cost (in relative terms) and how significant the impact of these orders is in terms of absolute costs for the hedge fund. Table 7 shows the orders with the 1% highest trading cost (in basis points).

Name	Code	Side	Contracts	Size Class	P_0	Trade Value ($)	Cost (bps)	Cost ($)
Sugar No 11	SBH1	Buy	19	1	16.04	341,225	25.76	879.20
Cotton No 2	CTN1	Buy	10	1	84.08	420,375	25.45	1070.00
Cocoa	CCN1	Sell	19	1	2394.50	454,955	24.07	1095.00
Cotton No 2	CTZ1	Buy	8	1	88.55	354,180	23.15	820.00
Lean Hogs	LHZ1	Sell	17	2	79.98	543,830	20.04	1090.00
Soybeans	SU1	Sell	7	1	1323.00	463,050	17.82	825.00
Corn	CZ1	Sell	13	1	531.13	345,231	15.03	518.75

Table 7: 1% most expensive orders in terms of relative trading costs

From Table 7, we find the average trading cost for the 1% most expensive orders to be 21.6 bps, which seems to have limited implications at first glance. However, taking into account the dollar value of the order, the average dollar cost of these orders is approximately $900 per order. Hence, the magnitude of trading cost for the 1% most expensive orders could have significant implications for the profitability of the trading strategy. Analyzing Table 7, two other interesting observations stand out. Firstly, except for the Lean Hogs order, all orders belong to size class 1. This confirms the conjecture in Section 7.3.3: it is not necessarily the largest orders (in relative terms) that lead to the highest trading costs. We will come back to this in Section 8. Secondly, all orders in Table 7 belong to the commodity futures class, which suggests a potential difference in costs across the different asset classes. The next subsection elaborates on this.

7.3.5 Cost Distribution

Another question yet unanswered is how the trading costs are distributed and whether there is indeed a substantial difference in costs across the various asset classes, as suggested in the previous subsection. From the results in Table 6, we concluded that the average relative trading cost is indistinguishable from zero, indicating that the average Varick order seems to have little or no effect on prices. The distribution of trading costs, on the other hand, reveals that the costs of the trades vary substantially. As mentioned before, Fraenkle et al. (2011) state that the externally induced market movement r_e does not contribute to the average impact $\mathbb{E}[I]$, but it significantly raises the variance of the relative price change (and thus of the trading costs). As a result, the externally generated price movements dominate the width of the trading cost distribution.

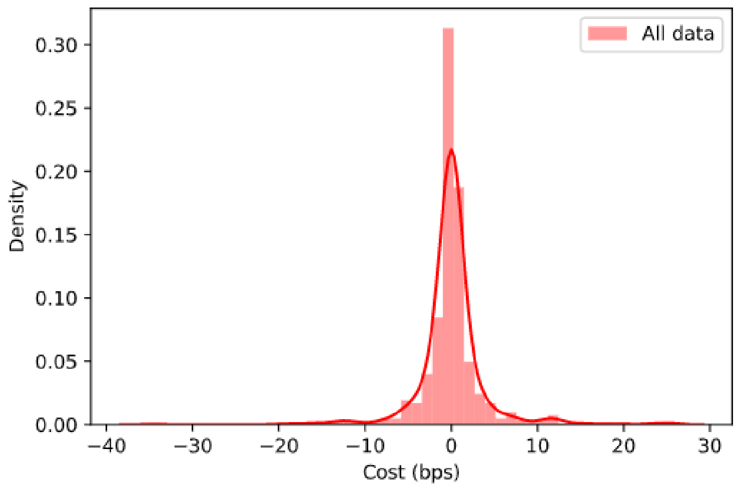

Figure 6: Distribution of relative trading costs (in bps) for the entire data sample

Figure 6 shows the distribution of relative trading costs (in bps) for the full data set. We observe a relatively high peak around 0, confirming our conjecture that the average impact of Varick's trades on prices seems small, but we also detect outliers. Just like we observed in Table 6, the graph shows that a substantial part of the orders has incurred negative costs (left tail), apart from the orders with positive costs in the right tail. This indicates that a substantial part of the executed orders has resulted in profits rather than losses. However, the orders in the right tail of the distribution are the ones that really matter in terms of trading costs and play an essential role in cost management (Bikker et al., 2009). These contribute significantly to the dispersion of trading costs around the median (mean) and play an important role in the quantile regression of Section 8.

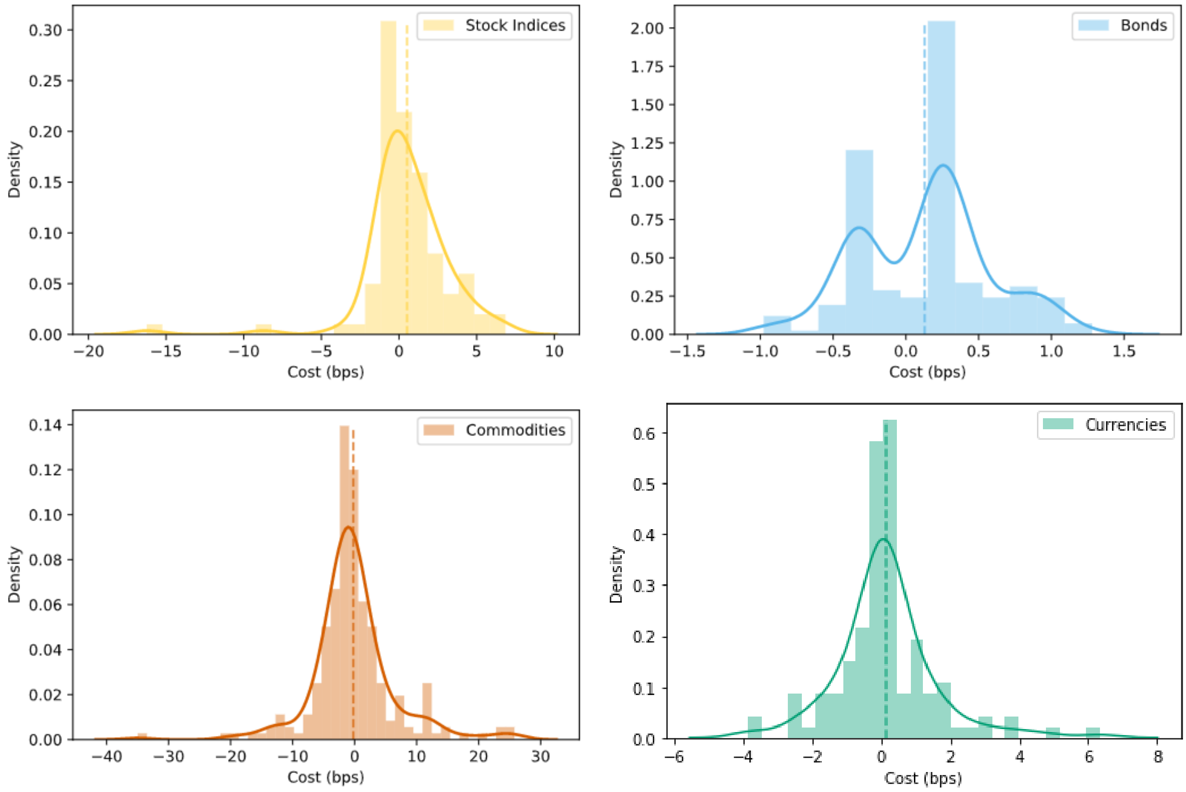

Figure 7: Distribution of relative trading costs (in bps) for different asset classes

Figure 7 shows a more detailed distribution of the relative trading costs for each separate asset class. The vertical dashed line indicates the mean cost (in basis points) for each asset class. Clearly, there is a significant difference between these figures; although the mean trading costs in the graphs are similar and close to zero for all of the asset classes, there is a substantial difference in terms of outliers. For instance, the cost distribution for bonds seems to be less centered around the mean at first sight, but its variation in costs is substantially smaller than for the other asset classes, especially when compared to commodities.

Linking this to market microstructure theory, we know that the price formation process is dependent on the distinctive characteristics of each (futures) market. Moreover, in Table 15, each asset class was shown to have its own characteristics in terms of, for instance, volatility and liquidity. These features seem to have a considerable influence on how expenses are dispersed. The more intriguing question, though, is *which* of these factors is most essential in explaining these outliers. This will be the focus of Section 8, which employs quantile regression to evaluate which variables increase the likelihood of incurring high trading costs.

7.4 Benchmarking

Using the methodology above to measure costs in terms of basis points, spreads and ticks already provides a quantification of the implicit trading costs of Varick Capital's orders. In fact, the trading cost measure defined earlier is the most comprehensive and complete as the aggregate of implicit trading costs is captured. However, this measure is not without shortcomings; it generally has a relatively large variance, since it incorporates all price changes during the execution period, some of which may be completely unrelated to the specific execution. For instance, if a buy order is executed while the market as a whole is rising, the trading cost measure will exhibit a higher cost than is attributable to the price change caused by the order itself. In contrast, if the market as a whole is falling while a buy order is being executed, the cost measurement could be negative as well.

Hence, in evaluating execution performance, it seems advisable to consider a broad range of cost measures, rather than a single number, in order to obtain a more reasonable determination of the transaction costs incurred by Varick Capital. In this research, we compare the execution price to *two* additional benchmarks, which we can apply across the entire sample of available trades. The idea behind these benchmarks is that by combining the considered trades with all other trades over the corresponding time period, we can assess what contribution the execution algorithm is making underneath.

