

ABSTRACT

Title of dissertation: High Frequency Market Dynamics
An Analysis of Market Depth and Quoting
Behaviors in Crude Oil Futures Markets

John Spencer Roberts
Doctor of Philosophy, 2018

Dissertation directed by: Professor Albert “Pete” Kyle
Department of Finance

Most derivative and equity transactions occur in electronic order driven markets and depend on a limit order book. Yet many questions remain regarding the way traders interact with the limit order book, especially the role of algorithmic and high frequency trading. This dissertation investigates how the limit order book evolves over time. We study the nature of fleeting liquidity and flash quotes to deepen our understanding of the way modern markets operate.

This research is based on raw message data sold by the exchange and contains every update to the limit order book linked to the top ten levels. We rebuild the limit order book and define quote segments to divide the day into non-overlapping intervals based on observed changes in the best quotes and the bid-ask spread. We propose a novel way to visualize dynamics of the limit order book by combining changes in best quotes and visible depth. Using the limit order book and quote segments, we define a measure for offered liquidity and then a measure to capture the responsiveness on both sides of the market during sub-second intervals. Flash quotes are identified and combined with measures of offered liquidity to study why such behavior is observed in the market.

We find empirical evidence that changes in market depth explains movement in the bid-ask spread. We show how combining movements in best quotes and visible depth provides a clearer picture of market direction. Evidence is presented that breaks down the dynamics of offered liquidity into both trade response and prior movement of depth. We find standard measures of market liquidity, such as the bid-ask spread, can appear normal while responsiveness can remain elevated following a major market movement.

Depth data assists with best execution, but this research highlights alternative uses that are important to consider when participating in modern markets. The observed dynamics of the limit order book contain relevant information that needs to be captured in a full discussion of market liquidity.

HIGH FREQUENCY MARKET DYNAMICS
AN ANALYSIS OF MARKET DEPTH AND QUOTING
BEHAVIORS
IN CRUDE OIL FUTURES MARKETS

by

John Spencer Roberts

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2018

Advisory Committee:
Professor Albert "Pete" Kyle, Co-Chair
Professor Anna Alberini, Co-Chair
Professor Peter Cramton
Professor Jim Hanson
Professor Lars Olson

© Copyright by
John Spencer Roberts
2018

Dedication

To my parents

Acknowledgments

I would like to thank my parents, Bruce and Amy Roberts, for their unyielding support and encouragement, both before and during graduate school. Thank you for encouraging me to ask questions and then seek out answers.

I would like to thank my husband, Michael Kaiser, for his continuous support during these last few years. Thank you for supporting my desire to finish and the gentle nudges required to keep me on track. I would not have been able to finish this without you.

I would like to acknowledge the important role of my advisor, Pete Kyle, in assisting me along the way. I can not think of a better person to guide me. Thank you for the conversations and supporting my research interests.

I need to include a special thank you to Jeff Cunningham, Director of Information Systems at AREC, who built and maintains the technology that has made this research possible. Also, I need to extend a thanks to Pin-Wuan Lin, at AREC, who was always willing to help with department questions.

Finally, I would like to thank the dedicated people in the Office of the Chief Economist and in the Division of Market Oversight at the Commodity Futures Trading Commission. Working with you has provided me with a deep understanding of these important markets. I have been very lucky to work along side you and learn together.

Contents

Dedication	ii
Acknowledgements	iii
1 Introduction	1
2 Literature Review	10
3 Market Depth Data	16
3.1 Introduction	16
3.2 Processing Raw Message Data	19
3.2.1 Raw Message Data	20
3.2.2 Book Update Data	24
3.2.3 Transactions Data	29
3.2.4 Processed Data Adjustment — Reshape Book	36
3.3 Analysis of Best Quotes	38
3.3.1 Quote Improvements & Promotions	39
3.3.2 Bid-Ask Spread Analysis	45
3.4 Conclusions	53
4 Quote Segments	55
4.1 Introduction	55
4.2 Identification of Quote Segments	57
4.2.1 Processed Market Depth Data	58
4.2.2 Definition of Quote Segments	60
4.2.3 Quote Segment Sample Summary	68
4.3 Depth, Updates, and Volume Analysis	77
4.3.1 Quote Segments & Visible Depth	77
4.3.2 Quote Segments & Quote Updates	81
4.3.3 Quote Segments & Transactions	83
4.4 Alternative Identification Methods	89
4.5 Conclusions	91
5 Visualizing Market Depth Data	93
5.1 Introduction	93
5.2 Data & Methodology	96
5.3 Single Visualizations	98
5.3.1 Visualization of Best Quotes	99
5.3.2 Visualization of Best Quotes & Transactions	101

5.3.3	Visualization of Market Depth	102
5.4	Combining Depth and Quotes	108
5.4.1	Combining Depth Changes with Best Quotes	108
5.4.2	Combining Depth Changes & Trades with Best Quotes	111
5.5	Adding Depth at Deeper Levels	112
5.6	Conclusions	122
6	Fleeting Liquidity	125
6.1	Introduction	125
6.2	Offered Liquidity & Responsiveness	130
6.2.1	Offered Liquidity (OL)	130
6.2.2	Responsiveness of Offered Liquidity (ΔOL)	132
6.3	Data & Methodology	134
6.3.1	Data	134
6.3.2	Measuring OL	139
6.3.3	Measuring ΔOL	140
6.4	Summary of OL and ΔOL	142
6.4.1	Intraday & Daily Summary	143
6.4.2	Further Study of OL , ΔOL , Volume, & Volatility	145
6.5	Quote Segment Start	151
6.5.1	Single Quote Segment Example	151
6.5.2	Event Based Measurement vs Random Sample	153
6.5.3	Event Analysis — Six Types of Events	156
6.5.4	Event Compare/Contrast	158
6.5.5	Regression Analysis	160
6.6	Directional Movement	167
6.6.1	Observed Imbalance in OL During Directional Moves	167
6.6.2	OL and ΔOL at Start & End of Quote Segments	169
6.6.3	Predictive Modeling — Logistic Regression Analysis	171
6.7	Conclusions	176
7	Flash Quotes	178
7.1	Introduction	178
7.2	Data & Methodology	183
7.2.1	Data	183
7.2.2	QIP Event — Sequence of QI QP Events	188
7.2.3	QIP Identification	189
7.3	QIP Summary	190
7.3.1	Distribution of QIP Characteristics:	191
7.3.2	QIP Number of Updates & Switches:	193
7.3.3	Intraday Investigation:	195
7.4	QIP Events and Liquidity Provision	201
7.5	QIP Events, Volume & Volatility	204
7.5.1	Conditional Expectations: ΔOL	205
7.5.2	Conditional Expectations: Bid-Ask Spread	206
7.5.3	Information Conditional on Number of QIP Events	208
7.5.4	Regression Analysis of QIP Volume Share	208
7.5.5	Minutes Before and After Unexpected Market Events	212

7.6	QIP Events & Expected News	214
7.6.1	Sample of Events	215
7.6.2	Analysis	215
7.7	Conclusions	220
8	Conclusion	222

List of Tables

3.1	Sample of Book Updates — No Trades	25
3.2	Summary of Book Updates — by Book Level	27
3.3	Summary of Limit Order Arrival & Cancellations	28
3.4	Sample of Book Updates — with Trades	32
3.5	Summary of Trades and Trade Groups	36
3.6	QI QP Event Summary	44
3.7	Using Depth Movement to Explain Wide Bid-Ask Spreads	51
3.8	Cause of Wide Bid-Ask Spreads and Next Book Updates	52
4.1	Quote Segments — Distribution Summary	69
4.2	Quote Segments — Autocorrelation Study	71
4.3	Quote Segments — Directional & Bounce Movement	73
4.4	Summary Statistics of Visible Depth Levels	80
4.5	Summary Statistics of Time Weighted Visible Depth	81
4.6	Distribution of Observed Depth Change by Quote Segment Level	82
4.7	Frequency of Non-Zero Depth Changes by Quote Segment Level	82
4.8	Distribution of Time Between Book Updates by Quote Segment Level	83
4.9	Volume and Quote Change Event Conditional on Quote Segment Start	86
6.1	Measured Responses: Random Sample vs Quote Segment Start	154
6.2	OL and ΔOL : By Event Type	159
6.3	Regression Results	165
6.4	Regression Results — Different Measurement Intervals	166
6.5	OL : Buy and Sell Side with Directional Movement	168
6.6	OL and ΔOL : Buy and Sell Side with Directional Movement	170
6.7	Logistic Regression Results	174
6.8	Logistic Regression Results — Predicted Probabilities	175
7.1	QIP Summary Table	192
7.2	QIP Characteristics Conditional on QP Event	194
7.3	Summary Statistics of QIPs by Number of Updates & Switches	196
7.4	Correlations — QIP Counts to QIPs with Trades	203
7.5	Regression 1 — QIP Volume Shares Per Minute	211
7.6	Regression 2 — QIP Volume Shares Per Minute	212

List of Figures

3.1	Examples of Two Raw Messages	21
3.2	Quote and Volume Compare	31
3.3	Movement in the Best Quotes	39
3.4	Bid-Ask Spread (BAS) Examples	46
3.5	Volume Shares Conditional on Bid-Ask Spread: Daily and Intraday	47
3.6	Duration of Wide Bid-Ask Spreads	48
4.1	Quote Segment Examples	75
4.2	Counts of Quote Segments by Date and Intraday Intervals	76
4.3	Time Weighted Spread, Volume and Depth for Quote Segments	87
4.4	Percent of Total Volume by Quote Segment Level	88
5.1	Visualizing Best Quotes & Bid-Ask Spread	101
5.2	Visualizing Best Quotes & Trades	103
5.3	Visualizing Depth — Best Quotes	105
5.4	Visualizing Depth	107
5.5	Visualizing Adjusted Quotes	110
5.6	Visualizing Adjusted Quotes with Trades	111
5.7	The Complete Visual: Adjusted Quotes with Depth Information	113
5.8	Additional Examples: Calm and News Announcements	115
5.9	Additional Examples: Market Movement	116
5.10	Complete Visual: Narrow Range	119
5.11	Complete Visual: Price Discovery	120
5.12	Complete Visual: Price Wave	121
6.1	Six Examples of Fleeting Liquidity	129
6.2	Balanced and Imbalanced Offered Liquidity	131
6.3	Dynamics of Offered Liquidity	133
6.4	Measurement of OL and ΔOL at t_0	143
6.5	Intraday Volume, Volatility, OL , and ΔOL	147
6.6	Daily Volume, Volatility, OL , and ΔOL	148
6.7	One Minute Averages: Ranked by Volume	149
6.8	One Minute Averages: Ranked by Net Volume	150
6.9	Selecting Start of Quote Segments	151
6.10	OL — Single Quote Segment	152
6.11	Visualization of Quote Segment Movement — Six Event Types	157
7.1	Six Examples of Quote Improvements	182
7.2	Example of three QIPs	189
7.3	QIPs by Minute Intervals	199

7.4	Time Weighted Information Conditional on Number of QIPs per Minute	200
7.5	Balance between QIPs Arrival on Both Sides of the Market	202
7.6	Average Absolute Imbalance by Number of QIP Deciles	204
7.7	Analysis of One Minute Intervals — Ranked by ΔOL	206
7.8	Analysis of One Minute Intervals — Ranked by Bid-Ask Spread	207
7.9	Distributions of Market Activity Conditional on Observed QIP Events	209
7.10	Activity Conditional on Proximity to Unexpected Event	214
7.11	Analysis Before & After Expected News — 10 Minutes After	218
7.12	Analysis Before & After Expected News — 20 Minutes After	219

Chapter 1

Introduction

This dissertation is focused on the study of a marketplace that exists to ensure that the production, distribution, and consumption of goods takes place at the best possible price. While the goods in this case are commodity futures contracts — specifically, a single futures contract for a fixed number of barrels of crude oil — we argue that the insights can apply to all markets that share similar features in how they operate.

The market is a centralized electronic marketplace composed of buyers and sellers who are able to submit limit orders to a matching engine that collects and organizes these orders into an electronic limit order book. Buyers submit bids and sellers submit ask quotes with specified price and quantities. In contrast to equity markets, there is not a defined fixed supply of futures contracts, so if a new buyer and seller both meet to trade, they produce a new futures contract and open interest increases by one contract. This is a true order driven market place that is open close to 24 hours of the day.

These markets are made up of many different types of participants with dif-

ferent objectives. Some choose to participate in the market because they have an economic relationship to the underlying commodity. For example, major oil production firms participate given their exposure to price movements in crude oil. Similarly, many firms which consume crude oil also choose to participate in futures markets. Market participants without an underlying connection to the commodity, but still wish to get exposure to the commodity are also welcome to trade — this set of market participants is historically referred to as speculators.

Holding horizons of the contracts vary along with sensitivities to price movements. A subset of speculators are day traders and a subset of day traders are market makers. The market maker plays a key role as the intermediary between non-market makers. Their importance is highlighted in the numerous theoretical models that make explicit reference to this subset of market participants and their role in establishing the bid-ask spread, which is defined as the observable price difference between the best ask and the best bid quote.

While much research exists on these markets, over the past ten years these markets have experienced a significant change that has demanded a fresh look at how these modern markets operate. The introduction of electronic trading has destroyed the prior model of human traders shouting and tossing paper around in a pit. Most trading pits have been closed. Server farms and terms such as microwaves, algorithms, and co-locating dominate the conversation surrounding the quest for survival in these modern electronic markets. This change has had the most impact on how market making is performed, especially since, in many of these markets, there is not a designated role assigned to someone (or some algorithm).

High frequency trading firms compete with each other to process and respond to relevant information the fastest. This relevant information is not so focused on major public news events. To these firms, the most valuable information might be that a large trade just occurred or that the price in the current market is slightly mis-priced relative to a closely correlated market. Much of this comes down to identifying and using whatever relevant information exists.

When a single new limit order is placed at the best bid, firms able to process this information can place this depth change in context. They can answer questions such as: How has depth recently changed on both sides of the market? Is this a new movement or part of an expected trend? When a single trade arrives relevant questions include: Was it unusually large? Did the trade kill the quote? Was this trade part of a cluster of trades? If so, do we observe new limit orders arriving to re-establish depth?

This dissertation research contributes to our understanding of these important markets by utilizing the most detailed set of information publicly available. The information is exactly the source processed in near real-time by high frequency trading firms and other market participants. The primary interest in order level data is the ability to study market dynamics related to information that exists prior to realized trades — market dynamics that exist in the limit order book due to the flow in and out of resting limit orders.

Historically, empirical studies incorporated very high level summaries of market transactions. Increases in computing power, along with the better collection of relevant information, have allowed us to move from monthly to daily data. Continued advances then allowed research to incorporate intraday information. We

make use of ultra-high frequency data since it contains every single recorded event and represents the final stage of data availability. This dissertation contributes to the literature, not only by the insights discussed in later chapters, but also by increasing our understanding of how to use such data.

This dissertation is timely since there have been recent changes to what information is made available to the subset of market participants with the ability to read and react to this information almost instantly. The available data now includes more information that allows the trader to monitor limit orders that are resting in the book. This will allow the trader to know exactly their place in line to get their order filled. This changes the traders ability to measure the probability their order will be filled — one of the key factors in deciding at what quote level to place a limit order — and also affects the decision to place a limit order that rests in the book versus the use of a marketable limit order that immediately results in a trade.

This research is focused on understanding how visibility of the limit order book and observed dynamics can be used to make sense of market performance. For example, heavy emphasis is often placed on traded volume and how traders can learn from observing signed volume. But trades occur at established quotes and observing a buyer-initiated trade, which is defined as an aggressive trade submitted by a buyer that matches with a passive order resting at the best ask quote, is supposed to contain information that there exists buying interest in the market. How does the story change if the best ask quote was just established from a trader wishing to sell? Similarly, the interpretation of the trade should change based on the level of observed depth at the best ask quote. If depth is close to

zero or if there exists an unusually large depth, representing a lot of interest to sell, does this change the way we view the buyer-initiated trade? These questions require the use of market depth observed at each change of the electronic limit order book.

We focus on two themes relevant to making sense of modern markets: first, we focus on an analysis of fleeting liquidity, or the sudden entry and/or exit of market depth or offered liquidity at quotes closest to the market; second, we focus on the occurrence of flash quotes and flickering interests of liquidity provision.

Fleeting liquidity requires first a definition of the offered liquidity and then a method to measure the rate at which this offered liquidity changes over time. Once the limit order book is rebuilt, we have complete information and can watch both sides adjust to a market event. We find there are certain periods of the day that have elevated responsiveness, implying fleeting liquidity changes over the day. Understanding such dynamics becomes relevant for the market participant as the dynamics will influence the cost of trading as well as the market impact of placing limit orders.

Should we view fleeting liquidity as a negative aspect of modern markets? Speed and technological improvements allow for market participants to make sense of a market and better manage risk, but also make decisions to increase their returns. Certain market participants, who can monitor market dynamics, can watch for a build up of depth at the best bid and use such information to expect pressure for the market to move upward — expectations that should result in cancellations of resting limit orders at the best ask and moving them upward.

Flash quotes are defined as offered liquidity which exists for a very short du-

ration in the limit order book. We choose to focus on the subset of events that create a change in the best quotes. We focus on events where the flash quote improves the market on one side, for example the arrival of a new best bid, but then quickly disappears, due to either a cancellation or trade. These are of particular interest because a non-trivial share of passive liquidity provision can be linked to resting quotes that have a very short resting time. This implies that a limit order is placed into the limit order book and very soon after results in a trade.

Why do we observe flash quotes and should we be concerned with their presence in the limit order book? If the majority of flash quotes result in cancellations, should we view flash quotes as limit orders placed without the intention of trade?

One of the main questions market regulators have concerns identifying the intent of the action, especially when the action is placing a new non-marketable limit order. At the time the order is established, was there a true intent for the order to result in a trade? A more difficult question relates to how this intent changes over time as the limit order book changes. Is the trader able to modify the order as their expectations of future market movement changes due to observing recent dynamics of the limit order book? If so, should there be a limit on how fast this change is allowed to take place? Is an order placed with a cancellation 20 milliseconds later a sign the order never intended to be part of a trade, or does it simply indicate the market participant has a fast reaction to an updated expectation? What if the order is placed simply to see if one or both sides of the market responds to the initial move? Should this be against the rules of the market?

Chapter 2 provides a review of related literature. There exists a large liter-

ature composed of papers, both theoretical and empirical, which focus on topics related to limit order markets. While the questions have evolved over time as the markets have changed, there exist a set of underlying questions such as: How do market participants make decisions under uncertainty (traditional dealer markets vs anonymous electronic limit order books)? What is the role of transparency in a limit order book and how does it impact market quality? How does technology (increased automation) change the way the market operates?

Chapter 3 introduces the raw data and outlines steps required to rebuild the limit order book. The information was individually purchased from the Chicago Mercantile Exchange (CME) and is made available for public use. While a few research papers exist that also use these data, none go into the detailed use presented in later chapters. In this chapter, we highlight aspects of the data and show how the data contains very useful information. We review the two main components of the data: the limit order book updates and the transaction records.

Chapter 4 introduces our method of breaking up the sample into intervals — termed quote segments — allowing for improved analysis while retaining many important features of the market activity. Quote segments allow us to keep track of book levels and properly account for order flows in and out of the market. We argue these quote segments are properly defined based on a review of market activity. This chapter investigates visible depth, traded volume, and book updates within the context of the defined quote segments.

Chapter 5 presents a number of visualizations of the processed limit order book data. Working with huge data sets presents a number of challenges; one of these challenges involves how one chooses to summarize and review the contents.

Visualization has become a main tool in working with lots of data and we incorporate many visualizations of order book activity in later chapters to motivate our research questions. Once individual data elements are discussed, we present a visualization that combines quotes and changes in depth on both sides of the market. We show how this visualization can be enhanced by incorporating depth at deeper levels of the book on both sides of the market.

Chapter 6 focuses on the dynamic nature of visible depth at various levels of the book. We define a measure of offered liquidity and study how this changes over very short intervals of time — termed the responsiveness of offered liquidity. This chapter focuses on both the arrival of new limit orders that increase visible depth and the act of cancelling previously placed limit orders, resulting in decreases in visible market depth. We show there exist certain predictable periods of both sides adjusting quotes conditional on certain well defined events. This chapter addresses the fleeting nature of liquidity and shows how the use of limit orders, and the ability to watch the market depth adjust, can provide a lot of information about future events.

Chapter 7 presents analysis of the high frequency limit order book data focusing on flickering quotes and flash liquidity provision. With the increased use of technology and increases in market speed and complexity resulting from the arrival of high frequency and algorithmic trading, a number of market characteristics have emerged that have become the focus of academics and market practitioners. One of these relates to the flickering nature of resting quotes at the best bid or ask (we refer to these as flash quotes). We focus analysis to observations of very short durations for passively placed limit orders that provide liquidity to the electronic

limit order book. We then explore the relationship between this activity and visible market depth and market volatility. Furthermore, this chapter will explore the relationship between traded volume and flickering quotes.

Chapter 8 concludes the dissertation and summarizes the main findings and lists a number of areas for future research.

Chapter 2

Literature Review

This research contributes to our understanding of how modern electronic high frequency markets operate. There are a number of different lines of related research that focus on market participants' decisions to use market or limit orders, concerns regarding liquidity and how to best measure such liquidity, studies that are focused on the information content of individual orders/trades along with the complete limit order book, and high frequency trading. For excellent papers that survey the large literature related to limit order markets and high frequency trading, see Parlour and Seppi (2008) and Jones (2013). Gould et al. (2013) provides a detailed review of research related to limit order books.

There exists a large literature that builds mathematical models of order book dynamics. Cont et al. (2010) and Cont and De Larrard (2013) are two examples of papers that seek to better understand the relationship between order flow, liquidity and price dynamics. These models also are designed to reflect documented empirical regularities. An and Chan (2017) is a recent example that builds on these models. One concern with these models is their assumption that the process

driving order arrival and cancellations on both sides of the market is an independent process. This fails to capture the feedback influences that result from market participants responding to observed order flow.

A large literature has developed that explores market participants deciding between a limit order or a market order. The use of a limit order introduces costs related to non-execution and adverse selection (see Chakravarty and Holden (1995); Parlour (1998); Foucault (1999); Foucault et al. (2005); Goettler et al. (2005); Kaniel and Liu (2006); Roşu (2009)). Handa and Schwartz (1996), in particular, shows that limit orders are more attractive during periods of temporary price fluctuations. This motivates the use of high frequency data to track how liquidity provision changes within the day as volatility changes.

Theory often does not make room for the market participants' willingness or ability to cancel previously placed limit orders. Given the fact that the majority of high frequency firms acting like market makers incorporate high cancellations rates in their strategies, some recent theoretical work has explored what happens when market participants have the option to cancel limit orders.

One example of a model allowing for cancellations is a paper by Ait-Sahalia and Sağlam (2016); their model includes a strategic high frequency trader allowed to exploit its speed and informational advantage to place quotes. In this model, the market maker is able to revise quotes based on the flow of information. Cancellations are allowed here as the market maker updates its expectation of future order flow (expectation of facing a patient or impatient trader). The bid-ask spread is endogenously determined here based on the quote placement of the single market maker. This model is then extended in Ait-Sahalia and Sağlam (2017) to allow

for competition between market makers.

Another example is found in Rosu (2016) who sets up a dynamic model of order-driven markets that allows informed investors to adjust their limit orders by either modifying or cancelling them. This research finds that a larger share of informed traders improves liquidity. Market participants make decisions on acquiring information, order placement, and selecting a market or limit order.

Along with providing market participants with more options, recent developments have allowed these decisions to depend on the dynamics of supply and demand. Obizhaeva and Wang (2013) shows that static properties are not as important as considerations of how the market responds over time to trades. My research is related as we study the dynamics of depth instead of simply the static information.

There exist many empirical studies that characterize limit order markets. Biais et al. (1995) was one of the first papers to conduct a detailed analysis of the limit order book. This study focused on how observed order flow changes depending on the state of the limit order book. They are able to make a number of statements related to observed order flow. In particular, a great deal of focus was paid to the side of the bid-ask spread and how this signals to the market the need for liquidity provision. One of the important insights from this work is the ranking of limit orders, which expanded the concept of an aggressive limit order. Hamao and Hasbrouck (1995) is another early paper that analyzes intraday behavior of trades and quotes on the Tokyo Stock Exchange. The focus here was on understanding how an order driven market operated without designated dealers — the same market set up that is the focus of this dissertation.

Ahn et al. (2001) expands on the analysis of order flow by investigating the relationship between transitory volatility and limit order placement. They find that it is important to account for what market side is responsible for the increased volatility. This is relevant for this dissertation since it relates to limit order choice and the arrival process for limit orders. Biais et al. (2010) is a more recent example that studies the competitive behavior in order placement strategies on INET and Nasdaq.

A recent line of the literature has raised questions regarding the price impact of limit orders. This questioned the prior view that only market orders contained information related to prices. Now, with the increased use of limit orders that rest in the book, along with the increased use of market data to determine order routing, limit orders should be expected to have some influence on the market. Hautsch and Huang (2012) conducts an empirical analysis on permanent price impacts and compares market orders to limit orders; they find both have an impact, but the impact of market orders is four times larger. Rosu (2016) sets up a theoretical model that also shows limit orders have a reduced permanent price impact.

Early research into the role of market transparency, specifically related to the order book, can be found in Boehmer et al. (2005), which presents an empirical exploration of the consequences of moving to an open limit order book. This is also an early example of a paper that addresses limit order cancellations. These authors set up a duration model of limit order cancellations to study trading behavior given the new access to information. The results suggest increased transparency can increase the number of orders as traders are better able to determine

available liquidity. Other papers that address cancellations of limit orders include Chakrabarty et al. (2006) and Cho and Nelling (2000); Lo et al. (2002).

Other papers that explore the role of market transparency include: Flood et al. (1999), who sets up an experiment and find that increased transparency decreases trading cost and price efficiency. Bloomfield and O'Hara (1999) include results showing that transparency helps prices become more efficient. Hendershott and Jones (2005) investigates the increase of transparency by exploring how the limit order book can be used to collect information prior to trade. This paper, along with Madhavan et al. (2005), are relevant to this dissertation since we focus on the exact information market participants study when making trading decisions.

Market participants are often assumed to have access to information that guides them in how they trade. Weller (2016) points to the conflict between acquiring information and how this becomes reflected in prices. This raises the question if the market is better off as individuals are in constant search for better information.

A group of other papers present evidence that high frequency trading firms are able to forecast price changes (e.g., Carrion (2013); Hirschey (2013); Brogaard et al. (2014)). This raises a number of questions about how this is possible. One answer is access to information. We are able to explore what useful information exists in the visible order book. Information is often costly to acquire and other research has pointed to the processing costs and highlight important tradeoffs that occur between waiting for a better signal (see Dugast and Foucault (2016)).

There are also segments of the literature that discuss specific events observed in modern high frequency electronic markets. One example concerns flash quotes

which represent flash liquidity provision. In these cases, market participants submit limit orders that have very short durations and are often cancelled without trade. Hasbrouck and Saar (2009) define fleeting orders as those that are cancelled within two seconds after the order arrives. The authors suggest that these fleeting orders result from high-frequency order-submission strategies. These strategies are designed to signal the desire to trade and suggests a new economic role for limit orders. They model the probability of cancellations and show how this can be explained by changes in the market. This approach is similar to Smith (2000); Ellul et al. (2007); Ranaldo (2004).

Skjeltorp et al. (2016), extends the analysis of flash events and finds evidence consistent with Hasbrouck and Saar (2009). Using a detailed data set, they focus on actionable IOIs (indications of interest) and find that these orders are used to advertise liquidity needs. They also highlight an important concern related to access since algorithmic traders are the only participants able to participate in this activity.

Chapter 3

Market Depth Data

3.1 Introduction

Modern electronic order driven markets depend on market participants submitting limit orders that do not result in immediate trade, but are collected and organized into the limit order book. The market, prior to trade, exists and is defined by the presence of these anonymous resting limit orders, where each individual order is associated with a defined price level and quantity.

At any point in time, the market is defined by the best bid and ask quote. The best bid (ask) quote represents the resting quote submitted by the market participant wishing to buy (sell) at the highest (lowest) price among all the other buyers (sellers) who have resting limit orders. Together they define the two prices where trade will occur as other market participants decide to either hit the bid or lift the offer.

Modern electronic markets produce a lot of information as new orders flow in and out of the market. With the increased interest in big data and the pursuit of

developing trading strategies that use information to make decisions, market data has become increasingly valuable. Exchanges sell this information and market participants compete over the speed in processing and reacting to such information.

Once the raw data are processed and we have rebuilt the limit order book, we provide an overview of the ultra-high frequency order book information used in this dissertation. The information represents the best possible data to study questions related to how these important markets function. We focus on two different aspects of the data: the quote information and the transaction information.

Quote information relates to changes to the visible electronic limit order book. These changes result from the arrival or cancellation of limit orders submitted by market participants and are referred to as book updates since the flow of orders results in changes to the limit order book. Market participants (or pre-programmed algorithms) will be interested in how the book changes over time and this information is directly observable if one is able to track the flow of book update messages.

Apart from the quote information, these data contain the full set of realized trades. Each of these transactions result from an ‘aggressive’ marketable limit order that arrives and either hits the bid or lifts the offer. The richness of the data comes from the ability to sort the full set of trades and place them in context of the state of the electronic limit order book that existed just prior to the trade and how the book then changes after the trade occurs.

For this research, we focus on the commodity futures market for crude oil, currently traded on the Chicago Mercantile Exchange (CME) under the ticker

symbol CL. This futures market is important for a number of reasons: first, West Texas Intermediate (WTI), a light sweet crude stored and distributed in Cushing, Oklahoma, is considered a benchmark for global oil prices in the US; second, the market for WTI and this related futures market are some of the most active and liquid markets for energy; and third, this futures market is one of the most closely followed markets and is used as an indicator of the larger US economy.

Each futures contract represents 1,000 barrels of WTI crude oil and the contract's price quotation is U.S. dollars and cents per barrel. There is a defined minimum price fluctuation of \$0.01 per barrel, which we will refer to as the minimum tick size of the market. This exchange-defined minimum tick size becomes relevant as this will determine the minimum distance between the two sides of the market as the best bid is often \$0.01 below the best ask quote in the limit order book.

While the focus of this dissertation is on a specific futures product, similar analysis can be carried out across all futures markets at the CME since the underlying information is recorded using the same procedure. The electronic matching engine (GLOBEX) produces the information and therefore, any futures product traded here will contain a similar source of information. Therefore, similar studies can be performed across the product groups: equities, interest rates, foreign exchange, energy, agricultural, and metals.

The remainder of the chapter is organized as follows: Section 3.2 outlines the required steps to process the raw data, which results in the rebuilt limit order book for both sides of the market. We then provide a review of the limit order book as observed during our sample period. Section 3.3 provides an analysis of events

that impact the best quotes. Specifically, we study book updates that result in changes to the best quote price level on one or both sides of the market. We then study the bid-ask spread and show how book updates can be very useful to explain future events. Section 3.4 concludes and summarizes the chapter.

3.2 Processing Raw Message Data

In this section, we introduce the raw data and discuss the steps required to process the message data and build tables that reflect the limit order book. Since the limit order book represents a stock variable, the raw data provides us with instructions for how the book changes. The limit order book then exists until the next recorded message arrives with details on how to update the limit order book. The information is appropriately termed ‘ultra-high frequency’ because it contains the entire record of the limit order book resulting from thousands of individual traders actively submitting both limit and market orders to trade futures contracts. The data is also appropriately termed ‘ultra-high frequency’ given the messages are marked to the millisecond.

We start with a detailed discussion of the raw data. While similar types of data have been used in the literature, there are not many studies that make use of this particular data source. We then highlight the steps required to rebuild the limit order book. Once the raw data is processed and the limit order book rebuilt, we explore the two main elements of the data: the limit order book updates and observed trades.

3.2.1 Raw Message Data

This dissertation is based on market depth data purchased from the CME DataMine service that allows for access to historical information. Market depth files contain all data messages that are required to rebuild the order book. We purchased one year of information related to crude oil futures (CL) — one of the most popular futures products traded on the CME — covering a time range from August 1, 2012 to the end of July 2013.

The information contains depth information at ten levels on each side of the order book. Thus, focusing on the bid side, at any point in time during the market open hours, once we rebuild the book, we can see ten quoted price levels sequenced from the best bid (the maximum of the bids) followed by the second-best and third-best, down to the tenth-best bid quote. At each quoted price level, we see the depth — the total quantity of all outstanding limit orders at that quoted price. If this depth is ten for the best bid, then we know at least a trade for ten contracts can be done at that single price.¹

Before we can study the order book, we must process the raw information delivered from the exchange. The raw data is made up of a series of messages that the electronic matching engine, operating the electronic limit order book, generates when there is an event that either changes volume or changes the visible order book.² These messages are sequenced in the order they are generated by the electronic matching engine. The exchange allows for ten quotes to be visible on both sides of the market, and therefore, any event that causes a change at

¹The exchange does allow for options to hide some or all of the quantity associated with an order. This is commonly referred to as iceberg orders. An example is an order for 25 contracts is placed at the best bid, but only shows at most 5 contracts.

²In our case, GLOBEX is the electronic matching engine operated by the CME.

these quotes will be packaged up into a message. This message is placed for public consumption on a server for third parties to pick up and process.

```
[M1] "1128=9 9=977 35=X 49=CME 34=46149 52=20130728220024438 75=20130729 268=10 279=2 22=8 48=603475 83=209 107=CLV3 269=0 270=10391
271=1 273=220024000 276=K 336=2 1023=1 279=0 22=8 48=603475 83=210 107=CLV3 269=0 270=10390 271=1 273=220024000 276=K 336=2 1023=1 279=2
22=8 48=353903 83=198 107=CLX 269=0 270=10264 271=1 273=220024000 276=K 336=2 1023=1 279=0 22=8 48=353903 83=199 107=CLX 269=0 270=10263
271=1 273=220024000 276=K 336=2 1023=1 279=2 22=8 48=176500 83=254 107=CLZ3 269=0 270=10123 271=1 273=220024000 276=K 336=2 1023=1 279=0
22=8 48=176500 83=255 107=CLZ3 269=0 270=10122 271=1 273=220024000 276=K 336=2 1023=1 279=2 22=8 48=176501 83=150 107=CLZ4 269=0 270=9215
271=1 273=220024000 276=K 336=2 1023=2 279=0 22=8 48=176501 83=151 107=CLZ4 269=0 270=9214 271=5 273=220024000 276=K 336=2 1023=2 279=2 22=8
48=176502 83=121 107=CLZ5 269=0 270=8755 271=1 273=220024000 276=K 336=2 1023=1 279=0 22=8 48=176502 83=122 107=CLZ5 269=0 270=8754 271=1
273=220024000 276=K 336=2 1023=1 10=105"
```

```
[M2] "1128=9 9=255 35=X 49=CME 34=46150 52=20130728220024438 75=20130729 268=2 279=0 22=8 48=53213 83=1253 107=CLU3 269=2 270=10469
271=1 273=220024000 451=-1 1020=82 5797=1 5799=1 279=1 22=8 48=53213 83=1254 107=CLU3 269=1 270=10469 271=16 273=220024000 336=2 346=1 1023=1 10=133"
```

Figure 3.1: Examples of Two Raw Messages

Notes: Figure contains two messages from the raw market depth data. CME message specifications provide details for each tag (numeric ID before each equal sign). Information in red corresponds to details of the entire message and includes a unique message ID and timestamp marked to the millisecond.

Within a single message, numeric identifiers called tags contain all the necessary elements to read the message and collect the information.³ A set of tags contain product details. Another set of tags contain book level-specific information. Since there are many levels made available for public consumption, a single message can have one or multiple sets of tags that make reference to specific book levels within a single market side. Each set of information related to a specific book level and market side is referred to as a data block. Prior to processing a single message, it is necessary to determine how many data blocks the message contains. Within the message, there is a sequence ID to properly order the data blocks for processing.

Once we determine the number of data blocks within the message, we reshape to transform each data block into an individual observation. For example, an order modification occurs when a trader adjusts a previously submitted limit order by changing the price and leaving quantity unchanged. If the modification is from the second-best bid (book level 2) to the best bid (book level 1), then we will see

³See the Message Specification page under CME Market Data on the CME Group Client Systems Wiki for more information.

a message containing two data blocks: one showing the decrease in visible depth at the second-best bid followed by one showing the increase in visible depth at the best bid. If a best quote is removed, then all nine quoted prices and depths will need to move positions and a new quote for the tenth level will need to be added, which results in a series of ten data blocks within a single message sorted from new information at book level 1 and proceeding to new information at book level 10.⁴

The raw message data was delivered by the exchange as weekly text files. Within each week, the messages were not sorted in any way, which meant we had to process all messages for the week, then sort them into day files, then sort each day according to the message and data block ID. For each week, we processed the large set of messages by extracting one million messages at a time, then creating subsets of messages based on the number of data blocks. For each subset of messages with the same number of data blocks, we generated empty rows for each data block (number of messages times the number of data blocks), placed the message specific information in each row, then filled in the content from each data block. Since each message has a sequence ID and each data block has a sequence ID within the message, we then sort the full table of all single data blocks.

There are specifics about futures markets that are different than equity markets. For example, within defined product such as WTI crude oil futures traded on the CME, there are multiple expirations actively traded; each expiration is con-

⁴There appears to be a set pattern for how data blocks are organized within a message. For example, if a message contains data blocks updating information on both sides of the market for multiple quote levels, one side will be listed first and the first data blocks will relate to quote levels closest to the first book level. This means if one is processing messages, one side of the book will be fully processed before the other side can be checked.

sidered a different product. Furthermore, traders can trade the calendar spread between two (or more) expirations (e.g., there is a market for the difference between the December 2013 and the December 2014 expiration) which are also considered as different products. Each message might contain information related to a number of different products. To complicate matters, the electronic limit order book for one product can interact with the electronic limit order books for other products that include that product via the concept of implied matching. This happens when the electronic matching engine determines a match between a single calendar spread and two single expiration limit orders resting in each order book. One can also build the consolidated order book that combines depth across various books. We do not do that in this research, but one needs to be aware that this information is also contained in the purchased data.

After processing all messages, we build the final table of book updates. Each row in this table contains two sort IDs (one to sort messages and another to sort data blocks within each message), a time stamp, visible depth at levels one to ten, and quoted prices at levels one to ten. If trades occur, then we have rows for each trade that include the traded price, quantity, and an indicator that informs us if the trade is buyer- or seller-initiated.⁵ Trades are always immediately followed by a sequence of book updates, which allow us to determine the exact impact of the trade. For example, if the trade removes all resting depth at the best bid, then information will immediately follow within the message containing the trade of

⁵Labeling each trade as buyer or seller-initiated is referred to in the literature as signing the trade. Historically, data sets of just trades did not include this information and therefore had to be estimated. A popular method was the Lee & Ready method. The advantage of using order book data with trades allows us to verify the signed trades are appropriately marked by comparing the trade price with the best bid and ask quotes that existed immediately prior to trade. If the trade occurs at the best bid, then the trade is marked as a seller-initiated trade. Likewise, a trade occurring at the best ask is defined as a buyer-initiated trade.

the adjustment in all book levels, which existed prior to trade, and the addition of a new book level for the tenth spot. If a trade is large enough to remove all resting depth at the first two book levels, then prior book levels 3 to 10 are moved up to 1 to 8 and new information is defined for levels 9 and 10.

Once the book is recreated from the raw message data it is possible to reduce the number of observations by restricting the number of book levels. For example, if one is simply interested in the best quotes, then we drop all information related to book levels deeper in the order book on both sides of the market. At this point, we keep only those observations with an observed change at either the price level or visible depth.

The same logic can be applied to different subsets of book levels. Along with the table of all ten book levels, we also define a table with book levels one to five. Note, this adjustment only impacts the number of observations related to the order book and therefore does not reduce the number of trade records.