7.4.1 Interval VWAP

We first compare the execution price with the VWAP of the corresponding order execution period. This benchmark is often referred to as the *interval VWAP* and is defined as the volume-weighted average of all transaction prices in the same futures contract during the life of the order.

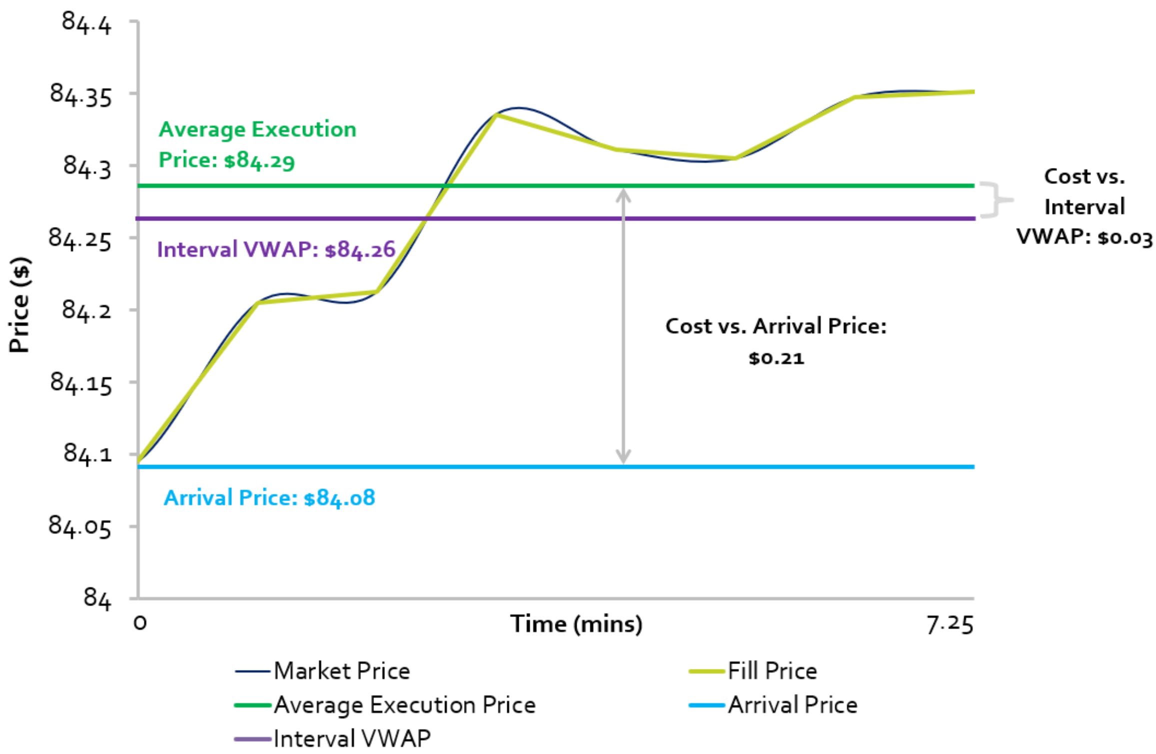

Figure 8: Analysis of a real Cotton order: benchmarking vs. interval VWAP and arrival price. *Based on Hedayati et al. (2018)*

Figure 8 shows the execution of a Cotton futures buy order on 3 June 2021. The parent order consists of buying ten July 2021 contracts and is split into seven separate trades (child fills). The order accounts for 0.08% of the daily trading volume and takes 7 minutes and 15 seconds to be executed. Based on order book data, the graph plots the various prices at which the execution algorithm has submitted each separate trade and the volume-weighted average market price throughout the trading interval. Aggregating all trades, the volume-weighted average of the executed prices determines the average execution price. By comparing this to the arrival price ($84.08) we measure the total trading cost, which in this case is $0.21 or 25.5 bps in relative terms.

In addition, the graph shows the execution price of the order compared against the interval VWAP. As Figure 8 illustrates, the execution price is closer to the interval VWAP: it is $0.03 away from this benchmark price, or 3.6 bps in relative terms. Generally, for an order which executes in a relatively short period of time we would expect cost measured against the arrival price to be similar to cost measured against the interval VWAP. However, as an order takes longer to execute, market risk may cause the arrival price to significantly deviate from the interval VWAP and cost measured against the arrival price will mostly reflect this risk.

Conclusion
In conclusion, using different benchmarks clearly results in different views on trading costs. While the interval VWAP accounts for all price changes over the execution period, the arrival price does not take into account subsequent market movements, since it is a pre-trade benchmark (Hedayati et al., 2018). In the case of Varick Capital, where the orders make up a small percentage of total trading volume, benchmarking against the interval VWAP seems to be appropriate as well.

7.4.2 Aggregate Results

For the sake of completeness, we also benchmark against the *interval TWAP*, the time-weighted average mid price during the life of the order.[23] Using the metric below we can compare the execution price to the different benchmarks for each order:

$$\text{Performance}_i \text{ (spreads)} = d_i \cdot \frac{p_i^b - p_i^{exe}}{S_i},$$

where p_i^b is either of the two benchmark prices for order i and S_i represents the average bid-ask spread during the life of order i. Note that we use units of spreads here, since different futures contracts may have different absolute spreads. Hence, if we are averaging performance across different contracts we need to account for this.

In contrast to our trading cost definition, when evaluating performance, a *negative* number indicates that an order has been executed against a *worse* price than the benchmark, both for buy orders and sell orders. Calculating the two benchmark prices across all of the orders, we obtain an aggregate performance, as reported in Table 8. The table specifically reports the volume-weighted average, mean, standard deviation and quantiles of the performance metric for all orders.

[23]Each mid price is prevailing for a certain period of time, up until the order book data quotes a new mid price. Hence, in order to obtain a time-weighted average mid price, we use the length of time until the mid-price is updated as weights.

	Arrival Price (spreads)	Interval TWAP (spreads)	Interval VWAP (spreads)
All data			
Volume-weighted	-0.22	-0.15	-0.15
Mean	-0.16	-0.05	-0.09
Standard deviation	1.52	0.76	0.79
5% Quantile	-2.43	-1.06	-1.00
95% Quantile	1.64	0.80	0.84

Table 8: Performance measured against interval TWAP and interval VWAP (in spreads)

The table shows that the volume-weighted average performance measured against the arrival price is about -0.22 spreads, whereas immediate execution at market prices would generally have lead to a performance of (at most) -0.5 spreads. As a result, the algorithm has reduced costs when compared to placing a market order at the time of arrival. Moreover, we see that, as expected, the performance measured against the arrival price has a larger standard deviation than the performance measured against the interval benchmarks.

Moreover, from Table 8, we find that the volume-weighted average performance against both the interval TWAP and interval VWAP benchmark are similar, which is generally to be expected when considering a limited time period and a full and balanced order book.

Specifically, the results in Table 8 imply that, on average, execution has lead to an underperformance of 0.15 spreads relative to the VWAP of all trades executed within the trading interval. Moreover, we observe that, on average, the order execution has lost 0.15 spreads relative to the average mid price during the execution period (the interval TWAP). Broadly speaking, an average performance of -0.5 spreads would imply aggressive execution and an average performance of +0.5 spreads would imply passive execution. Hence, the average performance of -0.15 spreads we observe here would imply

$$-0.15 = -0.5 \cdot p + 0.5 \cdot (1 - p),$$

which is equivalent to $p = 0.65$, or an average of 65% aggressive execution. This indicates that, on average, the algorithm managed to pick up 35% of passive execution instead of buying (selling) at the ask (bid).

Discussion

It is worth noting that performance measured against the interval VWAP and interval TWAP will generally be smaller (closer to zero on either side) than performance measured against the arrival price. However, performance measured against the arrival price is based on the actual trades and measures the actual trading costs, while performance measured against the interval benchmarks is based on all trades over the trading interval and tells us how the algorithm has behaved relative to other trades. It would be quite possible to have a strong performance of +0.5 spreads versus the interval TWAP, indicating passive execution, and have a performance of -10 spreads versus the arrival price, due to adverse price movements during the trading interval. The converse is possible as well. Hence, in analysing performance, both types of benchmarks should be taken into account.

8 Quantile Regression

From Section 6.3, we observed a significant variation in characteristics of the markets in our sample. A question yet unanswered, however, is *which* of these variables are most important in explaining differences in costs between the 681 orders in our sample. This section employs quantile regression to evaluate which factors increase the likelihood of incurring high trading costs. It starts by briefly introducing the concept of quantile regression, followed by estimation results, formal testing and goodness-of-fit. In addition, the interpretations of the estimation results are discussed. Finally, the relative importance of the regression variables is examined.

8.1 Introduction

Following the definition of Koenker and Hallock (2001), the (observed) trading cost for a specific order is at the θ^{th} $(0 < \theta < 1)$ quantile of the trading cost distribution if its value is higher than the proportion θ and lower than the proportion $1 - \theta$. Hence, exactly half of the orders have a higher cost than the median and half have a lower cost. The quantiles, or percentiles, refer to the general case: if the number of quantiles is n, then the quantiles divide the population into $n + 1$ equally sized groups in a data set. For a more formal explanation of quantiles and the differences between unconditional and conditional quantiles, we refer to Appendix B.1.

Quantile regression, as shown by Koenker and Bassett (1978), aims to extend the concept of quantiles to the estimation of conditional quantile functions, models in which the conditional distribution of the response variable is expressed as a function of covariates (Koenker & Hallock, 2001). Compared to the traditional homoskedastic OLS, which focuses on the conditional mean, quantile regression may be used to examine the variables influencing the orders with the $100\theta\%$ highest cost, where $\theta \in (0, 1)$. Hence, it provides a more complete picture of the conditional distribution, since it does not produce a single estimate but rather provides a series of estimates covering the whole range of trading costs from low to high.

Formally, using the notation by Bikker et al. (2009), we can state that, given a K-dimensional vector of covariates X, the classical least squares (homoskedastic OLS) model assumes that the conditional expected trading cost C equals $X\alpha$ (where α is a K-dimensional vector of coefficients). On the other hand, the quantile regression model assumes that, given the same covariates X, the θ^{th} conditional quantile of C equals $X\beta_\theta$ (where β_θ is a K-dimensional vector of coefficients).

Aside from the fact that the quantile regression method allows for direct examination of the complete range of trading costs, it is also more flexible than the homoskedastic OLS model in capturing how market- and trade-specific characteristics impact the distribution of trading costs. According to the homoskedastic OLS model, any trading cost determinant impacts these costs exclusively through the mean. The quantile regression technique, on the other hand, allows the effect of market- and trade-specific characteristics on trading costs to be dependent on the level of trading costs. Lastly, Koenker and Hallock (2001) conclude that another benefit of quantile regression is its robustness to outliers in the dependent variable (i.e., trading costs), which are prevalent in our data set. Appendix B.1 provides a more technical discussion on quantile regression.

8.2 Estimation Results

The quantile regression process is estimated by applying iteratively reweighted least squares. Following the procedure in Greene (2008, p. 414-416), heteroskedasticity robust standard errors are used to calculate the variance-covariance matrix.[24] The estimation results are displayed in Figure 9. These charts show the impact of trade- and market-specific variables on the trading cost distribution as a function of the quantile θ, where θ ranges from 0.05 to 0.95 (solid red curve) and 95% confidence bands (shaded red area). In addition, to enable visual comparison of the quantile regression with the homoskedastic OLS estimates, these graphs plot the OLS coefficients (solid black line) and corresponding 95% confidence intervals.

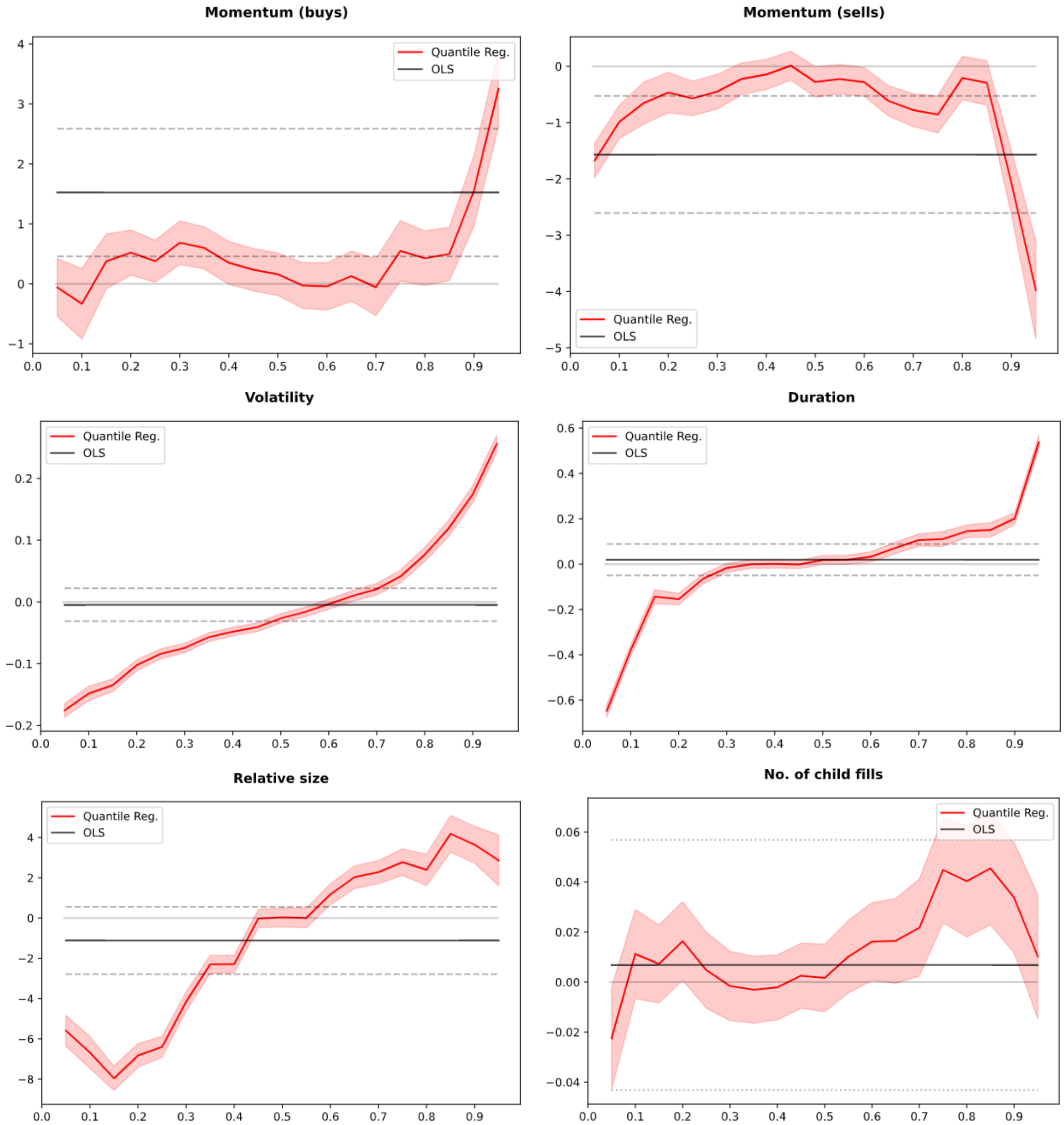

Figure 9: Estimated quantile regression coefficients with 95% confidence bands

[24]This is the default setting in statsmodels' QuantReg class in Python, which is used to estimate the quantile regression here. As Greene (2008) proposes, we use kernel density estimation for estimating the asymptotic covariance matrix. For the density estimation, we use the Epanechnikov kernel and apply the bandwidth selection method proposed by Hall-Sheather (1988). These are again the default settings.

While existing equity studies typically evaluate buy and sell orders separately, we do not make that distinction in our research, since we found no clear evidence of buy-sell asymmetry. Analyzing Figure 9, our first observation is that quantile regression better explains the relationship between trading costs and the various factors than the homoskedastic OLS. It particularly illustrates the effect of the considered variables on the risk of incurring significantly higher trading costs. Overall, we can conclude that the effect found in the cost distribution's tails differs considerably from the mean effect; except for the momentum and child fills coefficients, the estimated quantile regression coefficients generally lie outside the confidence intervals of the homoskedastic OLS regression in Figure 9. This suggests that the effects of these characteristics may not be constant across the conditional distribution of the dependent variable. Formal testing of this hypothesis is done in the next subsection.

8.3 Formal Testing

According to the graphs in Figure 9, the relationship between trading costs and most of its determinants varies significantly over the range of low-cost to high-cost trades. The majority of quantile regression estimates fall outside of the classical regression model's confidence intervals, which demonstrates that the homoskedastic OLS model is insufficiently robust to reflect the distribution of trading costs across the entire range of low-cost to high-cost trades. To support this finding, we use the Khmaladze (1982) test[25] provided by Koenker and Xiao (2002) to compare the quantile regression process to the homoskedastic OLS (*location shift*) model. For a more extensive explanation of why this model is referred to as the location shift model and a derivation of the Khamaladze (1982) test statistic, see Appendix B.2.

Variable	**Test Statistic**
Volatility	2.50
Duration	0.44
Relative size	1.82
Child fills	0.32
Momentum	0.57
Joint Effect	8.44

Table 9: Tests of the Location-Shift Hypothesis

Table 9 displays the results for testing the location-shift hypothesis. As expressed in Koenker and Xiao (2002), the null hypothesis of the location shift model is rejected when the joint test statistic exceeds the joint critical value. The bottom of Table 9 reports a joint test statistic value of 8.44. This indicates a rejection of the null hypothesis of the location shift model at any reasonable significance level, since for $p = 5$ covariates the 5% and 1% critical values are 6.64 and 7.64, respectively.[26] We can further investigate which variables contribute most to the joint significance of our test statistic and thus to the rejection of the location shift model; Table 9 also reports *marginal* test statistics corresponding to the same null hypothesis. In a similar way as before, a variable's coefficient is of a significantly different form than in the location shift model if the marginal test statistic exceeds the marginal critical value (Koenker & Xiao, 2002). However, Koenker and Xiao (2002) note that these marginal statistics should only be interpreted as formal tests, due to the possible dependence of covariates. Hence, formally, volatility contributes most to the joint significance of the test statistic.[27]

[25]This test is implemented in the Quantreg package in R.

[26]This test statistic is based on the asymptotic critical values in Koenker and Xiao (2001), where we set the truncation interval equal to $\mathcal{T} = [0.05, 0.95]$.