3.2.2 Book Update Data

Each row in the processed message table relates to a specific event that has changed the visible depth associated with at least one quoted price level. Updates can result either from mechanics of the order book display (e.g., updates after trades) or from direct decisions of market participants. The majority of these book updates occur at timestamps unrelated to transactions — referred to as no-trade book updates — and correspond to actions related to the arrival of new limit orders or the cancellations of previously placed limit orders. Due to the method of message creation, each row in this table corresponds to a single change in the visible book.

Table 3.1 contains a sample of sequenced book updates which occur without the presence of trades. This particular example corresponds to the ask side of the market since the best quote ($p1 = 88.56$) is lower than the second best quote ($p2 = 88.57$). As we can see, each row represents a point in time where the visible electronic limit order book underwent some change so the first row represents the visible depth that lives from this timestamp to the time found in the next book update — just two milliseconds. The second message (message ID = 30025530) has three data blocks and corresponds to a decrease of visible depth by one contract at the second ask (14 to 13), a decrease of two contracts at the third ask (16 to 14), and an increase of one contract at the fourth ask (14 to 15). The next message arrives, again two milliseconds later, and increases visible depth at the best ask by one contract (4 to 5). This is followed in the next millisecond by the arrival of a new best ask — resulting in a shift of prices and corresponding visible depths. In this example, each message occurs at a unique time stamp and the majority of updates do not actually adjust quoted prices.

Table 3.1: Sample of Book Updates — No Trades

Notes: Table contains a sample of book updates on the ask side without trades, therefore, any changes result from the arrival of new limit orders or cancellations. The first three columns contain information related to the update. The next ten columns show the first five quoted price levels, where $p1$ represents the best ask quoted price level; for each price level, we show changes in red. The last five columns contain the visible depth at each quoted price level, where $q1$ represents the visible depth at $p1$.

Book Update Information			Visible Quotes by Book Level (with quote changes in red)										Visible Depth by Book Level				
Message ID	Data Block ID	Timestamp	p1	p1c	p2	p2c	p3	p3c	p4	p4c	p5	p5c	q1	q2	q3	q4	q5
30025526	8012978	8:00:13.921	88.56	0	88.57	0	88.58	0	88.59	0	88.6	0	4	14	16	14	21
30025530	8012982	8:00:13.923	88.56	0	88.57	0	88.58	0	88.59	0	88.6	0	4	13	16	14	21
30025530	8012983	8:00:13.923	88.56	0	88.57	0	88.58	0	88.59	0	88.6	0	4	13	14	14	21
30025530	8012984	8:00:13.923	88.56	0	88.57	0	88.58	0	88.59	0	88.6	0	4	13	14	15	21
30025532	8012986	8:00:13.925	88.56	0	88.57	0	88.58	0	88.59	0	88.6	0	5	13	14	15	21
30025535	8012988	8:00:13.926	88.55	-0.01	88.56	-0.01	88.57	-0.01	88.58	-0.01	88.59	-0.01	1	5	13	14	15
30025553	8012995	8:00:13.931	88.55	0	88.56	0	88.57	0	88.58	0	88.59	0	1	5	14	14	15
30025553	8012996	8:00:13.931	88.55	0	88.56	0	88.57	0	88.58	0	88.59	0	1	5	14	15	15
30025555	8012998	8:00:13.932	88.55	0	88.56	0	88.57	0	88.58	0	88.59	0	1	5	15	15	15
30025559	8013001	8:00:13.948	88.55	0	88.56	0	88.57	0	88.58	0	88.59	0	1	4	15	15	15
30025595	8013009	8:00:14.031	88.55	0	88.56	0	88.57	0	88.58	0	88.59	0	1	5	15	15	15
30025615	8013017	8:00:14.071	88.55	0	88.56	0	88.57	0	88.58	0	88.59	0	1	5	14	15	15
30025618	8013021	8:00:14.073	88.55	0	88.56	0	88.57	0	88.58	0	88.59	0	1	5	14	14	15
30025657	8013029	8:00:14.315	88.55	0	88.56	0	88.57	0	88.58	0	88.59	0	1	5	14	13	15
30025657	8013030	8:00:14.315	88.55	0	88.56	0	88.57	0	88.58	0	88.59	0	1	5	14	13	16
30025680	8013036	8:00:14.416	88.55	0	88.56	0	88.57	0	88.58	0	88.59	0	1	5	14	14	16

We can use the table of sequenced book updates to infer a number of market

events that might be of interest. For example, the arrival of a new limit order, identified by comparing two neighboring rows and looking for any increases of visible depth at the same quoted price level (regardless of position in the book). Likewise for limit order cancellations — represented as a decrease in visible depth at the same quoted price level without observed trades. Order modifications can also be found by comparing information within the same message. For example, if a single message contains a data element showing an increase at book level one along with a decrease at book level two, both for five contracts, then it is likely that this message results from a market participant modifying a previously placed order.

It is possible for some actions to not show up in the message data. This occurs if there are hidden orders. The exchange offers a number of options for traders to hide their trading intentions. One example here is an iceberg order. In this situation, the trader places a limit order, say at level two, for 100 contracts, but tells the matching engine to only show five at a time. Once these five are removed from trades, the matching engine places five more in the visible depth. Stop orders are another aspect of hidden orders, but will not be the focus of this research. While we are not able to determine the amount of hidden orders if such orders are entered and then cancelled without resulting in trades, we can observe instances of hidden depth resulting from trades by comparing visible depth before and after the trade.

We find 97 percent of book updates are unrelated to observed trades; of these, just three percent result in changes to the quoted price vector. Of those updates related to trades, approximately two thirds occur at the best quotes immediately

prior to trades and do not result in changes to the best quotes.

Table 3.2 shows book update counts related to different levels of the book for the first day of our sample. We find the majority of book updates occur at the best quotes (21 percent of book updates) or second-best quotes (19 percent); only 16 percent of all book updates occur at levels six to ten. The high concentration of book updates at levels one to five allow us to focus in later chapters on this subset of book levels closest to the market. We find the majority of traded volume occurs at the best quotes as well and the volume share decreases dramatically as we move to levels deeper in the book, which indicates that single trade events are hardly ever interacting with quotes deep in the book. The last two rows of the first panel show that it is very uncommon for book updates to change the entire price vector.

Table 3.2: Summary of Book Updates — by Book Level

Notes: Table shows a breakdown of the total number of bid updates according to if they correspond to trades or a change in the quoted price vector and by book level. For example, the first row contains book updates unrelated to trades that do not change the quoted prices and just correspond to the best quote level. Updates related to messages on 8/1/2012.

Trade	PVC	Book Level w/ Change Observed										N Updates	% Updates	
		1	2	3	4	5	6	7	8	9	10			
		1											172,021	21%
			1										148,867	19%
				1									117,762	15%
					1								102,039	13%
						1							76,771	10%
							1						51,696	6%
								1					26,286	3%
									1				21,929	3%
1		1											17,733	2%
										1			15,748	2%
											1		15,125	2%
	1	1	1	1	1	1	1	1	1	1	1	1	13,120	2%
1	1	1	1	1	1	1	1	1	1	1	1	1	6,512	1%

Focusing on the subset of no-trade book updates allows us to measure the frequency of depth removed or added conditional on the quotes' book level. Table 3.3 provides a summary of no-trade book updates that change visible depth. The

majority of cancellations correspond to the removal of a single contract near the market. As we move to levels deeper in the book, the number of observed changes decreases quickly. For example, at book level two, close to seven percent of no-trade book updates are related to a decrease of one contract compared to book level ten with just 0.4 percent of the total sample of book updates.

Table 3.3: Summary of Limit Order Arrival & Cancellations

Notes: This table contains information from bid updates that are unrelated to trades, so changes in visible depth reflect the arrival or departure of limit orders. The first panel shows counts of observed depth changes, focusing on changes between -5 and $+5$, by book level; percentages in red reflect the share of total updates unrelated to trades. Updates related to messages on 8/1/2012.

Depth Change	Book Level 1		Book Level 2		Book Level 3		Book Level 4		Book Level 5	
-5	196	0.1%	193	0.1%	60	0.0%	37	0.0%	98	0.0%
-4	508	0.2%	511	0.2%	143	0.0%	119	0.0%	80	0.0%
-3	855	0.3%	1,044	0.3%	258	0.1%	112	0.0%	64	0.0%
-2	3,603	1.2%	7,391	2.4%	2,056	0.7%	1,138	0.4%	613	0.2%
-1	40,925	13.3%	21,780	7.1%	14,234	4.6%	15,695	5.1%	10,189	3.3%
1	48,723	15.9%	23,865	7.8%	16,925	5.5%	19,632	6.4%	13,944	4.5%
2	3,413	1.1%	6,140	2.0%	2,636	0.9%	1,477	0.5%	707	0.2%
3	1,194	0.4%	1,014	0.3%	264	0.1%	113	0.0%	73	0.0%
4	869	0.3%	456	0.1%	186	0.1%	115	0.0%	100	0.0%
5	425	0.1%	214	0.1%	91	0.0%	54	0.0%	115	0.0%
	Book Level 6		Book Level 7		Book Level 8		Book level 9		Book Level 10	
-5	177	0.1%	169	0.1%	80	0.0%	33	0.0%	34	0.0%
-4	87	0.0%	83	0.0%	85	0.0%	50	0.0%	71	0.0%
-3	36	0.0%	27	0.0%	14	0.0%	18	0.0%	595	0.2%
-2	436	0.1%	321	0.1%	445	0.1%	506	0.2%	400	0.1%
-1	6,098	2.0%	2,300	0.7%	1,756	0.6%	1,165	0.4%	1,377	0.4%
1	7,905	2.6%	3,161	1.0%	2,167	0.7%	1,200	0.4%	1,478	0.5%
2	462	0.2%	404	0.1%	531	0.2%	796	0.3%	705	0.2%
3	58	0.0%	51	0.0%	21	0.0%	28	0.0%	430	0.1%
4	90	0.0%	105	0.0%	79	0.0%	57	0.0%	68	0.0%
5	299	0.1%	477	0.2%	192	0.1%	77	0.0%	57	0.0%

Studying the distance between quoted prices across all ten levels of the order book shows that all quoted prices are separated by the minimum ticks size ($\$0.01$ in crude oil futures). For example, focusing on the first day of the sample on the bid side, we find 99.22% of observations display a sequence of quoted prices with a difference between neighboring levels of the minimum tick size. The next most common sequence, with 0.31%, has a difference of two-ticks between the best and second-best quote and a one-tick difference separating the remaining book levels. This results when the best quote moves toward the market before the rest of

the book can adjust. The third most common sequence, with 0.09%, shows the best and second-best quotes with a one-tick difference, but a two-tick difference between the second and third best quote — which would follow as the second-best quote is the first to respond to a move in the best quote.

3.2.3 Transactions Data

Besides updates to the limit order book, the raw messages contain the complete record of transactions. We are able to correctly sequence the trade records with book update messages to recreate the limit order book immediately before and after transaction information. For example, if we observe a transaction for one futures contract at \$88.50 and, prior to trade, the best bid is also \$88.50 with a depth of 10, then we know the trade was initiated by a seller and we expect an update immediately following the trade to revise the depth at the best bid down to 9.

Trades are initiated when a market participant decides to submit an order that GLOBEX is able to match immediately with another market participant. The trader who submits the initiating order is often referred to as the aggressive side since their decision results in ‘crossing the spread’ — submitting an order to buy that trades at the best ask quote (or higher) or submitting an order to sell that trades at the best bid quote (or lower). Crossing the spread is often considered an action that extracts or consumes liquidity from the book.⁶ Buyer-initiated trades result from aggressive orders to buy entering the book and interacting with the best ask quote (extract liquidity from the sell side); seller-initiated trades result

⁶There is some disagreement here since practitioners view crossing the spread to ‘aggressively’ buy in a downward moving market as supplying liquidity.

from aggressive orders to sell entering the book and interacting with the best bid quote (extract liquidity from the buy side).

Volume represents the total quantity of futures contracts traded during some interval of time, often displayed in daily frequencies, although it is now common to see intraday volume totals by hour or lower frequencies. Volume is an important metric since this is often considered a proxy for information flows. Volume can be different from the number of trades depending on the average size of each trade. The data contains the most precise record of individual trades available for public use as other sources fail to include milliseconds.

Trades can also result from the submission of a limit order if the limit order price to buy is submitted at the best ask quote. This type of order guarantees trades at a single price. If the order quantity is larger than the total depth, then the order results in a trade for the total depth, then converts into a resting limit order to buy at this same quoted price level. The use of a marketable limit order, as opposed to a traditional market order, allows the market participant precise control over price slippage.

Figure 3.2 shows the close connection between order book updates and traded volume. The first row contains daily total book update count (in blue) and volume (in red) by hour and minute intervals. The full day shows the majority of volume occurs between 8am and 3pm. Focusing on the one minute intervals (subfigure 3.2(b)) shows spikes of increased book updates and traded volume starting at 8:30am and then occurring each half hour. The afternoon calms down until the large spike right before 2:30pm, which corresponds to the daily settlement period for this futures contract. The second row shows minute intervals zoomed in around

10:30am, when there is a major news event each Wednesday, and then around 2:30pm.

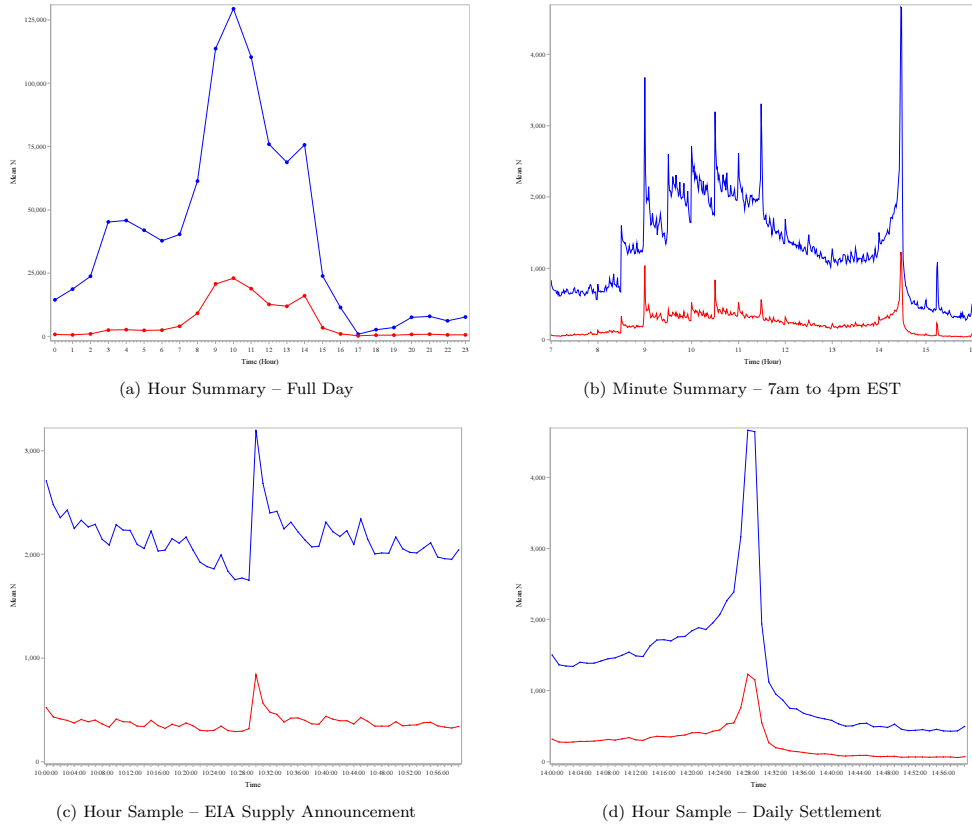


Figure 3.2: Quote and Volume Compare

Notes: Figure contains four subfigures which highlight the relationship between book updates (blue) and realized volume (red); for updates, just the bid side is shown here. Subfigure (a) shows hourly averages. Subfigure (b) shows averages by minute. Subfigures (c) and (d) show minute averages for the hour surrounding 10:30am — when the EIA Supply Announcement is released — and the hour surrounding the daily settlement, respectively.

Table 3.4 contains rows of book updates, corresponding to the ask side of the book, that includes a number of trades. The set of three trades arrive just one millisecond after the prior book update message. Comparing the traded price of these buyer-initiated trades with the prior best quoted prices, we find the first two trades have a total quantity of six and extracted the visible depth of six at the best quote. The last trade occurs at the second ask quote. These three trades are followed, within the same message, by a book update that reflects the adjusted ask quote prices and new visible depth of the new best ask. The book is unchanged

over the next eight milliseconds, until the arrival of another set of buyer-initiated trades that remove two contracts from the best ask quote.

Table 3.4: Sample of Book Updates — with Trades

Notes: Table contains the same set of columns as discussed in Table 3.1, but now contains two additional columns related to the transaction price and quantity. Comparing the arrival of book updates that reflect how the trades change visible depth, the table shows two groups of trades and a single trade. Since the book information relates to the ask side, these are buyer-initiated trades.

Book Update Information				Visible Quotes by Book Level (with quote changes in red)										Visible Depth by Book Level					Trade Activity	
Message ID	Data Block ID	Timestamp	p1	p1c	p2	p2c	p3	p3c	p4	p4c	p5	p5c	q1	q2	q3	q4	q5	Price	Qty	
30024645	8012728	8:00:10.948	88.54	0	88.55	0	88.56	0	88.57	0	88.58	0	6	6	8	19	16			
30024645	8012729	8:00:10.948	88.54	0	88.55	0	88.56	0	88.57	0	88.58	0	6	6	8	18	16			
30024645	8012730	8:00:10.948	88.54	0	88.55	0	88.56	0	88.57	0	88.58	0	6	6	8	18	17			
30024657	8012731	8:00:10.949																88.54	1	
30024657	8012732	8:00:10.949																88.54	5	
30024657	8012733	8:00:10.949																88.55	1	
30024657	8012739	8:00:10.949	88.55	0.01	88.56	0.01	88.57	0.01	88.58	0.01	88.59	0.01	5	6	18	17	12			
30024681	8012741	8:00:10.957																88.55	1	
30024681	8012742	8:00:10.957																88.55	1	
30024681	8012751	8:00:10.957	88.55	0	88.56	0	88.57	0	88.58	0	88.59	0	3	6	18	17	12			
30024681	8012752	8:00:10.957	88.55	0	88.56	0	88.57	0	88.58	0	88.59	0	3	6	17	17	12			
30024689	8012763	8:00:10.966	88.56	0.01	88.57	0.01	88.58	0.01	88.59	0.01	88.6	0.01	5	14	17	13	20			
30024712	8012768	8:00:10.974																88.56	1	
30024714	8012773	8:00:10.974	88.56	0	88.57	0	88.58	0	88.59	0	88.6	0	4	14	17	13	20			
30024714	8012774	8:00:10.974	88.56	0	88.57	0	88.58	0	88.59	0	88.6	0	4	13	17	13	20			
30024714	8012775	8:00:10.974	88.56	0	88.57	0	88.58	0	88.59	0	88.6	0	4	13	16	13	20			

Along with the sign of the trade, we distinguish between trades that kill a quote from those trades that simply extract a share of resting depth without resulting in a change in the best quotes. For example, if resting depth at the best bid and ask is ten and one, respectively, then a buyer-initiated trade for one contract kills the best ask quote resulting in the second-best ask promotion to the best ask and a larger difference between the best bid and new best ask quote. Conversely, a seller-initiated trade for one contract results in a change of depth from ten to nine — the trade has no impact on the best bid quote level. This is important to note since the former trade has a price impact while the latter trade does not. Therefore, the presence of a price impact is a function of resting depth at the best quotes.

In general, we see an immediate book update following the trade records as the book updates itself based on how the trades impact visible depth at each level. The number of trade records is a function of the number of market orders

submitted at the same point in time and the number of individual orders resting at the best quotes. For example, if a single market order to buy five contracts arrives and there are three distinct orders resting at the best ask for a total depth of five, then we will observe a group of three trade messages. We define trade groups as sets of individual trade records uninterrupted by book updates.⁷

It is possible for a trade group to include multiple traded price levels. This occurs when a market order is larger than the available depth at the best quote. From the data, we can determine exactly when this occurs by comparing each trade price to the prior best quotes. Buyer-initiated trades can then take place at the best ask quote level or at higher quote levels. Similarly, seller-initiated trades can take place at the best bid quote level or at lower quote levels.

Table 3.5 provides a summary of trades and quantity for both individual trades and trade groups for all transactions during the first two months of our sample. There are 4.2 million trade records, 1.9 million trade groups, and a total volume of 5.3 million futures contracts during August and September of 2012. To build this table, we collect all trade records as well as limit order book updates immediately before and after each trade record. Using book updates, we define trade groups and any sequence of trade records uninterrupted with a book update. For each trade record, we compare the trade price with both best quotes from the limit order book prior to trade and determine the price impact by taking the difference between prior best quote and the trade price. For example, if the trade price is equal to the best bid, then the price impact is zero, but if the trade price is below the best bid, then we will find a negative price impact.

⁷The ability to classify trades into trade groups can be used to study the role of trades as a proxy for information flows — especially when many trades contain the same timestamp.

The upper portion of Table 3.5 shows a breakdown of total trades, quantity, and trade groups by quantity of the individual trade or aggregate quantity of the trade group. We find 88% of individual trades are for one contract which represents 71% of volume. Furthermore, it is very uncommon to observe trades with more than five contracts, which tells us something about resting orders and/or marketable orders — the vast majority of these are for one contract. Just based on individual trades, it is not possible to say which is the driving force here.

Trade groups do tell us something about the difference between resting orders and marketable orders. Based on upper right table, we find the share of trades and total quantity is much less when quantity is equal to one. Trade groups are more likely to have an aggregate size larger than one. In fact, we can see that 1/4 of trades are found in trade groups with a total quantity greater than 10 contracts. Based on the comparison between individual trades and trade groups, and assuming it is not common for marketable orders to arrive at the same exact millisecond, we can state that resting limit orders are most often a small size — close to one contract in our market during the sample period. These conclusions would not be so clear if one only had trades marked to the second — we would need a lot of assumptions to draw similar conclusions based on the comparison of individual trades and trade group information.

The lower portion of Table 3.5 shows a breakdown of total trades, quantity, and trade groups by observed price impact. Regarding individual trades, we find 82% of total trades and quantity have no price impact — this includes individual trades that extract a share of visible depth as well as those trades that kill a best quote. We find 12% of total trades and quantity are linked to individual trades

with a one-tick price impact. This occurs when an arriving trade is for more than what is available at the best quote and the marketable order allows for trades away from the best quotes. Based on individual trades, close to 100% of the sample has a three-tick or less price impact. This evidence shows how it is very uncommon to see aggressive trades at quotes 4 or more ticks from the best quotes.

While the majority of trade groups have no price impact, we find a larger share of total trade count and volume occurs with a price impact. Where there is no price impact, total trade count and quantity are lower in trade groups because it is possible for a portion of trades within a trade group to show no price impact. We find close to 96% of trade groups have no price impact — based on size of trade groups, this implies that it is likely to see trade groups with size occur without an observed price impact. As expected, we find a clear linear positive relationship between size of the trade group and average expected price impact (using an absolute value of price impact).

In future work, we only use trade group information which will be associated with book update information that reflects the change in visible depth immediately following the trade group. This ensures for the exact millisecond the trade is recorded, we know exactly how the visible book is changed immediately following the trade. We make sure and note if a trade group has volume that is related to quotes either above the best ask quote or below the best bid quote. We also keep track of the trade count for each trade group. Comparing the total quantity of the trade group with the limit order book that existed just prior to trade allows us to determine if the trade extracts a portion of available depth or kills the best quote.

Table 3.5: Summary of Trades and Trade Groups

Notes: This table shows information related to all trades during August and September 2012, the first two months of our sample. The upper table shows trade count and total quantity (Qty) by size; the lower table shows the same information by price impact. Both show information specific to individual trade records, on the left, and information related to trade groups, on the right. Trade groups are defined as sequences of trade records uninterrupted by book updates. Price impact is defined as the distance from the prior best quote of the side providing liquidity; for trade groups, we define the price impact as the smallest (for seller-initiated) or largest (for buyer-initiated) observed difference among trades in the group.

Quantity	Individual Trades				Trade Groups					
	N Trades		Total Qty		N Trade Groups		N Trades		Total Qty	
1	3,722,840	88%	3,722,840	71%	1,075,570	57%	1,075,570	25%	1,075,570	20%
2	330,193	8%	660,386	13%	327,909	17%	581,520	14%	655,818	12%
3	82,794	2%	248,382	5%	140,544	7%	352,180	8%	421,632	8%
4	47,819	1%	191,276	4%	97,464	5%	305,749	7%	389,856	7%
5	25,766	1%	128,830	2%	72,644	4%	276,502	7%	363,220	7%
6	9,025	0%	54,150	1%	36,214	2%	161,471	4%	217,284	4%
7	4,347	0%	30,429	1%	21,347	1%	111,071	3%	149,429	3%
8	3,993	0%	31,944	1%	17,013	1%	96,655	2%	136,104	3%
9	2,670	0%	24,030	0%	12,183	1%	77,239	2%	109,647	2%
10	5,661	0%	56,610	1%	23,160	1%	158,159	4%	231,600	4%
gt 10	6,576	0%	123,885	2%	60,217	3%	1,045,568	25%	1,522,602	29%
Total	4,241,684		5,272,762		1,884,265					

Price Impact	Individual Trades				Trade Groups					
	N Trades		Total Qty		N Trade Groups		N Trades		Total Qty	
lt -0.1	2,794	0%	4,992	0%	53	0.0%	7,384	0.2%	12,406	0.2%
-0.1	651	0%	1,074	0%	15	0.0%	1,701	0.0%	2,512	0.0%
-0.09	847	0%	1,411	0%	29	0.0%	3,428	0.1%	4,574	0.1%
-0.08	1,694	0%	2,510	0%	46	0.0%	5,687	0.1%	7,739	0.1%
-0.07	2,281	0%	3,313	0%	56	0.0%	5,143	0.1%	6,930	0.1%
-0.06	3,339	0%	4,553	0%	104	0.0%	7,852	0.2%	10,260	0.2%
-0.05	5,869	0%	8,360	0%	176	0.0%	11,921	0.3%	15,688	0.3%
-0.04	11,622	0%	15,795	0%	454	0.0%	27,711	0.7%	35,928	0.7%
-0.03	27,401	1%	35,012	1%	1,247	0.1%	53,040	1.3%	66,949	1.3%
-0.02	78,928	2%	99,033	2%	4,894	0.3%	121,395	2.9%	153,671	2.9%
-0.01	251,667	6%	315,397	6%	35,512	1.9%	298,344	7.0%	373,382	7.1%
0	3,490,407	82%	4,308,198	82%	1,798,589	95.5%	3,178,823	74.9%	3,912,200	74.2%
0.01	243,583	6%	311,438	6%	36,106	1.9%	298,708	7.0%	378,050	7.2%
0.02	69,882	2%	91,132	2%	4,322	0.2%	108,914	2.6%	140,747	2.7%
0.03	24,725	1%	32,632	1%	1,064	0.1%	44,719	1.1%	58,955	1.1%
0.04	11,074	0%	15,173	0%	449	0.0%	26,192	0.6%	35,001	0.7%
0.05	5,353	0%	7,503	0%	169	0.0%	12,073	0.3%	16,489	0.3%
0.06	2,997	0%	4,361	0%	97	0.0%	7,845	0.2%	10,965	0.2%
0.07	1,776	0%	2,902	0%	53	0.0%	4,418	0.1%	6,299	0.1%
0.08	1,022	0%	1,616	0%	36	0.0%	3,568	0.1%	5,300	0.1%
0.09	772	0%	1,427	0%	24	0.0%	2,992	0.1%	4,487	0.1%
0.1	456	0%	824	0%	12	0.0%	1,092	0.0%	1,535	0.0%
gt 0.1	2,544	0%	4,106	0%	43	0.0%	6,306	0.1%	9,882	0.2%

3.2.4 Processed Data Adjustment — Reshape Book

At this point we have processed the raw messages by breaking down each message into the individual pieces of information and then properly sorting to rebuild the limit order book for each side of the market. This processed table is divided into three sets of information: the order book related to the buy side, the order book related the sell side, and signed trades. As the examples above show, a single

message can produce multiple rows in this final table, which can be related to all three sets. When there are multiple rows related to the same message ID for a single market side, they are ordered in a sequence that retains the prior update. This allows us to only focus on the last observation for each message ID.

Once the raw message data is processed and we have rebuilt the order book, we can extract information just related to the best quotes — the bid and ask quote each at the first level of the book. The best bid is defined as the quoted price associated with a resting order to buy with the highest limit price among all other resting orders to buy. Similarly, the best ask is defined as the quoted price associated with the resting order to sell with the lowest limit price among all other resting orders to sell. The market depth associated with these best quotes reflect the quantity available to either buy or sell at these quotes.

To extract just the best quote information, we simply keep just the information at the first book level for both sides of the market. This will allow us to keep only unique timestamps that contain a change in the best quote on either side of the market. A change can take the form of a price change or a change in visible depth. Along with the limit order book related to just the best quotes, we also can select the subset of trade records with prices equal to the best quote displayed in the market immediately prior to the trade arrival.

At this point, we have rebuilt the order book and have information related to the best quotes along with trades that interact with the best quotes. But there exists more information related to book levels deeper in the limit order book. To incorporate this information, we add snapshots of visible depth at levels deeper in the book, where the snapshot time reflects some change, either quote level or

depth change, in one or both of the best quotes.

To do this, we extract a list of unique times from our table of best quote information. We then find the last book update for each unique timestamp and then add this information into our best quote table. Now for each observation, reflecting some change at the best quotes, we have corresponding depth at this millisecond for book levels 2 through 10.

3.3 Analysis of Best Quotes

First, in subsection 3.3.1, we define terms related to movement in the best quotes. According to evidence shown in the prior section, we know a very small share of book updates actually change the quoted price vector. Based on this, we argue that a subset of book updates are meaningful based on how they change the information set and reflect movement on one or both sides of the market — movement that we can define, according to the direction, as movement toward the market or away from the market.

Second, in subsection 3.3.2, we analyze the bid-ask spread. This is an important measure of any order driven market since it is often used as a measure of market liquidity. One of the advantages of using this source of information is the ability to carefully track the bid-ask spread and collect detailed information related to why the spread changes. We highlight the advantage of monitoring the limit order book with a number of examples which progressively add information from the limit order book to show how expected patterns emerge.

3.3.1 Quote Improvements & Promotions

Figure 3.3 contains a conceptual example of movement in the best quotes — the defined prices reflect available price levels. At t_0 , the market is displaying a one-tick bid-ask spread (\$0.01 for crude oil futures) with the best ask quote at $Price_1$ (in red) and the best bid quote at $Price_0$ (in blue). Based on analysis of spacing between quoted price levels, we can expect the remaining visible quotes two through ten are spaced out with one-tick between each quote (e.g., the second best ask quote, at t_0 , is at $Price_2$ and the second best bid quote is at $Price_{-1}$). At t_1 , an event occurs that causes the best ask quote to move upward away from the best bid quote; the movement indicates a shift in all visible quotes on the ask side where the prior second-best ask quote level is now the first-best ask quote level — a **promotion** has just occurred on the sell side. Starting at t_1 , the market is now displaying a two-tick bid-ask spread, which exists until t_2 at which point a new limit order to buy arrives at the open quote level. The arrival of a new best bid quote represents an **improvement** in the quoted level on the buy side.

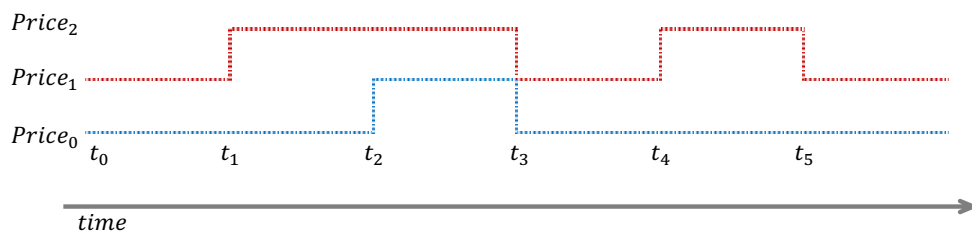


Figure 3.3: Movement in the Best Quotes

Notes: This figure shows a hypothetical sequence of book updates creating movement in the best quotes, defined across three possible prices. The best ask quote is shown in red; the best bid quote is shown in blue.

We define two terms based on possible movement up or down for each best quote:

- A **Quote Improvement (QI)** is defined as an event where a new book update creates a new best quote and the prior best quote becomes the current second-best quote.
- A **Quote Promotion (QP)** is defined as an event where a book update removes the current best quote and the prior second-best quote becomes the current best quote.

Completing the discussion of Figure 3.3, between t_0 and t_3 , the market has moved up by one-tick. This movement is accomplished with an initial movement of the best ask, then followed after an interval of time by a movement in the best bid quote. At t_3 , however, we find a directional movement can occur at the same moment in time in both quotes. Here, the best bid quote is removed and simultaneously replaced with a new best ask quote. This occurs when a marketable limit order to sell at the best bid is for a larger quantity than is resting at the best bid quote. We denote the type of changes in the best quotes with notation such as $(bid - QI, ask - NC)$ at t_2 and $(bid - QP, ask - QI)$ at t_3 . The activity following t_3 shows a sequence $(bid - NC, ask - QP)$ at t_4 followed by $(bid - NC, ask - QI)$ at t_5 . In this case, only the best ask quote is observed to change while there is no change in the best bid quote.

We will pay close attention to QI and QP events throughout the remaining chapters of this dissertation for two reasons. First, we have shown that the majority of book updates do not, in fact, change the visible quote vector on either side of the market and, therefore, these events represent a change in the visible book that contains potentially more information than a non-quote changing book update. Second, these events are meaningful changes that indicate a market side's

willingness to move toward or away from the market.

While a QI event can only occur with the arrival of a new limit order, a QP event can occur due to trades or from limit order cancellations. In either situation, a QP event occurs when the current best quote, on one side of the market, experiences an event that pushes depth to zero. A bid-QP can either happen from seller-initiated trades extracting all available depth, or if participants with resting bids at the best quote decide to cancel them. In both cases, the current second-best bid, existing prior to the event, is promoted to the best bid and prior book levels two through ten become levels one through nine and a new quoted depth level is included as level ten.⁸

A QP event is more complicated given the two different types of events that create the QP event. Buyer-initiated trades create the ask-QP which widens the spread and is then soon followed by the bid-QI. Without further analysis, it is difficult to associate this with the ask side moving away from the market. But if the Ask-QP was instead associated with resting ask limit order cancellations, then it would be a direct sign of the ask side moving away from the market. In fact, the expectation of the next event changes once we identify the reason for the QP.

A QI event should always be interpreted as a bullish move on the associated side of the market. A bid-QI indicates the buy side is willing to improve the market, and therefore, should be interpreted by anyone (or anything) watching the market update in real time, as an aggressive move by the buy side. Similar logic applies to an ask-QI — the market observers should take this to mean the sell side is aggressively moving toward the market. The market side that caused

⁸It is possible that a QP event results in a level deeper than two becoming promoted to the best quote. Consider a case where depth at the best and second-best is equal to one. If a trade for two arrives, then the QP results in the third-best being promoted to the best quote.

the wide spread is more likely to be related to the soon to follow QI event (if a buyer-initiated trade widens the spread, then we expect a Bid-QI to soon follow causing the spread to decrease; the bid quote vector moves upward toward the new market level).

During an interval where best quotes are monotonically pushing higher, it makes sense to observe a sequence of an ask-QP event followed by bid-QI event — a sequence such as ask-QP, bid-QI, ask-QP, bid-QI would represent a market moving higher. Similarly, when best quotes are monotonically moving downward, it makes sense to observe a sequence of a bid-QP event followed closely by ask-QI event.

Including trade information, buyer-initiated trades, attacking the best ask quote and extracting available liquidity, should be associated with a soon to follow ask-QP event as the depth at the current best quote is eventually fully removed. If this is not the case, then either the available depth was large, or the sell side continues to replenish liquidity. If the buying interest (representing an upward pressure on the market) is strong enough and eventually overtakes the sell sides willingness to supply liquidity at that quote level, then an ask-QP will result. If there still exists buying interest, then it is reasonable to expect a bid-QI to soon follow. If the buyer is willing to wait, then it is best to place a resting new best bid than cross the spread and trade at the now higher ask quote level.

Table 3.6 contains a summary of QI and QP events related to the first two months of our sample. We build the first table by collecting all trade groups and, for each group, assigning the resulting event based on changes in the best quotes due to trades. For each trade group that doesn't change the best quotes, we

determine its proximity to the prior and future event; we then group these trade groups based on proximity using 50 and 100 millisecond thresholds. The purpose of this upper table is to investigate where volume lives relative to observed QI or QP events.

We find 44% of volume at the best quotes is linked to a QI or QP event on either side of the market. We find all volume that occurs away from the best quotes is linked to a QI or QP event. Of the 56% of volume at the best quotes unrelated to an event, we find significant volume occurs just prior to an event — 12% of volume at the best quotes happens within 50 milliseconds of a future event and greater than 100 milliseconds from the prior event — this suggests trades tend to cluster just prior to an event. The share drops for trades that follow an event. We do find a similar percentage linked to trades that are both within 50 milliseconds from the prior event and future event. Studying the share of volume between 51 and 100 milliseconds from the past and future event shows the clustering of trades takes place over a very short interval of time (within 50 milliseconds). We find approximately 1/4 of volume at the best quotes happens at least 100 milliseconds before a future event — arrival of a trade that simply removes a fraction of available depth and leaves the best quotes unchanged. We find similar shares for either buyer-initiated or seller-initiated.

The lower portion of Table 3.6 is based on the sample of all QI or QP events. Here, we allow for events to happen on one or both sides of the market. For each event, we determine if the event returns the market to a minimum bid-ask spread and if the event has at least one trade (some low probability event pairings here have been removed, for example, a QI event on both sides of the market). The

first two rows, shows the majority of events are a single QI event which returns the market to a minimum bid-ask spread without trades — approximately 40% of the sample; this is due to new limit order arrivals that change the best quotes. Rows 3 and 4 shows QP events linked to trades. These events result from the arrival of a marketable limit order that kills the best quote and results in a bid-ask spread larger than the minimum allowable value. We find 2/3 of volume at the best quotes and away from the best quotes is related to these events. Rows 5 and 6 again show QP events, but unrelated to trades. These events reflect limit order cancellations that adjust the bid-ask spread — the cancellation opens up the bid-ask spread. A large share of events are unrelated to trades (85% of either a QI or QP events are on one side of the market; but only 25% are linked to trades).

Table 3.6: QI QP Event Summary

Notes: This table shows information related to all book updates during August and September 2012, the first two months of our sample. The upper table shows volume shares according to proximity to a QI or QP event for a defined time interval prior to and/or following the event; the lower table shows frequency of QI or QP events on both sides of the market conditional on the if the bid-ask spread is equal to the minimum tick size (Min BAS) and presence of a trade.

QI or QP Event	Prior	Future	Time Interval	SI Qty Best Quotes	SI Qty Deeper	BI Qty Best Quotes	BI Qty Deeper
1			0 ms	44%	99%	43%	99%
0		1	0 - 50 ms	12%		13%	
0	1		0 - 50 ms	5%		5%	
0	1	1	0 - 50 ms	5%		5%	
	1	1	51 - 100 ms	3%		3%	
0	1		51 - 100 ms	2%		2%	
0		1	51 - 100 ms	2%		2%	
0	1	1	gt 100 ms	26%		26%	

Buy Side	Sell Side	Min BAS	Trade	Frequency	SI Qty Best Quotes	SI Qty Deeper	BI Qty Best Quotes	BI Qty Deeper
	QI	1		20.9%				
QI		1		20.8%				
	QP		1	12.4%	1%	0%	68%	67%
QP			1	12.2%	68%	68%	1%	0%
QP				10.5%				
	QP			10.4%				
QP	QI	1	1	3.9%	26%	21%	0%	0%
QI	QP	1	1	3.9%	0%	0%	26%	21%
QI				1.5%				
	QI			1.5%				
QI		1	1	0.5%	0%	0%	2%	0%
	QI	1	1	0.5%	2%	0%	0%	0%
QI	QP		1	0.2%	0%	0%	2%	11%
QP	QI		1	0.2%	1%	10%	0%	0%

3.3.2 Bid-Ask Spread Analysis

The bid-ask spread is often used in market microstructure research. Historically, the bid-ask spread was used to model the decision process of a dealer or market maker. The placement of the best quotes, and therefore the spread, calculated as the difference between the best bid and ask quote, is thought to reflect an outcome of the market makers desire to offer the best quotes but also used to protect against informed traders.