[27]The critical values for these coordinate-wise tests are 2.140 at 5%, and 2.721 at 1% significance level. Hence, volatility

8.4 Goodness-of-fit

In traditional regression models, the R^2 is a frequently used goodness-of-fit statistic. For quantile regression models, Koenker and Machado (1999) introduce a general goodness-of-fit process that is similar to the traditional R^2 statistic, which they refer to as $R^1(\theta)$. A special case of this statistic is the so-called pseudo R^2. The pseudo R^2 assesses the relative goodness-of-fit of two quantile regression models at a certain quantile in terms of an "appropriately weighted sum of absolute residuals" (Koenker & Machado, 1999), while the traditional R^2 measures the relative goodness-of-fit of the respective models for the conditional mean function. For a more technical definition of the pseudo R^2 and its differences to the traditional R^2, we refer to Appendix B.3. Overall, the pseudo R^2 constitutes a *local* goodness-of-fit measure, rather than applying a global measure to the whole conditional distribution, such as the R^2.

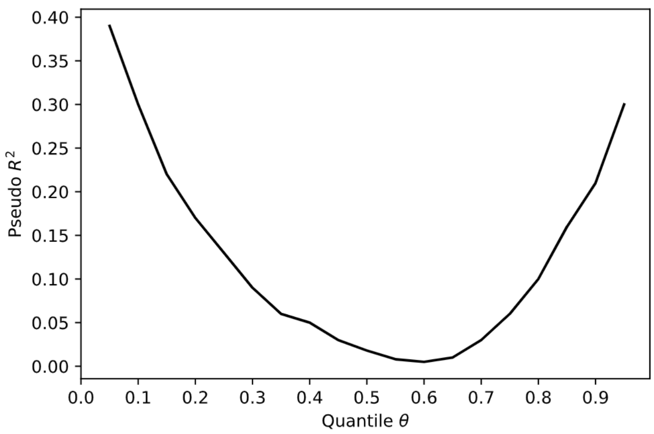

Figure 10: Pseudo R^2 for different quantiles θ

To acquire an indication of the quantile regression process' goodness-of-fit, we compute the pseudo R^2 as a function of θ for the estimated quantile regression model. Interestingly, Figure 10 shows a parabola-shaped curve, with a minimum of 0.005 attained at approximately the 0.6th quantile. The pseudo R^2 reaches its maximum value of approximately 0.39 when θ approaches 0. Evidently, the best fit is achieved in the tails of the distribution, suggesting that a larger part of the variation in trading costs is explained by the trade- and market-specific variables as the quantiles get more extreme. In comparison, using the same variables as in the quantile regression model, the R^2 of the location shift (OLS) model equals 0.004, implying that the variables used here explain only 0.4% of variation in mean trading costs. This confirms our conjecture that the homoskedastic OLS model is insufficiently robust to capture the trading cost distribution across the entire range of low-cost to high-cost orders.

8.5 Interpretation of Results

This subsection provides the (economic) implications of the quantile regression model's estimation results. We start by briefly explaining how to interpret the plots in Figure 9. For each covariate, the point estimates in the graphs may be interpreted as the impact of a one-unit change of the variable on trading costs, starting on the left with the lowest quantile and ending on the right with the highest quantile, ceteris paribus. In the subsection that follows, our focus will be on the covariates that impact the right tail, since our main objective is examining the features that influence the probability of incurring relatively high trading costs.

is the only variable whose marginal Khmaladze statistic exceeds the critical value at 5% significance level.

Volatility
First of all, market-specific volatility has a strong/substantial effect on the trading costs: the tight confidence bands around the volatility coefficients indicate a high significance level. Secondly, whereas volatility has a moderate effect on the quantiles in the middle of the distribution, it has a considerable impact on its lower and upper tails. Specifically, an increase in volatility implies a reduction in costs in the bottom half of the cost distribution while the opposite is true in the upper third. Furthermore, market-specific volatility has the largest effect on the most extreme quantiles of the cost distribution. A one-unit increase in volatility can result in as much as -0.17 bps in trading cost in the lower tail and more than 0.2 bps in the upper tail, ceteris paribus. Overall, we agree with Bikker et al. (2009) by concluding that volatility seems to have a scale effect on trading costs; the cost distribution is more spread in more volatile markets/periods. More volatile futures contracts typically have more idiosyncratic price fluctuations, resulting in larger trading cost dispersion.

Duration
The tight confidence bands around the duration coefficients imply a high significance level. Following the reasoning by Bikker et al. (2009), we can conclude that duration, like volatility, seems to have a scale effect on trading costs; when orders take longer to execute, the distribution of trading costs becomes more dispersed. A one-unit increase in duration implies a decrease in trading costs in the lower three-tenths and a decrease in costs in the upper third of the cost distribution. Again, trade duration has the greatest impact on the highest and lowest quantiles in these areas of the cost distribution. A one-unit increase in duration can result in as much as -0.4 bps in the lower tail and around 0.6 bps in the upper tail, ceteris paribus. This pattern can be explained from the theory in Sections 5.1 and 5.2, where we observed the trade-off between trading quickly (aggressively) and trading slowly (passively), as expressed by Kissell and Malamut (2005). Executing an order quickly is generally associated with a higher expected trading cost, but lower intrinsic price risk, whereas the opposite is true when an order takes longer to be completed. The longer the order takes to be completed, the more likely events unrelated to the orders' own market impact will influence the price of the contract being traded, resulting in more risk and, as a result, a more dispersed cost distribution.

Relative Trade Size
The tight confidence bands around the relative trade size coefficients of the quantile regression estimates indicate a high level of significance. As stated in Section 6.2.1, orders in size class 1 are generally not expected to move the market, while orders in size class 2 are expected to cause significant impact even if executed gradually. The omitted category is size class 1, so the quantile coefficients may be interpreted relative to this category. The difference in trading costs between orders in size class 2 and orders in size class 1 is substantial; at the quantiles in the left tail of the distribution, the coefficients of the relative trade size variable are significantly negative, while they are significantly positive in the right tail. This result suggests that, when compared to orders in size class 1, 'large' orders have a significant risk of incurring higher trading costs as well as a significant risk of incurring lower trading costs. In turn, this implies that size mainly exerts a scale effect: the variance of trading costs increases with relative trade size.

Interestingly, the OLS coefficients show an insignificant effect of relative size on trading costs. As stated in Section 4.2.2, usually a positive relationship between trading costs and trade size is observed from literature. For the quantile regression results, this would imply that trade size would mainly have a location effect on the conditional cost distribution.

However, Jones et al. (1994), for instance, conclude that "the size of trades (or volume) reflects the extent of disagreement among traders about a security's value". This would suggest that the variance of trading costs may also increase with trade size. Furthermore, and perhaps more importantly, we stated before that larger orders will only incur a higher cost than smaller orders *in the same asset* and with the same strategy. In this case, however, we are considering a wide variety of futures markets from different asset classes. Evidently, when considering a variety of futures from different asset classes, it is not necessarily true that larger orders (in relative terms) lead to higher costs. Furthermore, even though we observed that size has a significant effect on the cost distribution, the information contained in size is possibly already contained in other variables, such as the order's duration, making these variables inter-correlated. This would imply that the proportion of variance in costs purely explained by relative size is small when compared to duration. We will come back to this conjecture in Subsection 8.6.

Number of Child Fills

Clearly, the confidence bands around the coefficients of the number of child fills are substantially wider than the confidence bands of the other coefficients. This indicates a relatively low significance level. Although one would generally expect the number of child fills to be a proxy for the difficulty of executing an order and thus expect the number of fills to have a positive effect on trading costs, the effect is found to be insignificant in this analysis.

Momentum

From Section 4.2.2, we would expect momentum to have a positive effect on the trading costs of buy orders and a negative effect on the trading costs of sell orders. For this reason, contrary to the previous coefficient interpretations, we perform a separate analysis for buy and sell orders here. Firstly, the effect of a one-unit increase in momentum is substantially overestimated by the homoskedastic OLS model, both for buy and sell orders. This is most likely due to extreme observations in the right and left tails having a disproportionately large influence. By analyzing the quantile regression results we obtain some additional insights.

For buy orders, the effect of momentum on trading costs is found to be insignificant over the major part of the distribution, except at the highest quantiles. From approximately the 70% quantile, the momentum coefficient tends to become more significant: in the upper tail, a one-unit increase in momentum shows an effect of more than 3 bps, ceteris paribus. The plot's convex shape suggests that during periods of strong positive momentum (i.e., a buying trend), buy orders could potentially incur higher trading costs, while lower (negative) costs are less likely.

For sell orders, momentum has a significant negative effect over a larger part of the cost distribution, especially at the highest quantiles. In the upper tail, the negative effect is the strongest: a one-unit decrease in momentum increases trading costs by more than 4 bps, ceteris paribus. The graph's concave shape illustrates that, for sell orders, during times of strong negative momentum (a selling trend), relatively high trading costs are more likely than relatively low (negative) costs.

Overall, our findings show that when momentum prior to execution (as measured by the 5-day average return) increases, the probability of incurring significant trading costs tend to increase as well. This is consistent with findings in other studies, such as Bikker et al. (2009).

Conclusion
In conclusion, volatility, duration and relative size have a significant effect on a substantial part of the trading cost distribution, as we would expect from the literature in Section 4.2.2. Moreover, although momentum was shown to have an insignificant effect over the major part of the distribution, especially for buy orders, we conclude that the probability of incurring significant trading costs is higher during periods of strong (positive or negative) momentum.

8.6 Relative importance of variables

The charts in Figure 9 show which variables have a significant impact on the distribution of trading costs. However, an unsolved question is *which* of these variables best explains (variation in) trading costs. What complicates addressing this issue, is the fact that not all factors affect each fraction of the trading cost distribution in the same way. The literature on classical regression models provides several methods to measure the relative importance of explanatory variables. For instance, Frino and Oetomo (2005) calculate the Adjusted R^2 of the model with and without a specific variable and conclude that the most explanatory power comes from the variable that leads to the greatest reduction in Adjusted R^2. We will use a similar approach here, except for the fact that we evaluate the relative importance of a variable over different quantiles of the trading cost distribution. The squared partial correlation (SPC) tells us which fraction of variance in the dependent variable (Y) has been left unexplained by other variables, but is explicitly explained by X_i. For a more detailed description of the theory behind SPC, we refer to Appendix B.4.

Variable	**SPC**		
	left tail	*median*	*right tail*
Volatility	**0.1959**	0.0167	**0.1685**
Duration	**0.1131**	0.0002	**0.1439**
Relative size	0.0561	0.0001	0.0168
No. of child fills	0.0008	0.0000	0.002
Momentum	0.0031	0.0005	0.0066

Table 10: Most important determinants based on squared partial correlations

The SPCs for each variable used in the quantile regression model are reported in Table 10. These squared partial correlations show that the variables barely affect the median, but volatility and duration substantially influence the tails of the cost distribution.[28] Furthermore, the SPCs show that some variables like momentum and relative size - although both statistically significant in the tails - have limited explanatory power.

Conclusion
Despite the fact that relative size and momentum increase the dispersion of trade costs, these variables alone cannot explain a substantial part of the variation in costs between the observations in our sample. The two key factors in explaining differences in trading costs for a portfolio of futures contracts from different asset classes are shown to be market-specific volatility and duration as a proxy for liquidity. Furthermore, none of the variables has a significant impact on the median of the cost distribution.

[28] By calculating the average SPC over the quantiles $\theta = \{0.01, \ldots, 0.1\}$ as a proxy for the left tail and $\theta = \{0.9, \ldots, 0.99\}$ as a proxy for the right tail, we get an indication of the contribution of each variable to the left and right tail of the cost distribution.

In addition, as the plots in Figure 7 demonstrate, the mean trading costs were shown to be similar and close to zero for all asset classes, with a significant difference in terms of cost dispersion. Given the results above, we can now deduce that this is mostly due to the differences in volatility and trade duration between orders in different asset classes.

9 Conclusion & Discussion

9.1 Conclusions

Combining a unique data set of futures order book data and trade execution data from a hedge fund across 50 global futures markets over a 9-months period with a total transaction value of 2.3 billion dollars, this research measures and benchmarks the actual transaction costs incurred by a large trader. Order executions for futures contracts in four different asset classes are examined: stock indices, currencies, commodities and bonds. We have obtained several interesting results.

Firstly, we have redefined total transaction costs specifically for futures trading based on Perold (1988). This research calculates total transaction costs as the sum of delay costs and trading costs. To the best of our knowledge, this is the first study to measure delay costs. We find that for the buy orders, due to prices moving away from the decision price, the average cost associated with waiting to trade (i.e., the delay cost) is approximately 32 basis points. Surprisingly, we observe the opposite when it comes to sell orders. Waiting to trade has actually resulted in an average savings of approximately 18 basis points relative to the decision price for these orders.

Secondly, we find the average impact of Varick Capital's trades on prices to be rather small; average trading costs equal 0.15 bps for buy orders and 0.16 bps for sell orders. We find trading costs in our study to be of a similar magnitude to the corresponding costs documented in previous studies of S&P500 index futures and SFE futures markets. However, we find the magnitude of costs to be substantially lower than previously documented in equity studies. Thirdly, in line with Frino and Oetomo (2005), we conclude that there is no clear evidence of an asymmetry between the trading costs of buy and sell orders. Our findings confirm conjectures in previous research that i) we expect low information asymmetry and therefore low adverse selection costs in futures markets and ii) futures markets are generally liquid in nature and trading costs are inversely related to market liquidity.

Furthermore, we find a substantial difference in trading costs between trades in the left and right tails of the cost distribution. Although average trading costs are shown to be relatively small in terms of market disruption, this is not necessarily true in terms of absolute costs for the hedge fund.
In addition, the results show substantial differences in trading costs across the different asset classes. Even though the mean trading cost for the different asset classes is shown to be similar and close to zero, we observe a substantial difference in terms of outliers. As we would expect from market microstructure theory, the wide variation in characteristics among the markets in our sample seems to have a considerable influence on the dispersion of costs.

Moreover, since we have access to order book data, we can extract intraday data for all of the executed trades. This specifically enables us to evaluate the performance of the execution algorithm used, by comparing the execution prices against two additional (intraday) benchmarks. We particularly find that the execution algorithm has on average outperformed immediate execution at market prices by 0.28 spreads and managed to pick up 35% of passive execution instead of buying (selling) at the ask (bid).

Furthermore, we analyze to which extent trade- and market-specific variables from previous research impact trading costs using the quantile regression approach. We find the majority of variables to impact the conditional cost distribution in ways that are poorly reflected by the homoskedastic OLS model, which is typically used in transaction cost analysis. The quantile regression model is shown to better capture the relationship between trading costs and the various factors, especially at higher quantiles. Moreover, by estimating the impact of variables on the orders with the $100\theta\%$ highest trading cost, where θ can take any value in the interval $(0, 1)$, our study particularly finds volatility and trade duration to be important risk factors. Furthermore, we find momentum to have an amplifying effect on the likelihood of incurring significant trading costs.

Overall, this thesis demonstrates that quantile regression may significantly contribute to transaction cost management, particularly when applied to a real-world portfolio of futures from several asset classes. It is a useful tool for analyzing the determinants of high-cost trades and understanding trading costs, which is particularly beneficial for transaction cost management.

9.2 Discussion & Future Research

Similar to Bikker et al. (2009), we would like to start by emphasizing that the current research is limited to a particular hedge fund during a specified time period. As a result, the findings of this study may or may not apply to other funds. However, we argue that the framework we use is generalizable to other funds and managers trading futures or other instruments. The measured costs are exogenous to the portfolios being traded (Frazzini et al., 2018), since our data analyzes trade execution data without making any assumptions on the portfolio construction process.[29]

In terms of data restrictions, our data sample spans a limited time period and includes a relatively small number of orders as compared to other studies. The current data sample, for instance, does not enable us to fit a different regression model for each separate market and hence does not allow us to test which variables are the most important in explaining differences *within* a specific market. Moreover, it may be interesting to investigate to what degree results remain consistent over a longer time period. Also, we would expect an increase in assets under management to result in higher trading costs, because of higher trading volumes. In addition, since all of the data is from 2021, we are unable to stratify our sample into bullish and bearish markets, as proposed by Chiyachantana et al. (2004). According to Chou et al. (2011), an asymmetric pattern in trading costs of large buy and sell orders depends on the number of bullish versus bearish periods in the entire sample period. Chou et al. (2011) find that in bearish markets, the trading costs of sell orders are greater than the costs of buy orders, and in bullish markets, and observe the opposite pattern in bullish markets.

Our findings suggest a number of possible future research directions. Firstly, while the focus of our research is on trading costs, we discovered that, in our sample, delay costs are actually the larger component of overall transaction costs. Besides, while the average delay cost for buy orders was positive and the average delay cost for sell orders was negative, we observed a significant variation in delay costs between orders in the left and right tails of the cost distribution. Hence, identifying which markets contribute the most to these costs might be an interesting future research topic.

[29] The universe of futures included in our data set is the *only* endogenous asset selection decision made with regard to trading costs: Varick Capital excludes futures with low daily trading volume.

In addition, we analyzed the performance of the execution algorithm using a variety of benchmarks. While benchmarking against the interval TWAP and interval VWAP provides us an indication of the performance of the algorithm relative to trades executed in the same time interval, it does not tell us whether the algorithm might have performed better with different algorithmic settings. This gives potential to more research in the future. Furthermore, by analyzing performance using benchmark prices, we solely concentrate on the absolute performance of the algorithmic orders, rather than comparing those orders with samples containing orders completed by alternative means such as other algorithms.

Finally, one of the major incentives for developing transaction cost estimation models is to use them to predict future costs. As a consequence, future research might investigate the quantile regression model's accuracy in predicting trading costs using the variables defined in this study. According to Bikker et al. (2009), quantile regression significantly outperforms the homoskedastic OLS model in terms of predictive power, which seems a promising addition to our research.

10 References

Almgren, R., & Chriss, N. A. (1997). Optimal liquidation. *Available at SSRN 53501.*

Almgren, R., & Chriss, N. A. (2001). Optimal execution of portfolio transactions. *Journal of Risk, 3*, 5-40.

Almgren, R., Thum, C., Hauptmann, E., & Li, H. (2005). Direct estimation of equity market impact. *Risk, 18*(7), 58-62.

Baltas, N. (2017). Optimizing cross-asset carry. In *Factor Investing* (pp. 317-364). Elsevier.

Berkman, H., Brailsford, T., & Frino, A. (2005). A note on execution costs for stock index futures: Information versus liquidity effects. *Journal of Banking & Finance, 29*(3), 565-577.

Berkowitz, S. A., Logue, D. E., & Noser Jr, E. A. (1988). The total cost of transactions on the NYSE. *The Journal of Finance, 43*(1), 97-112.