The modern market place with algorithmic and high speed traders fundamentally alters the market and participants whose combined decisions determine the market's bid-ask spread. Recent studies have used the measured bid-ask spread to show how markets have improved with the movement to a fully electronic limit order book.

We present a study of the bid-ask spread, as observed during our sample, to get a sense of how a market participant might make use of the information available from the limit order book. We show how flows of orders change depth, which then help us see where the market is going. We show how the processed data can make sense of observed sequences of book updates. We also know exactly when the bid-ask spread changes based on the arrival of book updates that create a quote promotion or improvement.

Summary of Observed Bid-Ask Spread:

Figure 3.4 shows two examples of the observed bid-ask spread over the course of five seconds. Futures markets have a defined minimum tick size, which is set by the exchange, and for crude oil futures, this minimum tick size is \$0.01. The first

subfigure here shows the market where the bid-ask spread is equal to the minimum tick size. We observe a series of seller-initiated trades hit the best bid (denoted as circles where the size reflects the traded quantity) followed by trades that lift quantity from the best offer. The second subfigure shows how trades can change the bid-ask spread. Here we see three seller-initiated trades push the best bid quote down, indicating that the trades removed all resting limit orders at the best and second-best bid quotes which existed prior to trades. The example shows how the spread then continues to widen as buyer-initiated trades push the best ask quote higher.

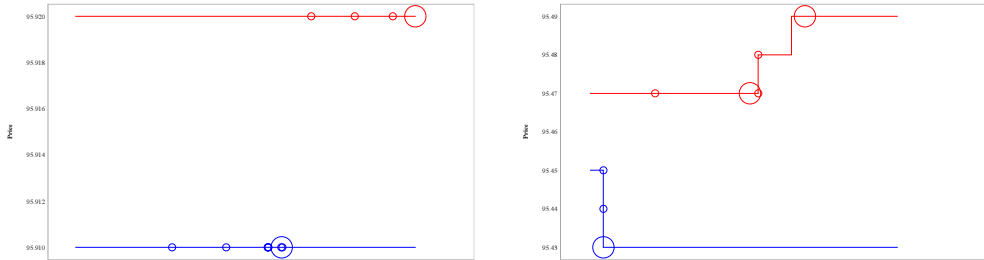


Figure 3.4: Bid-Ask Spread (BAS) Examples

Notes: Figure contains two subfigures which show the best bid (blue) and ask (red) quotes combined with trades. Best quotes are represented as horizontal lines; trades are denoted by circles. The first figure, on the left, shows a constant BAS of one tick — \$0.01 for crude oil futures. The second figure, on the right, shows a widening BAS resulting from trades.

Since buyer-initiated trades were the most recent activity, we might expect other market participants to place new bid limit orders inside this spread. Observing these trades given a wide bid-ask spread also raises a number of questions, including: Why did these market participants decide to lift the offer when there was the option to place a resting bid limit order inside the spread? This is related to the execution risk associated with each non-marketable limit order.

Figure 3.5 shows that the majority of volume occurs when the bid-ask spread is equal to the minimum tick size. We do observe some days during the sample where

the volume associated with a two-tick bid ask spread is noticeably higher. During the day, there are clear drops in the volume share during the minutes surrounding 8:30am, 10am, and 10:30am. The bid-ask spread is affected by periods where expected news is released.

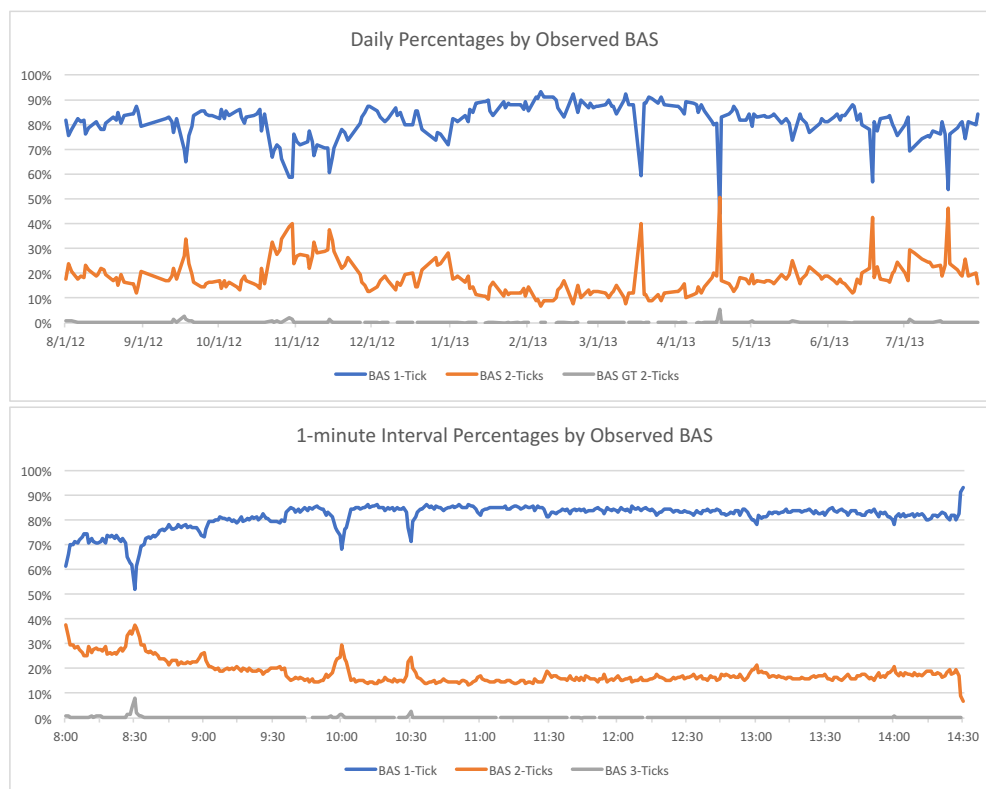


Figure 3.5: Volume Shares Conditional on Bid-Ask Spread: Daily and Intraday

Notes: Figure contains two subfigures which show volume shares conditional on a minimum bid-ask spread (blue; BAS 1-Tick), a two tick bid-ask spread (orange; BAS 2-Ticks) or greater than two ticks (grey; BAS GT 2-Ticks). The top subfigure shows how the daily volume percentages change across days in the sample; the bottom subfigure shows intraday variation based on one minute intervals.

When the bid-ask spread does open up to something larger, how long does it take to return back to the minimum tick size? We address this question by scanning the high frequency data and marking when the spread is larger than one cent and then capturing the required time until the spread equals one cent. Figure 3.6 provides a summary of the duration observed between a wide bid-ask spread and the eventual return to a one-tick spread during the first month of our

sample. We find that the count and expected duration of wide bid-ask spread measurements do exhibit strong intraday patterns. For bid-ask spreads larger than two ticks, the intraday pattern changes. As the spread increases, we find the observations cluster within minutes just after major expected news events. This figure shows expected durations of a wide bid-ask spread all are less than one second, with the majority less than 100 milliseconds.

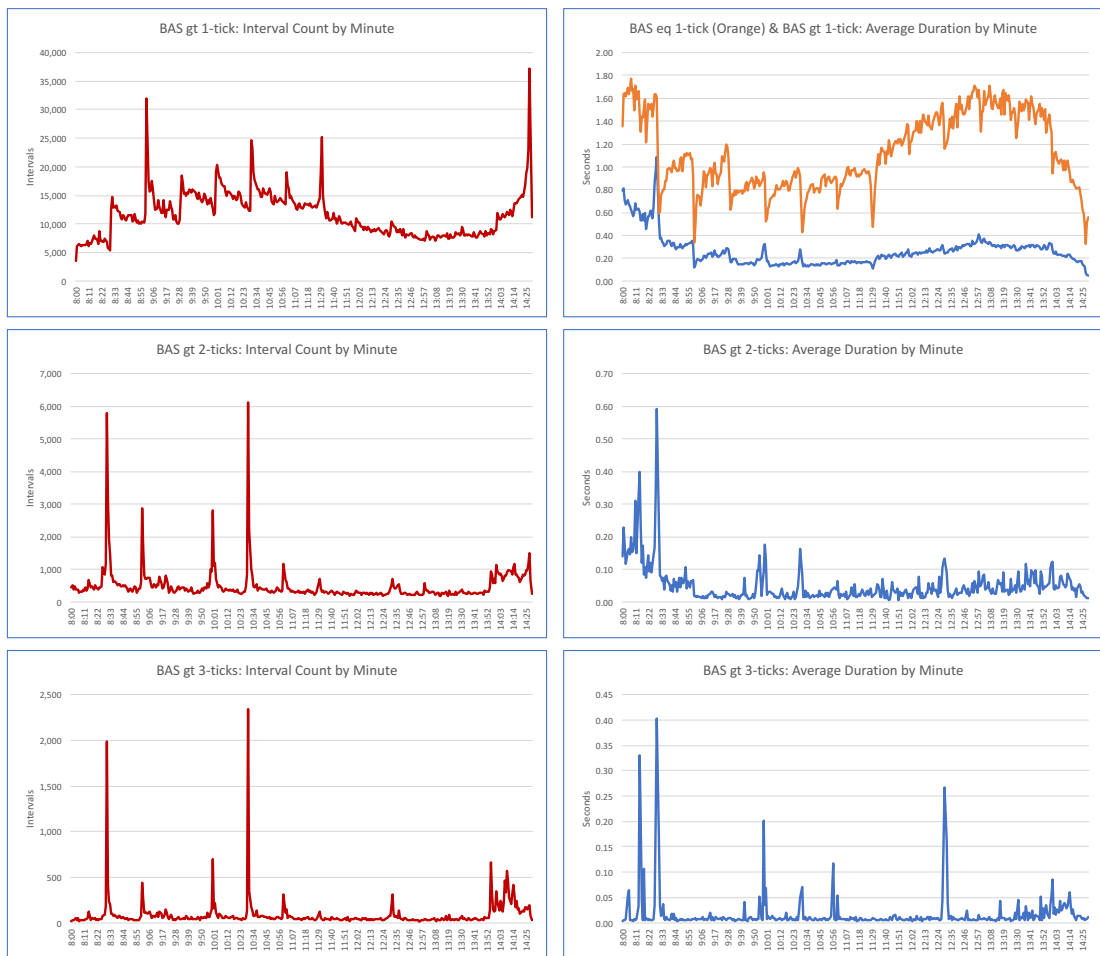


Figure 3.6: Duration of Wide Bid-Ask Spreads

Notes: This figure shows counts, on the left, and average duration, on the right, of one minute intervals conditional on the observed bid-ask spread. The first row contains information for bid-ask spreads larger than one-tick; the orange series plot captures the duration of bid-ask spreads equal to one-tick. The second and third rows summarize events where the bid-ask spread is larger than two-ticks and three-ticks, respectively.

Use of Book Updates to Explain Spread:

The findings from Figure 3.6 indicate there are many opportunities to observe a wide bid-ask spread. With the market depth data, we can make use of the surrounding book updates to identify the action that causes a change in the spread along with future updates that return the market to a one-tick bid-ask spread.

Table 3.7 presents the analysis and shows the sample shares where the bid-ask spread initially widens and then returns to a minimum bid-ask spread. For each observation of a wide bid-ask spread, we collect information regarding the initial event, changes in depth on both sides of the market during the wide spread, and the eventual event that returns the market to a one-tick bid-ask spread.

There are four possible sequences that start and end a wide bid-ask spread. These include: (1) ask-QP followed by bid-QI; (2) bid-QP followed by ask-QI; (3) ask-QP followed by ask-QI; and (4) bid-QP followed by bid-QI. We find observations are equally divided into one of these types of movement. The first two result in a movement in the market, while the last two are actions specific to a single market side and therefore results in no actual movement in the market.

Regarding total sample shares, we find the largest share (11%) is linked to the following: starting bid-QP event that widens the bid-ask spread, followed by depth flowing out on the new best bid and depth flowing in at the best ask, ending in an ask-QI event. Similar story and share linked to an ask-QP followed by bid-QI event sequence where depth is flowing into the best bid and out of the best ask. The next largest share (5%) is associated with event sequences that do not move the market up or down, but instead temporary widen the bid-ask spread on one side, followed by a same side QI event that returns the market to the prior

best quotes with a minimum tick size. This is associated with the opposite side building depth during the wide bid-ask spread. One side appears to back away from the market, but eventually someone moves to return the market back to a one-tick spread.

The next largest share (4%) is linked to directional movement in the best quotes (e.g., a downward movement due to a Bid-QP followed by Ask-QI sequence). This occurs when there is a single side flow out of depth on the side associated with the initial QP event. For example, the bid-ask spread is the minimum tick size, then a Bid-QP event occurs, we then observe depth flowing out of the new best bid and no change in depth at the best ask. (Slightly less here at 3.7%, is a similar sequence of events, but no change in depth on either side — No Flow). Comparing the first to the third, we find a sizable difference between when there is observed changes in depth on both sides (opposite flows) vs on a single side.

The percentage related to shares within each DM allows us to ask what are the expected chances of observing the second book event conditional on the initial event and depth movement on both sides of the market. For example, for the first row, we can see these events have a flow out of depth on the best bid and flow in of depth at the best ask. We find 60% of event sequences are bid-QP followed by ask-QI. We can see it is very unlikely to observe this kind of movement in depth during a wide bid-ask spread that is associated with an ask-QP starting event. The shares show certain types of movement can be associated with each of the four groups: consider the case of no movement on either side of the market. While the majority are associated with same side sequences, there is a significant share of events related to sequences that result in a market move.

Table 3.7: Using Depth Movement to Explain Wide Bid-Ask Spreads

Notes: This table is based on a sample of events from August and September 2012 — the first two months of the sample. An event is defined by a change in the best quotes that creates a wide bid-ask spread followed by an event that returns the market to the minimum bid-ask spread. We divide these events into four groups according to the start and ending best quote movement: [1] Bid-QP; Ask-QI, [2] Ask-QP; Ask-QI, [3] Bid-QP; Bid-QI, and [4] Ask-QP; Bid-QI. For each event, we note depth changes on both sides of the market, following the starting movement, and then assign each event to a group based on flow in or out of depth on both sides of the market. The table then shows event frequency (Total Sample %) according to movement in the best quotes and visible depth. For each type of depth movement (DM), we also show how events are distributed across the four groups (DM %).

Depth Movement (DM)	Bid-QP; Ask-QI (Down Move)		Ask-QP; Ask-QI (No Change)		Bid-QP; Bid-QI (No Change)		Ask-QP; Bid-QI (Up Move)	
	Total Sample %	DM %	Total Sample %	DM %	Total Sample %	DM %	Total Sample %	DM %
Bid out, Ask in	10.9%	59%	1.4%	8%	5.2%	28%	0.6%	3%
Ask in, no Bid	2.4%	31%	1.4%	18%	3.1%	40%	0.7%	9%
Both in, Ask more	0.5%	27%	0.5%	24%	0.6%	32%	0.3%	15%
Bid in, Ask out	0.6%	3%	5.4%	29%	1.4%	8%	10.9%	59%
Bid in, no Ask	0.7%	9%	3.1%	40%	1.3%	18%	2.4%	32%
Both in, Bid more	0.3%	14%	0.6%	33%	0.5%	24%	0.5%	28%
Both out, Bid less	0.4%	18%	0.6%	28%	0.5%	24%	0.6%	29%
Both out, Ask less	0.6%	29%	0.5%	25%	0.6%	27%	0.4%	18%
No Flow	3.7%	22%	4.9%	28%	4.8%	28%	3.7%	22%
Bid out, no Ask	3.9%	41%	2.7%	28%	2.1%	22%	0.6%	7%
Ask out, no Bid	0.6%	7%	2.3%	24%	2.7%	28%	3.9%	41%
Both out equally	0.3%	21%	0.5%	29%	0.4%	27%	0.3%	22%
Both in equally	0.3%	21%	0.4%	29%	0.4%	28%	0.3%	21%

Table 3.8 extends Table 3.7 by now conditioning on the cause of the initial QP event. We find it is much less clear what side re-establishes a one tick spread in situations where the QP event occurs without the presence of trades. Here we show trade information related to starting event (signed trades; aggressive side of the trade is identified) — expect ask-QP events to be linked to buyer-initiated trades; bid-QP events to be linked to seller-initiated trades. Each set, linked to a depth movement (DM) on both sides of the market, contains four rows for possible sequence of movement in best quotes. The first two sets are linked to opposite movement in depth; the third set shows no flow on either side; the fourth and fifth sets show one side flow out and the other show no flow; the last two sets show one side flow in and the other show no flow. The first two sets: out flow of depth is associated with the first QP event — if bid-QP, then shows majority events linked to a bid out flow of depth. Traded volume is linked to the side not experiencing the QP event — if bid-QP event, then majority of volume is seller-initiated. We find the number of events linked to no trades also are related to the QP side with

outflow of depth. Regarding the no flow row: High concentration (21%) linked to no trade at the initial event and no resulting change in best quotes — cancellations more common to be replaced with a new improved quote (Ask-QP event followed by Ask-QI event)

Table 3.8: Cause of Wide Bid-Ask Spreads and Next Book Updates

Notes: This table is based on the same information as found in Table 3.7. Here we divide up events according to the reason for the initial wide bid-ask spread, which can be due to either trades or quote cancellations (No Trade). For trades, we further distinguish based on the aggressive side (Seller-Initiated & Buyer-Initiated).

Depth Movement (DM)	First; Second	Market Move	No Trade		Seller-Initiated Trade		Buyer-Initiated Trade	
			Total Sample %	DM %	Total Sample %	DM %	Total Sample %	DM %
Bid out, Ask in	Ask-QP; Ask-QI	No Change	0.9%	5%	0.0%	0%	0.5%	3%
	Ask-QP; Bid-QI	Up Move	0.2%	1%	0.0%	0%	0.4%	2%
	Bid-QP; Ask-QI	Down Move	3.1%	17%	7.7%	42%	0.0%	0%
	Bid-QP; Bid-QI	No Change	2.2%	12%	2.9%	16%	0.0%	0%
Bid in, Ask out	Ask-QP; Ask-QI	No Change	2.3%	12%	0.0%	0%	3.1%	17%
	Ask-QP; Bid-QI	Up Move	3.2%	17%	0.0%	0%	7.7%	41%
	Bid-QP; Ask-QI	Down Move	0.2%	1%	0.4%	2%	0.0%	0%
	Bid-QP; Bid-QI	No Change	1.0%	5%	0.5%	3%	0.0%	0%
No Flow	Ask-QP; Ask-QI	No Change	3.6%	21%	0.0%	0%	1.2%	7%
	Ask-QP; Bid-QI	Up Move	0.7%	4%	0.0%	0%	3.1%	18%
	Bid-QP; Ask-QI	Down Move	0.6%	4%	3.1%	18%	0.0%	0%
	Bid-QP; Bid-QI	No Change	3.6%	21%	1.2%	7%	0.0%	0%
Bid out, no Ask	Ask-QP; Ask-QI	No Change	1.9%	20%	0.0%	0%	0.8%	9%
	Ask-QP; Bid-QI	Up Move	0.2%	2%	0.0%	0%	0.4%	4%
	Bid-QP; Ask-QI	Down Move	1.0%	10%	2.9%	31%	0.0%	0%
	Bid-QP; Bid-QI	No Change	1.3%	13%	0.8%	9%	0.0%	0%
Ask out, no Bid	Ask-QP; Ask-QI	No Change	1.4%	14%	0.0%	0%	0.9%	9%
	Ask-QP; Bid-QI	Up Move	1.0%	11%	0.0%	0%	2.9%	30%
	Bid-QP; Ask-QI	Down Move	0.2%	2%	0.4%	4%	0.0%	0%
	Bid-QP; Bid-QI	No Change	1.9%	19%	0.8%	8%	0.0%	0%
Bid in, no Ask	Ask-QP; Ask-QI	No Change	1.5%	20%	0.0%	0%	1.5%	20%
	Ask-QP; Bid-QI	Up Move	0.7%	9%	0.0%	0%	1.7%	23%
	Bid-QP; Ask-QI	Down Move	0.2%	3%	0.4%	6%	0.0%	0%
	Bid-QP; Bid-QI	No Change	1.0%	13%	0.3%	4%	0.0%	0%
Ask in, no Bid	Ask-QP; Ask-QI	No Change	1.1%	14%	0.0%	0%	0.3%	4%
	Ask-QP; Bid-QI	Up Move	0.2%	3%	0.0%	0%	0.4%	6%
	Bid-QP; Ask-QI	Down Move	0.7%	8%	1.7%	22%	0.0%	0%
	Bid-QP; Bid-QI	No Change	1.6%	20%	1.5%	19%	0.0%	0%

Bid-Ask Spread Summary:

We started with the the findings that given one observed a QP event, it is equally likely to observe a QI event on either side of the market. Once we start to explore how depth on both sides of the market is changing while there exists a wide bid-ask spread, it becomes more clear what the next event will be. We then further our analysis by including information related to the initial quote movement. These

findings indicate how the ability to monitor the market depth data in near-real time can allow the market participant to make better predictions regarding future market events.

The analysis above reveals important empirical facts that will be used in the next chapter to motivate our defined method of segmenting the data. First, while the bid-ask spread is commonly equal to the minimum tick size, there are many instances where the spread opens up to two cents (two-ticks). Second, when the bid-ask spread is two-ticks or wider, the market returns to a one-tick spread very quickly. Finally, knowledge of prior book activity is very helpful in explaining how the bid-ask spread returns to the minimum tick size.

3.4 Conclusions

This chapter has introduced the raw data and described steps required to process messages with the ultimate goal of rebuilding the limit order book. The information is the best available ultra-high frequency data which is ideal to investigate modern order-driven markets. While this research is focused on the market for crude oil futures, the steps required to process and analyze the information can be applied to any product traded on the CME.

We have documented a number of relevant facts using the rebuild limit order book related to book update information and trade records. We find a significant number of book updates do not influence the quoted prices of the limit order book and instead simply adjust visible depth at quotes closest to the market. This indicates a large number of new orders flowing into the market as well as a large number of limit order cancellations. The frequency of book updates raises a

number of questions related to how modern electronic quote driven markets now operate — questions that will be explored in later chapters.

Focusing exclusively on the best bid and ask quote, we have defined two key terms that indicate a market in movement. We have defined QP (quote promotion) events to reflect the movement from the second-best quote to the best quote. We have defined QI (quote improvement) events to reflect the entry of a new quote that betters the market.

Focusing on the bid-ask spread, we have presented an analysis that demonstrates how monitoring book updates, or the dynamics of the limit order book, can provide significant information regarding future market events. These results raise a number of questions regarding how a subset of market participants utilize speed to collect information and then use such information to adjust expectation and adjust their placement of resting limit orders in the electronic limit order book.

Chapter 4

Quote Segments

4.1 Introduction

Many empirical studies in market microstructure rely on methods to divide up the trading day into segments that are then analyzed as a sequence of observations. This decision regarding how to identify segments is often motivated by either the available data or the desire to hold some unobserved feature of the market fixed. As an extreme, early research only had daily information which defined each day as a single segment. Since more data has become available, researchers often divide up the day into segments based on either fixed time (e.g., 5-minute or 30-minute segments) or fixed volume (e.g., 500 contracts). Related to the use of fixed volume, studies also define segments based on the number of events such as trades.

Moving away from fixed time intervals is motivated by the desire to hold constant the amount of information flowing into the market. The implicit assumption is that each unit of volume or event carries the same amount of information. But is it true that all trades are equally informative? Does a buyer-initiated trade

which removes one unit of depth at the best ask quote with visible depth of 30 carry the same amount of information as a similar trade for one unit that takes out the best quote?

The purpose of this chapter is to define a new method to divide the day into segments using details from the limit order book. This assists us in how we use multiple quote levels which contain information that we want to study. For example, working with order book data with multiple levels requires one to keep track of both the visible depth and the quote level. Our desire is to keep track of depth only when we fix the quote level. So along with the desire to control for information flows, we also desire a method that helps us manage so much information.

For this analysis, we use ultra-high frequency limit order book information related to crude oil futures traded on the CME, rebuilt from raw message data purchased from the exchange. This rich source of information contains all visible updates to the limit order book across ten book levels on both sides of the market. One advantage of access to such data comes from the opportunities to specify rules regarding how the day is divided up into segments not available with simple trade data. After studying many different types of rules, we select a method that is based on observed best quotes. Using information on how each book level changes over time, along with the bid-ask spread, we divide up each day into a series of segments — referred to as quote segments.

Our defined method of identifying quote segments is most closely related to the use of price durations in market microstructure research. Price durations are defined as intervals where observed prices are within a well defined upper and

lower threshold; once a trade is observed above or below, a new interval starts with new upper and lower thresholds. Instead of realized trade prices, we use information associated with the best quotes. We also define a very narrow range of available quotes within each segment.

One of the primary advantages is the ability to keep track of changes in depth at specific quotes instead of the reference location within the limit order book. For example, once quote segments are defined, we no longer need to refer to a depth as associated with the best bid. Visible depth now is associated with a level defined by the current quote segment. The same logic applies to observed traded volume. We no longer assign it to volume at the best bid, but to a specific quote segment level; buyer-initiated volume at the best ask quote can now be distributed across multiple quote segment levels.

Our contribution is to show how quote segments can be defined and how they are a useful way to study market depth data.

The remainder of the chapter is organized as follows: Section 4.2 defines our method for segmenting the data into intervals that will assist with analysis. Section 4.3 provides a summary of market activity within the context of our defined segmented data. Section 4.4 contains a discussion of alternative ways to define non-overlapping segments using the limit order book data. Section 4.5 concludes and summarizes the chapter.

4.2 Identification of Quote Segments

This section starts with a discussion of our processed market depth data. We then proceed to introduce a method to break up the day into intervals referred to as

quote segments. Once the method is introduced, we present a high level summary of the identified quote segments.

4.2.1 Processed Market Depth Data

Raw CME Message Data

Purchased one year of CME market depth data for crude oil futures. This is the same information sold by the exchange to market participants and allows us to rebuild the limit order book for the top ten levels of the book. The data include timestamps precise to the millisecond and, besides the limit order book, the data contains the complete set of transactions fully integrated with the observed changes to the limit order book.

The raw information, delivered by the exchange, represent a collection of messages created by the exchange whenever there is an change in the visible limit order book. Each individual message contains details for how the limit order book has changed. For example, if a market participant places a new limit order for one contract at the second-best bid, then, if prior visible depth is 4, the action will force the exchange to issue a message that an updated depth at the second-best bid is now 5.

A single message is can have many updates. For example, if a market participant places a market order to buy and if the requested quantity is equal to the resting depth at the best ask quote, then the trade event removes the best ask quote. The exchange will issue a message with trade details followed by all ten ask levels with updated price levels and visible depth for each book level one through ten.

Each message needs to be broken down into individual pieces of information (referred to as data blocks), which adds to the challenge of using such data. Information within the message is defined using TAGs. Within each message there is an ID to correctly sort components of the message.

Rebuilt Limit Order Book

To rebuild the limit order book, it is necessary to breakdown each raw message and then sort the set of book updates using both the message ID and the data block ID. We select all information connected to the nearby expiration with the largest observed daily volume.

We rebuild the limit order book for each side of the market separately. Focusing on a single side, we rebuild the limit order book by starting with the first update and place the information in a grid of ten levels. We retain information at each level if updates do not adjust the level. We keep the last update within a related set of updates. For example, if all ten levels change, then there will be ten rows with updates starting with an update at level one only followed by and updated level one with an update to level two, etc. We only need to keep the last one here, which contains the updates to all book levels in a single row.

For a single side of the market, the rebuilt limit order book contains two sort IDs, a timestamp, product information, ten quote levels and ten visible depth numbers corresponding to each book level one to ten.

Best Quotes Table

After processing the raw message data to rebuild the limit order book, we want to combine information from both sides of the market into a single observation

linked to the millisecond the update was recorded.

Focusing on the best quotes from each side of the market, we pull all rows displaying a change in either the price level or visible depth for the best bid and ask quotes. We then build a table where each row corresponds to a unique timestamp that can be linked to some observed change at one or both of the best quotes. Then, for each row, we fill in the observed updates and pull down price and depth over time if one side doesn't change. This process results in a table with observations linked to a change at the best quotes and allows us to observe changes over time on both sides of the market.

To add information related to quotes above the best ask or below the best bid, we collect snapshot information of visible depth for the next four book levels following the best quote. Since differences between observed quotes within a single side are almost always the minimum allowed tick size (\$0.01 in crude oil futures), we only keep visible depth information for each book level.

The use of snapshot information related to book levels away from the best quotes means we are not including timestamps linked to an observed change away from the best quotes. For example, if a market participant places an order to buy at the second best bid, we will not see this update reflected in our table until there is an observed change at the best quotes. We have documented evidence that it is most common to see updates at the best quotes, so we are sampling depths deeper in the book often.

4.2.2 Definition of Quote Segments

Here we define the steps to identify quote segments for each day in our sample.

Steps to Identify Quote Segments:

- **Step 1:** Start with a sorted table of observed best quotes — keep a sort ID to properly sort information
- **Step 2:** From best quotes, compute the bid-ask spread
- **Step 3:** Identify the first occurrence of a two-tick bid-ask spread, drop all prior observations
- **Step 4:** Breakout all observations that have bid-ask spreads larger than two-ticks — save for later
- **Step 5:** The first observation, by design, has a two-tick bid-ask spread — define the lower quote segment equal to the best bid quoted price; define the upper quote segment equal to the best ask quoted price
- **Step 6:** Move down the table keeping the defined lower and upper quote levels from the prior observation, for each new observation, compare the best quotes to the lower and upper quote levels inherited from the prior observation
 - if the best bid quoted price is below the lower quote level, then redefine the lower quote level equal to the current best bid quoted price and define the upper quote level equal to the current best bid quoted price plus two-ticks;
 - if the best ask quoted price is above the upper quote level, then redefine the upper quote level equal to the current best ask quoted price and

define the lower quote level equal to the current best ask quoted price minus two-ticks.

- If neither condition is true, then the best bid must be equal to the lower quote level or the inside quote level and the best ask must be equal to the upper quote level or the inside quote level.
- Proceed until we have reached the last observation
- **Step 7:** Add back in the observations with bid-ask spreads larger than two-ticks, resort the table by the sort ID
- **Step 8:** Drag down the defined lower and upper quoted price levels — assigns this information to the observations added in the prior step
- **Step 9:** Mark instances when the lower/upper quoted price levels change
- **Step 10:** Assign a unique ID to each interval where the first observation contains a change in the lower/upper quoted price levels — a change starts a new quote segment
- **Step 11:** Define a quote segment reference table where each observation relates to a unique quote segment ID — define the start sort ID and time for each quote segment. Define the end time of the quote segment as equal to the start of the next quote segment.

Quote Segment Levels:

For each defined quote segment, we have three defined quotes that are fixed for the duration of the segment. Without defined quote segments, we need to use the defined book locations. For example, a single quote might alternate between book

location one and book location two. With defined segments, a quote's location is fixed. This allows us to no longer refer to locations of the quote using the book locations. Within a quote segment, we have a defined inside quote along with a lower and upper quote -1 and $+1$ tick difference from the inside quote, respectively. Given a two-tick BAS, the best ask will be defined as the upper quote level, Q_i^{upper} , and the best bid will be defined as the lower quote level, Q_i^{lower} . The open quote level between the best bid and ask quote will be defined as the inside quote level, Q_i^{inside} . These levels are fixed for the entire life of the i^{th} quote segment, QS_i .

Once we fix these quote levels, we can also define quote levels that live above the upper quote level and below the lower quote level. At the moment the quote segment starts, for example, the second-best ask quote level becomes the first quote level above the upper quote level, $Q_i^{upper+1}$, and the third-best ask (99 percent confident this is $+2$ ticks from the best ask) becomes the second quote level above the upper quote level, $Q_i^{upper+2}$. We define similar quote levels for the bid side, but these become the first and second quote levels below the lower quote level, $Q_i^{lower-1}$ and $Q_i^{lower-2}$, respectively. Similar to the lower/inside/upper quote levels, these are fixed for the entire life of the i^{th} quote segment.

Examples of Quote Segments:

Figure 4.1a highlights how we define our first quote segment. Assuming the day starts at t_0 , the distance between the best quotes is \$0.01, which is equivalent to one tick or the minimum bid-ask spread in the crude oil futures market. According to our defined method, we wait for the first instance of a two tick bid-ask spread,

which we see occurs at t_1 . At this point, we start the first quote segment, QS_1 , by defining the upper, inside, and lower quote segment levels, represented in the figure as solid red, green, and blue lines, respectively. Since our identification method makes no reference to trades, we exclude them from this discussion; later analysis will summarize the location of trades within the quote segment.

In Figure 4.1a, the first quote segment starts once there is a drop in the best bid quote (the second best bid quote at t_0 is promoted to the first best bid quote at t_1) while the best ask quote is unchanged. Initially, the inside quote segment level is open, but after some time, we see the buy side establish a position at this open quote level, which returns the market to a one tick bid-ask spread. As long as the best quotes are contained within the quote segment, the bid-ask spread can be either one or two ticks; in this example, the best bid quote moves back and forth between the lower and inside quote segment level. Eventually, the sell side establishes a position at the inside quote segment level and remains there until t_2 when there is a drop in the best bid quote (now with a defined quote segment, the best bid at t_2 is below the lower quote segment level). Since the conditions for a new quote segment are satisfied at t_2 , we define a new quote segment, QS_2 , and define new upper, inside, lower quote segment levels; this new quote segment exists until conditions are met for a new quote segment to be defined.

Figure 4.1b shows an example of directional movement in both best quotes and the corresponding movement in quote segments. This subfigure shows two different ways a quote segment can change. First, as observed at the start of QS_2 , QS_4 , QS_5 , and QS_7 , just prior to the change, the market displays a one tick spread and immediately after the change, the market displays a two tick spread. In each

case, the large spread is eventually reduced to a one tick spread. Alternatively, as observed at the start of Q_3 and Q_6 , there can be a change such that the market displays a one tick spread just before and after the change. This results when a marketable limit order arrives that removes all offered liquidity at the best quote and then results in a new resting limit order at the same quote level (e.g., at QS_3 , an aggressive marketable limit order to sell extracts all resting depth at the best bid and transforms into a resting best ask quote with the remaining unfilled quantity).

One of the conditional requirements for defining a new quote segment places restrictions on the allowed size of the bid-ask spread not being larger than two ticks, which allows for the possibility for the spread to become larger than two ticks within a quote segment. Figure 4.1c shows an example of such a situation. Here we see the initial quote segment start with a two tick spread. We observe a shift upward of the best ask quote and the market opens up to a three tick bid-ask spread. The best ask quoted price is equal to the quote segment level one above the upper quote segment level, but we cannot yet define a new quote segment since the bid-ask spread is too large. Eventually, the best bid quote moves upward, which brings the market to a two tick spread and the next quote segment, QS_2 , is defined. Of course, it is possible for the best ask quote to move down prior to any change in the best bid quote, at which point the quote segment would continue on until conditions were met to establish a new quote segment.

Quote Segments, Best Quotes, and Bid-Ask Spread:

A quote segment represents an interval of time where, conditional on the bid-ask spread of 1- or 2-ticks, best quotes are restricted to live within a narrow range of quoted price levels — at most one tick removed from the inside quote segment level. A quote segment represents a market level with both sides in agreement, while allowing for small price movements around this agreement. We view the inside quote segment level as the best estimate for the market. A quote segment allows us to now focus attention to the inside quote segment level and market activity 1-tick above and below this reference level. We expect small market movements around the inside quote segment. The defined quote segment allows us to focus attention on larger movements that signal the market is working toward a new agreement regarding the level of the inside quote segment level.

Based on the restriction on the bid-ask spread, we are essentially watching the bid-ask spread and each time the bid-ask spread is less than three ticks, we check to see if the best quotes are still at most one-tick away from current inside quote segment level. If they are, then we continue with the currently defined inside quote segment level. If not, then we need to define a new quote segment by defining a new inside quote segment level.

The restriction on the bid-ask spread prevents us from defining new quote segments, thereby shifting the inside quote segment level up or down, due to periods of extreme volatility or due to temporary widening of the quotes from the arrival of aggressive orders that remove one or more quoted price levels. During periods of extreme volatility, the bid-ask spread can open up and stay wide for an extended period of time. Our identification method waits to define a new

quote segment until the bid-ask spread returns to a normal level. For example, following major news releases, it is possible to observe the best ask live above the upper quote segment level and the best bid to live below the lower quote segment level. In these instances, our identification method waits to see where the market eventually settles, which we assume happens when one side issues new quotes to move the bid-ask spread back to a reduced level. If the best ask is above the upper quote segment level and the buy side moves upward, then the market has agreed to re-establish an expected bid-ask spread and the market has moved up, which at this point we define the new inside quote segment level.

Outside of extreme volatility, it is also possible to observe one of the best quotes outside of the defined quote segment if a market order arrives and wipes out some book levels. For example, the market might show the best ask at the inside quote segment level and the best bid as the lower quote segment level. A large seller-initiated market order arrives and afterward the best bid is two below the lower quote segment level. It is possible that soon after, new bid quotes arrive to return the best bid to the lower quote segment level. In this case, we see a temporary movement of one quote outside the quote segment range, but there is a response that returns the market to our current quote segment. Waiting to for the bid-ask spread to return to a normal level prevents us from preemptively defining a new quote segment when the market has not really changed.

Associated Market Information:

We process the entire set of information, processing one date at a time, resulting in a table of observations corresponding to unique quote segments. For each

quote segment, we retain the start/end time (used to define the quote segment duration, QS_i^{dur}), the start/end update ID, a count of the number of three or larger spread occurrences, the lower/inside/quote levels, and a summary of volume observed during the quote segment. Retaining the start/end update IDs allows us to identify all book updates and trade information that occurs within the quote segment. The volume summary includes volume according to buyer- or seller-initiated at specific levels within the quote segment (e.g., the number of buyer-initiated contracts traded at the inside quote level or upper quote level). Using visible depth at fixed levels of the quote segment, we also define a time-weighted measure of visible depth at each level of the quote segment.

According to our method, it is possible for the inside quote to jump by more than one tick either up or down once a new quote segment is defined. For example, if a large seller-initiated market order arrives that wipes out the best four bid levels, as the large BAS here will not initiate a new quote segment, we are forced to wait to see how both quotes respond. If the bid side is the first to respond by re-establishing position with new quotes, then it is possible for the current quote segment to live on. On the other hand, if the ask side is the first to respond by establishing new lower quotes at the open levels, it is possible to see a five tick movement when we compare the prior inside quote level to the newly established inside quote level.

4.2.3 Quote Segment Sample Summary

Applying the identification method defined above, we find a total of 1.7 million quote segments across all trade dates in the sample.

Table 4.1 contains summary statistics of the quote segment sample. We find, on average, quote segments last three seconds with an observed average volume and trade count of 20 and 16, respectively. Comparison of the median shows these averages are much larger than the median value, especially for the duration which has a median value of 1.4 seconds. The majority of the sample has a duration of less than ten seconds.

Table 4.1: Quote Segments — Distribution Summary

Notes: Table summarizes the sample of 1.7 million quote segments identified according to the defined method. The upper table contains distribution information related to the number of quote moves, the number of wide bid-ask spreads, the number of trades, total quantity, and the duration; the lower table shows distribution information conditional on the observed balance between signed volume during the quote segment.

	Mean	P1	P5	P10	Q1	Median	Q3	P90	P95	P99
Number of Quote Moves	45	2	5	7	14	29	56	99	135	239
Number of BAS gt Min	0	0	0	0	0	0	0	0	1	5
Number of Trades	16	0	1	2	5	10	20	34	47	86
Total Quantity	20	0	1	2	6	12	24	44	62	119
Time Duration (in Seconds)	3.34	0.01	0.03	0.06	0.34	1.38	3.93	8.59	13.03	26.93

	Mean	P1	P5	P10	Q1	Median	Q3	P90	P95	P99
Number of Trades										
Balanced	18	0	0	0	5	13	25	41	55	91
Mixed	19	3	4	5	8	14	24	38	51	88
Imbalanced	12	1	1	2	3	7	14	25	38	83
Total Quantity										
Balanced	22	0	0	0	5	15	30	51	69	120
Mixed	24	3	4	5	9	16	30	49	67	123
Imbalanced	16	1	1	2	4	9	17	34	53	115
Time Duration (in Seconds)										
Balanced	5.05	0.01	0.03	0.09	0.61	2.39	6.32	12.82	18.85	36.72
Mixed	4.64	0.03	0.09	0.24	0.89	2.48	5.78	11.21	16.26	31.26
Imbalanced	1.74	0.01	0.02	0.04	0.14	0.71	2.05	4.52	6.90	14.18

The second panel of Table 4.1 shows how quote segment characteristics vary based on the balance between buyer and seller-initiated trades. Quote segments with balanced volume from both sides of the market tend to have a longer duration and have more trades and quantity. Indicates some possible reasons why there exists variation in quote segment characteristics — depends on both sides willingness to cross the spread. Nothing here regarding quote segment level, but expect most of this volume to extract a share of visible depth on both sides of the

market.

Figure 4.2 shows both daily and 1-minute frequencies of observed quote segments. The top figure shows how the count of quote segments vary across dates in our sample. The lower figure shows how quote segments are distributed within the trading day. Regarding daily variation, the number of quote segments can be a proxy for intraday volatility and we find days with the largest number of quote segments are those with noticeable intraday market movement. Days in January and February display a low number of quote segments with just 5,000 per day.

The plot of one-minute intervals, found in Figure 4.2, is very similar to the volume and book update intraday plot of Figure 3.2, which indicates a close direct relationship between the number of quote segments and both volume and number of book updates. The large spikes occur just about every 30-minutes between 8:30am and 11:30am. The intraday plot of quote segments is also very similar to the plot of bid-ask spread counts by minute that are larger than the minimum allowed bid-ask spread (see Figure 3.5), indicating a close relationship to market volatility.

Autocorrelation

Table 4.2 contains an analysis of autocorrelation in the quote segments using a log return series based on changes in the sequence of inside quote segment levels. The upper portion of the table shows results for ten lags. The lower portion includes figures of daily measured autocorrelations for the first four lags.

We find positive significant autocorrelation at lag 1 and 2 with other lags below the significance level. The four figures below show variation in the estimated lag

autocorrelations by day. We find the majority of dates have significantly large positive autocorrelations in the first two lags; only a small number of days with large values for lags 3 and 4.

Table 4.2: Quote Segments — Autocorrelation Study

Notes: Upper table shows the estimated autocorrelation, inverse autocorrelation, and partial autocorrelation function for the first ten lags of a log return series based on the sequence of inside quote segment prices. The lower section shows four figures corresponding to lags 1 to 4 for daily autocorrelation estimates.

LAG	ACF	IACF	PACF	-1*2*STDERR	2*STDERR
1	0.0587	-0.0536	0.0587	-0.0015	0.0015
2	0.0527	-0.0493	0.0494	-0.0015	0.0015
3	-0.0019	0.0077	-0.0078	-0.0015	0.0015
4	0.0075	-0.0063	0.0055	-0.0015	0.0015
5	-0.0066	0.0063	-0.0068	-0.0015	0.0015
6	-0.0011	0.0007	-0.0010	-0.0015	0.0015
7	-0.0045	0.0035	-0.0037	-0.0015	0.0015
8	-0.0003	-0.0002	0.0002	-0.0015	0.0015
9	-0.0017	0.0013	-0.0012	-0.0015	0.0015
10	0.0013	-0.0015	0.0014	-0.0015	0.0015



Directional & Bounce Movement

In this section, we investigate directional and bounce movement observed in the sequence of quote segments. To do this, we define directional movement groups based on the observed change in the inside quote segment level between neighbor-

ing quote segments. For example, based on an observed sequence of one downward move followed by three upward moves followed by a downward move in the inside quote segment level, we would define a directional movement of three quote segments ($DM = 3$) starting with the quote segment following the first upward move.

For all quote segments with a directional movement (DM) of one, we identify uninterrupted sequences of these quote segments to identify groups of quote segments that display a bounce behavior that is similar to the observed bid-ask bounce. If we find a sequence of three quote segments all associated with $DM = 1$, then this set of quote segments will be assigned a bounce count of three.

Table 4.3 presents details related to both directional movements and bounce sets. The upper table contains a summary, based on all identified quote segments, of the directional movement counts. For each directional move, we show the number observed using the set of identified quote segments. We also include a 95% confidence interval based on a simulation of a fair coin flip and the observed sample shares of a given directional movement. For each directional movement, we also show the total number of quote segments along with the number of trades and quantity observed. The same information is then shown for the set of bounce groups with the totals corresponding to $DM = 1$.

Regarding directional movement information in Table 4.3: We find close to 75% of quote segments are part of some directional movement. There are close to 1/5 of quote segments that represent directional movement of two, indicating an “Up-Up” or “Down-Down” pattern. Over the sample, we find many observations of directional movement in the quote segments. For example, we find 32 thousand

Table 4.3: Quote Segments — Directional & Bounce Movement

Notes: Upper table shows the frequency of observed directional movement in the ordered sequence of quote segments. We assign each quote segment to a directional movement based on prior and future quote segment moves, where DM represents the number of quote segments in each movement. For each movement group, we show the total number of directional movements observed, the total quote segments, trades and quantity. The lower table focused on quote segments where $DM = 1$, which combines these into bounce sets.

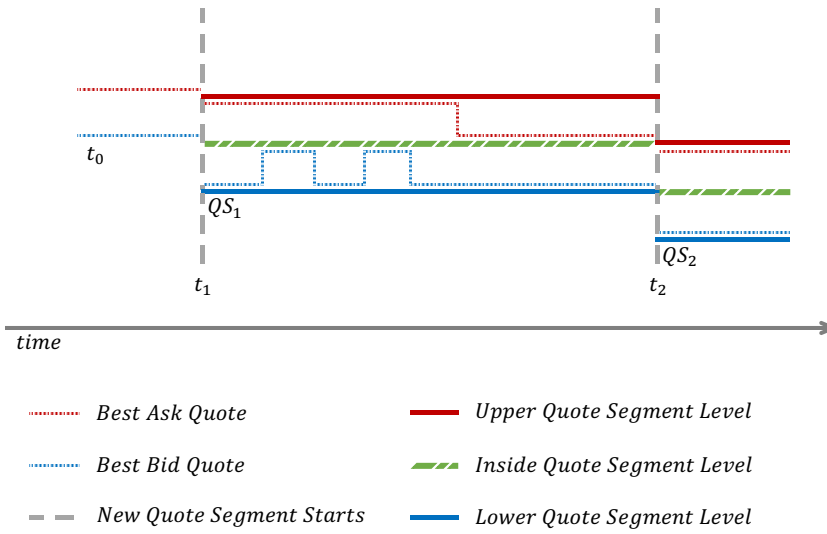
DM	N DMs		CI Runs Test		N Quote Seg		Total Trades		Total Qty	
1	387,886	49.0% *	49.4%	50.6%	387,886	23%	5,212,954	20%	6,541,214	19%
2	178,192	22.5% *	24.5%	25.6%	356,384	21%	4,976,532	19%	6,251,622	19%
3	96,584	12.2%	12.1%	12.9%	289,752	17%	4,412,159	17%	5,582,960	17%
4	55,663	7.0% *	6.0%	6.6%	222,652	13%	3,616,398	14%	4,592,887	14%
5	32,385	4.1% *	2.9%	3.3%	161,925	9%	2,767,383	10%	3,522,444	10%
6	18,559	2.3% *	1.4%	1.7%	111,354	7%	2,006,134	8%	2,556,042	8%
7	10,262	1.3% *	0.7%	0.9%	71,834	4%	1,353,014	5%	1,728,732	5%
8	5,461	0.7% *	0.3%	0.5%	43,688	3%	841,713	3%	1,077,328	3%
gt 8	6,433	0.8%			65,387	4%	1,371,372	5%	1,756,241	5%
Total:	791,425				1,710,862		26,557,659		33,609,470	

Bounce Count	N		CI Runs Test		N Quote Seg		Total Trades		Total Qty	
1	93,655	49.53%	49.37%	50.62%	93,655	24%	1,376,535	26%	1,729,166	26%
2	47,710	25.23%	24.46%	25.55%	95,420	25%	1,318,150	25%	1,652,020	25%
3	22,484	11.89% *	12.09%	12.91%	67,452	17%	884,735	17%	1,105,155	17%
4	11,944	6.32%	5.95%	6.56%	47,776	12%	618,719	12%	775,084	12%
5	5,939	3.14%	2.91%	3.34%	29,695	8%	373,212	7%	469,356	7%
6	3,292	1.74% *	1.41%	1.72%	19,752	5%	249,033	5%	313,325	5%
7	1,753	0.93% *	0.67%	0.89%	12,271	3%	145,385	3%	183,774	3%
8	1,030	0.54% *	0.31%	0.47%	8,240	2%	98,529	2%	125,888	2%
gt 8	1,287	0.68%			13,625	4%	148,656	3%	187,446	3%
Total:	189,094				387,886		5,212,954		6,541,214	

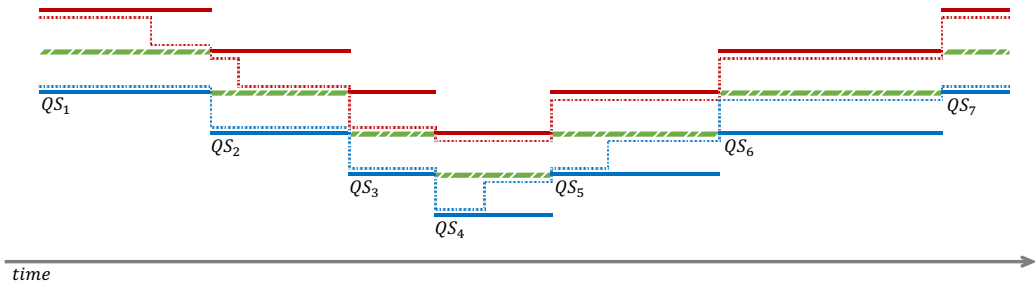
examples of a sequence of five quote segments either moving the market upward or down. Added new CI based on a runs test simulation — this allows us to show our counts exists outside of the CI based on a fair coin toss. We find longer runs of directional movement than would otherwise be expected based on runs using a fair coin.

Regarding bounce movement information in Table 4.3: We find close to 1/4 of quote segments have a directional movement of one, which indicates they are not part of a trend. But these can be isolated events, e.g., Up-Up-[Down]-Up-Up, or part of a bounce in the quote segments, e.g., Up-Up-[Down-Up-Down]-Up-Up. If the majority of these are part of bounce sets, then might indicate we have incorrectly defined the identification of our quote segments. Added new CI based on a runs test simulation — this allows us to show our counts exists outside of the

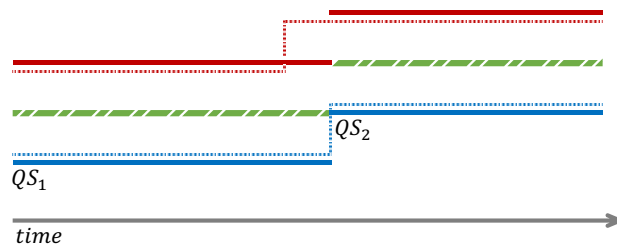
CI based on a fair coin toss. We find longer runs of bounce than would otherwise be expected based on runs using a fair coin.



(a) Quote Segment Introduction



(b) Quote Segment Directional Movement



(c) Example of Large Bid-Ask Spread

Figure 4.1: Quote Segment Examples

Notes: This figure contains three examples of quote segments as observed in the sample. Each example highlights the main components of a quote segment: the lower, inside, and upper quote segment levels. Subfigure (a) is used to to introduce related terminology; subfigure (b) shows an example of quote segments moving directionally down three times followed by movement up three times; subfigure (c) highlights a case when the best ask quote moves above upper quote segment level.

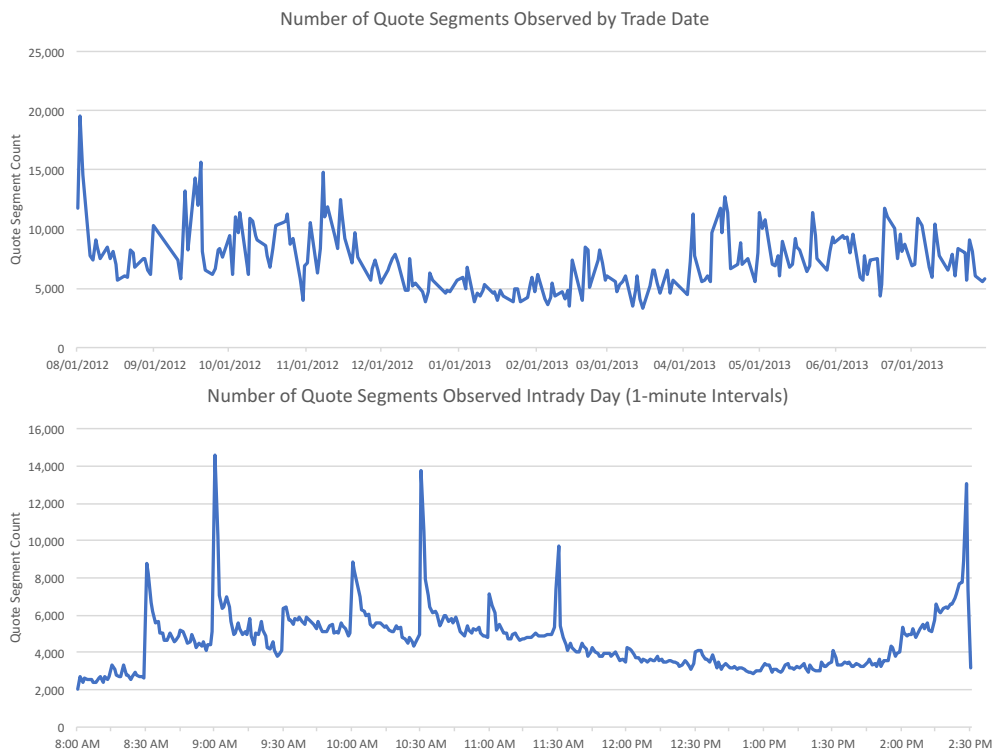


Figure 4.2: Counts of Quote Segments by Date and Intraday Intervals

Notes: This figure shows variation of identified quote segments across days and within the day. The upper figure shows quote segment counts per day across the sample. The lower figure shows quote segment counts within the trading day.

4.3 Depth, Updates, and Volume Analysis

This section investigates visible depth, quote updates, and transactions now in the context of our defined quote segments.

For this analysis, we use the quote book update table and join in the quote segment ID corresponding to the update ID that initiates the quote segment, pull this ID along with the lower, inside, and upper quote segment quote level down until the start of the next quote segment ID. We then redefine visible depth according to the quote segment levels: now both bid and ask have a first quote segment level that is the inside quote segment level. Possible for the inside quote segment level to be empty in the case of a two-tick bid-ask spread. Once the quote segment information is added to the table, we are able to summarize depth levels and changes, along with signed trade volume, according to the exact quote segment where the activity took place.

One of the primary advantages of the quote segments is the ability to summarize accurately information occurring within a given segment. Within a quote segment, observed transaction prices and quotes are able to move around within our narrow range, but we are able to use their location within the quote segment that is fixed for the duration of the segment. Otherwise, without quote segments, one is forced to deal with changing book levels as the book moves in small direction up or down.

4.3.1 Quote Segments & Visible Depth

The book updates discussed in Chapter 3, relate to changes in visible depth for specific quotes. The ultimate value of the data to this research exists in our

ability to track how visible depth close to the market adjusts. This adjustment process contains information regarding how recent market events are interpreted and adjust expectations of future market movement. The depth data also help us interpret realized trades and the decision process to add or remove offered liquidity to the market.

Prior to the identification of quote segments, the processed market depth data contains rankings of quotes from one to ten on both sides of the market. Pausing the market would allow us to get a sense of market depth at these ten quote levels. Prior studies focused mostly on depth at the best bid and ask quotes.

Our perspective on market depth fundamentally changes once we allow quote levels to be defined according to the current quote segment. We no longer simply compare market depth at the best quotes, but now we also have information regarding the exact location. For example, interpreting a visible depth of five at the best bid and 10 at the best ask might differ if we know the best bid is currently established at the lower quote segment level or the inside quote segment level. Furthermore, if we know the best bid quote is established at the lower quote, it is important to know if the best ask quote is established at the inside or upper quote segment level.

This subsection starts with snapshots of observed depth at quote segment levels and contrasts this with observed depth making use of the standard order book location. Next, we provide information regarding depth at various levels of the quote segment; due to symmetry, we focus on the bid side of the market. Finally, we explore intraday patterns in visible depth to investigate differences that can be explained by time of day.

Snapshots of Visible Depth

A scan of the prior literature focused on incorporating information from the limit order book would show a heavy emphasis on using snapshots of offered liquidity at the best available quotes. The location of these snapshots often are selected according to their proximity to realized trades or the desire for fixed time differences between each snapshot (e.g., selecting a snapshot every 30 minutes). This provides limited information for how the book actually evolves over time without trade activity.

Table 4.4 compares visible depth at different book levels, based on a sample of randomly selected timestamps, for standard book levels (upper panel) and based on quote segment levels (lower panel). For each randomly drawn timestamp, we find the most recent book update prior to the random timestamp to assign visible depth. We find there is substantially more variation using the standard book level ordering relative to the quote segment levels, at quotes closest to the market. This is especially clear for the first location which represents the best quote under the standard book level and the inside quote segment level; we find the median drops from seven to zero. Similar differences exist for levels two and three. Interestingly, the distributions are very similar for levels deeper in the book. Since we are primarily focused on book activity near the market, this table provides an indication that quote segments help control for changes in depth that arise simply from changes in the quoted levels.

Table 4.4: Summary Statistics of Visible Depth Levels

Notes: Table shows distribution information of market depth for multiple levels of the book. The top panel shows depth for book levels one to five, where book level one is equal to the best quote. The bottom panel shows depth for book levels redefined according to the quote segment levels, where one represents the inside quote segment level.

	Mean	StdDev	P1	P5	P10	Q1	Median	Q3	P90	P95	P99
Book Level											
1	8.8	8.6	1	1	2	4	7	11	17	21	39
2	20.7	12.6	5	8	10	13	18	25	33	41	64
3	29.3	14.3	9	14	16	21	27	35	44	53	78
4	32.5	15.3	10	15	18	24	30	38	48	57	84
5	32.4	15.3	10	15	18	23	30	38	48	57	85
Quote Segment Level											
1	2.5	4.9	0	0	0	0	0	4	8	11	20
2	13.7	10.9	1	3	4	7	12	17	24	30	51
3	24.7	13.7	6	10	12	16	22	30	39	47	71
4	31.1	14.8	10	14	17	22	29	37	47	55	81
5	32.7	15.4	10	16	18	24	30	38	49	58	85

Time-weighted Measures of Visible Depth

Figure 4.3 provides a high level overview of time weighted bid-ask spread, volume, and visible depth based on the quote segments. We find similar patterns of volume and spreads that jump at specific times during the day. The last subfigure shows market depth according to the level within the quote segment. We can see the depth at the inside quote is very close to zero. Depth increases as we move farther away from the market. We find the depth at the fourth and fifth quote segment levels are very similar.

To investigate visible depth patterns within a day, we compute time-weighted visible depth measures for short intervals of time. Similar to bid-ask spread analysis each observation of depth in our processed data exists due to the arrival of a book update, possibly unrelated to the single quote segment level we are focused on studying. By computing time-weighted values of depth, we are able to get a better sense of what depth existed in the book. Table 4.5 shows distribution information for quote segment levels one to five. For each level, we show distribution

information for three different time intervals.

Table 4.5: Summary Statistics of Time Weighted Visible Depth

Notes: Table contains distribution information for quote segment levels one to five for fixed time intervals of 500 milliseconds, one minute and five minutes. Depth values are computed as time weighted averages across each time interval.

	Mean	StdDev	Min	Max	P1	P5	P10	Q1	Median	Q3	P90	P95	P99
Quote Segment Quote Level 1													
500ms	2.5	4.6	-	563.7	-	-	-	-	0.0	3.8	7.9	10.9	18.7
1min	2.5	1.8	-	66.1	0.2	0.5	0.8	1.3	2.1	3.2	4.5	5.6	8.8
5min	2.5	1.2	0.0	17.1	0.6	1.0	1.2	1.7	2.4	3.1	3.9	4.5	6.3
Quote Segment Quote Level 2													
500ms	13.7	10.5	-	566.5	1.8	3.1	4.5	7.3	12.0	17.2	24.0	29.2	49.0
1min	13.7	5.7	1.0	221.7	4.8	6.7	7.9	10.1	12.9	16.2	19.8	22.8	31.9
5min	13.7	4.3	1.9	97.3	5.7	7.6	8.7	10.8	13.4	16.1	18.9	20.9	25.8
Quote Segment Quote Level 3													
500ms	24.7	13.3	-	581.6	6.1	10.0	12.2	17.0	22.3	29.6	38.8	46.0	69.5
1min	24.7	9.1	1.6	215.2	8.7	12.6	14.8	18.8	23.5	29.1	35.8	40.8	53.4
5min	24.8	7.8	3.7	102.2	10.0	13.7	15.7	19.7	24.0	28.9	34.7	38.6	48.0
Quote Segment Quote Level 4													
500ms	31.2	14.5	-	631.0	10.0	15.0	17.3	22.4	29.0	37.0	46.9	55.0	79.9
1min	31.2	10.6	2.0	207.9	11.7	16.6	19.3	24.3	29.9	36.6	44.4	49.8	63.8
5min	31.3	9.2	3.6	88.5	13.1	17.6	20.3	25.2	30.5	36.4	43.2	47.7	57.8
Quote Segment Quote Level 5													
500ms	32.8	15.0	-	633.7	11.0	16.0	19.0	24.0	30.0	38.6	49.0	57.0	83.6
1min	32.9	11.1	2.9	335.5	12.1	17.7	20.6	25.7	31.5	38.5	46.4	51.9	66.4
5min	32.9	9.6	4.8	99.4	13.3	18.7	21.6	26.6	32.2	38.3	45.3	49.7	60.3

4.3.2 Quote Segments & Quote Updates

Depth Changes by Quote Segment Level:

Table 4.6 shows how depth changes vary according to quote segment levels. For all levels here Q1 to Q3 show zero — indicates depth is changing at single book levels. We can see that book level 2 and 3 show the most change — supports the decision to focus on these later on when studying offered liquidity. Table is based on a single day — we find little variation as we change the date.

Non-Zero Depth Changes by Quote Segment Level:

Table 4.7 shows information related to non-zero depth changes by quote segment level. We find majority of changes are 1 lot increases and decreases — this confirms

Table 4.6: Distribution of Observed Depth Change by Quote Segment Level

Notes: Table shows distribution information for observed changes in depth conditional on the quote segment level: level 1 marks the inside quote; level 2 marks either the lower or upper quote segment; levels 3 to 5 then move away from the inside quote segment level. This excludes book updates associated with trades to only show changes due to arrival of new limit orders and cancellations. Based on the first day of the sample, August 1, 2012.

	P1	P5	P10	Q1	Median	Q3	P90	P95	P99
Ask Level 5	-2	-1	0	0	0	0	0	1	2
Ask Level 4	-3	-1	-1	0	0	0	1	1	3
Ask Level 3	-5	-2	-1	0	0	0	1	2	5
Ask Level 2	-4	-2	-1	0	0	0	1	2	5
Ask Level 1	-2	-1	0	0	0	0	1	1	3
Bid Level 1	-2	-1	0	0	0	0	1	1	3
Bid Level 2	-4	-2	-1	0	0	0	1	2	5
Bid Level 3	-5	-2	-1	0	0	0	1	2	5
Bid Level 4	-3	-1	-1	0	0	0	1	1	3
Bid Level 5	-2	-1	0	0	0	0	0	1	2

our findings with trade and trade groups — the majority of resting depth is made up of limit orders with a single contract. Difference between +1 and -1 at level 1 can reflect the fact that trades eventually pick up the resting order, but also can reflect the market in motion. If the market is moving upward, order entry at the inside quote will eventually move down deeper into the book on the buy side.

Table is based on a single day — we find little variation as we change the date.

Table 4.7: Frequency of Non-Zero Depth Changes by Quote Segment Level

Notes: Table shows counts and percentages of observed non-zero depth changes by quote segment level: level 1 marks the inside quote; level 2 marks either the lower or upper quote segment; levels 3 to 5 then move away from the inside quote segment level. This excludes book updates associated with trades to only show changes due to arrival of new limit orders and cancellations. Based on the first day of the sample, August 1, 2012; for the buy side only

Depth Change	Level 1		Level 2		Level 3		Level 4		Level 5	
lt -5	359	1%	1,575	1%	1,932	2%	490	1%	307	1%
-5	152	0%	1,009	1%	1,570	2%	319	0%	144	0%
-4	327	0%	2,224	1%	2,766	3%	854	1%	488	1%
-3	668	1%	4,416	3%	4,926	6%	2,113	3%	1,042	2%
-2	1,941	3%	11,993	8%	10,469	12%	5,710	8%	4,091	7%
-1	16,820	26%	53,914	35%	21,715	25%	22,899	34%	21,941	40%
1	34,668	53%	55,315	36%	21,937	26%	25,028	37%	20,882	38%
2	5,570	8%	11,025	7%	9,519	11%	6,033	9%	4,036	7%
3	1,942	3%	4,585	3%	4,672	5%	2,402	4%	1,144	2%
4	1,101	2%	2,531	2%	2,847	3%	1,094	2%	521	1%
5	565	1%	1,374	1%	1,488	2%	365	1%	122	0%
gt 5	1,456	2%	2,421	2%	1,882	2%	487	1%	217	0%

Time Between Book Updates by Quote Segment Level:

Table 4.8 shows the distribution of time between book updates by quote segment level. We find the second quote segment level (corresponds to the upper or lower quote segment level) has the shortest time between book updates — a median of just 12 milliseconds and average of 154 milliseconds. We find the time between updates increases as we move deeper into the book. Table is based on a single day — we find little variation as we change the date.

Table 4.8: Distribution of Time Between Book Updates by Quote Segment Level

Notes: Table shows distribution of time between non-zero book updates by quote segment level: level 1 marks the inside quote; level 2 marks either the lower or upper quote segment; levels 3 to 5 then move away from the inside quote segment level. This excludes book updates associated with trades to only show changes due to arrival of new limit orders and cancellations. Based on the first day of the sample, August 1, 2012; for the buy side only

	Count	Mean	P1	P5	P10	Q1	Median	Q3	P90	P95	P99
Level 1	65,568	0.357	0.001	0.001	0.002	0.005	0.021	0.203	0.787	1.600	5.397
Level 2	152,381	0.154	0.001	0.001	0.002	0.004	0.012	0.102	0.399	0.749	2.071
Level 3	85,722	0.273	0.001	0.002	0.004	0.008	0.036	0.230	0.720	1.247	3.289
Level 4	67,793	0.345	0.001	0.003	0.005	0.013	0.069	0.337	0.921	1.547	3.659
Level 5	54,934	0.426	0.001	0.004	0.007	0.018	0.096	0.426	1.127	1.878	4.434

4.3.3 Quote Segments & Transactions

Our perspective on volume also changes once we have defined quote segments. For a specific quote segment, we expect the majority of buyer-initiated trades to take place at either the inside or upper quote segment levels; similarly, the majority of seller-initiated trades to take place at either the inside or lower quote segment levels. According to Table 3.5, we know the vast majority of trades only interact with the best quotes, but should we think of a buyer-initiated trade at the inside quote differently from a buyer-initiated trade at the upper quote segment level? It might become especially relevant to differentiate a buyer-initiated trade at the inside quote segment level from one that takes place at one (or two) above the

upper quote segment level.

A subset of the book updates found in the processed depth data contain trade information. At a minimum, we have aggregated trade quantity from an individual trade group (see Section 3.2 for a discussion of how trade records are combined into trade groups) that is identified as buyer- or seller-initiated. We can add further information that identifies the quote segment level associated with the trade price. Every trade in our sample either takes place within a quote segment or creates a book update that meets the criteria for the start of a new quote segment, which associates the trade with the newly defined quote segment.

Volume by Quote Segment Level

Where does volume live within a quote segment? This is an important question and serves also as a check on our defined method of segment identification. For example, if the majority of volume takes place outside of the quote segment (above the upper quote segment or below the lower quote segment), then this would indicate our allowable quote levels need to be expanded.

Figure 4.4 provides information on the distribution of volume by quote segment level. We find the majority of volume occurs at quote segment levels one (the inside quote level) and two (the lower quote segment level); approximately 10 percent of volume is found at levels deeper in the book. The figure shows a stable pattern regarding how daily volume is distributed across the quote segment levels — especially at the extreme upper and lower quote segment levels. One might expect to see a larger share of volume at the upper (lower) quote segment level for days that display large price increases (decreases); but we do not find such

evidence here.

While we do not differentiate between buyer- or seller-initiated volume in Figure 4.4, we expect the majority of volume at the upper quote level to be associated with buyer-initiated trades. In fact, we find that 99% of upper quote level volume can be linked to aggressive buy-side use of marketable orders that extract liquidity from the best ask quote located at the upper quote segment level. Volume at the inside quote segment is equally shared between the two sides of the market.

Volume by Quote Segment Start/End

According to our defined method of segment identification, we should expect some amount of traded volume to be associated with the start of a new quote segment. For example, assume the current best bid is sitting at the inside quote and the best ask is sitting at the upper quote level — indicating the market is trading at a one tick bid-ask spread. If aggressive buyer-initiated trades are depleting depth at the best ask quote, then once the depth is completely exhausted, the best ask quote shifts to the second-best ask quote. At this exact moment (the timestamp of the last trade) the bid-ask spread opens up to two ticks and a new quote segment is defined to start. Another option exists where the bid side acts to kill immediately the best ask and establish a new best bid at the same quote level. In either case, the new best ask quote is above the prior best ask quote and as long as the bid-ask spread is at most two ticks, then a new quote segment is defined to start.

Analyzing volume according to the end of a quote segment is related to the summary of volume and its location of a current best quote. But here we are able to divide end of quote activity to that which occurs within the quote segment from

that occurring at the end of a quote segment.

Table 4.9 shows a clear subset of those QI QP events — some start a new quote segment, but others do not. Means those that do not are related to movement at the inside quote segment level. There are three sets related to events associated with new quote segments: [,ask-QP] [bid-QP,] have the most activity. If this happens, then it must be true that the best ask is already at the inside quote segment level. Above, in the QI QP section we also showed about 1/2 of volume at the best quotes is not related to a QI or QP event, therefore, cannot be associated with a new quote segment.

Table 4.9: Volume and Quote Change Event Conditional on Quote Segment Start

Notes: This table shows how book updates and volume is distributed according to changes in the best quotes accounting for when the changes result in a new quote segment. Table is based on the book updates on 8/1/2012, the first day of the sample.

Buy Side	Sell Side	N Updates	BI Deeper	BI Best Quote	SI Best Quote	SI Deeper
Activity within a Quote Segment -- No Change in Best Quotes						
		324,243	73 0%	35,831 54%	35,146 53%	103 1%
Activity within a Quote Segment -- Change in Best Quotes						
QI		12,344	0 0%	625 1%	128 0%	32 0%
	QI	11,793	116 1%	200 0%	608 1%	4 0%
QP		9,701	6 0%	287 0%	11,827 18%	9,541 56%
	QP	9,073	8,541 50%	10,973 17%	273 0%	8 0%
QI	QP	1,824	1,891 11%	4,296 7%	48 0%	0 0%
QP	QI	1,725	0 0%	35 0%	3,861 6%	1,517 9%
QI	QI	114	2 0%	11 0%	3 0%	0 0%
QP	QP	94	39 0%	93 0%	127 0%	72 0%
Quote Segment Change -- Change in Best Quotes						
	QP	3,421	1,388 8%	8,135 12%	153 0%	0 0%
QP		3,177	1 0%	117 0%	7,932 12%	1,212 7%
QP	QI	1,541	8 0%	51 0%	5,971 9%	4,588 27%
QI	QP	1,472	5,042 29%	5,277 8%	103 0%	21 0%
	QI	672	12 0%	28 0%	92 0%	0 0%
QI		607	0 0%	109 0%	6 0%	0 0%
QI	QI	82	0 0%	3 0%	4 0%	0 0%



Figure 4.3: Time Weighted Spread, Volume and Depth for Quote Segments

Notes: First subfigure shows bid-ask spread for one minute intervals, where the minimum bid-ask spread is \$0.01. The second subfigure shows average volume for one minute intervals. The third subfigure shows average depth at five quotes segment levels on the bid side.

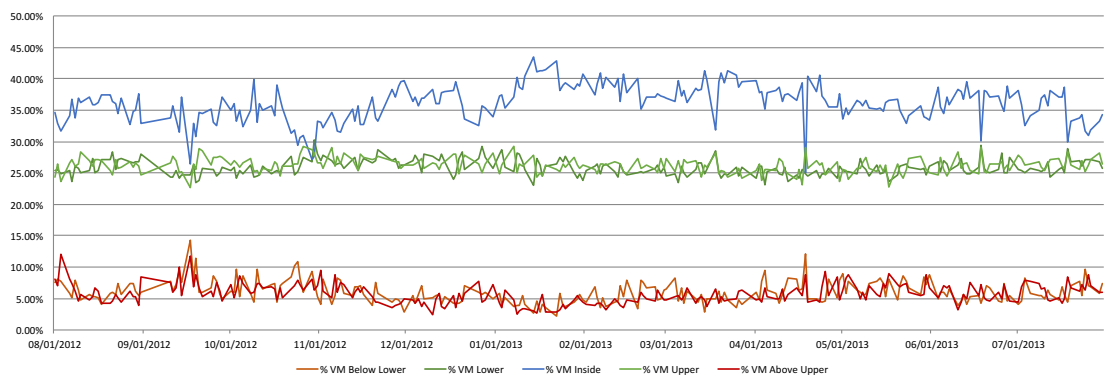


Figure 4.4: Percent of Total Volume by Quote Segment Level

Notes: This figure shows daily volume shares associated with quote segment levels. For each day, we divide the volume based on where this volume takes place within the quote segment, including at the inside quote, at the upper or lower quote segment levels, or below the lower or above the upper quote segment levels.

4.4 Alternative Identification Methods

The identification method discussed in Section 4.2 is based only on observed changes in the best quotes along with a requirement on the observed bid-ask spread. Given the reliance on best quote information, we refer to the identified non-overlapping intervals as quote segments. The richness of the data, although, allows for many other options for specifying rules to break up the day into segments of market activity.

Initial research defined 14 different methods that differed in the conditional requirements to start and end an interval. One primary difference between these methods is the decision of how much to incorporate from quote updates, visible depth, and/or realized trades. This section reviews alternative methods.

Trade Based Identification:

This identification method starts by selecting the first observed trade and defining an interval above and below this price. We considered different multiples of ticks, where one tick represents \$0.01 for crude oil futures. For example, we can specify one-tick, two-ticks, and three-ticks for the price band. The first interval extends through time as long as trades are within these defined price band. Once there is a trade that is either above the upper limit or below the lower limit, a new segment is defined and this trade price is used to define a new price band. This has a close connection to price durations — prior research that only uses trade data defines intervals based on observed price movement. We focus on a very small price change resulting in many more intervals. We investigated resulting segments: one-tick, two-tick, three-tick. Issue came up with just trades comes

from market order events that push deeper into the limit order book. A trade only based approach would need to consider how the market reacts to such events and their frequency.

Quote Based Identification:

Our selected identification method discussed in Section 4.2 only uses information from the best bid and ask quotes. This also allows us to focus on a smaller subset of the full data, but relative to just trades, there is a lot more information here as quotes change often without trades.

We explored a number of alternatives just using best quote information. For example, it is possible to expand the set of quotes within a single segment. One alternative method is to define two inside quotes; another method would keep a single inside quote, but allow for two quotes above and below this defined inside quote.

Combining Trade & Quote Based Identification:

After exploring methods based only on trades or quote information, we then explored various ways to combine both types of information to define non-overlapping intervals. We used trade information to add conditions on the start of a new quote segment. For example, instead of just a change in the best quotes, require a trade to occur at the new best quote before we allow for a new quote segment. Other methods extended the trade based approach to require trades occur at best quotes. We also extended the trade based approach to require trades to occur within a range of quotes (e.g., best and second best quotes).

Possible Extensions:

The richness of the data opens up many other possible extensions. For example, above there is no reference above to visible depth — we can extend any method that uses quote level information to condition on the size of visible depth. Considered extending the quote segment method to require visible depth be larger than five contracts at a new quote level. Introducing information on visible depth allows us to only consider new quote levels with visible depth above some defined level. Assumes a lower visible depth is likely a temporary event.

The exact decision rules would need to vary according to the specific market characteristics of visible depth. This rule can even be allowed to change according to the time of day or prior market volatility. As one increases the required depth, it becomes more difficult to establish a new quote segment.

4.5 Conclusions

Large data presents a number of challenges. One of these challenges is how to best collapse the information into a smaller table while retaining the important components of the data. Historically, the selection of a method to collapse the data has been a function of data availability: from daily to intraday and from fixed time intervals to flexible time intervals.

We have introduced a new method to divide up the trading day into intervals determined by the movement of best quotes and the displayed bid-ask spread. By defining a method to clearly identify the moment a new interval begins, we are able collapse the large data table into a set of observations of individual quote

segments.

The advantage of defining a quote segment comes from our ability to redefine the location of quotes in the limit order book on both sides of the market. Instead of only using the location in the book (e.g., the best or second-best or third-best bid quote), we are able to make reference to quotes depending on the distance from single quote defined by the quote segment — the inside quote segment level. Moving away from the inside quote segment level, for the buy side, we have the lower quote segment level that has the majority of resting depth and traded volume due to seller-initiated transactions. Moving deeper in the book, we defined the one-below lower quote segment level followed by the two-below lower quote segment level.

We find a number of results related to identified quote segments that demonstrate the value of collapsing the information. We find that visible depth is more steady at defined quote segment levels than standard book levels. Regarding traded volume, we find the majority of traded volume occurs at the lower and upper quote segment levels. This allows us to link traded volume to more than simply occurring at the best bid quote.

This chapter has also discussed a number of ways to define different identification methods. These methods differ in the selection of information used to define a new quote segment. Our defined method only makes use of the best quotes and the observed bid-ask spread, but there can be many other methods that incorporate transaction information or visible depth at quotes near the market.

Chapter 5

Visualizing Market Depth Data

5.1 Introduction

The importance of data visualizations is continually increasing with the need to study complicated data. The information used in this research is an excellent source for building visualizations given the rich dynamics of the limit order book. In this chapter, we build a visualization of order book data that will be used to identify patterns and establish a foundation for later chapters. While there exists many other examples of how to visualize the limit order book, this chapter introduces a novel way to capture dynamics of quote movement and how market depth dynamically adjusts at fixed quote levels.

The market depth data is difficult to work with because of the large number of observations and the amount of information each observation contains. Within a single row, we have a defined time when the update to the limit order book was recorded and a price vector assigned as quote levels one to ten for each side of the market. With each quote level, we have a quantity, representing the market

depth. Each update reflects a change to stock information that lives over time until the next book update.