Bertsimas, D., & Lo, A. W. (1998). Optimal control of execution costs. *Journal of Financial Markets, 1*(1), 1-50.

Bikker, J. A., Spierdijk, L., & van der Sluis, P. J. (2007). Market impact costs of institutional equity trades. *Journal of International Money and Finance, 26*(6), 974-1000.

Bikker, J. A., Spierdijk, L., & van der Sluis, P. J. (2009). What factors increase the risk of incurring high market impact costs?. *Applied Economics, 42*, 369-387.

Breen, W. J., Hodrick, L. S., & Korajczyk, R. A. (2002). Predicting equity liquidity. *Management Science, 48*(4), 470-483.

Chan, L. K., & Lakonishok, J. (1993). Institutional trades and intraday stock price behavior. *Journal of Financial Economics, 33*(2), 173-199.

Chan, L. K., & Lakonishok, J. (1997). Institutional equity trading costs: NYSE versus Nasdaq. *The Journal of Finance, 52*(2), 713-735.

Chiyachantana, C. N., Jain, P. K., Jiang, C., & Wood, R. A. (2004). International evidence on institutional trading behavior and price impact. *The Journal of Finance, 59*(2), 869-898.

Chou, R. K., Wang, G. H., Wang, Y. Y., & Bjursell, J. (2011). The impacts of large trades by trader types on intraday futures prices: Evidence from the Taiwan Futures Exchange. *Pacific-Basin Finance Journal, 19*(1), 41-70.

Easley, D., & O'Hara, M. (1995). Market microstructure. *Handbooks in Operations Research and Management Science, 9,* 357-383.

Ederington, L. H., & Lee, J. H. (1993). How markets process information: News releases and volatility. *The Journal of Finance, 48*(4), 1161-1191.

Fleming, J., Ostdiek, B., & Whaley, R. (1996). Trading costs and the relative rates of price discovery in stock, futures, and option markets. *Journal of Futures Markets, 16,* 353-387.

Fraenkle, J., Rachev, S. Z., & Scherrer, C. (2011). Market impact measurement of a VWAP trading algorithm. *Journal of Risk Management in Financial Institutions, 4(3),* 254-274.

Frazzini, A., Israel, R., & Moskowitz, T. J. (2018). Trading costs. *Available at SSRN 3229719.*

Frino, A., & Oetomo, T. (2005). Slippage in futures markets: Evidence from the Sydney Futures Exchange. *Journal of Futures Markets: Futures, Options, and Other Derivative Products, 25*(12), 1129-1146.

Frino, A., Lepone, A., & Wang, G. H. (2007). The Price Impact of Institutional Trades on SP 500 Index Futures. *Available at SSRN 1009071.*

Gatheral, J., & Schied, A. (2013). Dynamical models of market impact and algorithms for order execution. *Available at SSRN 2034178.*

Greene, W. H. (2008). *Econometric Analysis* (6th ed.). Pearson.

Gorton, G. B., & Pennacchi, G. (1991). Security baskets and index-linked securities. *Journal of Business, 66*(1).

Hedayati, S., Brian, K. & Hurst, E. S. (2018). *Transactions costs: Practical application.* AQR Capital Management. https://www.aqr.com/Insights/Research/White-Papers/Transactions-Costs-Practical-Application

Hull, J. C. (2014). *Options, Futures and Other Derivatives* (9th ed.). Pearson.

IMARC Group. (2021). *Algorithmic trading market: Global industry trends, share, size, growth, opportunity and forecast 2021-2026.* https://www.mordorintelligence.com/industry-reports/algorithmic-trading-market

Johnson, B. (2010). *Algorithmic Trading & DMA: An introduction to direct access trading strategies.* London: 4Myeloma Press.

Jones, C. M., Kaul, G., & Lipson, M. L. (1994). Information, trading, and volatility. *Journal of Financial Economics, 36*(1), 127-154.

Keim, D. B., & Madhavan, A. (1997). Transactions costs and investment style: an inter-exchange analysis of institutional equity trades. *Journal of Financial Economics, 46*(3), 265-292.

Khmaladze, E. V. (1982). Martingale approach in the theory of goodness-of-fit tests. *Theory of Probability & Its Applications, 26*(2), 240-257.

Kissell, R., & Glantz, M. (2003). *Optimal Trading Strategies: Quantitative Approaches for Managing Market Impact and Trading Risk.* Amacom, USA.

Kissell, R., Glantz, M., & Malamut, R. (2004). A practical framework for estimating transaction costs and developing optimal trading strategies to achieve best execution. *Finance Research Letters, 1*(1), 35-46.

Kissell, R., & Malamut, R. (2005). Algorithmic decision-making framework. *The Journal of Trading, 1*(1), 12-21.

Kissell, R. (2006). *Algorithmic Trading Strategies.* Fordham University.

Kissell, R. (2013). *The Science of Algorithmic Trading and Portfolio Management.* Academic Press.

Koenker, R., & Bassett, G. (1978). Regression quantiles. *Econometrica, 46*(1), 33-50.

Koenker, R., & Machado, J. A. (1999). Goodness of fit and related inference processes for quantile regression. *Journal of the American Statistical Association, 94*(448), 1296-1310.

Koenker, R., & Xiao, Z. (2001). *Inference on the quantile regression process: Appendices.* http://www.econ.uiuc.edu/~roger/research/inference/khmal6ap.ps

Koenker, R., & Hallock, K. F. (2001). Quantile regression. *Journal of Economic Perspectives, 15*(4), 143-156.

Koenker, R., & Xiao, Z. (2002). Inference on the quantile regression process. *Econometrica, 70*(4), 1583-1612.

Korajczyk, R. A., & Sadka, R. (2004). Are momentum profits robust to trading costs?. *The Journal of Finance, 59*(3), 1039-1082.

Labadie, M., & Lehalle, C. A. (2010). Optimal algorithmic trading and market microstructure. https://hal.archives-ouvertes.fr/hal-00590283v1

Lillo, F., Farmer, J. D., & Mantegna, R. N. (2003). Master curve for price-impact function. *Nature, 421*(6919), 129-130.

Lillo, F. (2016). *Market impact models and optimal execution algorithms.* http://www.ceresaig.com/wp-content/uploads/2017/09/Market-impact-models-and-optimal-execution- algorithms-Lillo-Imperial -Lecture1.pdf

Mandes, A. (2016). Algorithmic and high-frequency trading strategies: A literature review. *Working paper.*

Pedersen, L. H. (2018). Sharpening the arithmetic of active management. *Financial Analysts Journal, 74*(1), 21-36.

Perold, A. F. (1988). The implementation shortfall: Paper versus reality. *Journal of Portfolio Management, 14*(3), 4.

Rydberg, T. H., & Shephard, N. (2003). Dynamics of trade-by-trade price movements: Decomposition and models. *Journal of Financial Econometrics, 1*(1), 2-25.

Subrahmanyam, A. (1991). A theory of trading in stock index futures. *The Review of Financial Studies, 4*(1), 17-51.

The TRADE. (2021). *The 2021 algorithmic trading survey: Hedge funds 21.* https://www.thetradenews.com/wp-content/uploads/2021/06/Algo-survey.pdf

Vyetrenko, S., & Xu, S. (2019). Risk-sensitive compact decision trees for autonomous execution in presence of simulated market response. *arXiv preprint arXiv:1906.02312*

Wagner, W. H., & Edwards, M. (1993). Best execution. *Financial Analysts Journal, 49*(1), 65-71.

A Appendix A: Definitions

A.1 Definition order-size classes

Orders may be split into various size categories based on their executed size in the following way:

Size Category	**Definition**
LessThanTCH	Executed sizes equal to or smaller than the aggressive touch size at the start time for the order
MoreThanTCH	Executed sizes larger than the aggressive touch size at the start time for the order but smaller or equal to half the sweep book size at the start time for the order
Medium	Executed sizes between half of the sweep size and the sweep size
Large	Executed sizes larger than the sweep size

Table 11: Size categories and their definitions

Here, the sweep size is defined to be the total volume on the visible order book at the start time of the order. Typically 10 order book levels are stored historically, so the total volume is calculated based on the first 10 levels of the book only. For some particular contracts, there may be fewer book levels available. Furthermore, the *aggressive touch size* is defined to be the quantity available to take at the first price level on the aggressive side. Hence, for a buyer that would be the size at the lowest ask price and for a seller that would be the size at the highest bid price.

Overall, these size categories are a simple attempt to separate orders into categories based on their expected impact on prices; a LessThanTCH order is generally not expected to move the market even if executed as a market order; a Large order might be expected to cause significant impact even if executed quite gradually. Based on these categories, we define two relative size classes in our analysis: size class 1 includes orders that belong to size category *LessThanTCH* or *MoreThanTCH*; size class 2 includes trades that belong to size category *Medium* or *Large*. These two size classes allow us to test in which way trading costs are related to relative trade size in futures markets.

B Appendix B: Quantile Regression

This appendix on quantile regression briefly addresses some issues that are relevant to this thesis. It uses the theory by Koenker and Bassett (1978), Koenker and Hallock (2001) and Koenker and Xiao (2002) to mathematically explain the main properties behind the quantile regression technique. We focus on trading costs as the dependent variable and trade and market-specific characteristics as the covariates to simplify the explanation.