The difficulty increases once we try to study the dynamics of the limit order book. For example, within a single update, a single quote is assigned a number from one to ten. The next update might include this quote at the same assigned number or there might be a movement up or down. If one is concerned about depth at a fixed quote, it is necessary to account for the quote level and not the assigned number within the quoted price vector.

Visualizations that simply show depth at the best quote are influenced by movement up or down in the quoted price. For example, if we observe an increase of 20 at the best book level, there are three possible ways for this to happen. First, if there is no change in the quoted price, then this event results from a new order arrival at the best quote. Second, if there is also an increase in the quote price, then not only did the new order add depth, but the order also improved the market. Finally, if there is a decrease in the best quote price, then this change simply results from the prior best quote being removed from the market. This last case is clearly different from the first two — the observed increase might simply reflect the fact that depth should be larger as we move to lower bids.

Based on our exploration of the data presented in Chapters 3 and 4, we are aware of features of the data that we want to highlight in the visualization. First, we know the majority of book updates do not change the quoted price vector. Depth changes over time are often associated with a single quote and do not change the distance to the market. Second, we know the majority of book updates occur at quotes closest to the market. We find arrival and/or cancellations most often

occurs at quotes closest to the market. Finally, we know the majority of book updates add or remove a single contract from the visible depth.

A number of sections provide us with insights that can be used to define the best visualization. These include:

- Section 3.2: No-Trade Book Updates — shows depth changes, frequency, and differences by book level — finds patterns in update location relative to the market
- Section 3.2: Volume — shows proximity to quote changes, also size and tendency to just extract depth from level one; proximity of book updates following trades
- Section 3.3: Bid-Ask Spread Analysis — shows how observed changes in the best quotes (represents the market), distances (reflected by the bid-ask spread) from each other, and how these distances tend to move together. Also studied initial and next market movement to find tendency for one side to move away and then move back toward the market. Shows tendency for one quote to move toward the market while one moves away — evidence why the depth data matters
- Section 4.3: Quote Segments — shows book updates by quote segment level, traded volume by quote segment level, depth by quote segment level. Differences in observed depth between quote segment levels. Based on evidence presented here, we expect depth to be larger at the second book level and the second quote segment level (upper or lower quote segment).

The final visualization allows us to see market dynamics that capture direc-

tional changes in best quotes, depth at multiple levels of the book, and signed trades. After studying the visualization, a number of observations will become the source of motivation for later chapters. First, the visualizations provide evidence that changes in visible depth is concentrated across the first three quote segment levels. We find less response and movement in depth at levels deeper in the book. This will allow us to define offered liquidity, based on visible depth at quotes closest to the market, and study how this quantity changes over time on both sides of the market. Second, the visualization provides evidence of flash quoting, or quote improvements with a very short duration, which will become the subject of the last chapter and represent a different kind of fleeting liquidity.

In Section 5.2, we review the underlying data and the methodology to process the raw information. In Section 5.3, we present some fairly standard visualizations of the main parts of the data: the best quotes, the bid-ask spread, transactions, and visible depth. We then begin to combine the different features of the market into a single plot. In Section 5.4, we extend the visualization by making adjustments to depth and quote information that help us see exactly what is taking place in the market over very short time scales. In Section 5.5, we add additional depth information at levels above the best ask and below the best bid.

5.2 Data & Methodology

CME Market Depth Data

The visualizations make use of raw message data reflecting updates to the limit order book purchased from the Chicago Mercantile Exchange (CME). Raw mes-

sages are processed and used to recreate the limit order book for both sides of the market for ten quote levels. Once the raw messages have been processed, we use two sequence IDs to properly sort the information, which results in the order book. Each observation, marked to the millisecond, contains the visible market depth, or the quantity associated with resting limit orders for the highest ten bid quotes and lowest ten ask quotes. Each observation reflects some update to the limit order book.

Quote Movement Terminology

As discussed in Section 3.3, we define a Quote Improvement (QI) occurring when the change in the best quote results in a higher quote (on the bid side) or a lower quote (on the ask side). This event happens when there is an arrival of an order that betters the market, thereby shifting the prior book levels one to nine down to two to ten.

A Quote Promotion (QP) is defined to happen when there is a removal of a best quote. This happens whenever depth is completely removed from the best quote either due to trades or cancellations.

Quote Segment Identification

As discussed in Chapter 4, we define quote segments based on movement of the best quotes. An individual quote segment reflects an interval of time and lasts until specific conditions are met to end the current quote segment and start a new quote segment that is shifted up or down. A single quote segment is defined by the inside quote segment level. The upper quote segment level is defined as one tick above the inside quote; the lower quote segment level is defined as one tick

below the inside quote. With quote segments defined, we no longer make reference to the bid based on the book location (where location one means the best quote), but instead we make reference to the location of a quote either within the quote segment or above/below the quote segment.

Event Time vs Clock Time

These visualizations are all shown in event time and, therefore, do not reflect the timestamps attached to each book update or trade record. This is done to give each book update equal weight in the visualization. Given the nature of the data, updates tend to arrive in clusters so plotting information by calendar time, would make it difficult to identify patterns when the market is moving quickly.

Unless specifically noted otherwise, any visualization showing changes in depth from the rebuilt limit order book represent a fixed number of book updates. We start with figures representing 200 book updates, then expand this to 500 updates. Fixing the number of updates means the duration changes depending on the arrival of book updates. For example, during periods of high volatility following a major scheduled news release, it is possible to observe 200 updates over a few seconds. On the other hand, during slow periods, 200 updates might represent 30 seconds.

5.3 Single Visualizations

We start with some simple visualizations of the primary aspects of the data. These aspects include best quotes, trades, and visible depth. Best quotes refer to the best bid and ask quoted price level, which exists at a fixed point in time. Trades are marked as buyer- or seller-initiated, where buyer-initiated (seller-initiated) trades

take place at the best ask (bid) quote and extracts provided liquidity from the best quote. Visible depth refers to the number of contracts linked to limit orders that are resting in the limit order book on both sides of the market.

5.3.1 Visualization of Best Quotes

We start with a standard visualization of best quotes. Together, the best quotes define the bid-ask spread (BAS) as the best ask quote minus the best bid quote. We expect this bid-ask spread to be equal to the minimum allowed price fluctuation as defined by the exchange — \$0.01 for crude oil futures. A narrow bid-ask spread is commonly used to define a liquid market. We start with a simple view of the best quotes since these define the market at any point during the day.

Figure 5.1 contains four examples which show the best quotes and the computed bid-ask spread. Each example is comprised of 200 book updates and shows movement in the best quotes in the upper panel (the best ask quote in red and the best bid quote in blue) and corresponding movement in the bid-ask spread directly below in the lower panel (in green). Instead of plotting information according to the timestamp associated with each book update, we plot the information in event time. Therefore, each example can correspond to a different duration of time. In our set, example 1 is 12 seconds; example 2 is 10 seconds; examples 3 and 4 represent 16 seconds. Note that updates might only adjust visible depth, which explains why we see periods of no movement in the best quotes.

Focusing on the best quotes, we can spot instances of quote promotion (QP) and quote improvement (QI) events on both sides of the market. QP events on the ask side occur when there is a vertical shift up; for the bid side, QP events

occur when there is a vertical shift down. QI events are identified by the opposite movement and reflect movement of the quotes toward the other side of the market.

The majority of updates observed in Subfigure 5.1a are located on three quote levels and there is a one-tick move up during the last 50 update observations. While there exist long intervals of a minimum bid-ask spread, often we observe a two-tick bid-ask spread following a QP event that lasts until the arrival of a QI event. We can see around book update 50, there is an adjustment of best quotes without a jump in the bid-ask spread, which occurs from a bid-QI and an ask-QP event occurring at the same time. The figure shows a few cases of flickering in the best quotes — observed when the best ask quote has a QI event followed closely, in event time, by an ask-QP event. Toward the end, we see the best quotes shift upward by one-tick — this also corresponds to a new quote segment as we redefine the location of the inside quote segment level.

Subfigure 5.1b, shows a directional movement downward broken up into two different movements. Between each downward move, we find an interval, between book update 50 and 150, with the best quotes located on three quote levels. We find the bid-ask spread tends to be one-tick during directional movements and then, once the market calms down, the bid-ask spread tends to open up to two-ticks with a greater frequency.

Subfigure 5.1c shows a sequence of updates that commonly display a two-tick bid-ask spread and we observe one occurrence with a market displaying a three-tick bid-ask spread. The three tick bid-ask spread arrives after a sequence of ask-QP events. Just from the plot of best quotes, it is impossible to say much about why the bid-ask spread is often showing a two-tick spread.

Subfigure 5.1d shows a clear directional movement upward in the best quotes. The example shows a slower movement up in the first half relative to the second half. Observing just the movement in the best quotes, we would expect this to be associated with buyer-initiated trades that extract available liquidity from the best ask quote. The slower movement should be associated with less traded volume.

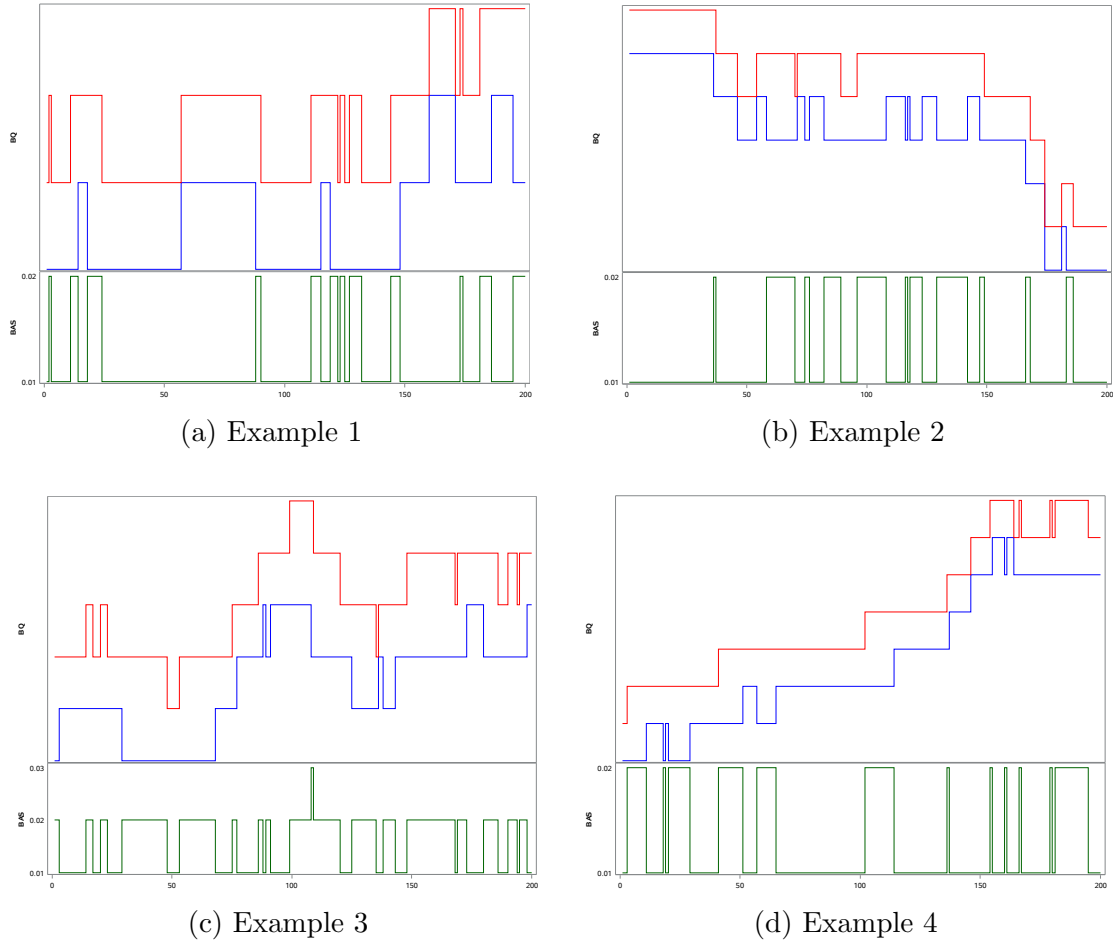


Figure 5.1: Visualizing Best Quotes & Bid-Ask Spread

Notes: Each example contains two panels: the upper panel shows the best bid quote (in blue) and best ask quote (in red); the lower panel shows the computed bid-ask spread (BAS) in green, defined as the best ask quote minus the best bid quote. Each example corresponds to 200 book updates and is displayed in event time.

5.3.2 Visualization of Best Quotes & Transactions

The examples displayed in Figure 5.1 allowed us to talk about the market based on observed changes in the best quotes, but we were able to say very little about

why these changes took place. We know that a lot of resting limit orders get cancelled and do not result in trades. Since we have trade data, we can add this information into the visualization.

Figure 5.2 shows the same four examples from Figure 5.1, but now includes trade information. Each trade is marked on the figure according to the price, direction, and whether the trade resulted in partial or complete depth extraction from the quote. Trades are color coded depending on whether the trade is buyer-initiated (blue) or seller-initiated (red). Furthermore, we distinguish between trades that extract all available liquidity at the best quote level (these trades kill the quote level), represented as filled in squares, from trades that extract a partial amount of available liquidity, represented as open circles. Accordingly, an open circle must always occur within a quote horizontal line, while a filled square must be followed by a QP event.

With the new trade information, we are able to get a better sense of what is taking place in the market. In all subfigures, we can see the majority of trades occur at or just prior to QP events. We also see that volume is one-sided during periods of directional movement.

5.3.3 Visualization of Market Depth

Including trade information into the best quote visualization allowed us to tell a more complete story of market events which helped explain the observed changes in best quotes. But there are still some observed quote changes without corresponding trade information. How is visible depth at or near the best quote able to explain the observed changes in best quotes or the occurrence of trades?

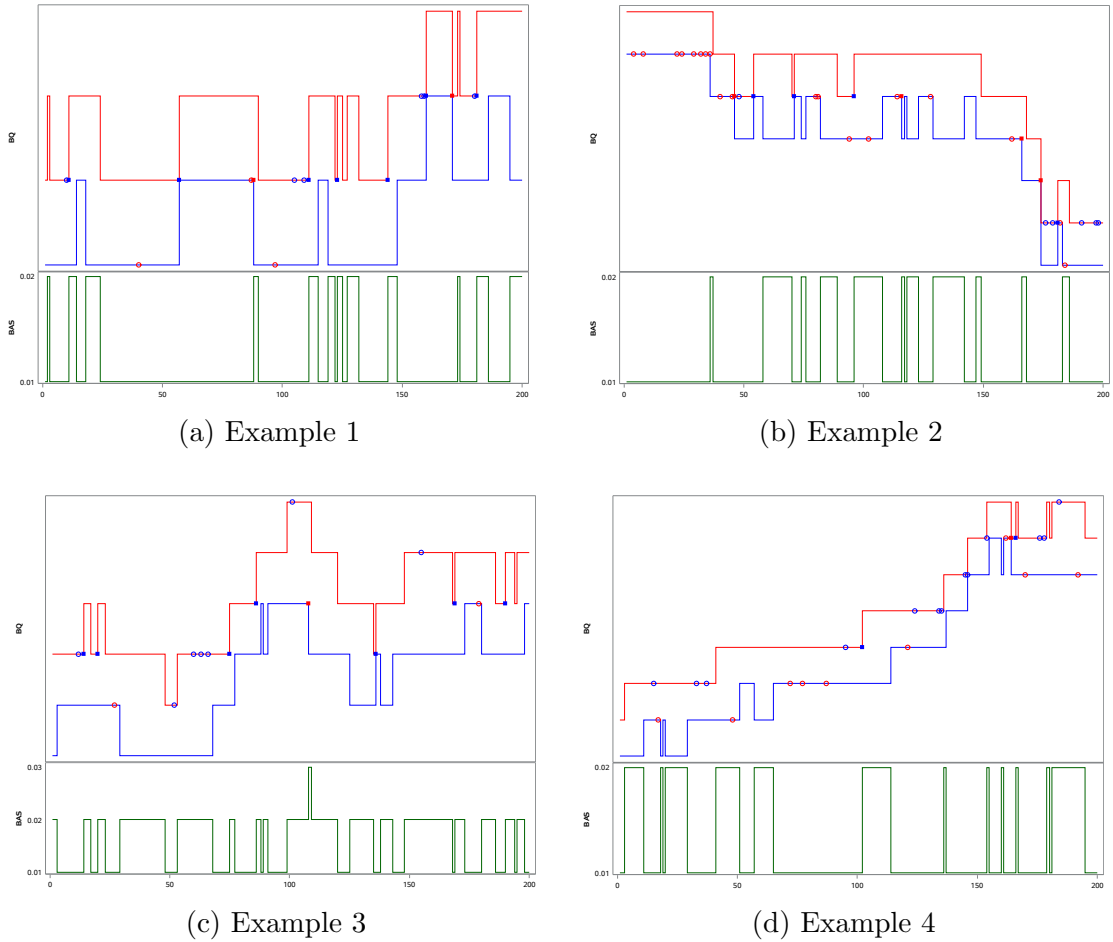


Figure 5.2: Visualizing Best Quotes & Trades

Notes: The examples contain the same best quote and bid-ask spread information as presented in Figure 5.1, but now includes signed trades. Blue marks for buyer-initiated and red marks for seller-initiated. Open circles mark trades that extract partial depth from the quote; filled squares mark trades that extract complete depth from the quote.

In this subsection we attempt to answer this question by showing two alternative ways to include visible depth — one based on levels within the order book and the other based on quote segment levels. We are still plotting information in event time and each event in our data corresponds to something that changes visible depth or quote levels at one or both of the best quotes. The purpose of adding depth information is to explore what is taking place at quote levels that define the market. For example, in plots of the best quotes, the unchanging quotes in our data reflect that something is changing with visible depth. Right before a bid-QP, we want to see if depth was adjusting on both sides of the market.

In the visualization, we present depth information above and below the best quotes to reflect the different sides of the market. Visible depth related to the sell side is placed above the best quotes plot and visible depth related to the buy side is placed below the best quotes plot. Since we want the visual to reflect both sides' willingness to move toward or away from the market, we want to display an increase of depth associated with the sell side as a movement down, or toward the buy side. Therefore, we flip the sign of a change in the ask visible depth to reflect a steady increase of depth at the best ask to move lower, toward the buy side.

Using the same four intervals of book updates, Figure 5.3 now shows visible depth at the best quotes associated with both sides of the market along with the best quotes. Each example contains three panels. The first panel contains depth information on the ask side. The second panel shows the best quotes (the best ask quote in red and the best bid quote in blue). The third panel contains depth information on the bid side.

At first glance, the inclusion of depth information looks very messy, but after some inspection a few general patterns emerge. One of the most clear patterns is related to depth activity following a QP event. This is clearly demonstrated in subfigure 5.3a just after 50, subfigure 5.3c between 50 and 100, and in subfigure 5.3d near 50 and 100. Another pattern is the jumps in depth created from QI and QP event sequences on the same side of the market. This is particularly clear in subfigure 5.3b. We can also see that there are build ups of depth just prior to QI events.

As we see above, there are a number of issues related to plotting depth information corresponding to the best quotes. Figure 5.4 now shows depth information

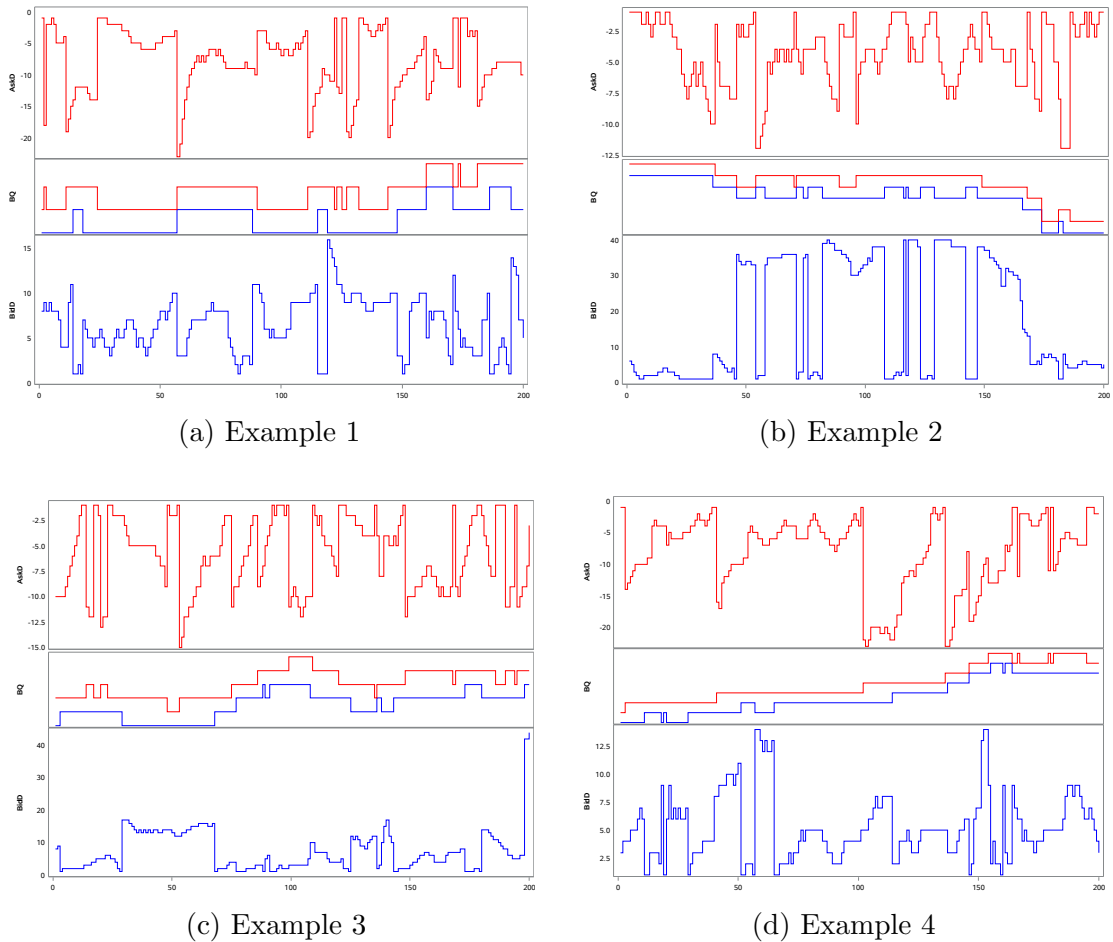


Figure 5.3: Visualizing Depth — Best Quotes

Notes: Each example is made up of three panels. The first panel shows visible depth on the ask side related to the best ask. The second panel shows the standard best bid and ask quotes. The third panel shows visible depth on the bid side related to the best bid. The ask depth is shown with a sign flip to have a depth increase move down, toward the bid side.

corresponding to the first two quote segment levels. As in the prior visual of depths at the best quote on both sides of the market, here we show ask depth information above the best quotes and bid depth information below the best quotes. Instead of just depth at the best quotes, we show depth at the first two quote segment levels.

Quote segments are defined by the location of the upper, inside and lower quote segment levels. For each quote segment, both sides have the option to locate the best quote at the inside quote segment level. In figure 5.4, the inside quote segment level is shown in black. The green line in the depth chart is related to the second

quote segment level — the upper quote segment for the ask side and the lower quote segment level for the bid side. This is most clear in subfigure 5.3a where the first 150 updates all occur within a single quote segment. The visualization highlights the relationship between observed depth across various quote segment levels. We see clear patterns of opposite movement in depth between both sides — when increasing, or moving toward the market, on the buy side, we find depth is decreasing, or moving away from the market, on the sell side.

Subfigure 5.3b highlights the advantage of plotting depth information according to quote segment levels. Starting just before 50, we see a large jump in depth at the lower quote segment level, which results from a newly defined quote segment. This large depth likely was already in place at the one-below lower quote segment level immediately prior. We can see that depths associated with the bid-QI events are very small. We can see there exists some clear movement toward and then away from the market at the upper quote segment level for the sell side.

Subfigures 5.3c shows how ask depth responds to ask-QP events — depth leaves the upper quote segment level, which also corresponds to the best ask quote. We can see the decrease is slow implying depth changes tend to be one contract in this futures market. The bid depth charts demonstrate how when markets are moving upward, toward the end of the quote segment, the buy side has established a position at the inside quote segment level and this quote level becomes the lower quote segment level when conditions are met to shift upward the quote segment. In our visual, this occurs when the black line switches to green.

At the start of subfigure 5.3d we see an ask-QP and the depth at the upper quote segment level moves away from the market. During this time, depth at the

lower quote segment is moving toward the market and eventually, the buy side establishes control of the inside quote segment level. With the next ask-QP, a new quote segment is defined and ask depth continues to move away from the market and the bid depth continue to move toward the market. At the next ask-QP, ask depth is initially large but we see this again move away from the market. At the moment, we are not able to tell if this decrease is due to trades or limit order cancellations, but the reduction in depth does make it easier to allow for further ask-QP events, which is exactly what we see.

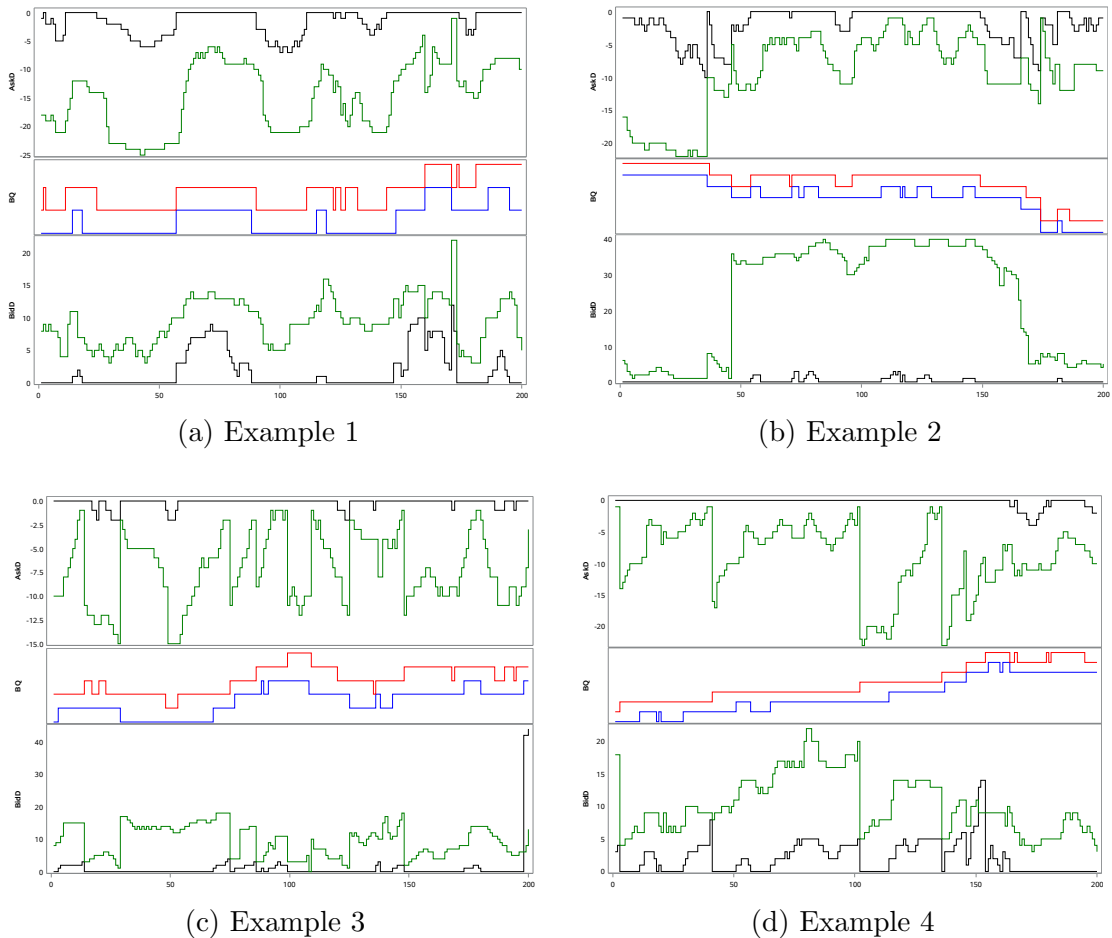


Figure 5.4: Visualizing Depth — Quote Segment Levels

Notes: These figures correspond to the same sets of book updates as shown in Figure 5.3, but here, instead of using depth associated with the best quote, we show depth associated with the first two quote segment levels for both sides of the market. Within each depth panel, black corresponds to the inside quote segment level; green corresponds to the upper quote segment level for the sell side or the lower quote segment level for the buy side. The ask depth is shown with a sign flip to have a depth increase move down, toward the bid side.

5.4 Combining Depth and Quotes

This section describes a method to visualize high frequency limit order book information that combines the three data elements (quotes, depth, and trades) into a single figure which allows for a more complete visualization of market dynamics. First, we describe the process to combine depth and best quotes. Second, we complete the visual by including signed trades. The end result is a visual that allows the viewer to observe the sequence of events that unfolds due to the order routing/management and trading decisions of the complete market.

5.4.1 Combining Depth Changes with Best Quotes

The goal here is to combine the depth changes with quote level information on both sides of the market. We do this by accounting for incremental changes in visible depth resulting from the arrival of book update messages. For example, focusing on the bid side, we start the visual by starting the best bid quote level at zero, then quote improvements result in the quote level moving to $+0.01$ for a one tick improvement. If prior to the quote improvement, we observe an increase of one contract in the visible depth, we add $+0.001$ to our current quote level. This remains fixed at the new level until we observe another change in visible depth; a second increase of one contract would result in the quote level moving to $+0.002$. Once the quote level improves and moves upward by one tick, we jump from $+0.002$ to $+0.01$. Similarly, an increase of depth by two contracts results in a new quote level of $+0.012$.

We do not account for visible depth that initiates a quote improvement or initial depth at the second-best quote once it is promoted to the best quote level

— we only consider changes to the visible depth once the best quote is established. For example, if a new best quote arrives for ten contracts one tick above the prior best quote, then we still only shift up the adjusted bid quote level by $+0.01$. Likewise, once the best quote is removed and there is a quote promotion, we shift the quote level down -0.01 , regardless if the resting depth is 100 or 15, all we consider are changes from this initial depth.

The visualization needs to capture individual market sides either moving toward the market or away from the market; movement toward the market indicates the ask quote moving toward the bid quote or ask depth increasing at the best ask quote. For the bid side, no adjustments are necessary — the bid side moves toward the market (adjusted quote level increases) as either the quotes move up or depth increases. But for the ask side, we need to flip the direction of depth such that the ask adjusted quote level moves downward when there is an increase in visible depth.

Figure 5.5 shows the same four intervals of updates as shown in prior visualizations. Within each subfigure, the first and third panels reflect the same information as those observed in figure 5.4. Here, in the second panel, we have the best quotes adjusted for changes in depth at the first book level; the fourth panel shows the unadjusted best quotes for both sides of the market.

From this figure set, we can see how the adjusted best quotes captures dynamics on both sides of the market — as depth is flowing out of the best ask, depth is flowing into the best bid. Depth flows out of a best quote prior to a QP event. Subfigure 5.5a highlights the tendency for flows to reverse after market movements — depth is leaving the best ask and increasing at the best bid, then this directional

5.4.2 Combining Depth Changes & Trades with Best Quotes

Visible depth can decrease either from the cancellation of resting limit orders or due to transactions, resulting from the arrival of market orders that hit the best quote. Therefore, it is important to note any trades in the visual of depth changes to allow the viewer to distinguish between these two type of events. Figure 5.6 focuses on just the adjusted best quotes and now includes marks when trades occur similar to the marks defined in figure 5.2. We adjust the actual traded price to reflect the prior movement in depth.

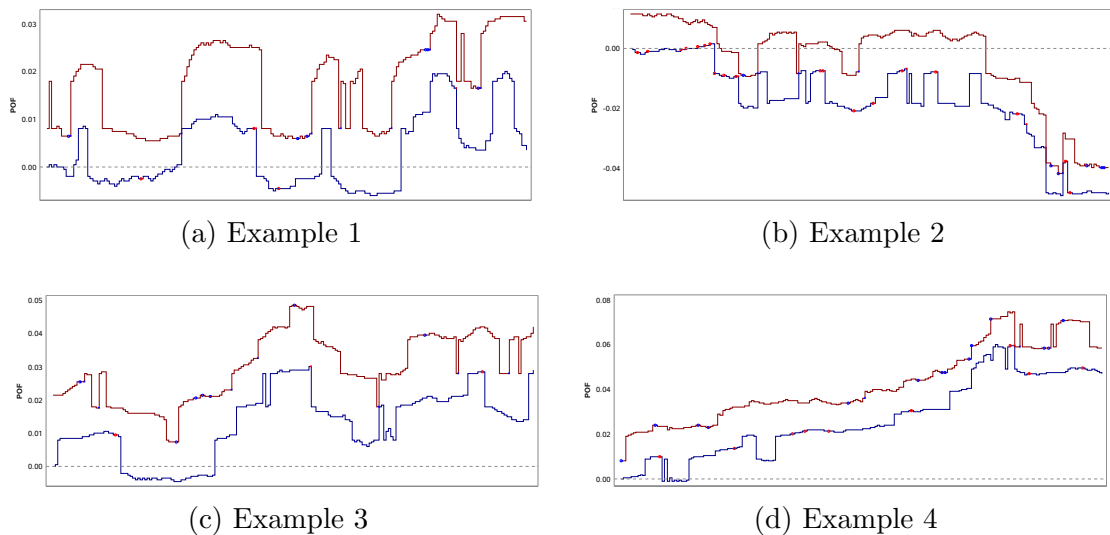


Figure 5.6: Visualizing Adjusted Quotes with Trades

Notes: These examples focus the second panel found in Figure 5.5 and includes trade information. As in Figure 5.2, buyer-initiated trades are marked in blue and take place on the adjusted ask quote; seller-initiated trades are marked in red and take place on the adjusted bid quote. Open circles indicate trades that remove a subset of visible depth; closed squares indicate trades that remove all available depth from the best quote.

Figure 5.6 shows how much of the observed decreases in visible depth are not related to transactions, but result from limit order cancellations.

5.5 Adding Depth at Deeper Levels

The sections above have set up the adjusted best quote visualization that captures changes in best quotes, adjustments to visible depth at the best quotes, and the arrival of signed trades focusing only on trades that interact with the best quotes. Since this only captures information at the best quotes, we want to add additional information that captures visible depth levels deeper in the book.

We accomplish this by adding two panels. The first new panel is placed above the adjusted best quote visualization and shows visible depth at five ask quote segment levels. These levels are based on our quote segments, so the first quote segment level represents the inside quote, the second quote segment level represents the upper quote, the third represents one above the upper quote, etc. We also reverse the sign of these depth levels for the ask (sell) side, which allows for larger depth levels to extend toward the bid (buy) side. The second new panel is placed directly under the adjusted quote visualization and shows the corresponding bid side information.¹

Figure 5.7 shows an example of the complete visualization. There are three panels corresponding to ask depth quote segment levels, adjusted best quotes with best quote level and depth changes and signed trades, and bid depth quote segment levels. In these four examples, we find there is not the same kind of response in visible depth as we move to quote segment levels farther from the market. This observation corresponds with our findings in prior chapters: depth changes are less concentrated at book levels farther from the best quotes. In the first example,

¹For the bid side, the first quote segment level corresponds to the same inside quote from the ask side. The second quote segment level represents the lower quote, the third represents one below the lower quote, etc. We do not flip the sign of visible depth on the bid (buy) side since increases in visible depth move toward the ask (sell) side.

the red and orange lines are close to being horizontal, which is similar to all the other intervals as well. In Figure 5.7c we do see similar movement in the green and blue quote segment levels as the market ticks upward — this makes sense as during this time, the blue line represents the second-best ask book level.

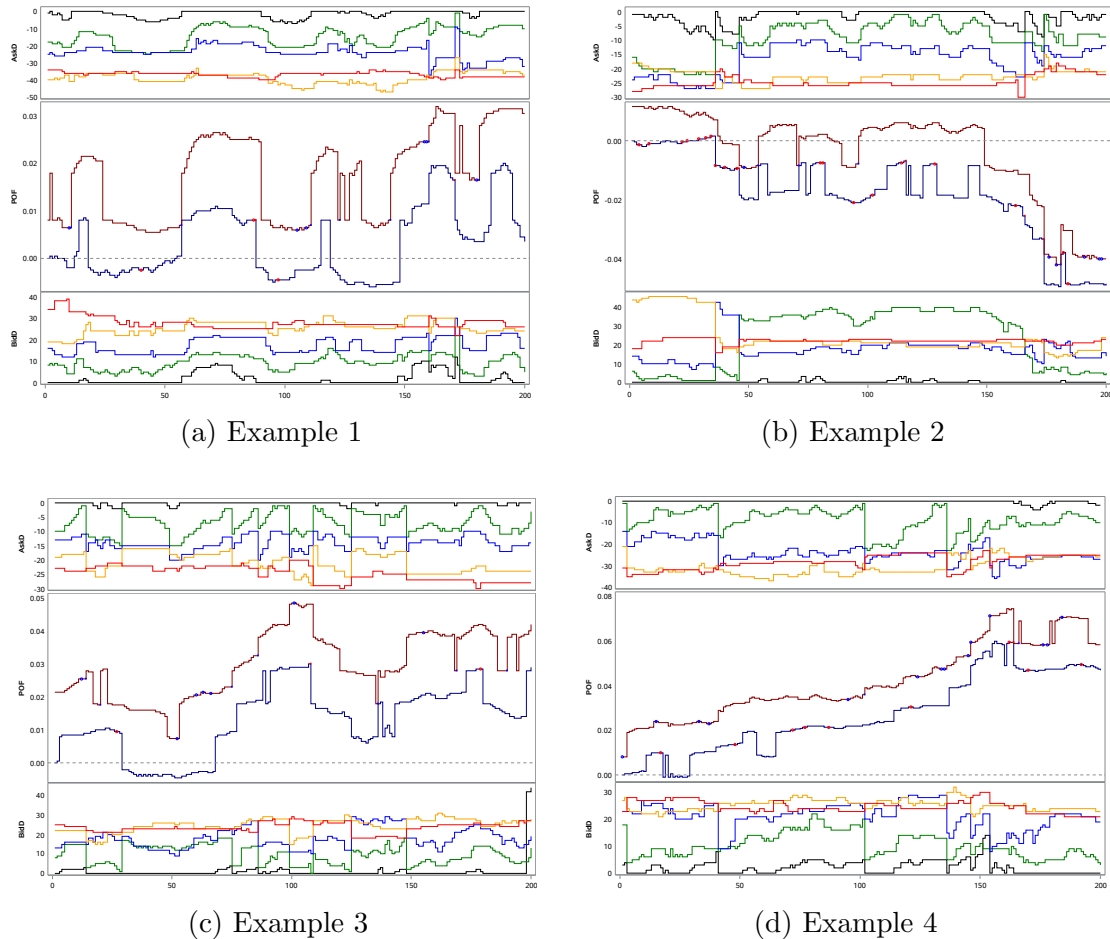


Figure 5.7: The Complete Visual: Adjusted Quotes with Depth Information

Notes: This figure is composed of four example sets of 200 best quote book updates. Each example has three panels. The upper panel corresponds to the visible depth on the sell side for the first five quote segment levels: black marks the inside quote segment level, green marks the upper quote segment level, blue marks the one-above upper quote segment level, red marks the two-above upper quote segment level, and orange marks the three-above upper quote segment level. Signs of depth changes are flipped for the ask quotes to show movement toward the market for depth increases. The middle panel shows the adjusted best quotes with trade information. The lower panel shows visible depth on the buy side with similar ordering as found in the upper panel, but now we focus on lower quote segment levels.

Examples of Calm and Volatile Markets:

Figure 5.8 contains two examples of book updates and highlights the difference between a calm market and one just after an expected public news announce-

ment. Immediately following a major news release, the market experiences elevated volatility and a wide bid-ask spread. For each example, we show one visual of the unadjusted best quotes with trades and the measured bid-ask spread; a second visual, placed just below, shows the adjusted best quote visual with visible depth across multiple quote segment levels.

During the calm period, the set of updates are spread across 18 seconds and we see similar movements in depth at quote segment levels closest to the market with occasional intervals of a two-tick bid-ask spread. The second example, which starts one second after 10:30am inventory news release and lasts for 1.1 seconds, shows many trades with lots of variation in the bid-ask spread — observed as high as ten ticks.

Comparing unadjusted with adjusted best quote plots, we find they are most similar just after an expected news event where the market is quickly trying to incorporate information and is true since quotes levels do not exist very long if we focus on event time (or clock time for that matter). Visible depth across all quote segment levels on both sides of the market are very low indicating a low amount of available liquidity in the order book following an expected news event.

Examples of Directional Movement in Best Quotes:

Figure 5.9 contains two examples of book updates and highlights a market in movement. In both examples, we see there are clear subsets of a market in motion: in the first example, directional movement upward starts close to update 150, while the second example starts with directional movement downward. During these moves, we find realized trades take place toward the end of the quote and prior to

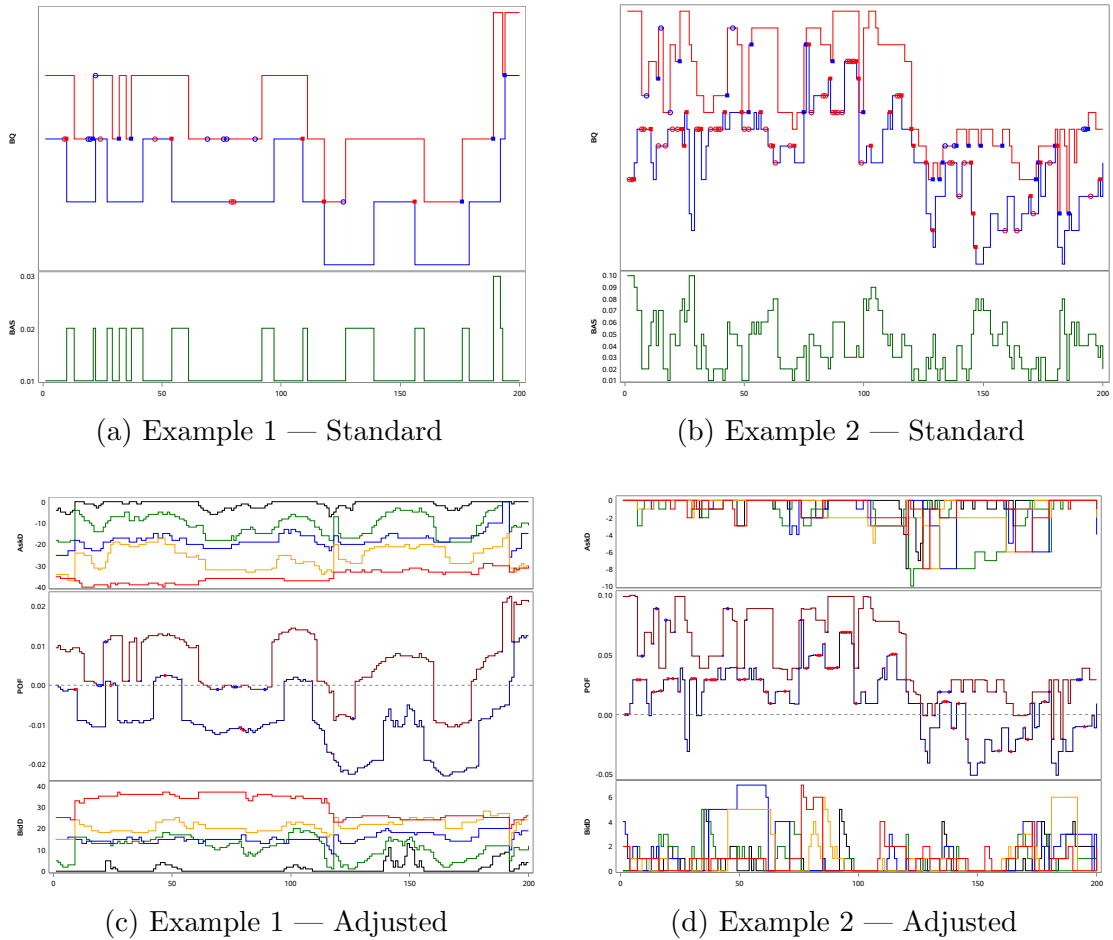


Figure 5.8: Additional Examples: Calm and News Announcements

Notes: This figure contains two sets of 200 best quote book updates where columns display the same interval in two ways. The first column corresponds to a set of 200 book updates during an interval with low volatility; the second column corresponds to a set of 200 book updates following the release of a major news announcement at 10:30 a.m. Within each column, the first subfigure shows standard best quotes with bid-ask spread and trade information; the second subfigure shows our adjusted best quotes with visible depth across the first five quote segment levels.

trades, depths are reflecting the same market movement — upward (downward) markets contain the departure of ask (bid) depth and the increase of bid (ask) depth.

According to the way we adjust the best quotes with observed changes in depth, it is possible for the adjusted best ask quote to touch or move below the adjusted best bid quote. We see an example of this in Figure 5.9c as the depth at the upper quote segment level steadily increases and moves closer and closer to the adjusted bid quote. This suggests a measure, such as the difference between

the adjusted quotes, can be used to determine the ‘closeness’ of the market. If depths are increasing on both sides, then we can imagine cases where this measure can take on negative values. We would expect such cases to be associated with a mixture of buyer and seller-initiated trade volume as both sides have an interest to trade and support the quotes as reflect by the willingness to provide liquidity at the best quotes.

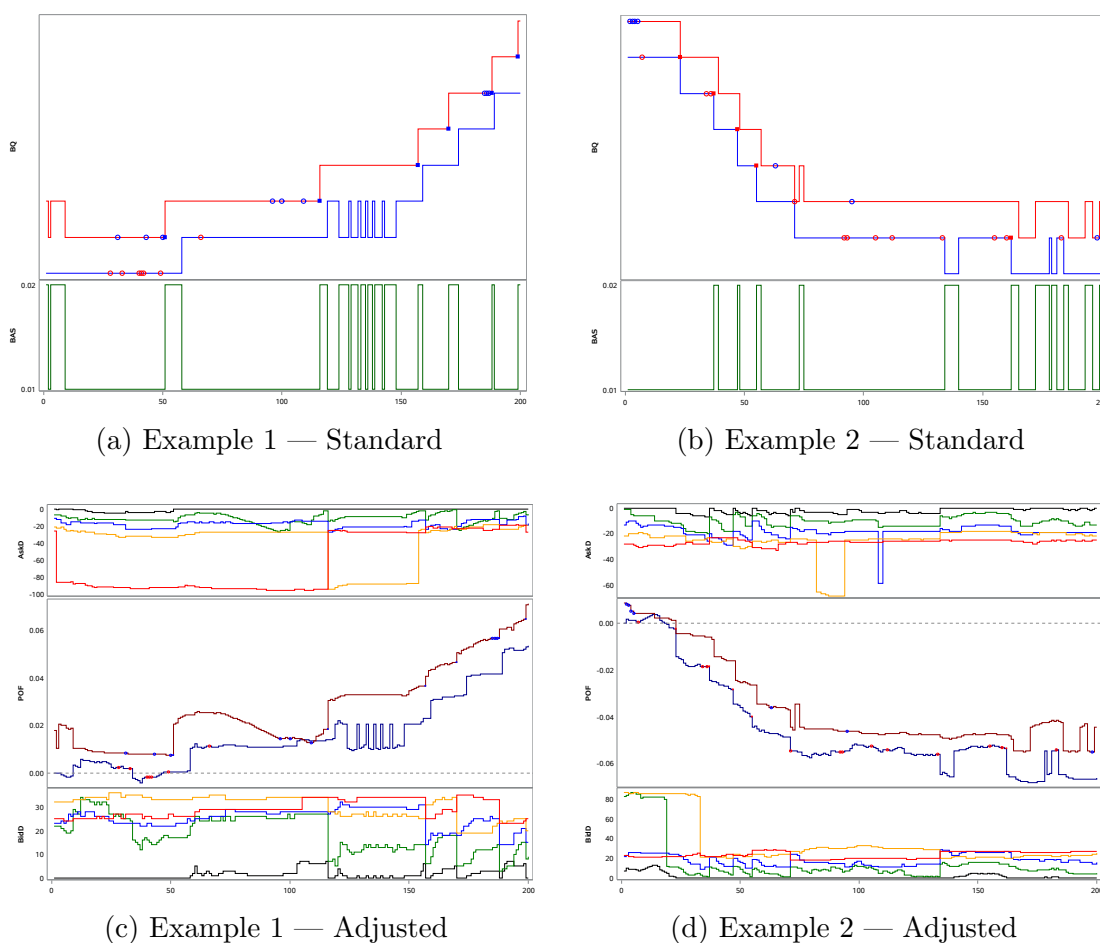


Figure 5.9: Additional Examples: Market Movement

Notes: This figure contains two sets of 200 best quote book updates where rows display the same interval in two ways. The first row corresponds to information found in Figure 5.2; the second row corresponds to the information found in Figure 5.7. These two intervals are selected to highlight how the visual helps make sense of market movement.

Additional Examples of Market Dynamics — Narrow Range:

Figure 5.10 contains six examples of the complete visualization now covering 500 book updates. Each example highlights intervals of updates that focus activity within a narrow range of quotes. Each example shows many instances of QI events at both sides of the market. Explains why we observe sequences trade prices alternating between buyer- and seller-initiated at the same quote level — also helps justify the defined quote segments where the inside quote segment level alternates between the buy and sell side. Visualizations clearly show a regular pattern of flows in and out of the market at the first three quote segment levels

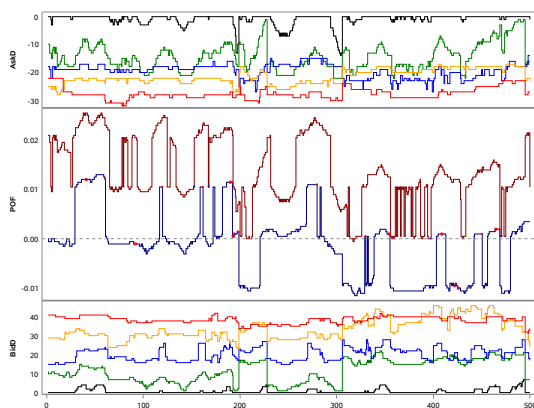
Additional Examples of Market Dynamics — Price Discovery:

Figure 5.11 contains six examples of the complete visualization now covering 500 book updates. Each example highlights intervals of updates that show price discovery. As markets move up or down, we find depth stays constant. The second row has two examples highlighting how large depths appear deeper in the book and the market doesn't move toward them. Last row contains two figures that can be combined to show 1,000 book updates. We see a market move downward which results in a large visible depth deeper in the book. The large buy support is never tested as the market later moves upward.

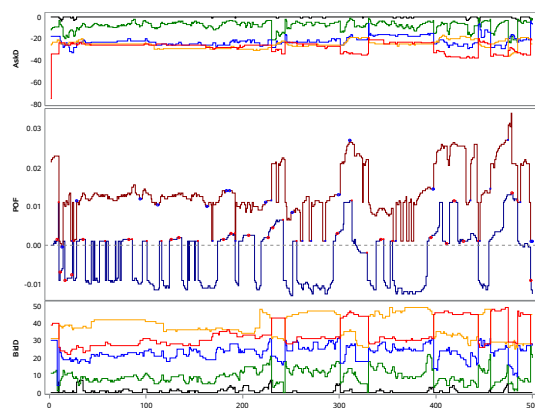
Additional Examples of Market Dynamics — Price Wave:

Figure 5.12 contains six examples of the complete visualization now covering 500 book updates. Each example highlights intervals of updates that show movement up and down in the prices. These examples touch on the positive autocorrelation

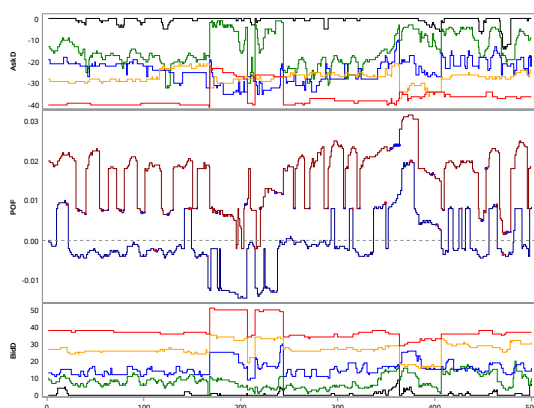
in the return series of the inside quote segment price level. Here we see many sequences of directional movement in the quote segments that often return back. Examples 1, 5, and 6 show large depths that show up but do not result in trades. Differences in observed depth provide some indication of prior market activity.



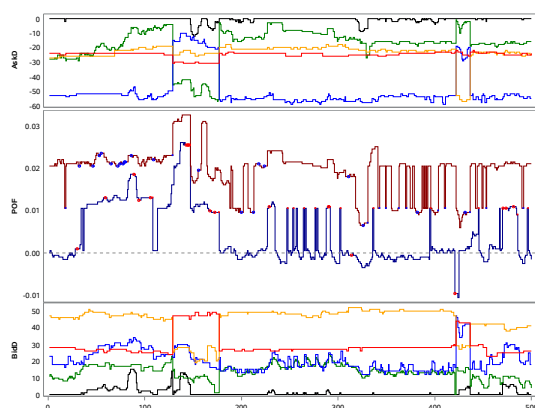
(a) Example 1



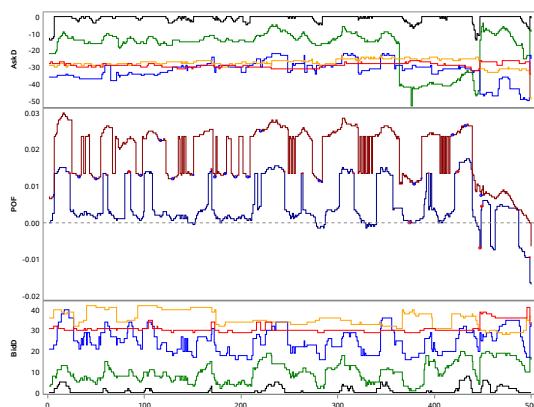
(b) Example 2



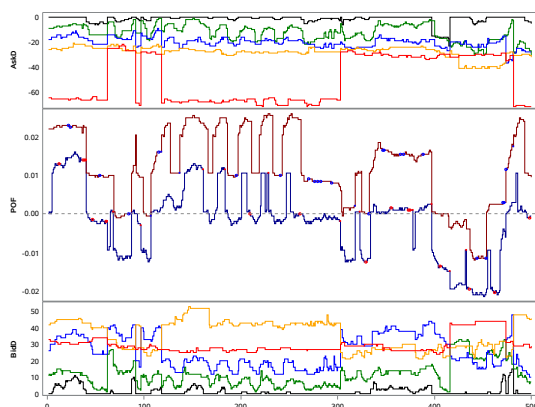
(c) Example 3



(d) Example 4



(e) Example 5



(f) Example 6

Figure 5.10: Complete Visual: Narrow Range

Notes: This figure is composed of six example sets of 500 best quote book updates. Each example has three panels. The upper panel corresponds to the visible depth on the sell side for the first five quote segment levels: black marks the inside quote segment level, green marks the upper quote segment level, blue marks the one-above upper quote segment level, red marks the two-above upper quote segment level, and orange marks the three-above upper quote segment level. Signs of depth changes are flipped for the ask quotes to show movement toward the market for depth increases. The middle panel shows the adjusted best quotes with trade information. The lower panel shows visible depth on the buy side with similar ordering as found in the upper panel, but now we focus on lower quote segment levels.

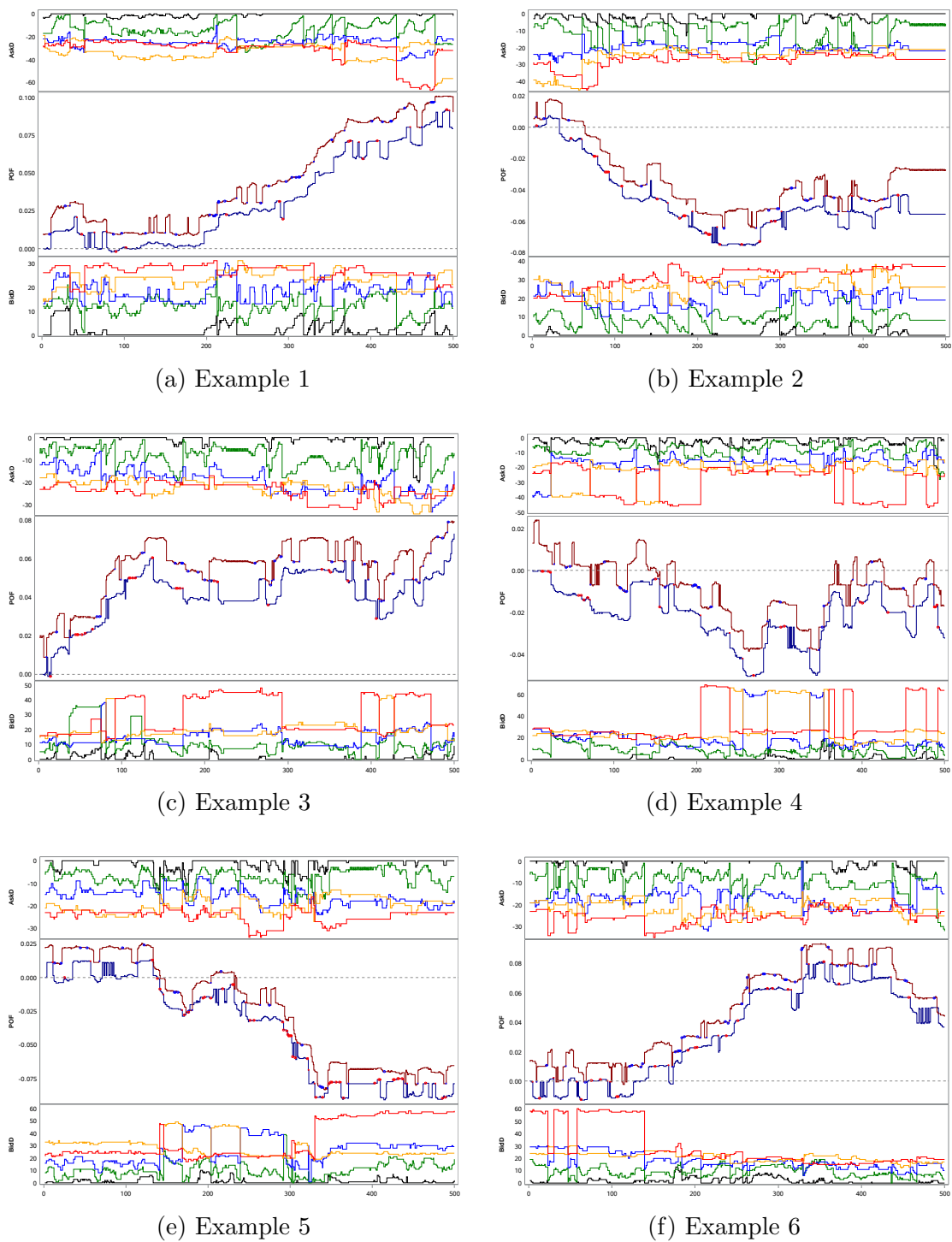


Figure 5.11: Complete Visual: Price Discovery

Notes: This figure is composed of six example sets of 500 best quote book updates. Each example has three panels. The upper panel corresponds to the visible depth on the sell side for the first five quote segment levels: black marks the inside quote segment level, green marks the upper quote segment level, blue marks the one-above upper quote segment level, red marks the two-above upper quote segment level, and orange marks the three-above upper quote segment level. Signs of depth changes are flipped for the ask quotes to show movement toward the market for depth increases. The middle panel shows the adjusted best quotes with trade information. The lower panel shows visible depth on the buy side with similar ordering as found in the upper panel, but now we focus on lower quote segment levels.

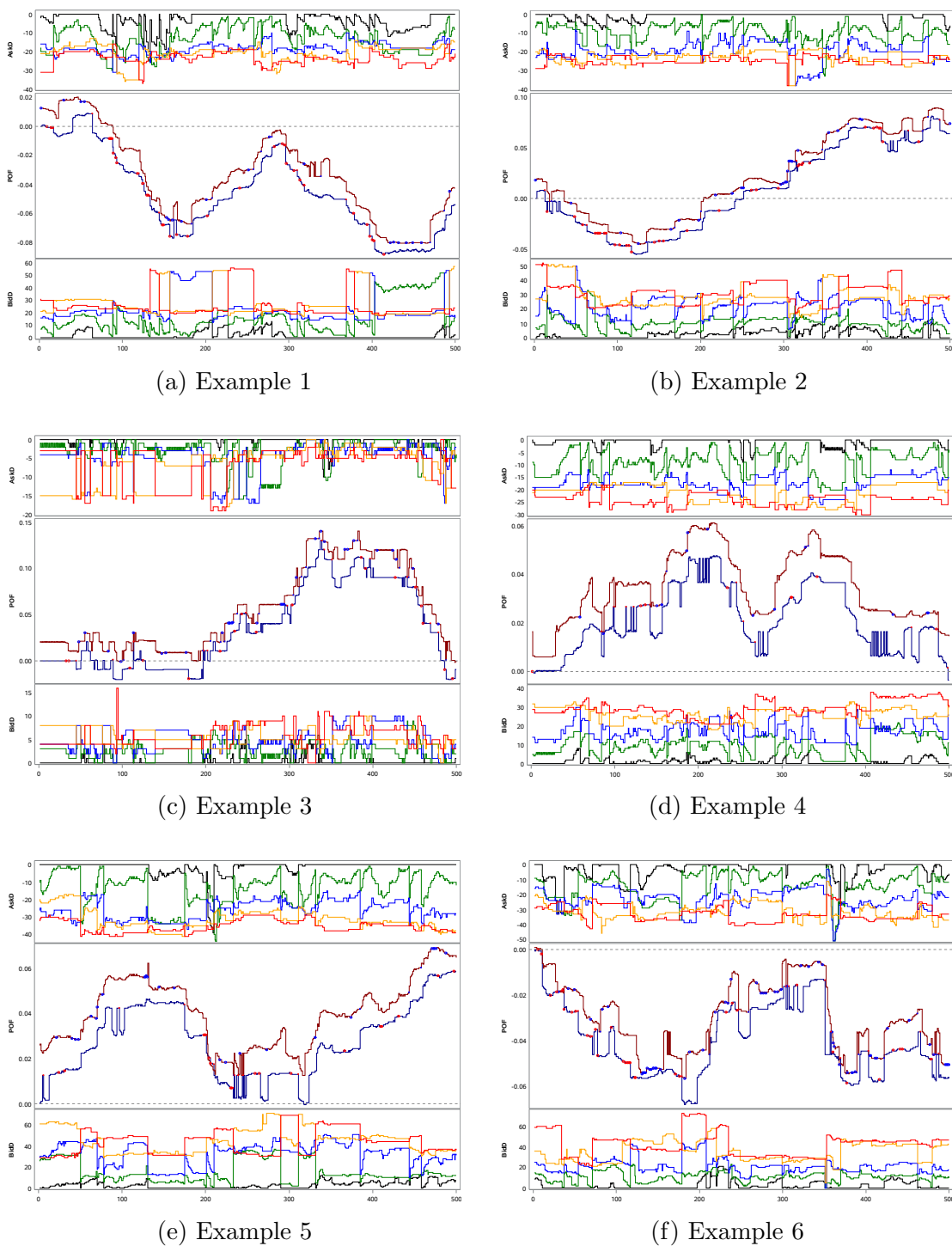


Figure 5.12: Complete Visual: Price Wave

Notes: This figure is composed of six example sets of 500 best quote book updates. Each example has three panels. The upper panel corresponds to the visible depth on the sell side for the first five quote segment levels: black marks the inside quote segment level, green marks the upper quote segment level, blue marks the one-above upper quote segment level, red marks the two-above upper quote segment level, and orange marks the three-above upper quote segment level. Signs of depth changes are flipped for the ask quotes to show movement toward the market for depth increases. The middle panel shows the adjusted best quotes with trade information. The lower panel shows visible depth on the buy side with similar ordering as found in the upper panel, but now we focus on lower quote segment levels.

5.6 Conclusions

This chapter has proposed a new way to visualize the dynamics of the limit order book using ultra-high frequency market depth data purchased from the CME. The rich data presents an opportunity to create a visualization that captures the various aspects of the market and allows us to learn how the market functions as participants submit market and limit orders. We expect as more data becomes available, visualizations such as this will become more valuable to study how the market operates.

Our primary contribution is combining the best quotes with changes in visible depth, which allow us to observe how both sides of the market interact and respond to each other. The construction of the visualization clearly shows how the sell side and the buy side adjust market depth and how this movement helps explain future market events. Along with the adjusted best quotes, the use of defined quote segments allows us to enrich the visualization by incorporating visible depth at levels deeper in the limit order book in an effective and accessible way.

The complete visualization confirms empirical findings from prior chapters. For example, the distribution of updates according to the distance from the best quotes. The visualization clearly shows how quotes closest to the market are most likely to adjust as new orders arrive and resting orders are cancelled. We also can confirm the existence of QI and QP events as the best quotes change according to trades that remove all resting depth or quotes being cancelled. We also can use the visualization to support our definition of quote segments based on the frequency of both sides tending to share the inside quote segment level along with the frequency of the bid-ask spread moving between a one-tick market to two-ticks.

Studying the complete visualization has allowed us to identify two main research questions that are studied in Chapters 6 and 7. First, the visualizations clearly show instances of fleeting liquidity — defined as intervals of time with changes in visible depth related to the arrival and cancellation of resting limit orders closest to the market. Observing depth suddenly leave across the first three bid quotes and, at the same time, depth suddenly enter across the first three ask quotes raises a number of questions related to how the market is adjusting. Is this simply a display of dynamic pricing with one side responding to the other? How can we measure this responsiveness of visible depth? These questions form the foundation for Chapter 6.

The second observation is related to flickering or flash quotes which is clearly displayed in the visualization. This appears to be very different from the observation of fleeting liquidity since flash quotes tend to occur on both sides of the market. How long do such quotes actually live in the limit order book and does the frequency of observing such action signal anything about prior or future volatility or volume? Are these flash quotes motivated by liquidity? How are flash quotes related to algorithmic trading since they seem to only exist for the fastest market participants? These questions form the foundation for Chapter 7.

The proposed method to visualize the dynamics of the limit order book can be extended in a number of ways. First, the current visualization is based on event time and while that allows for visual inspection of each update, there might be a desire to see market events unfold with some reference to the timestamp related to updates. The movement from event time to clock time would therefore be a natural extension. A second possible extension would focus on alternative ways

to change the size of depth to the adjusted best quotes. The selection here would depend on the objective of the visualization and how much emphasis is placed on book updates that adjust visible depth. Various other possible extensions relate to how visible depth is displayed at deeper levels in the limit order book.

Chapter 6

Fleeting Liquidity

6.1 Introduction

This chapter reveals the nature of fleeting liquidity in the electronic limit order book. There are many questions surrounding modern order driven markets. These include the movement toward a fully electronic limit order book and the rise of high frequency traders who design algorithms to execute orders and monitor the limit order book. Here we focus on questions surrounding visible liquidity in the limit order book.

At any moment in time, one can observe visible depth at quotes closest to the market. This visible depth provides information about prior decisions of market participants to allow limit orders to rest in the limit order book. This information is displayed to the subset of market participants with the ability to monitor the market depth data in near real time. By using the same market depth data that is sold to market participants, we are able to investigate the value of such information.

One aspect of the depth data, apart from being able to measure best execution, is the ability to monitor how the limit order book responds to market events such as trades or changes in visible depth at the best quotes or deeper in the limit order book. The limit order book data captures the dynamics of changes in visible depth. This allows participants to observe activity from both the buy and sell side.

In any market, there is a constant tension between the buy and sell side. This tension is intensified in a market such as futures when participants can easily switch sides depending on their expectations. For example, if the buy side observes a sudden expression of interest from the sell side, the buy side should respond by lowering their bids. Market expectations should adjust when new information arrives, but at what speed? Should we expect a large fraction of visible depth to suddenly disappear at the slightest hint of negative news? How soon should market participants be allowed to change their view of the market? Responding to information will lead to one side adding depth and the opposite side cancelling depth, but how can the limit order book be used to capture this response?

In this chapter, we propose a method for measuring responsiveness of the limit order book by using observed changes in the visible depth across quote levels closest to the market. We have shown evidence in Chapter 3 how directional flow of depth on both sides can be used to help explain the existence of wide bid-ask spreads. This raises questions about why we observe such patterns. Did the visible, available, offered liquidity suddenly exit? Is this part of some natural re-balancing of depth after an event that changes the best quotes?

Based on visual inspection of the market, it is evident that during certain periods, the responsiveness varies to a great degree. Figure 6.1 shows six examples

of directional movement of our adjusted best bid and ask quotes (See Chapter 3 for details on the visualization). In each example, we highlight a particular sequence that displays market movement due to a combination of trades, limit order cancellations and arrival of new limit orders. In Examples 1 and 5, the market moves upward and we see the bid side builds depth at each new bid level while the ask side moves away from the market due to a combination of buyer-initiated trades and cancellations of resting limit orders at the best ask. The rest of the examples show market movement downward, and likewise, show the ask side moving toward the market by establishing new limit orders at each new best ask quote.

The traditional view of the limit order book focuses on a dealer who places quotes and expects to earn the spread. Modern order driven markets, on the other hand, have no defined dealer and anyone can place limit orders or provide liquidity to the limit order book. If you can observe depth, then you can determine the cost of filling an order. But if the market is responsive, then you cannot trust the visible depth will remain if you slowly work the order. Responsiveness is a characteristic of visible depth. An understanding of what drives responsiveness will help us understand how the limit order book works.

This research makes use of a rich data set capturing the dynamics of the limit order book. Once we define a quantity representing offered liquidity at a fixed point in time, we then propose a method to measure the rate of change on both sides of the market per millisecond, which becomes our measure of responsiveness. The propensity of a market to shift suddenly, or the likelihood of seeing a large shift in the limit order book over a short time interval, should contain information

related to the expected level of informed trading. The propensity to shift should also depend on the markets location to true liquidity demand or observed quote levels that do not display such tendency to leave the book.

This research is relevant since it helps us understand how the limit order book functions. This line of research reveals the true intent to trade. If a trader always cancels expecting if they cancel at the best quote, they increase the odds of a lower bid quote being hit, are they manipulating the market? Or are they accurately judging the willingness of impatient, informed sellers to sell at a lower price? This is also related to how algorithms use the depth data to determine market reaction at quotes that define the market.

The remainder of the chapter is organized as follows: Section 6.2 provides a discussion of offered liquidity and responsiveness of offered liquidity. Section 6.3 reviews the data and methodology. Section 6.4 provides a summary of offered liquidity and responsiveness. Section 6.5 studies events related to the start of quote segments. Section 6.6 studies directional movement in quote segments. Section 6.7 concludes and summarizes the chapter.

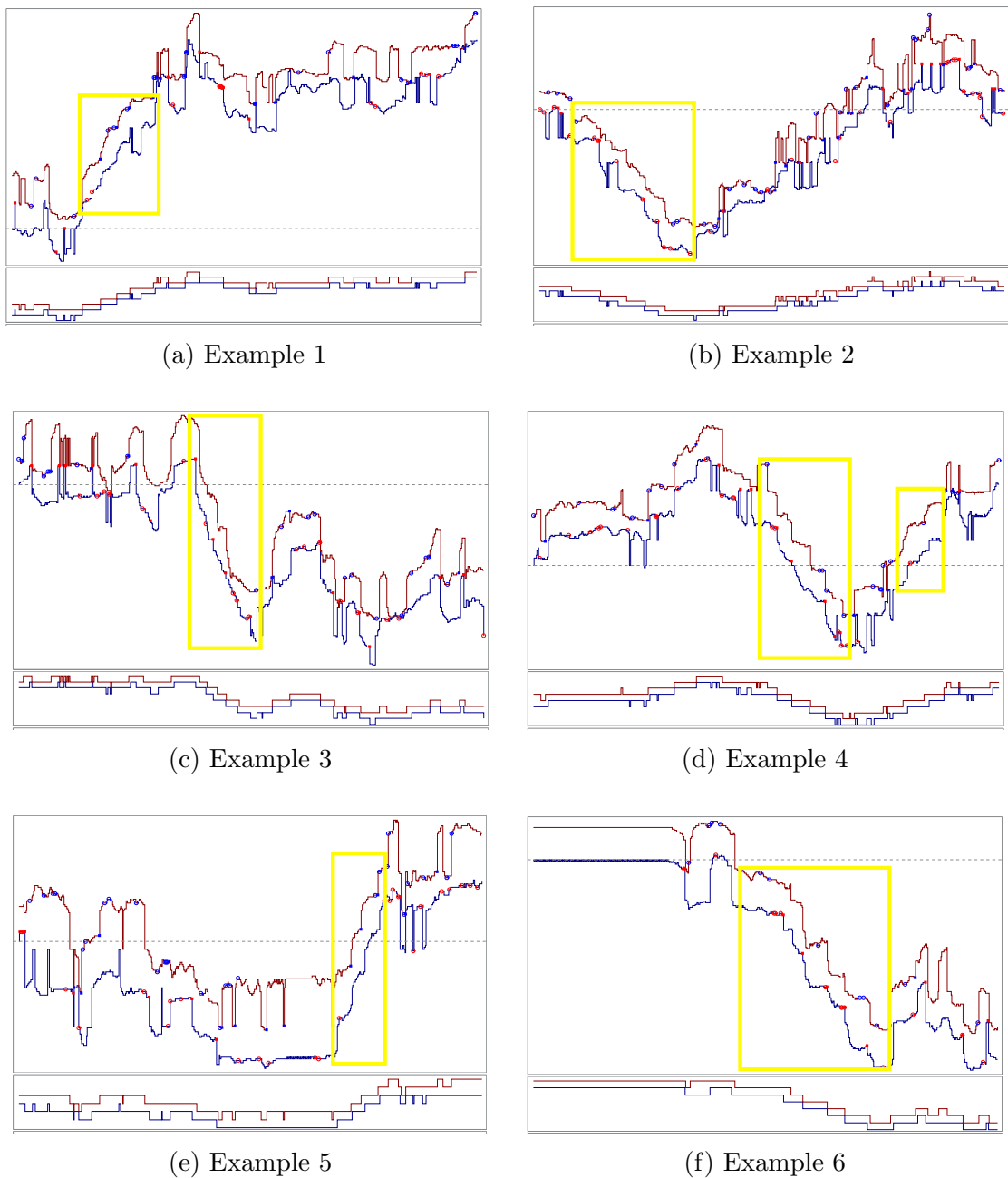


Figure 6.1: Six Examples of Fleeting Liquidity

Notes: Each figure shows a sample of order book data for a fixed interval of 500 book updates (updates not shown spaced out according to time) to the best bid (blue) and ask quote (red). Within each figure, the upper section shows the best bid and ask quotes adjusted to reflect changes in market depth. Ask quote is adjusted to move toward the best bid quote if the arrival of limit orders add depth to the quote. Trades are marked on the quote that provides liquidity — red for seller-initiated trades and blue for buyer-initiated; open circles indicate trades that did not extract all available liquidity, filled in squares indicate all liquidity removed from trades and the best quote expires. The lower section of each figure shows the standard quotes for both sides without incorporating information on market depth. Highlighted sections display the quote behavior of interest.

6.2 Offered Liquidity & Responsiveness

In this section, we provide a more detailed discussion of offered liquidity (OL) and a discussion of responsiveness of offered liquidity (ΔOL).

6.2.1 Offered Liquidity (OL)

Offered liquidity (OL) at a specific quoted price level is defined as the number of contracts that are available to initiate a trade. Offered liquidity results from the decisions of market participants to place limit orders which rest at this price level to either buy or sell — observed in our data as market depth. If this resting limit order eventually results in a trade, then it is matched with someone who removes this liquidity with a marketable order at the specific price level. Based on the available information, at any point during the day, we can pause the market and determine the amount of offered liquidity on both sides of the market.

We can define offered liquidity to span multiple book levels. By allowing offered liquidity to span multiple quotes, this captures the total number of visible contracts available with a well-defined price impact.

Figure 6.2 presents a conceptual visualization of offered liquidity, where each example reflects a quantity associated with resting limit orders, representing visible market depth at a fixed point in time for the sell side (in red) and the buy side (in blue). The vertical distance between the sell and buy offered liquidity reflects the bid-ask spread. The first set shows three different examples of balanced offered liquidity. The second set shows hypothetical examples with an imbalance between market sides.

Why might we expect the market to find itself in each state? First, considering

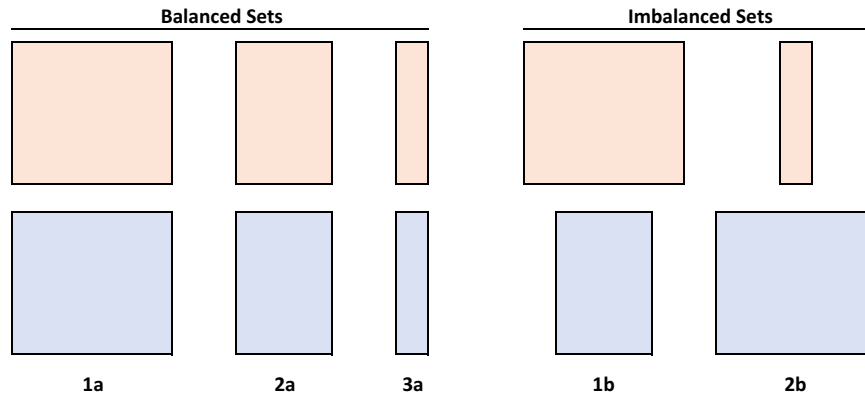


Figure 6.2: Balanced and Imbalanced Offered Liquidity

Notes: Figure shows two sets of offered liquidity on both sides of the market (sell side in red; buy side in blue). The first set on the left shows large balanced offered liquidity (1a), medium balanced offered liquidity (2a), and small balanced offered liquidity (3a). The second set of figures highlights situations where there exists an imbalance in offered liquidity between market sides.

different levels of balanced offered liquidity, there are naturally periods of the day where there is less uncertainty in the market that would explain why we find the market in large, balanced and small, balanced states. For example, during the daily settlement period there are large orders placed on both sides of the market; we would therefore expect to find large balanced, offered liquidity. Prior to expected news releases, the limit order book thins out; if there exists a great uncertainty then there is very little support for the current market price and we should expect both sides to symmetrically back away from submitting limit orders and a low, balanced state of offered liquidity would result. When we find OL is low on both sides, then there is less support for the current market price and it should indicate the increased chances of a possible movement up or down.

Different states of imbalance are fundamentally different from balanced states given one side has a size advantage over the other. A large imbalance favoring the buy side reflects a larger support for the market price from the buy side, relative to the sell side.

We can think of different reasons why an imbalance exists. Consider a story where the buy side has just aggressively moved toward the market. If that is the case, then the large buy offered liquidity means that the buy side is probably still willing to move the market upward; this also means it will take more effort from the sell side to push the market downward. In this case, it is also meaningful to question the likelihood of the arrival of ask quotes to re-establish the ask offered liquidity and re-establish support from the sell side for the current market price.

The interpretation of an imbalance in OL changes if the prior activity is due to aggressive sell side pushing the market down until we reach a large offered bid depth level. Here, the low offered liquidity on the ask side is simply a function of the market recently moving down and the newly established ask quotes have not had time to properly fill up. Regardless, given the state, the sell side has less support for the current market price. If the aggressive price movement from the sell side was unwarranted, the buy side will not face a lot of resistance for pushing the market higher.

This discussion highlights the challenges in interpreting snapshots of offered liquidity. It becomes necessary to have some prior information to better understand how the offered liquidity has developed.