B.1 Introduction

Following the definition of Koenker and Hallock (2001), the (observed) trading cost for a specific order is at the θ^{th} $(0 < \theta < 1)$ quantile of the trading cost distribution if its value is higher than the proportion θ and lower than the proportion $1-\theta$. Technically speaking, the θ^{th} quantile of trading costs (C) is defined as

$$Q_C(\theta) = \inf_c \{c : F_C(c) \geq \theta\}, \tag{1}$$

where $F_C(c) = \mathbb{P}[C \leq c]$ denotes the distribution function of C. The equation above states that the θ^{th} quantile is equal to the smallest value c for which $F_C(c)$ is larger than or equal to θ. Figure 9 illustrates this for the standard normal density, in which the point $y = 1.645$ indicates the 95% quantile.

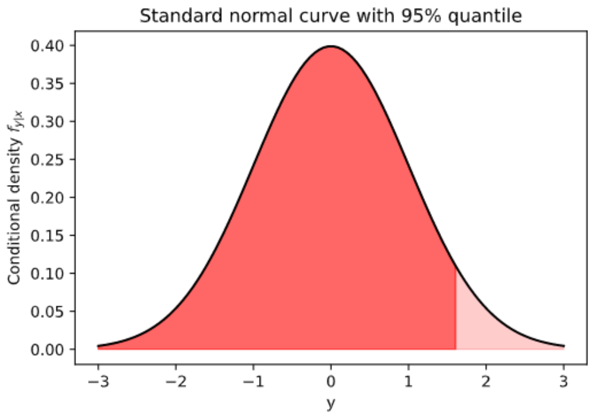

Figure 11: Standard normal curve with 95% quantile

However, in this specific case, our focus will be on conditional quantiles rather than unconditional ones, since we observe trading costs (the response variable) in combination with trade and market-specific characteristics (the covariates). For instance, we might want to know what the trading costs will be given a certain level of volatility. Hence, we may define the θ^{th} conditional quantile of C given $X = x$ as

$$Q_C(\theta \,|\, x) = \inf_c \{c : F_{C\,|\,X}(c \,|\, x) \geq \theta\}, \tag{2}$$

where $F_{C\,|\,X}(c \,|\, x) = \mathbb{P}\,[C \leq c \,|\, X = x]$ denotes the conditional distribution function of C given $X = x$. The quantile regression model assumes that the θ^{th} conditional quantile of C given the covariates X equals $X'\beta_\theta$. Using this information, we can formulate the quantile regression model as

$$C = X'\beta_\theta + \epsilon, \quad Q_\epsilon(\theta \,|\, X) = 0. \tag{3}$$

Hence, in the quantile regression model, the θ^{th} conditional quantile is given by

$$Q_C(\theta \,|\, x) = X'\beta_\theta. \tag{4}$$

Interpretation of β_θ coefficients

Given specification (4), the partial derivative of C with respect to regressor X_j equals β_θ^j and represents the change in the θ^{th} conditional quantile due to a one-unit change in X_j, ceteris paribus.

Relation with traditional models

Given trading costs (C) and its covariates (contained in the vector X of dimension $k+1$), the homoskedastic OLS or location shift model is formulated as

$$C = X'\alpha + \sigma\epsilon, \quad \mathbb{E}[\epsilon|X] = 0, \quad \mathbb{E}[\epsilon^2|X] = 1, \tag{5}$$

where α is a vector of coefficients of dimension $k+1$ and $\sigma > 0$. Since this model only allows covariates to affect the conditional mean (i.e., the term $X'\alpha$) of the trading costs, it is rather restrictive. As a result, a change in covariates primarily 'shifts' the conditional distribution of the dependent variable C, as shown in the left panel of Figure 12. For this reason, this type of model is often referred to as the *location shift* model. On the other hand, models like the heteroskedastic OLS allow covariates to influence the conditional variance of trading costs. This implies that the covariates are possibly stretching (increasing variance) or squeezing (decreasing variance) the distribution, as depicted in the right panel of Figure 12. However, in this research, we use the more flexible quantile regression approach to test whether trading costs are affected in more complex ways.

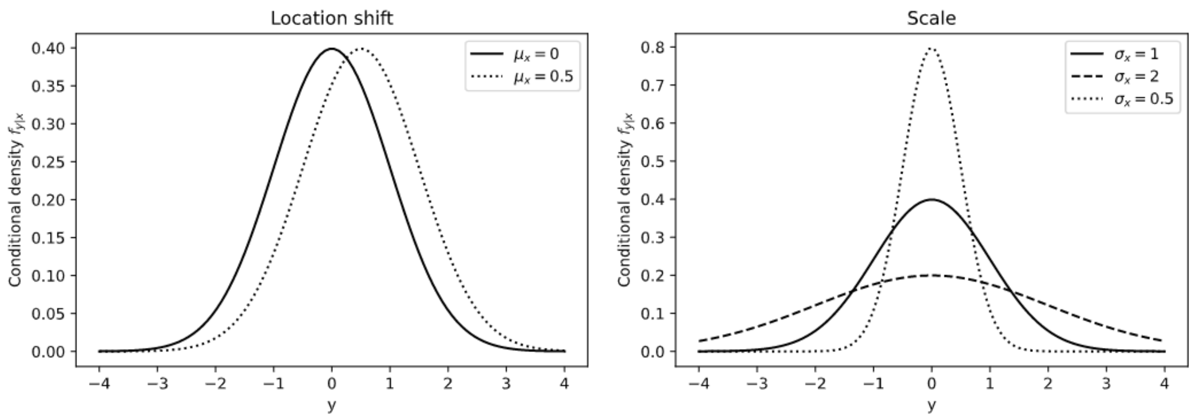

Figure 12: Location shift vs. scale

Now let us suppose the regression model contains an intercept. We may then write $X' = [1\ Z']$, $\alpha = (\alpha_0, \alpha_1)$ and $\beta_\theta = (\beta_{\theta,0}, \beta_{\theta,1})$. Applying equation (4) to the conditional quantiles in the location shift model, we find that the conditional quantiles are given by

$$Q_C(\theta \mid X) = X'\alpha + \sigma F_\epsilon^{-1}(\theta) = \alpha_0 + Z'\alpha_1 + \sigma F_\epsilon^{-1}(\theta), \tag{6}$$

where F_ϵ denotes the distribution function of ϵ. As a result, the homoskedastic OLS or location shift in equation (5) model imposes some constraints on the coefficient of the conditional quantiles:

$$\beta_{\theta,0} = \alpha_0 + \sigma F_\epsilon^{-1}(\theta) \tag{7}$$

$$\beta_{\theta,1} = \alpha_1 \tag{8}$$

From this, we can clearly see the difference between the location shift and the quantile regression model: in the former model, covariates have the same effect on all quantiles, but in the latter model, the effect of covariates on the θ^{th} quantile is dependent on θ.

B.2 Formal Testing

Khmaladze Test

In this subsection, we use the theory by Koenker and Xiao (2002) to shortly describe the Khamaladze (1982) test. We focus on the special case of the location-shift model

$$Q_C(\theta \,|\, X) = X'\alpha + \sigma F_\epsilon^{-1}(\theta),$$

which was shown to be equivalent to the classical homoscedastic OLS model

$$C = X'\alpha + \sigma\epsilon,$$

where ϵ has distribution function F_ϵ.

Although this model underlies much of classical econometric inference, it posits a very narrowly circumscribed role for the covariates in X. Hence, we want to test whether the sequence of linear quantile regression alternatives

$$Q_C(\theta \,|\, x) = X'\beta_\theta$$

takes the location shift form.

Specifically, the *location shift hypothesis* may be expressed in standard form as

$$R\beta(\theta) = r, \tag{9}$$

where $R = [0, I_{p-1}], r = (\alpha_2, \ldots, \alpha_p)^T$. This simply asserts that the quantile regression slopes are constant, independent of θ.

Test Statistics

We consider the test statistic

$$K_n = \sup_{\theta \in \mathrm{T}} \|\tilde{v}_n(\theta) - \tilde{v}_n(\theta_0)\| / \sqrt{\theta_1 - \theta_0}, \tag{10}$$

where $\tilde{v}_n$ is the Khmaladze transformation of the empirical quantile process. Understanding what empirical processes are and how they are transformed requires requires quite some involved theory, so we do not discuss it any further at this point. Note, though, that this test statistic is asymptotically distribution free.

Besides, it may be of some independent interest to investigate which of the covariates contribute most to the joint significance of the K_n statistic. Note that the covariates are not necessarily independent, but we plunge ahead nevertheless. In place of the joint hypothesis, we can consider univariate sub-hypotheses for each "slope" coefficient. In effect, this replaces the matrix standardization defined earlier by a scalar standardization. We construct the test statistics

$$K_{ni} = \sup_{\theta \in \mathrm{T}} |\tilde{v}_{ni}(\theta) - \tilde{v}_{ni}(\theta_0)| / \sqrt{\theta_1 - \theta_0} \tag{11}$$

for each of the covariates.

B.3 Goodness-of-fit

Following Koenker and Machado (1998), in the quantile regression model, the so-called pseudo R^2 can be constructed, a statistic that is similar to the classical R^2. For any quantile $0 < \theta < 1$, the pseudo R^2 is defined as

$$\text{pseudo } R^2_\theta = 1 - \frac{\min_{\beta_1,\beta_2} \sum_{i=1}^n \rho_\theta(C_i - \beta_1 - Z_i'\beta_2)}{\min_{\beta_1} \sum_{i=1}^n \rho_\theta(C_i - \beta_1)}, \tag{12}$$

where the function ρ_θ is the tilted absolute value function. Hence, for a specific quantile θ, the pseudo R^2 measures the success of the quantile regression model relative to the unconditional quantile in terms of an asymmetrically weighted sum of absolute residuals. The ordinary R^2, on the other hand, assesses the goodness-of-fit of a model in terms of residual variance:

$$R^2 = 1 - \frac{\min_{\beta_1,\beta_2} \sum_{i=1}^n (C_i - \beta_1 - Z_i'\beta_2)^2}{\min_{\beta_1} \sum_{i=1}^n (C_i - \beta_1)^2} \tag{13}$$

Just like the ordinary R^2, the pseudo R^2 lies between 0 and 1. However, the pseudo R^2 is a local goodness-of-fit measure for a specific quantile, while the R^2 only provides a global indication of the goodness-of-fit.

B.4 Relative Importance of Variables

One way to assess the relative importance of explanatory variables is the well-known partial R^2, which reflects the proportion of unexplained variation of the dependent variable that becomes explained by adding a covariate to the model. Because we examine a range of models over the quantiles in the interval (0,1), the problem is more challenging with quantile regression. Hence, it seems reasonable to measure the relative importance of a covariate over a range of quantiles of the trading costs. Hence, we might apply a generalized version of the squared partial correlation (SPC) to the quantile regression framework.

In a linear regression model, the SPC represents how much of the variation in the dependent variable that is not related with any other predictors is caused by the variance in a certain covariate. It is calculated as

$$SPC = \left(R^2 - R^2_{-i}\right) / \left(1 - R^2_{-i}\right), \tag{14}$$

where R^2 is the adjusted R^2 of the full model (containing all covariates) and R^2_{-i} is the adjusted R^2 corresponding to the model without covariate i.

By defining this measure in terms of the pseudo R^2 (denoted by $\tilde{R}^2$), it can be used to analyze the relative importance of variables in the quantile regression model. Specifically, the SPC is calculated as a function of the quantile $\theta \in (0, 1)$:

$$SPC(\theta) = \left(\tilde{R}(\theta)^2 - \tilde{R}(\theta)^2_{-i}\right) / \left(1 - \tilde{R}(\theta)^2_{-i}\right). \tag{15}$$

Using this equation, we can calculate the SPCs over a specific range of quantiles for each explanatory variable in the quantile regression model. By calculating the average SPC over the quantiles $\theta = \{0.01, \ldots, 0.1\}$ as a proxy for the left tail and $\theta = \{0.9, \ldots, 0.99\}$ as a proxy for the right tail, we get an indication of the contribution of each variable to the left and right tail of the cost distribution. Moreover, we calculate the SPC at the median ($\theta = 0.5$) to analyze the contribution of each covariate to the center of the cost distribution. The results for the specific trading cost example in this research can be found in Table 10.

Additional Tables

Study	Factors
Chan and Lakonishok (1997)	Size, Volatility, Trade Time, Log(Price)
Keim and Madhavan (1997)	Size, Mkt Cap, Style, Price
Bertsimas and Lo (1998)	Size, Market Conditions, Private Information
Almgren and Chriss (2001)	Size, Volume, Sequence of Trade
Breen, Hoodrisk, and Korajczyk (2002)	14 Factors - Size, Volatility, Volume, etc.
Kissell and Glantz (2003)	Size, Volatility, Mkt Conditions, Seq. of Trades
Lillo, Farmer, and Mantegna (2003)	Order Size, Mkt Cap

Table 12: Overview of cost estimation studies

Variable	Definition
Prices	
Arrival Price (P_0)	mid price of the contract absent any information about the incoming trade
Execution Price (P_{exe})	volume-weighted average price of all child fills of the specific order during the trading period
Decision Price (P_d)	futures closing price when the decision to buy or sell the asset was made
Benchmark Prices	
Interval TWAP	time-weighted average mid price during the life of the order
Interval VWAP	volume-weighted average of all transaction prices in the same futures contract during the life of the order
Market-specific Characteristics	
Momentum	5-day average return prior to trading (in %)
Volatility	30-day annualized volatility prior to trading (in %)
Trade-specific Characteristics	
Duration	time elapsed between the moment the order was initiated and the moment it was fully executed
Child Fills	number of the number of separate trades needed to fully execute a specific order
Relative Size	0/1 variable for trades belonging to size class 2

Table 13: Overview of variables and their definitions

Contract name	BLB Ticker	Currency	Exchange	Contract size	Tick size
Stock Indices					
AEX	EO	EUR	Euro Amst	200	0.05
CAC 40	CF	EUR	Euro Paris	10	0.50
DAX	GX	EUR	EUREX	25	1.00
eMini Dow ($5)	DM	USD	CME	5	1.00
eMini NASDAQ 100	NQ	USD	CME	20	0.25
eMini S&P 500	ES	USD	CME	50	0.25
Euro STOXX 50	VG	EUR	EUREX	10	1.00
FTSE 100	Z	GBP	ICE	10	0.50
FTSE/MIB Index	SW	EUR	IDEM	5	5.00
Hang Seng	HI	HKD	HKFE	50	1.00
Mini Russell 2000	RTA	USD	CME	50	0.10
Nikkei 225	NK	JPY	OSE	1000	10.00
S&P/TSX 60	PT	CAD	Montreal	200	0.10
SPI 200	XP	AUD	SFE	25	1.00
Stockholm OMX300	QC	SEK	OMX	100	0.25
Currencies					
Australian Dollar	AD	USD	CME	100000	0.00005
British Pound	BP	USD	CME	62500	0.0001
Canadian Dollar	CD	USD	CME	100000	0.00005
EUR/USD	EC	USD	CME	125000	0.00005
Japanese Yen	JY	USD	CME	12500000	0.0000005
Mexican Peso	PE	USD	CME	500000	0.00001
New Zealand Dollar	NV	USD	CME	100000	0.0001
Commodities					
Gold ($/ozt)	GC	USD	CME	100	0.10
Silver ($/ozt)	SI	USD	CME	5000	0.005
Lean Hogs ($/lbs)	LH	USD	CME	40000	0.025
Live Cattle ($/lbs)	LC	USD	CME	40000	0.025
Corn ($/bu)	C	USD	CME	5000	0.25
Soybeans ($/bu)	S	USD	CME	5000	0.25
Cocoa ($/mt)	CC	USD	ICE	10	1.00
Coffee ($/lbs)	KC	USD	ICE	37500	0.05
Cotton #2 ($/lbs)	CT	USD	ICE	50000	0.01
Sugar #11 ($/lbs)	SB	USD	ICE	112000	0.01
Brent Crude ($/bbl)	CO	USD	ICE	1000	0.01
Crude Oil ($/bbl)	CL	USD	CME	1000	0.01
Gasoil ($/mt)	QS	USD	ICE	100	0.25
Heating Oil ($/gal)	HO	USD	CME	42000	0.0001
Natural Gas ($/btu)	NG	USD	CME	10000	0.001
RBOB Gasoline ($/gal)	XB	USD	CME	42000	0.0001
Bonds					
US Ultra T-Bond	WN	USD	CME	100000	0.03125
10Y Australia T-Bond	XM	AUD	SFE	100000	0.005
10Y Canadian Govt Bond	CN	CAD	Montreal	100000	0.01
10Y Japan Govt Bond	JB	JPY	OSE	100000000	0.01
10Y T-Note	TY	USD	CME	100000	0.015625
2Y T-Note	TU	USD	CME	200000	0.0078125
5Y T-Note	FV	USD	CME	100000	0.0078125
Euro Bobl	OE	EUR	EUREX	100000	0.01
Euro Bund	RX	EUR	EUREX	100000	0.01
Euro Schatz	DU	EUR	EUREX	100000	0.005
Long Gilt	G	GBP	ICE	100000	0.01
US Long T-Bond	US	USD	CME	100000	0.03125

Table 14: Overview of traded futures contracts

Panel A: Volatility					
	All Data (%)	**Stock Indices** (%)	**Currencies** (%)	**Commodities** (%)	**Bonds** (%)
Mean	13.59	14.35	7.73	25.89	2.91
St. dev.	11.78	4.23	2.83	9.94	2.99
5% Quantile	0.37	8.88	4.60	12.59	0.23
95% Quantile	33.9	21.84	13.84	48.46	10.52

Panel B: Daily Volume					
	All Data ($\mathbf{10^5}$)	**Stock Indices** ($\mathbf{10^5}$)	**Currencies** ($\mathbf{10^5}$)	**Commodities** ($\mathbf{10^5}$)	**Bonds** ($\mathbf{10^5}$)
Mean	2.24	2.68	1.00	0.86	4.27
St. dev.	3.49	0.42	0.76	1.07	4.63
5% Quantile	0.14	0.15	0.23	0.01	0.24
95% Quantile	8.69	12.47	2.22	3.02	13.51

Panel C: Spreads					
	All Data (bps)	**Stock Indices** (bps)	**Currencies** (bps)	**Commodities** (bps)	**Bonds** (bps)
Mean	2.02	1.76	1.21	3.69	0.75
St. dev.	1.82	1.18	0.86	1.91	0.43
5% Quantile	0.35	0.69	0.45	1.04	0.35
95% Quantile	5.84	4.24	2.85	6.99	1.85

Table 15: Descriptive statistics of market-specific characteristics for each asset class