6.2.2 Responsiveness of Offered Liquidity (ΔOL)

Ultimately, we want to investigate the dynamics of OL : How does offered liquidity change? We point to three types of action that results in a change in OL : first, **trades** decrease offered liquidity (e.g., buyer-initiated trades reduce the offered liquidity on the sell/ask side); second, the arrival of limit orders increase offered

liquidity (e.g., the **arrival** of bid limit orders increase the offered liquidity on the buy/bid side); and finally, the **cancellation** of limit orders decrease offered liquidity (e.g., cancellations of bid limit orders reduce offered liquidity on the buy/bid side).

Each type of action sends a specific message to market participants who are able to observe such information in real time. Trades are naturally bullish conditional on the side initiating the trade. Buyer-initiated trades indicate elevated interest from the buy side. Similarly, the arrival of bid limit orders indicate elevated interest from the buy side. In both of these cases, the buy side is moving toward the sell side. Conversely, the cancellations of bid limit orders indicates the buy side has lost interest and the buy side is moving away from the sell side (or moving away from the market).

Figure 6.3 presents a hypothetical sequence of expected changes in offered liquidity. Initially, we start from a state where offered liquidity is balanced between the two sides of the market (as shown in Figure 6.3–1). The box size reflects the quantity of offered liquidity available to the market.

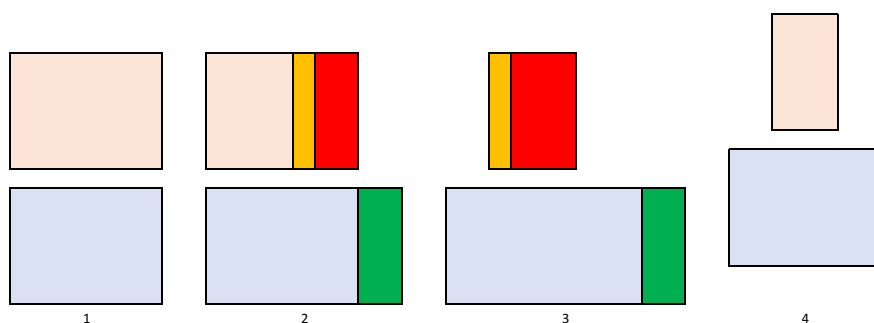


Figure 6.3: Dynamics of Offered Liquidity

Notes: Figure contains four snapshots of offered liquidity with events that change the offered liquidity — green represents the arrival of offered liquidity; red represents the removal of offered liquidity from order cancellation; and orange represents the removal of offered liquidity from trades.

Moving left from 6.3–1 to 6.3–2, we find a change in offered liquidity resulting

from each of the three events — the ask side has offered liquidity reduced due to both buyer-initiated trades and ask limit order cancellations, while the bid side has offered liquidity increased due to the arrival of bid limit orders. Does the order of these events matter? For example, does the interpretation change if trades occur first? If bid limit orders arrive first, then that would indicate the initial expression of buying interest is related to an increase in offered liquidity. If the ask limit orders are cancelled first, then that would indicate the buy side is following the ask side.

The remaining subfigures show the continuation of the initial change in offered liquidity. Of course, this is just one way the sequence can unfold. The next figure to the left shows the same three events, but here the ask offered liquidity is completely removed. The last figure shows an increase in the bid-ask spread, a new ask offered liquidity based on higher quote levels and the same large bid offered liquidity.

6.3 Data & Methodology

6.3.1 Data

In this section, we discuss the information used in the analysis and steps taken to rebuild the limit order book from the raw message data. We then provide a discussion of how we measure OL and ΔOL .

Starting Raw Data — CME Message Data:

This study uses individually purchased market depth data from the CME for one year related to crude oil futures. The sample period includes all trade dates between August 1, 2012 to July 31, 2013. This is the same information sold by the exchange to market participants and allows us to rebuild the limit order book for the top ten levels of the book. The data include timestamps precise to the millisecond and also contains the complete set of transactions fully integrated with the observed changes to the limit order book.

The raw information, delivered by the exchange, represents a collection of messages created by the exchange whenever there is a change in the visible limit order book. For example, if a market participant places a new limit order for one contract at the second-best bid, then, if prior visible depth is 4, the action will force the exchange to issue a message that an updated depth at the second-best bid is now 5.

A single message can have many updates. For example, if a market participant places a market order to buy and if the requested quantity is equal to the resting depth at the best ask quote, then the trade event removes the best ask quote. The exchange will issue a message with trade details followed by all ten ask levels with updated price levels and visible depth for each book level one through ten.

Each message needs to be broken down into individual pieces of information (referred to as data blocks), which adds to the challenge of using such data. Information within the message is defined using TAGs. Within each message there is an ID to correctly sort components of the message.

Rebuilt Limit Order Book for Both Sides:

To rebuild the limit order book, it is necessary to break down each raw message and then sort the set of book updates using both the message ID and the data block ID. We select all information connected to the nearby expiration with the largest observed daily volume.

We rebuild the limit order book for each side of the market separately. Focusing on a single side, we rebuild the limit order book by starting with the first update and place the information in a grid of ten levels. We retain information at each level if updates do not adjust the level. We keep the last update within a related set of updates. For example, if all ten levels change, then there will be ten rows with updates starting with an update at level one only followed by and updated level one with an update to level two, etc. We only need to keep the last row here, which contains the updates to all book levels in a single row.

For a single side of the market, the rebuilt limit order book contains two sort IDs, a timestamp, product information, ten quote levels and ten visible depth numbers corresponding to each book level one to ten.

Best Quote Combined Book Update Table:

After processing the raw message data to rebuild the limit order book, we want to combine information from both sides of the market into a single observation linked to the millisecond the update was recorded.

Focusing on the best quotes from each side of the market, we pull all rows displaying a change in either the price level or visible depth for the best bid and ask quotes. We then build a table with each row corresponding to a unique

timestamp that can be linked to some observed change at one or both of the best quotes. Then, for each row, we fill in the observed updates and pull down price and depth over time if one side does not change. This process results in a table with observations linked to a change at the best quotes and allows us to observe changes over time on both sides of the market.

To add information related to quotes above the best ask and below the best bid, we collect snapshot information of visible depth for the next four book levels. Since differences between observed quotes within a single side are almost always the minimum allowed tick size (\$0.01 in crude oil futures), we only keep visible depth information for each book level.

The use of snapshot information related to book levels away from the best quotes means we are not including timestamps linked to an observed change away from the best quotes. For example, if a market participant places an order to buy at the second-best bid, we will not see this update reflected in our table until there is an observed change at the best quotes.

Addition of QI/QP Events, Quote Segments, & Trades:

Once the best quote combined book update table is created from the processed raw message data, we add additional information based on observed changes in the best quotes and transaction information. Specifically, we use changes in the best quotes to define quote segments (see Chapter 4) to divide the day into disjoint intervals. We also incorporate transaction information to identify the reason for observed changes to the limit order book. Trade information is necessary to determine if an observed decrease results from the arrival of market orders or the decision to

cancel resting limit orders.

We define a **quote improvement (QI)** event when an update creates an improvement in the inside price. A Bid-QI event occurs when the buy side moves upward, toward the sell side; an Ask-QI event occurs when the sell side moves downward, toward the buy side.

A **quote promotion (QP)** event occurs when the second-best quote is promoted to the first-best quote. A Bid-QP event occurs when the buy side moves downward, away from the sell side; an Ask-QP event occurs when the sell side moves upward, away from the buy side.

We define quote segments as non-overlapping intervals of time which restrict quotes to live within a narrow range. For each day, we start by identifying the first occurrence of a two-tick bid-ask spread and define this at the starting point of the first quote segment; the best ask quote level is defined as the upper quote segment level and the best bid quote level is defined as the lower quote segment level. Since we have a two-tick spread, the open quote level between the best quotes is defined as the inside quote segment level. This quote segment exists as we encounter subsequent book updates until the market meets conditions for a new quote segment. We mark the start of a new quote segment if the bid-ask spread is two-ticks or less and either the best bid is below the lower quote segment level or the best ask is above the upper quote segment level. After we scan across all book updates, we are left with each observation belonging to a quote segment and we know exactly where the best quotes live within that quote segment.

Besides the updates to the limit order book, the raw message data contains the complete set of trades. For each trade, we can observe the timestamp with

millisecond precision, the quantity, and execution price. Since a book update message immediately follows trade information, we are able to correctly sequence observed trades within the evolving limit order book allowing us to determine the status of the book prior and immediately following each trade. This allows us to distinguish between buyer- or seller-initiated trades.

Individual trades are determined based on the market order and characteristics of resting limit orders. If a market order arrives to buy 10 contracts, the number of observed trades will depend on the resting limit orders first in line to trade. If all of these resting limit orders are for one contract, then we will observe ten individual trade records all for one contract — all occurring at the same millisecond and all within the same raw message. We define groups of individual transactions by combining quantity linked to trades uninterrupted by book updates. Since it is unlikely that market orders arrive at the same millisecond, we assume trades sharing the same millisecond are related to a single market order. For each trade group, we place this information at the book update reflecting the limit order book post-trade. Furthermore, we identify if the traded volume is buyer- or seller-initiated and if this traded volume occurs at the best quotes or away from the best quotes.

6.3.2 Measuring *OL*

At each point in time, we can use the limit order book to determine the amount of visible depth across various quote levels. Instead of making reference to quote location within the limit order book, we make use of the location according to the quote segment level.

Chapter 3 presented evidence that the majority of book updates correspond to the top three quote levels. Therefore, we focus exclusively on the top three quote levels on each side of the market. These quote levels are defined according to the quote segment. Therefore, the first quote level corresponds to the inside quote of the quote segment. The second quote level corresponds to the lower quote (for the buy side) or the upper quote (for the sell side). The third quote level corresponds to the quote directly above or below the upper or lower quote segment level.

$OL_{t_0}^i$ represents offered liquidity at t_0 , for market side i , where i can be equal to B for buy side offered liquidity or S for sell side offered liquidity. Offered liquidity is the sum of visible market depth at quote segment levels one to three:

$$OL_{t_0}^B = B_{t_0}^1 + B_{t_0}^2 + B_{t_0}^3 \quad (6.1)$$

$$OL_{t_0}^S = A_{t_0}^1 + A_{t_0}^2 + A_{t_0}^3 \quad (6.2)$$

6.3.3 Measuring ΔOL

We propose a method to measure the dynamics of OL for each side of the market over very short time intervals. Our method measures the rate of change in offered liquidity per millisecond (ΔOL). This measurement can take on positive or negative values. For example, a negative measurement related to the buy side would indicate a decrease in visible depth across the first three quote segment levels; a positive measurement would indicate an increase in visible depth across the first three quote segment levels.

Measurement Method 1: Observed Changes in OL

Unlike OL , which is a snapshot of cumulative depth across the first three quote segment levels, ΔOL requires a defined interval of time allowing us to measure how OL changes. For each selected timestamp, t_0 , we define an interval of time prior to t_0 and an interval of time immediately following t_0 following a defined parameter μ representing the number of milliseconds in the measurement interval. Our pre-interval covers a time range of $t_0 - \mu$ to t_0 (excluding t_0); our post-interval covers a time range of t_0 to $t_0 + \mu$.

For each interval we use all limit order book updates within the interval to define offered liquidity at each unique millisecond over the interval. We then measure ΔOL for each market side, as the slope parameter of a linear estimation of offered liquidity as a function of the millisecond count variable. This measure then is interpreted as the change in offered liquidity per millisecond.

$$OL = \alpha + \beta * t; \text{ where } \beta = \Delta OL \quad (6.3)$$

For each t_0 , there are four measurements associated with both sides of the market for the pre- and post-intervals: ΔOL_{pre}^S , ΔOL_{pre}^B , ΔOL_{post}^S , and ΔOL_{post}^B .

Measurement Method 2: Cancellations Only

It is possible for OL to decrease from trades or from cancellation of resting limit orders. Based on analysis in Chapter 3, we know a majority of trades never go past the best quotes. We want, therefore, a measure of fleeting liquidity that represents only cancellation activity. To accomplish this, we perform an adjustment to measure ΔOL excluding decreases in OL due to trades.

To accomplish this for each millisecond with trade information, we first determine the number of observed trade prices and the corresponding order book quote level. For each quote level we compare the traded quantity with the observed change in depth. If traded quantity is equal to the magnitude of the depth change, we remove the change in depth. If traded quantity is less than the magnitude of the depth change then we keep a portion of the depth change over the traded quantity to reflect cancellation activity. If traded quantity is greater than the magnitude of depth change then we use the difference to define the new arrival of offered liquidity.

Using the adjusted depth changes, which washes out decreases from trades, we solve for the adjusted offered liquidity at each millisecond during the interval and then proceed to run an OLS regression to determine the slope coefficient on the time variable. We perform this for both sides of the market to determine $\Delta OL^{B,adj}$ and $\Delta OL^{S,adj}$.

6.4 Summary of OL and ΔOL

We summarize OL and ΔOL using a random sample of timestamps. We sample 20 thousand timestamps per hour across all days in our sample. For each timestamp (t_0), we identify the pre- and post-measurement interval based on a measurement duration of 50 milliseconds. For each timestamp, we collect OL at the start of the pre- and post-measurement intervals and the measured ΔOL again for both measurement intervals (see Figure 6.4). For this section, we show the average absolute measure of responsiveness across the four different measurements of ΔOL .

Once we collect information related to OL and ΔOL for each selected times-

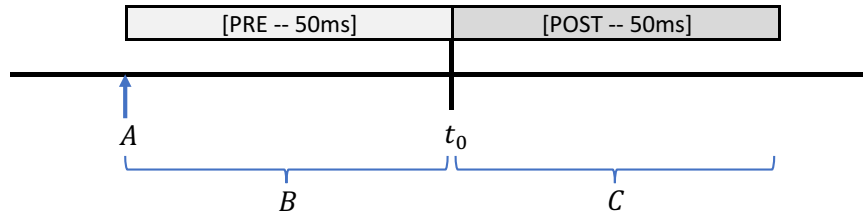


Figure 6.4: Measurement of OL and ΔOL at t_0

Notes: Here we show a selected timestamp (t_0) along with the pre- and post-measurement interval defined according to selected duration of 50 milliseconds. Marker A notes the moment we measure OL_{pre} for both sides of the market at the start of the pre-measurement interval. Marker B notes the interval used to define ΔOL_{pre} for both sides of the market. At t_0 , we measure OL_{post} for both sides of the market. Marker C notes the interval used to define ΔOL_{post} for both sides of the market.

tamp, we average information for each minute of the day. For each minute observation, we add additional information related to volume, net volume (defined as buyer-initiated volume minus seller-initiated volume), measured volatility (following a 2SRV procedure), and a time-weighted measure of the bid-ask spread.¹

6.4.1 Intraday & Daily Summary

Figure 6.5 shows intraday per-minute variation in volume, volatility, OL , and ΔOL . As shown in prior Chapters, this particular market contains a significant amount of intraday variation and we expect to find similar differences within the day related to OL and ΔOL . Here we show how these two measurements relate to observed intraday patterns in volume and volatility.

Comparing the top two subfigures, we notice the expected relationship between volume and volatility — minutes with high expected volume also have high expected volatility. We find a very similar relationship between volume, volatility, and ΔOL . The bottom subfigure showing ΔOL displays similar jumps, but there

¹To measure volatility, we use a two-scale realized volatility (2SRV) estimator using the returns from an evenly spaced measurement of the inside quote segment level. We select a distance of 100 milliseconds between measurements and then within each hour define 20 subgrids. For each sub-grid, we compute sum of squared log returns. We then estimate volatility within each minute by averaging these returns within each minute of the sample.

is a noticeable difference in activity at 8:30 a.m. and 10:30 a.m. when volatility is especially high due to expected major news events.

Regarding OL , we find three instances of drops at 8:30 a.m., 10 a.m., and 10:30 a.m. The primary measure of OL , as discussed above only uses visible depth from the first three quote segment levels. Here we show the intraday summary of the aggregate visible depth at levels four and five. Based on the figure, there is a strong relationship between these different measures of OL . The intraday plot shows how OL continues to build during the morning hours and then levels off with a slight decrease during the afternoon hours.

Along with exploring differences within the day, we are also interested in daily variation during our sample period. Figure 6.6 shows daily variation in volume, volatility, OL , and ΔOL . We find a few days in our sample with elevated volatility (one on the 2nd of August related to a major news event; a second group of dates occurring in the middle of September) and there is a visible decrease in volatility in November through February.

The decrease in daily volatility corresponds to an increase in OL and indicates variation across full days with differences in the ability or willingness of market participants to place resting limit orders within the top three quote segment levels. As volatility decreases, we should expect higher measurements of OL since there exists less risk of informed trading. We find less directional movement during the sample suggesting there are few differences across the sample once we look at daily averages across minutes.

6.4.2 Further Study of OL , ΔOL , Volume, & Volatility

Figure 6.7 shows how market activity, measured during one minute intervals, varies conditional on volume. For this work, we sort our set of minutes into 100 equally-sized buckets and arrange from low volume minutes to high volume minutes. For each bucket, we find average volume, volatility, net volume, ΔOL , OL , and a time-weighted bid-ask spread.

We find volume, volatility, and ΔOL exhibit the same general pattern — a steady increase as volume increases until there is a sharp increase in minutes with very large volume. The bid-ask spread shows a U-shaped pattern. A high bid-ask spread is not surprising during the minutes with the largest volume and volatility. We find that offered liquidity is generally very low during minutes of low volume and volatility, which will allow for the bid-ask spread to increase during these minutes as well. While there is a long established direct relationship between volume and volatility, the addition of ΔOL is new and displays the same relationship as found in volume and volatility.

Next we turn to net volume, defined as the difference between buyer-initiated and seller-initiated volume for each minute during the sample period. We sort all minutes during the sample from low to high net volume and define 100 groups using the rank procedure in SAS. For each group, we find average volume, volatility, net volume, ΔOL , OL (for quote segment levels 1–3 and for quote segment levels 4–5), and time-weighted bid-ask spread. We plot the sorted groups in Figure 6.8.

Again, we find a very similar relationship between volume, volatility, and ΔOL . The U-shape indicates minutes with larger imbalances between buyer- and seller-initiated volume are associated with higher volume, volatility and responsiveness

in offered liquidity. We also see related jumps in the average bid-ask spread, which can be expected as the spread widens from the potential of informed trading.

We also find an increase in the combined measure of OL across the first three quote segment levels when there exists a larger imbalance in signed volume. $OL^{(4,5)}$, on the other hand, shows a difference between market sides when net volume is large. Here we find that for minutes with a larger share of buyer-initiated volume, offered liquidity on the sell side is larger.



Figure 6.5: Intraday Volume, Volatility, OL , and ΔOL

Notes: This figure contains four subfigures each plotting intraday information. The first and second subfigures show average minute volume and volatility, respectively. The third subfigure shows average OL per minute aggregating visible depth at quote segment levels one, two and three as our defined measure of offered liquidity; we include a per minute average of total visible depth at quote segment levels four and five (shown in black). The final subfigure shows average per minute responsiveness measured by averaging across randomly selected timestamps within each minute.

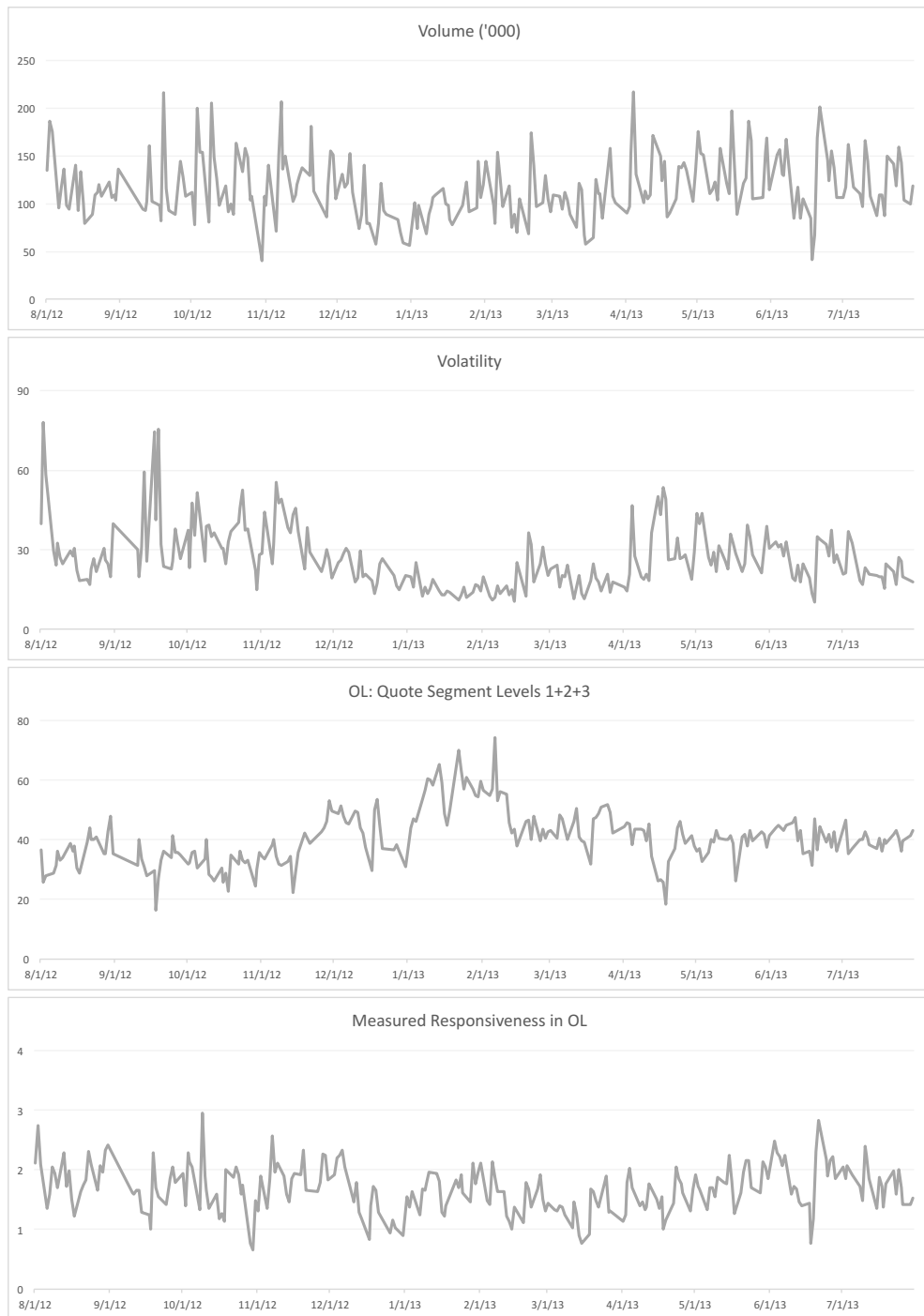


Figure 6.6: Daily Volume, Volatility, OL , and ΔOL

Notes: Figure contains four subfigures each plotting daily information. The first and second subfigures show average minute volume and volatility, respectively. The third subfigure shows average OL per minute aggregating visible depth at quote segment levels one, two and three as our defined measure of offered liquidity. The final subfigure shows average per minute responsiveness measured by averaging across randomly selected timestamps within each minute. All daily averages represent the time weighted value.

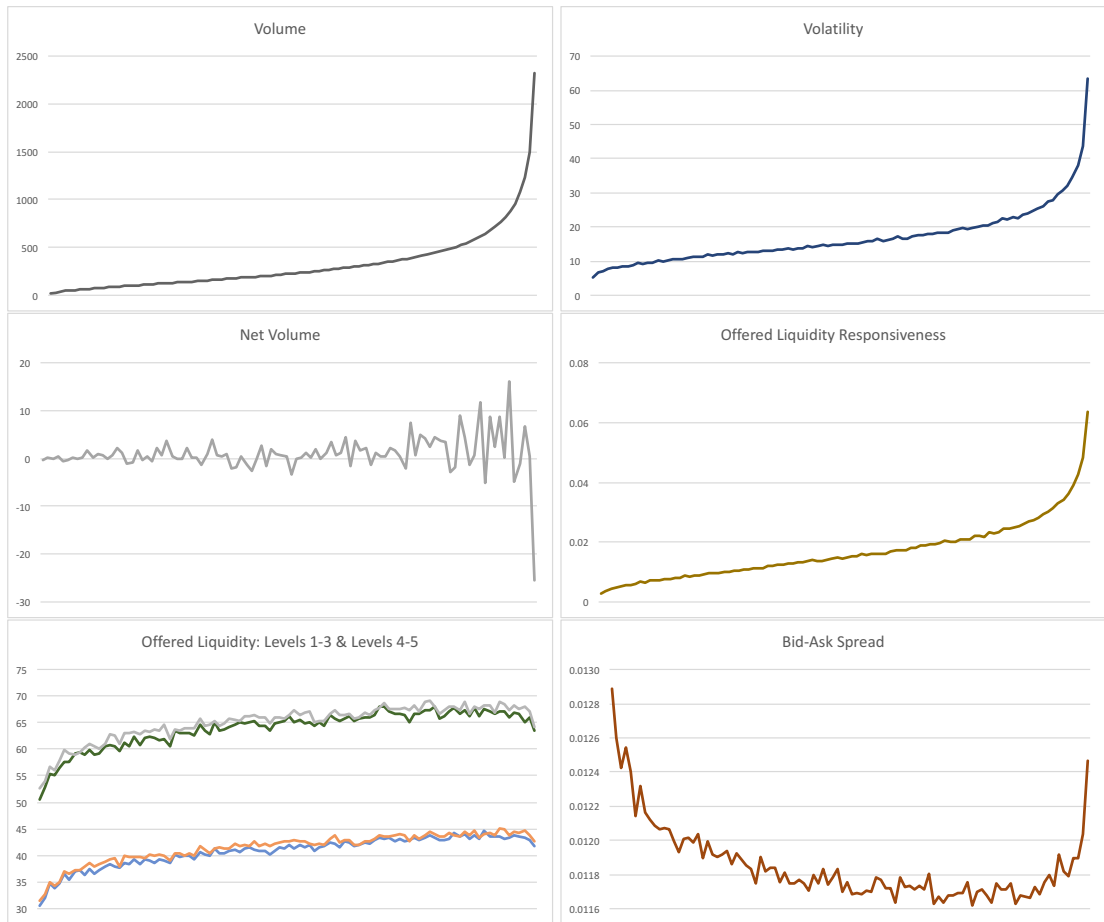


Figure 6.7: One Minute Averages: Ranked by Volume

Notes: Figure shows market information for one minute intervals averaged according to 100 equally sized groups according to volume. Displayed from minutes with the smallest volume to minutes with the largest volume during the sample. Volatility is defined here as the standard deviation of the inside quote level during each minute interval. Net volume is defined as total buyer-initiated volume less seller-initiated volume. Measured responsiveness is based on 50 millisecond windows. Offered liquidity is shown for levels 1-3 (in orange and blue) and for levels 4-5 (in green and grey). Bid-ask spread is the time-weighted value for each minute interval.

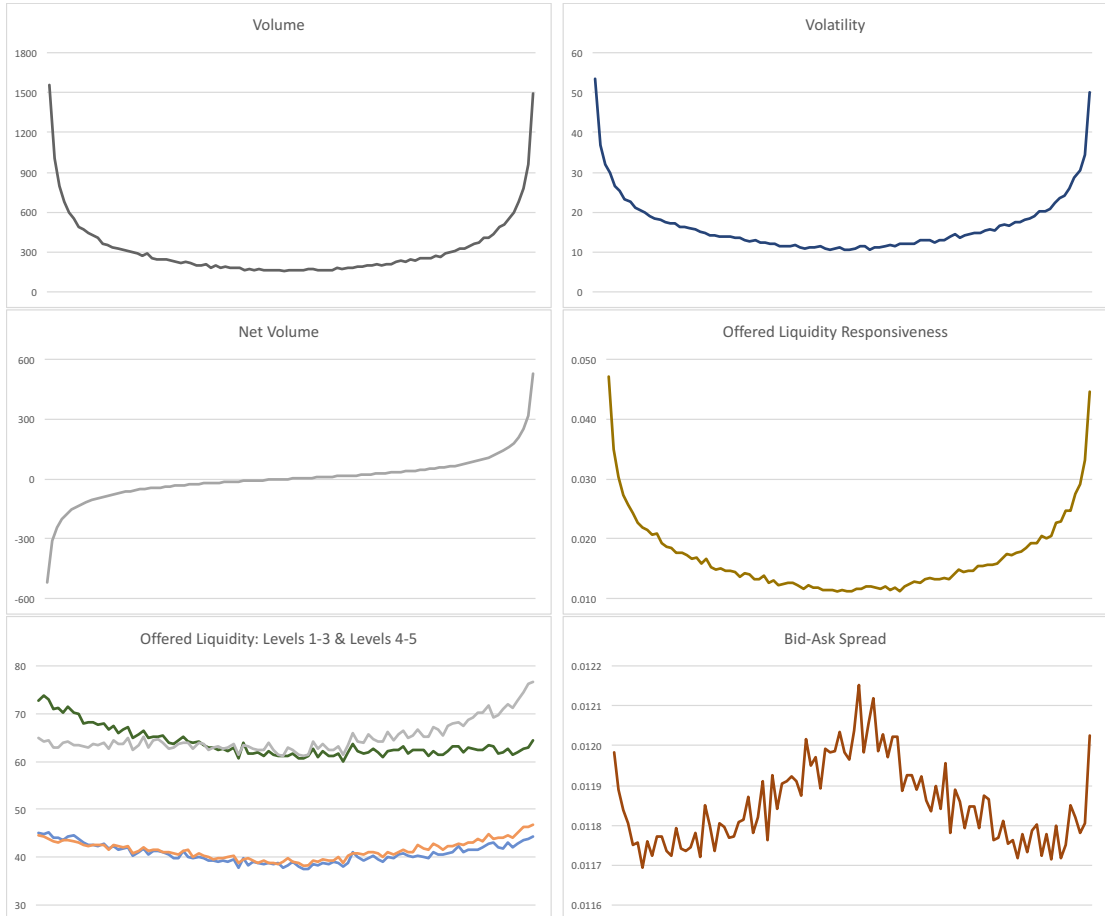


Figure 6.8: One Minute Averages: Ranked by Net Volume

Notes: Figure shows market information for one minute intervals averaged according to 100 equally sized groups according to net volume. Displayed from minutes with the largest negative net volume to minutes with the largest positive net volume during the sample. Volatility is defined here as the standard deviation of the inside quote level during each minute interval. Net volume is defined as total buyer-initiated volume less seller-initiated volume. Measured responsiveness is based on 50 millisecond windows. Offered liquidity is shown for levels 1-3 (in orange and blue) and for levels 4-5 (in green and grey). Bid-ask spread is the time-weighted value for each minute interval.

6.5 Quote Segment Start

In this section, we use observed changes in quote segments to define t_0 . For each observed quote segment start time, we measure OL and ΔOL during short intervals surrounding this moment in time. Figure 6.9 shows an example of two quote segments. For each quote segment, we can define a starting and ending interval based on a defined measurement interval. We use snapshots of visible depth to measure OL at specific points (Markers A, D, F, and H). Our measurement of ΔOL is defined over intervals (Markers B, C, E, and G). The example shows how variation in the time duration of quote segments changes the distance between markers within a single quote segment.

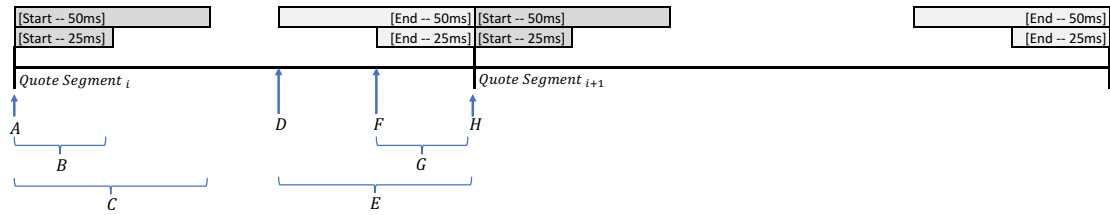


Figure 6.9: Selecting Start of Quote Segments

Notes: Figure shows two quote segments and highlights the various measurements. Marker A is the start of the i^{th} quote segment and marks the moment we measure the starting OL on both sides of the market. Markers B and C represents the measurement interval for the first 25 and 50 milliseconds after the start of the quote segment, respectively. Markers D and F notes the measurement OL at then end of the i^{th} quote segment for the 25 ms and 50 ms measurement interval, respectively. Markers E and G represent measurement intervals for ΔOL at the end of the i^{th} quote segment.

6.5.1 Single Quote Segment Example

Before we investigate the sample, we show how OL evolves over the course of a representative quote segment, which also highlights the dynamics we hope to capture with ΔOL . Figure 6.10 shows movement of OL during a single quote segment. The first subfigure, on the left, shows OL at each book update observed during the quote segment. The seconds subfigure, on the right, shows the buy side

individual quote segment levels contribution to OL .

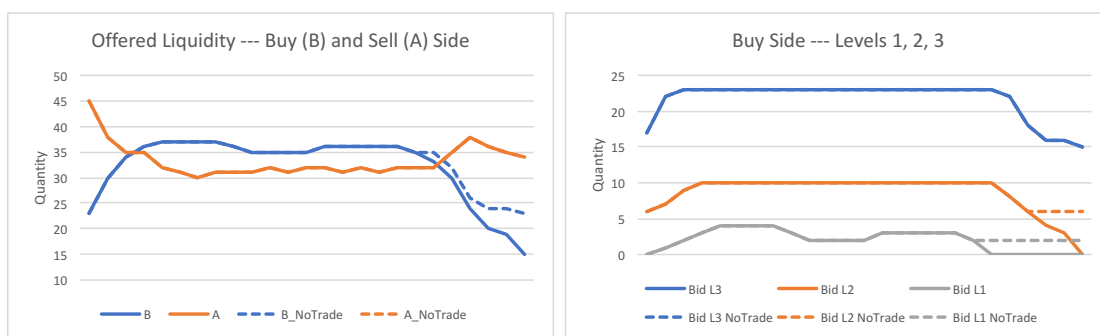


Figure 6.10: OL — Single Quote Segment

Notes: Figure, on the left hand side, shows OL during a single quote segment on both sides of the market; broken lines indicates movement not considering trades. Figure, on the right hand side, shows buy side activity by quote level; again, broken lines indicate movement not considering trades.

The visual contains three clear intervals within the quote segment. First, there is an initial interval with changes observed on both sides of the market. Second, there exists an interval with very little movement in OL on either side of the market. Finally, toward the end of the quote segment, there is a sudden decrease in OL on the buy side. We discuss each of these below in more detail.

Following the start of the quote segment, we see the sell side reduce and the buy side increase OL . In this example, the quote segment starts after an aggressive buyer-initiated trade removes the best ask quote thereby creating an Ask-QP event. The decrease in OL on the sell side results from market participants cancelling resting orders across the first three quote segment levels. Regarding the buy side movement, the initial increase in OL supports the prior movement toward the sell side. Based on the right hand side subfigure, we can see the increase in OL on the buy side results from an increase in visible depth across all three quote segment levels.

After the initial adjustments, OL is nearly balanced for the majority of the

quote segment. During this interval, we observe very small changes in OL on either side of the market. There exists a larger buy side OL which appears to be related to the quantity at the inside quote segment level.

Near the end of the quote segment, an imbalance develops in OL , primarily due to movement away from the market from the buy side. Comparing the solid blue line to the broken blue line, we see the movement is due to a combination of trades and limit order cancellations. There appears to be a sudden change initiated by the buy side moving away from the market and highlights how depth changes at the start and toward the end of the quote segment can be informative for predicting future market movement.

The subfigure on the right demonstrates how trades contributed, along with cancellations, to the reduction of OL at quote levels 1 and 2 for the buy side. The change in OL at quote level 3, on the other hand, resulted only from quote cancellations. If we draw response lines on the OL on the combined and cancellation only, we would find a difference in the slope parameter. This difference can be used to measure how much of the response is attributed to trades vs order cancellations.

6.5.2 Event Based Measurement vs Random Sample

The example above shows how there can be an adjustment, on both sides of the market, at the start of a quote segment as the market adjusts to the movement in quotes. Furthermore, we argued that such a response makes sense once we know the direction and reason for the movement. Here we summarize the distribution of OL and ΔOL where t_0 represents the starting time of each quote segment,

measured over a 50 millisecond interval. To test for significance, we compare to a sample comprised of t_0 selected at random according to a uniform distribution.

Table 6.1 shows information related to OL , at the start of the measurement interval, and ΔOL linked to the start of quote segments during the first month of the sample. The numbers in grey reflect the similar values summarizing a random selection of timestamps. All tests comparing distributions confirm a statistical difference between the measurements linked to the start of quote segments and those linked to randomly sampled timestamps.

Table 6.1: Measured Responses: Random Sample vs Quote Segment Start

Notes: Table shows distribution information for starting depth (OL) and measured response (dOL) related to a sample of random times drawn from the first month of the sample (numbers in grey) and information related to a sample of times linked to the start of quote segments (numbers in black). Regarding measured response, we show the unadjusted measure (combines trades and cancellations) with the adjusted measure (only showing cancellations) for OL defined as the total visible depth across the first three quote segment levels. The lower rows of the table show measured response by quote segment levels one to five. KS tests confirm there is a statistical difference between all distributions. Distributions are created based on 154,355 observations during August 2012.

	Mean	StdDev	P1	P5	P10	Q1	Median	Q3	P90	P95	P99
OL	33	21	5	11	14	21	30	41	54	65	99
	37	21	10	14	17	23	33	45	59	70	103
Trade & Cancellations:											
dOL Pre	-1	25	-68	-32	-19	-5	0	3	17	31	64
	0	7	-16	-3	0	0	0	0	0	3	19
dOL Post	0	26	-58	-39	-30	-15	0	14	32	44	70
	0	6	-17	-3	0	0	0	0	0	3	18
Cancellations Only:											
dOL Pre	2	23	-44	-19	-10	-2	0	3	18	32	65
	0	7	-14	-3	0	0	0	0	0	3	19
dOL Post	2	26	-55	-37	-29	-14	0	16	34	46	77
	0	6	-14	-3	0	0	0	0	0	3	18
Trade & Cancellations -- By Quote Segment Level:											
dOL--1	2	7	-8	-2	0	0	0	2	8	14	28
	0	3	-4	0	0	0	0	0	0	0	5
dOL--2	-1	13	-35	-23	-18	-8	0	6	14	19	31
	0	4	-9	-2	0	0	0	0	0	2	8
dOL--3	0	12	-31	-19	-14	-7	0	6	14	19	32
	0	3	-7	0	0	0	0	0	0	0	7
dOL--4	0	7	-22	-10	-6	-2	0	3	7	10	20
	0	2	-3	0	0	0	0	0	0	0	4
dOL--5	0	4	-11	-6	-4	-2	0	1	4	6	11
	0	2	-3	0	0	0	0	0	0	0	3

Comparing measured responses, we find a larger measured response in the post-measurement interval. For negative values, it is more likely to observe a

large decrease in offered liquidity immediately following the event. Likewise, we observe a similar situation with offered liquidity increasing at a faster rate immediately following the event. This indicates the event contains information that prompts the market to react by adjusting visible depth on both sides of the market. The non-zero expected response in the pre-measurement interval indicates the possibility of a rebalance going on prior to the event.

Comparing the distributions related to the pre-measurement interval based on the combined impact of trades and cancellations with just cancellations shows how trades do tend to influence a measured negative rate of change. We find smaller negative values at P10 and Q1 of the distribution in the measurements based only on cancellations. For example, Q1 for both trades and cancellations shows -19, while Q1 for just cancellations is -10. It is important that we still have a non-zero value for measurements only using cancellations. This indicates that cancellations are an important part of the measured response. We do not observe such a large difference in the post-measurement interval. Therefore, trades play a much smaller role in measured negative values immediately following an event that changes the quote segment.

The lower portion of Table 6.1 shows distribution information for measured responses for individual quote segment levels one to five. Comparing various points of the distribution shows the largest responses correspond to the second and third quote segment levels. Responsiveness is the lowest for the first quote segment level, which is linked to the inside quote segment level. The skew, at the first quote segment level, toward a positive response indicates it is more likely to observe a response that builds depth at the inside quote segment immediately

following the event.

6.5.3 Event Analysis — Six Types of Events

We can also evaluate how the characteristics of the event create the change in quote segments. For example, a quote segment change can occur due to the arrival of a Bid-QP. This can happen when the ask side is established at the inside quote segment level and either a seller-initiated trade kills the best bid quote or resting limit orders cancelling at the best bid result in the second-best bid quote get this promoted to the best bid quote. Another possibility is to observe a Bid-QP along with an Ask-QI. This should be viewed as a more aggressive movement in quotes.

Figure 6.11 shows six possible quote segment changes — three downward movements (first column) and three upward movements (second column). Within each example, we show the upper quote segment (solid red line), the inside quote segment (broken green line), and the lower quote segment (solid blue line) along with the best ask quote (dotted red line) and the best bid quote (dotted blue line). The event of interest occurs in the middle of each example, at the start of QS_i , and the various figures highlight the movement and/or non-movement of the best quotes.

The first row of Figure 6.11 corresponds to a single QP event at t_0 . This is then characterized by a one-tick bid-ask spread prior to the event; the QP event occurs on the market side related to the directional movement. For example, a Bid-QP event exists when the sell side has established a position at the inside quote segment level prior to the event and visible depth at the lower quote segment is completely removed from the market. This event starts with a two-tick bid-ask

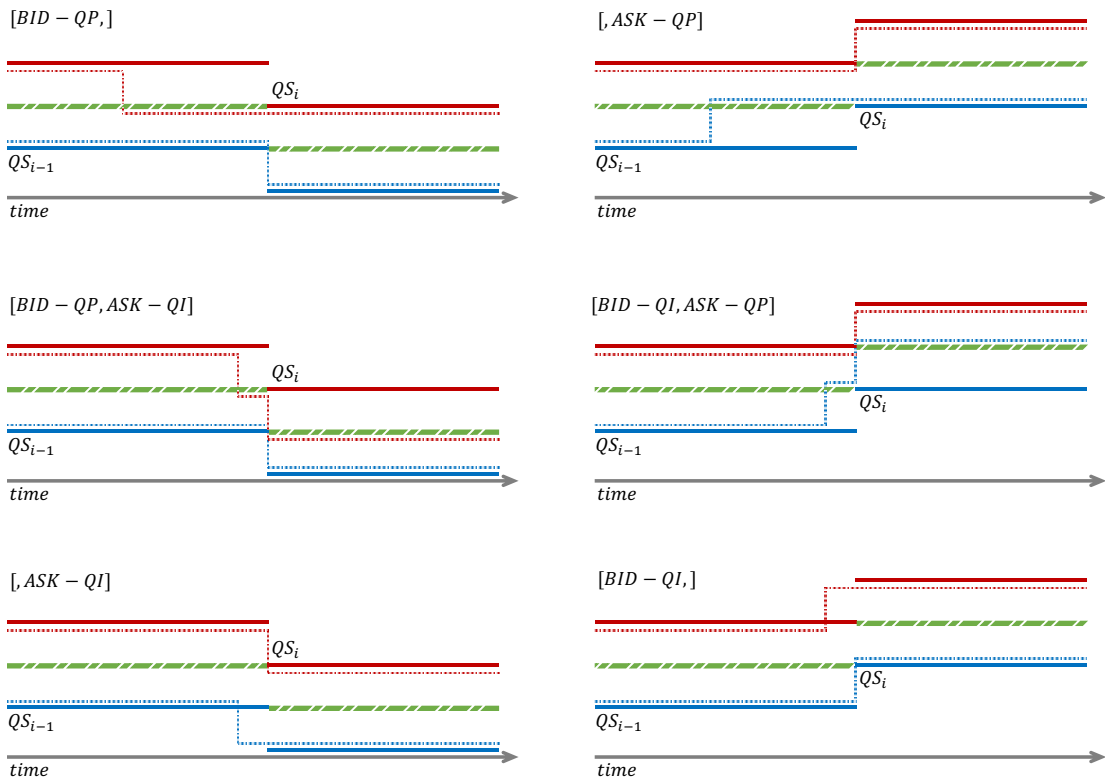


Figure 6.11: Visualization of Quote Segment Movement — Six Event Types

Notes: Figure shows six possible quote segment changes. The first column corresponds to downward shifts in quote segments; the second column corresponds to upward shifts in quote segments. The rows correspond to similar type movements in the best quotes which define the moment when rules are met to define a new quote segment: the first row corresponds to a single QP that opens the market to a two-tick bid-ask spread; the second row corresponds to a simultaneous QP and QI movement in both best quotes; and the third row corresponds to a single QI movement which returns the market to a two-tick bid-ask spread.

spread. Therefore, OL measured at t_0 is the cumulative visible depth across the second and third quote segment levels.

In the second row of Figure 6.11 we observe combined events consisting of a QP and QI event. In such cases, the bid-ask spread is one-tick at t_0 and OL measured at t_0 will include visible depth at the inside quote segment level, which is the residual demand of the marketable limit order that is responsible for the QP event. These events reflect a more aggressive movement in best quotes as there is an immediate establishment of visible depth at the inside quote segment level.

The third row of Figure 6.11 corresponds to a single QI event and is initiated by the arrival of new limit orders that return the market to a two-tick bid-ask spread. The immediate question regarding these events is why there exists a bid-ask spread larger than two-ticks. We can identify if aggressive market orders opened the spread or if this is relate to aggressive cancellations of visible depth.

While there exist other forms of movement in the best quotes that correspond to the start of a quote segment, the three types listed above account for 97% of all quote segments. We will refer to these three as movement groups (MG) 1, 2, and 3, respectively.

6.5.4 Event Compare/Contrast

Table 6.2 contains distribution information for OL and ΔOL for the three types of movement types discussed in Figure 6.11. We define measurement intervals using three time intervals: 25, 50, and 100 milliseconds to explore how the distribution changes as we change the amount of time surrounding the event linked to a change in the quote segment. We find a statistical difference between distributions across

the various time intervals and movement types.

Table 6.2: *OL* and ΔOL : By Event Type

Notes: Table shows distribution information for measured *OL* and ΔOL conditional on three different types of movement groups (MG) for three different measurement intervals: 25, 50, and 100 milliseconds for both pre- and post-measurements. Sample is the set of times associated with changes in the quote segment. MG1 corresponds to a single QP event; MG2 corresponds to the combination of a QP and QI event; and MG3 corresponds to a single QI event. Adjustments made to upward movements to be combined with downward movements. A KS test finds a statistical significance between distributions both across MGs within a fixed time interval and across time intervals within a fixed MG.

	Mean	p25	p50	p75	Mean	p25	p50	p75	Mean	p25	p50	p75
	25ms; MG=1				50ms; MG=1				100ms; MG=1			
OL S Pre	51	35	47	61	49	33	46	60	48	32	44	59
OL B Pre	26	17	23	32	28	17	24	34	29	18	25	35
dOL S Pre	15	0	3	20	11	0	2	12	8	0	1	7
dOL B Pre	-16	-22	-5	0	-10	-12	-3	0	-7	-6	-1	0
SI VM	3	0	2	4	3	0	2	4	3	0	2	4
BI VM	0	0	0	0	0	0	0	0	0	0	0	0
OL S Post	28	18	26	34	28	18	26	34	28	18	26	34
OL B Post	45	32	42	54	45	32	42	54	45	32	42	54
dOL S Post	32	6	23	50	16	1	9	24	7	0	3	9
dOL B Post	-27	-43	-23	-7	-15	-23	-11	-3	-7	-9	-3	0
	25ms; MG=2				50ms; MG=2				100ms; MG=2			
OL S Pre	51	34	48	63	49	32	46	61	48	31	45	60
OL B Pre	31	19	27	37	32	20	28	39	33	20	28	40
dOL S Pre	18	0	3	24	14	0	2	15	10	0	1	8
dOL B Pre	-22	-29	-6	0	-13	-16	-3	0	-8	-8	-1	0
SI VM	7	1	3	7	7	1	3	7	7	1	3	7
BI VM	0	0	0	0	0	0	0	0	0	0	0	0
OL S Post	34	21	31	42	34	21	31	42	34	21	31	42
OL B Post	48	33	44	57	48	33	44	57	48	33	44	57
dOL S Post	40	8	33	63	20	1	15	31	10	-1	5	13
dOL B Post	-37	-56	-34	-14	-23	-33	-19	-8	-12	-15	-7	-2
	25ms; MG=3				50ms; MG=3				100ms; MG=3			
OL S Pre	32	18	28	40	31	18	28	40	31	17	28	40
OL B Pre	26	11	23	36	27	12	23	37	27	12	23	37
dOL S Pre	4	0	0	5	2	-1	0	3	1	-1	0	2
dOL B Pre	-50	-66	-28	-8	-22	-24	-10	-3	-12	-9	-3	-1
SI VM	0	0	0	0	0	0	0	0	0	0	0	0
BI VM	0	0	0	0	0	0	0	0	0	0	0	0
OL S Post	16	8	13	20	16	8	13	20	16	8	13	20
OL B Post	38	20	34	50	38	20	34	50	38	20	34	50
dOL S Post	49	8	34	75	25	2	14	37	14	0	4	15
dOL B Post	-21	-37	-14	0	-13	-23	-8	0	-7	-11	-3	1

In the pre-measurement interval we find little difference in *OL* over the various time intervals. For example, the median is approximately 20 and 40, for the buy and sell side, respectively. Since the sample represents all down movement at the event, this imbalance in *OL* appears to indicate there is already a sufficient pressure from the sell side 100 milliseconds after the event time.

We find there is a large difference between ΔOL conditional on the defined measurement window. For example, we find a response of -20, -11, and -5 for ΔOL on the buy side for 25, 50 and 100 millisecond intervals, respectively. The

higher response across the 25 milliseconds indicates the majority of change is taking place very close to the event. The same finding exists on the sell side with the positive measurement indicating a larger rate of depth is arriving very close to the event. The post-measurements show a similar finding with the largest response in the 25 millisecond interval. This measured response moves to zero as we extend the measurement window.

In Chapter 5, we discussed a common occurrence with depth leaving the best quote and then returning. The results showing a large response within the first 25 milliseconds, and eventually moves toward zero as we extend the interval, can point to a similar situation with depth initially leaving the book and then returning. This type of movement in visible depth would explain why we are findings smaller measurements with longer measurement intervals.

The table reveals a difference between movement groups. The first movement group (MG1) corresponds to a single QP event which causes a new quote segment. Given how we identify quote segments, we can assume prior to these events, that the opposite side has depth established at the inside quote segment level. This explains why we see an *OL* imbalance at the start of the pre-measurement interval. MG2 corresponds to an immediate QP and QI event, which is more of an aggressive movement. Indeed, we can see from the table, measured responses in the post-interval are larger in MG2 relative to MG1.

6.5.5 Regression Analysis

In this section, we use a regression based approach to investigate what factors drive the observed responsiveness in the post-measurement interval.

We expect ΔOL , measured in a short time interval following the event to convey information related to the expectations of market participants with the technological power to observe when such events take place and react accordingly. As an example, consider a market participant with resting limit orders on both sides of the market. If a large unexpected order to buy arrives, the market participant might decide to cancel any resting limit orders at ask quotes near the market fearing the arrival of more aggressive buying representing bullish information arriving into the market. The speed and size of this response will likely vary according to how alarmed the market participant becomes with the new information. Along with information related to the event, we want to explore the relationship between measured responses before and after the event.

Model 1:

$$\Delta OL_{post,t}^B = \beta_0 + \beta_1 * \Delta OL_{pre,t}^B + \beta_2 * \Delta OL_{post,t-1}^B + \beta_3 * \Delta OL_{post,t-2}^B \quad (6.4)$$

$$\Delta OL_{post,t}^B = \alpha + \beta * \Delta \mathbf{OL}^B \quad (6.5)$$

The first model only considers same market side pre-response and two lagged values of post-response from the prior two quote segment changes. We include the measurement of ΔOL prior to the event ($\Delta OL_{pre,t}^B$) to capture the relationship between movement of visible depth before and after the event. The lagged values of post-response ($\Delta OL_{post,t-j}^B$ for $j = 1, 2$) are included to capture autocorrelation in the measured response. We collapse these three variables into a single vector ($\Delta \mathbf{OL}^B$).

Model 2:

$$\Delta OL_{post,t}^B = \alpha + \beta * \Delta OL^B + \gamma * \Delta OL^S \quad (6.6)$$

The second model expands the first model by including explanatory variables related to the opposite side of the market. The same set of explanatory variables shown above for the buy side (ΔOL^B) are now added for the sell side (ΔOL^S).

Model 3:

$$\Delta OL_{post,t}^B = \alpha + \beta * \Delta OL^B + \gamma * \Delta OL^S + \rho * \text{SignedVM} + \phi * \text{OL} \quad (6.7)$$

The third model expands the second model by including additional information related to signed volume (**SignedVM**) and OL (**OL**). The variables related to signed volume include information from the prior quote segment (Prior BI Qty) and information associated with the event at t_0 (BI Qty). We include a lagged value of prior quote segment volume along with a squared term to capture any non-linearity. The variables related to OL include measured offered liquidity at t_0 (OL B Post) and offered liquidity linked to the prior quote segment at levels 4 and 5 (OL B Prior L4, OL B Prior L5). We include variables for both signed volume and OL for both sides of the market.

Discussion of Regression Results:

Model 1: Model 1 shows how the ΔOL_{pre} on the same side as our dependent variable helps explain the direction. This indicates directional changes prior to the event at t_0 are useful in explaining the later market reaction. As ΔOL can be positive or negative, a positive movement prior to the event indicates a positive

movement immediately following the event.

Model 2: Model 2 now builds on Model 1 by including opposite market side information in the measurement interval prior to the event at t_0 . We find a negative coefficient for the pre-measurement and the first lag of the post-measurement; we find a very small positive coefficient for the second lag of post-measurement.

Model 3: Model 3 now builds on Model 2 by adding information related to traded volume and offered liquidity. The additional variables show how signed trades linked to the prior quote segment and at t_0 relate to the observed response during the post-measurement interval. The presence of buyer-initiated trades has a positive relationship to the response on the buy side — aggressive market orders are linked to the increase of offered liquidity following the event at t_0 . The presence of seller-initiated trades has a negative relationship to the response on the buy side. Aggressive sellers indicates the buy side will react stronger in a movement way from the market with a negative response following the event.

We find the size of OL on the buy side immediately following the event has a negative sign, which indicates as the size of offered liquidity increases, the response decreases. Switching market side, we find a larger OL on the sell side is associated with a larger response from the buy side indicating a desire from the buy side to move toward this liquidity. The signs associated with offered liquidity at quote segments four and five are associated with the quote segment prior to t_0 and show a reverse sign from the post-measurement interval. This is evidence that the response in the post-measurement reflects the movement prior to the event at t_0 .

Model Sensitivity to Measurement Interval:

To investigate how the model performs across different measurement intervals, we run model three across three different durations: 25, 50, and 100 milliseconds. Table 6.4 shows the results for each time interval. Comparing the R-square, we find the model is able to explain less of the variation as we increase the time interval. This result makes sense as the dependent variable becomes more removed from the independent variables as we increase the measurement interval. Estimated parameters retain their signs as we adjust the interval length; they move toward zero as the duration of the measurement interval increases.

Table 6.3: Regression Results

Notes: Table shows the estimation results of three models to explain the observed response on the buy side during the post-measurement interval. Events are selected according to our observed changes in quote segments. Results correspond to a measurement window of 25 milliseconds. Explanatory variables include response on both sides of the market in the pre-measurement interval (dOL B/S Pre), signed volume corresponding to the prior quote segment (Prior BI/SI Qty), signed volume (SI/BI Qty) and offered liquidity (OL B/S Post) linked to the event at t_0 , and information related to offered liquidity during the prior quote segment at quote segment levels 4 and 5.

Dependent Variable: dOL B Post (25ms Measurement Interval)

	Model 1				Model 2				Model 3			
	Parameter Estimate	Standard Error	t Value	Pr > t	Parameter Estimate	Standard Error	t Value	Pr > t	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	2.794	0.0374	74.8	<.0001	2.225	0.0373	59.7	<.0001	1.112	0.0928	12.0	<.0001
dOL B Pre	0.245	0.0008	296.0	<.0001	0.169	0.0009	196.6	<.0001	0.088	0.0008	117.2	<.0001
dOL B Post Lag1	0.109	0.0008	137.5	<.0001	0.058	0.0011	54.5	<.0001	0.032	0.0009	34.7	<.0001
dOL B Post Lag2	0.013	0.0008	16.4	<.0001	0.015	0.0011	14.5	<.0001	-0.002	0.0009	-2.6	0.0084
dOL S Pre					-0.217	0.0008	-260.4	<.0001	-0.126	0.0007	-172.2	<.0001
dOL S Post Lag1					-0.058	0.0010	-55.5	<.0001	-0.031	0.0009	-34.0	<.0001
dOL S Post Lag2					0.008	0.0010	7.3	<.0001	0.021	0.0009	24.2	<.0001
Prior BI Qty									0.660	0.0031	211.2	<.0001
Prior BI Qty Lag1									-0.084	0.0022	-38.1	<.0001
Prior SI Qty									-0.373	0.0030	-122.7	<.0001
Prior SI Qty Lag1									0.073	0.0022	33.3	<.0001
Prior BI Qty ^2									-0.002	0.0000	-158.7	<.0001
Prior SI Qty ^2									0.001	0.0000	84.3	<.0001
SI Qty									-0.833	0.0056	-148.7	<.0001
BI Qty									1.528	0.0055	276.3	<.0001
OL B Post									-0.508	0.0016	-321.0	<.0001
OL S Post									0.336	0.0016	215.5	<.0001
OL B Prior L5									0.071	0.0022	32.7	<.0001
OL B Prior L4									0.120	0.0023	53.0	<.0001
OL S Prior L5									-0.024	0.0021	-11.7	<.0001
OL S Prior L4									-0.045	0.0022	-20.5	<.0001
R-Square	0.07				0.11				0.38			

Table 6.4: Regression Results — Different Measurement Intervals

Notes: Table shows the estimation results of three models to explain the observed response on the buy side during the post-measurement interval. Events are selected according to our observed changes in quote segments. Results correspond to a measurement window of 25 milliseconds. Explanatory variables include response on both sides of the market in the pre-measurement interval (dOL B/S Pre), signed volume corresponding to the prior quote segment (Prior BI/SI Qty), signed volume (SI/BI Qty) and offered liquidity (OL B/S Post) linked to the event at t_0 , and information related to offered liquidity during the prior quote segment at quote segment levels 4 and 5.

Dependent Variable: dOL B Post

	Model 3 (25ms)				Model 3 (50ms)				Model 3 (100ms)			
	Parameter Estimate	Standard Error	t Value	Pr > t	Parameter Estimate	Standard Error	t Value	Pr > t	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1.112	0.0928	12.0	<.0001	0.845	0.0514	16.4	<.0001	0.413	0.0250	16.5	<.0001
dOL B Pre	0.088	0.0008	117.2	<.0001	0.072	0.0009	82.6	<.0001	0.023	0.0010	24.0	<.0001
dOL B Post Lag1	0.032	0.0009	34.7	<.0001	0.039	0.0010	38.0	<.0001	0.038	0.0011	34.5	<.0001
dOL B Post Lag2	-0.002	0.0009	-2.6	0.0084	-0.001	0.0010	-0.6	0.568	0.008	0.0011	7.2	<.0001
dOL S Pre	-0.126	0.0007	-172.2	<.0001	-0.105	0.0008	-124.9	<.0001	-0.061	0.0009	-64.7	<.0001
dOL S Post Lag1	-0.031	0.0009	-34.0	<.0001	-0.031	0.0010	-30.8	<.0001	-0.019	0.0011	-17.2	<.0001
dOL S Post Lag2	0.021	0.0009	24.2	<.0001	0.021	0.0010	21.8	<.0001	0.017	0.0011	15.5	<.0001
Prior BI Qty	0.660	0.0031	211.2	<.0001	0.337	0.0017	195.6	<.0001	0.117	0.0009	134.7	<.0001
Prior BI Qty Lag1	-0.084	0.0022	-38.1	<.0001	-0.043	0.0012	-34.6	<.0001	-0.011	0.0006	-17.6	<.0001
Prior SI Qty	-0.373	0.0030	-122.7	<.0001	-0.224	0.0017	-131.8	<.0001	-0.088	0.0008	-107.2	<.0001
Prior SI Qty Lag1	0.073	0.0022	33.3	<.0001	0.046	0.0013	37.1	<.0001	0.016	0.0006	25.8	<.0001
Prior BI Qty ^2	-0.002	0.0000	-158.7	<.0001	-0.001	0.0000	-122.4	<.0001	0.000	0.0000	-58.6	<.0001
Prior SI Qty ^2	0.001	0.0000	84.3	<.0001	0.000	0.0000	82.4	<.0001	0.000	0.0000	48.0	<.0001
SI Qty	-0.833	0.0056	-148.7	<.0001	-0.523	0.0032	-163.0	<.0001	-0.231	0.0017	-139.6	<.0001
BI Qty	1.528	0.0055	276.3	<.0001	0.723	0.0032	227.7	<.0001	0.273	0.0016	166.3	<.0001
OL B Post	-0.508	0.0016	-321.0	<.0001	-0.250	0.0009	-285.5	<.0001	-0.089	0.0004	-210.3	<.0001
OL S Post	0.336	0.0016	215.5	<.0001	0.148	0.0009	171.9	<.0001	0.046	0.0004	111.7	<.0001
OL B Prior L5	0.071	0.0022	32.7	<.0001	0.028	0.0012	23.3	<.0001	0.009	0.0006	15.7	<.0001
OL B Prior L4	0.120	0.0023	53.0	<.0001	0.070	0.0012	56.2	<.0001	0.032	0.0006	53.9	<.0001
OL S Prior L5	-0.024	0.0021	-11.7	<.0001	-0.016	0.0011	-14.1	<.0001	-0.005	0.0006	-9.2	<.0001
OL S Prior L4	-0.045	0.0022	-20.5	<.0001	-0.039	0.0012	-32.4	<.0001	-0.020	0.0006	-33.9	<.0001
R-Square	0.38				0.35				0.23			

6.6 Directional Movement

Here we investigate the relationship between OL and ΔOL during periods when the market is moving directionally up or down; we capture movement based on identified sequences of quote segments. As shown in Table 4.3, we observe many instances of directional movement — a sequence of N quote segments exist that move the market strictly up or down.

For each set of quote segments that define a directional move, the last quote segment of the sequence represents the turning point of the market movement. Therefore, if we have a sequence with $N = 3$, then the change in the inside quote segment level will be the same (either $-0.01, -0.01, -0.01$ or $+0.01, +0.01, +0.01$), but the next change will have the opposite sign. This implies the first quote segment, for each sequence, represents the continuation of a trend started, at some point, within the prior quote segment.

6.6.1 Observed Imbalance in OL During Directional Moves

Here we investigate how OL can help explain directional movement in quote segments. To do this, we compute a time-weighted measure of OL , on both sides of the market, for each quote segment: $OL_i^B = \text{time-weighted } OL \text{ on the buy side for quote segment } i$. We divide up each quote segment based on intervals with no change in OL , then for each interval within the quote segment, we have a start and end time used to compute the duration. Our final measurement is an average of the observed OL weighted by the duration.

Table 6.5 shows how this measure of OL on both sides of the market relates to directional movement. For the first month of the sample, we find all directional

movements of quote segments and then sequence each quote segment within the directional move (e.g., Move 1 is the first quote segment of the sequence). Next, focusing only on downward moving trends (upward trends contain the reverse findings), we find average OL on each side of the market conditional on the number of directional moves and the location within the movement.

Table 6.5: OL : Buy and Sell Side with Directional Movement

Notes: Table shows how OL varies conditional on directional movement of quote segments. Each row group contains a row for the buy side (bid) and the sell side (ask) and reflects the number of directional moves down. For each move within the group, we show offered liquidity on both sides of the market.

Directional Movement											
N	Side	Move 1	Move 2	Move 3	Move 4	Move 5	Move 6	Move 7	Move 8	Move 9	Move 10
10	bid	36	31	30	30	32	32	30	29	32	43
	ask	43	41	43	40	39	42	41	40	37	25
9	bid	30	34	31	30	31	31	28	31	40	
	ask	37	37	41	41	40	38	40	36	26	
8	bid	28	30	30	30	31	29	31	41		
	ask	38	38	38	40	39	39	36	27		
7	bid	30	32	31	30	29	32	44			
	ask	40	40	41	41	41	38	27			
6	bid	31	31	31	30	31	31	42			
	ask	38	37	40	39	37	27				
5	bid	31	31	29	30	41					
	ask	39	38	38	37	27					
4	bid	30	29	30	41						
	ask	38	37	37	28						
3	bid	29	30	40							
	ask	37	36	28							
2	bid	28	38								
	ask	36	26								
1	bid	38									
	ask	26									

Bounce Movement											
N	Side	Move 1	Move 2	Move 3	Move 4	Move 5	Move 6	Move 7	Move 8	Move 9	Move 10
10	bid	35	25	37	25	38	25	38	26	40	28
	ask	32	44	29	42	30	44	30	42	27	42
9	bid	41	31	44	32	45	31	46	31	46	
	ask	31	47	31	47	31	48	32	47	35	
8	bid	40	29	42	29	43	30	44	31		
	ask	34	49	33	51	33	51	34	52		

Based on the conditional averages displayed in Table 6.5, we find that during movements down, OL is larger on the sell side until the very last quote segment of the directional movement. This is significant since the last quote segment represents the interval when the directional movement stops. The last movement always flips and the buy side displays a larger OL . This shows that continued directional movement results from the build up of OL on the sell side — market depth places pressure on the market to keep moving in a direction of lower OL .

The lower portion of Table 6.5 shows similar information, but now for sequences of quote segments that bounce within a narrow range. For this figure, we focus on bounce movement with the first move representing a movement down, the second move representing a movement up, and so on. Comparing the bounce movement to the directional movement, we find a clear difference in the observed OL existing during each quote segment.

This exploration of OL raises a number of question. For example, if observed imbalance in OL provides information for future market direction, at what point during the quote segment is this imbalance observable? We are using a time-weighted measure of OL during the full quote segment, so it would be difficult for the imbalance to show up at the very end of a long quote segment.

6.6.2 OL and ΔOL at Start & End of Quote Segments

Next, we focus on a small interval at the start and just prior to the end of each quote segment. For each interval, we compute OL and ΔOL for both sides of the market. To measure ΔOL , we select an interval of 50 milliseconds for both the start and end interval of each quote segment. We use the snapshot of OL at the start of each interval. This results in four measurements of OL and ΔOL for each quote segment associated with the start and end for both sides of the market (see Figure 6.9).

This analysis requires knowing the exact time when the current quote segment ends. This is reasonable since we are simply using the information to study how OL and ΔOL relate during periods of directional movement in quote segments.

Table 6.6 shows offered liquidity and measured responses during the start and

end of a quote segment conditional on the quote segments location within a directional movement. Here we show a subset of directional movements from three movements to five movements down.

Table 6.6: *OL* and ΔOL : Buy and Sell Side with Directional Movement

Notes: Table shows *OL* and ΔOL from 50ms after the start of a quote segment and 50ms from the end of a quote segment. A subset of quote segments are selected that are included in directional movements down. Information is shown as averages conditional on the directional move and the location within the move.

Offered Liquidity (OL)													
		25ms				50ms				100ms			
Order		S Start	B Start	S End	B End	S Start	B Start	S End	B End	S Start	B Start	S End	B End
-5	1	31	45	49	30	31	45	48	32	31	45	46	32
	2	30	47	49	31	30	47	47	32	30	47	45	33
	3	32	45	49	29	32	45	47	31	32	45	46	32
	4	31	45	47	30	31	45	45	32	31	45	44	33
	5	28	50	26	53	28	50	27	51	28	50	28	51
-4	1	31	44	49	30	31	44	48	31	31	44	46	32
	2	30	45	49	29	30	45	47	31	30	45	45	32
	3	31	46	48	30	31	46	46	32	31	46	44	33
	4	28	49	26	52	28	49	27	51	28	49	28	51
-3	1	30	42	49	28	30	42	47	29	30	42	46	30
	2	29	45	46	30	29	45	44	31	29	45	43	33
	3	28	48	26	51	28	48	27	50	28	48	28	49
Responsiveness (dOL)													
		25ms				50ms				100ms			
Order		S Start	B Start	S End	B End	S Start	B Start	S End	B End	S Start	B Start	S End	B End
-5	1	43	-34	18	-24	23	-21	14	-14	13	-12	11	-9
	2	47	-39	19	-25	26	-24	16	-16	16	-14	13	-12
	3	46	-37	17	-24	26	-24	15	-16	16	-14	13	-12
	4	44	-35	15	-21	25	-22	14	-14	15	-14	12	-11
	5	28	-21	-20	14	11	-9	-10	9	2	-2	-5	5
-4	1	43	-35	18	-25	23	-21	15	-15	14	-12	12	-10
	2	46	-38	18	-25	26	-25	16	-16	16	-15	13	-12
	3	44	-36	15	-21	24	-23	13	-14	14	-14	11	-11
	4	29	-23	-19	13	11	-11	-10	8	2	-3	-5	4
-3	1	43	-34	19	-24	23	-21	15	-15	14	-12	12	-10
	2	44	-37	17	-22	25	-23	14	-14	15	-14	12	-11
	3	28	-25	-20	13	11	-12	-10	8	2	-3	-5	4

We find both *OL* and ΔOL are informative in determining when the movement will stop. Starting offered liquidity on both sides of the market do not show any change — for down movement, the ask side starts lower than the bid side. The end offered liquidity shows that the sell side is larger in all cases except for the very end. There is a clear flip in ΔOL at the last movement as well. This suggests that monitoring both *OL* and ΔOL is helpful in predicting when movements end.

6.6.3 Predictive Modeling — Logistic Regression Analysis

Here we test if information from OL and ΔOL can be used to predict the probability of a future upward movement in the quote segment. We use a logistic regression approach to test if information related to offered liquidity can help us predict a future market movement. As above, we use the change in quote segments to define t_0 . For each quote segment change, we measure OL and ΔOL during the pre- and post-measurement interval. For each observation, we use the next movement to define an indicator variable equal to 1 for an upward movement and 0 for a downward movement.

We define the logit equation to be a function of OL and ΔOL on both sides of the market during the post-measurement interval. We specify the following model:

$$\text{Logit}(Y) = \beta_0 + \beta_1 * \Delta OL_{post}^B + \beta_2 * \Delta OL_{post}^S + \beta_3 * OL_{post}^B + \beta_4 * OL_{post}^S \quad (6.8)$$

Table 6.7 contains the regression results. We run the model on three different measurement intervals with durations of 25, 50, and 100 milliseconds. In all measurement intervals, we find the estimated parameter values for ΔOL and OL are statistically significant in modeling the probability of an upward movement in the inside quote segment level. For each parameter estimate, we show the estimate along with a 95% confidence interval and we display the odds ratio with a corresponding 95% confidence interval. All confidence intervals for the odds ratio do not include 1. For each duration, we find the model is statistically significant and the c-statistic indicates an edge is gained by using these measurements to model

the probability of a future movement.

Comparing results across different measurement intervals shows a noticeable difference in the estimated parameter values for ΔOL on both sides of the market. While the signs remain the same, the levels increase significantly. For example, the estimated parameter value doubles in size as we move from a 25 millisecond duration to a 50 millisecond duration. We do not find the same change in the OL variables as they tend to stay the same size as we increase the measurement interval. We also find the prediction success rate also increases as we allow a longer duration to measure ΔOL as shown by the c-statistic.

The higher emphasis placed on ΔOL relative to OL as we extend the measurement interval makes sense given we are allowing the rate of change to include a longer history. For example, after an event, there might be an immediate movement away from the market by the buy side, which would indicate a higher probability of not observing an upward movement. If this movement occurs in the first 50 milliseconds, then a change happens and the buy side starts to increase visible depth, then we would expect a different measure of ΔOL as we expand the measurement interval. Furthermore, by extending the measurement interval, we are using information that becomes closer to the event we are modeling.

Table 6.8 shows predicted probabilities following the estimated models in Table 6.7. We show estimates related to two measurement intervals to explore how the change influences our prediction ability. The first six rows compares how the predicted probability of a future upward movement in quote segments depends on the size of ΔOL . We hold OL , for both sides of the market, constant at 30, which is very close to the expected level. We show results assuming a positive ΔOL for

the buy side and a negative ΔOL for the sell side. Therefore, it is not surprising to see an estimate larger than 50 percent for all six rows. We find there is a significant increase in the predicted probability as the magnitude of the response increases. All predicted probabilities are statistically significant.

The lower portion of Table 6.8 studies how the prediction changes as we vary OL and highlights the important role visible depth performs. Even with a sizable response that might indicate upward pressure on the market, when the sell side is much larger than the buy side, we have a low predicted probability of a future upward move.

Table 6.7: Logistic Regression Results

Notes: Table shows logistic regression results based on defined measurement intervals of 25, 50 and 100 milliseconds. We model the probability of an upward movement in the quote segment. For each measurement interval, we show the parameter and corresponding odds ratio estimates along with confidence intervals. These results are based on a random sample of 20 thousand quote segment changes during the first month of the sample (177 thousand for the total sample).

*** 25ms Measurement Interval**

Parameter	Estimate	95% Confidence Limits		Standard	Wald		Odds Ratio Estimates and Profile-Likelihood CIs		
				Error	Chi-Square	Pr > ChiSq	Effect	Estimate	95% Confidence Limits
Intercept	0.123	0.047	0.199	0.0388	10	0.0015			
dOL B Post	0.008	0.007	0.009	0.0005	262	<.0001	dOL B Post	1.008	1.007 1.009
dOL S Post	-0.007	-0.008	-0.006	0.0005	231	<.0001	dOL S Post	0.993	0.992 0.994
OL B Post	0.015	0.013	0.017	0.0009	270	<.0001	OL B Post	1.015	1.013 1.017
OL S Post	-0.018	-0.020	-0.016	0.0009	389	<.0001	OL S Post	0.982	0.981 0.984

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	1419	4	<.0001
Score	1283	4	<.0001
Wald	1191	4	<.0001
c-statistic	66%		

*** 50ms Measurement Interval**

Parameter	Estimate	95% Confidence Limits		Standard	Wald		Odds Ratio Estimates and Profile-Likelihood CIs		
				Error	Chi-Square	Pr > ChiSq	Effect	Estimate	95% Confidence Limits
Intercept	0.040	-0.037	0.118	0.0396	1	0.3069			
dOL B Post	0.015	0.013	0.017	0.0009	293	<.0001	dOL B Post	1.015	1.013 1.017
dOL S Post	-0.015	-0.017	-0.014	0.0009	317	<.0001	dOL S Post	0.985	0.983 0.986
OL B Post	0.017	0.015	0.018	0.0009	322	<.0001	OL B Post	1.017	1.015 1.018
OL S Post	-0.016	-0.018	-0.015	0.0009	327	<.0001	OL S Post	0.984	0.982 0.986

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	2164	4	<.0001
Score	1721	4	<.0001
Wald	1661	4	<.0001
c-statistic	69%		

*** 100ms Measurement Interval**

Parameter	Estimate	95% Confidence Limits		Standard	Wald		Odds Ratio Estimates and Profile-Likelihood CIs		
				Error	Chi-Square	Pr > ChiSq	Effect	Estimate	95% Confidence Limits
Intercept	-0.039	-0.118	0.041	0.0407	1	0.3443			
dOL B Post	0.035	0.032	0.038	0.0017	438	<.0001	dOL B Post	1.036	1.032 1.039
dOL S Post	-0.031	-0.035	-0.028	0.0016	386	<.0001	dOL S Post	0.969	0.966 0.972
OL B Post	0.018	0.016	0.020	0.0009	376	<.0001	OL B Post	1.018	1.016 1.020
OL S Post	-0.016	-0.017	-0.014	0.0009	312	<.0001	OL S Post	0.985	0.983 0.986

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	3140	4	<.0001
Score	2094	4	<.0001
Wald	1895	4	<.0001
c-statistic	72%		

Table 6.8: Logistic Regression Results — Predicted Probabilities

Notes: Table shows predicted probabilities associated with logistic regression results presented in Table 6.7 for 25 and 50 millisecond measurement intervals. Within each measurement interval, the upper portion fixed *OL* as ΔOL changes; the lower portion fixes ΔOL as *OL* changes. For each predicted probability estimate of a future upward movement in quote segments, we include the confidence interval.

*** 25ms Measurement Interval (Alpha = 5%)**

<u>dOL B Post</u>	<u>dOL S Post</u>	<u>OL B Post</u>	<u>OL S Post</u>	<u>Estimate</u>	<u>Standard Error</u>	<u>Confidence Limits</u>	<u>Wald Chi-Square</u>	<u>Pr > ChiSq</u>
10	-10	30	30	54%	0.4%	54% 55%	133.1	<.0001
20	-20	30	30	58%	0.4%	57% 59%	357.9	<.0001
30	-30	30	30	62%	0.5%	61% 63%	571.1	<.0001
40	-40	30	30	65%	0.5%	64% 66%	733.2	<.0001
50	-50	30	30	68%	0.6%	67% 70%	847.2	<.0001
60	-60	30	30	72%	0.6%	70% 73%	926.2	<.0001
30	-30	10	100	26%	1.2%	23% 28%	272.6	<.0001
30	-30	20	90	32%	1.2%	30% 35%	195.5	<.0001
30	-30	30	80	40%	1.0%	38% 42%	93.9	<.0001
30	-30	40	70	48%	0.9%	46% 50%	6.3	0.0123
30	-30	50	60	56%	0.8%	54% 57%	53.4	<.0001
30	-30	60	50	64%	0.8%	62% 65%	252.0	<.0001
30	-30	70	40	71%	0.9%	69% 73%	426.5	<.0001
30	-30	80	30	77%	0.9%	75% 79%	519.7	<.0001
30	-30	90	20	82%	0.9%	81% 84%	562.7	<.0001
30	-30	100	10	87%	0.9%	85% 88%	581.7	<.0001

*** 50ms Measurement Interval (Alpha = 5%)**

<u>dOL B Post</u>	<u>dOL S Post</u>	<u>OL B Post</u>	<u>OL S Post</u>	<u>Estimate</u>	<u>Standard Error</u>	<u>Confidence Limits</u>	<u>Wald Chi-Square</u>	<u>Pr > ChiSq</u>
10	-10	30	30	59%	0.4%	58% 60%	429.1	<.0001
20	-20	30	30	66%	0.5%	65% 67%	924.0	<.0001
30	-30	30	30	72%	0.6%	71% 73%	1221.4	<.0001
40	-40	30	30	78%	0.6%	77% 79%	1382.9	<.0001
50	-50	30	30	83%	0.6%	82% 84%	1473.9	<.0001
60	-60	30	30	87%	0.6%	86% 88%	1528.4	<.0001
30	-30	10	100	38%	1.5%	35% 41%	65.3	<.0001
30	-30	20	90	46%	1.3%	43% 48%	11.6	0.00
30	-30	30	80	54%	1.1%	52% 56%	12.0	0.00
30	-30	40	70	62%	0.9%	60% 63%	156.3	<.0001
30	-30	50	60	69%	0.8%	68% 71%	454.8	<.0001
30	-30	60	50	76%	0.8%	74% 77%	721.7	<.0001
30	-30	70	40	81%	0.8%	80% 83%	847.0	<.0001
30	-30	80	30	86%	0.7%	84% 87%	876.9	<.0001
30	-30	90	20	89%	0.7%	88% 91%	867.2	<.0001
30	-30	100	10	92%	0.6%	91% 93%	845.0	<.0001

6.7 Conclusions

There exist documented concerns related to the fleeting nature of liquidity — this chapter uses the limit order book data to investigate this aspect of liquidity. We investigate how visible depth changes on both sides of the market — fleeting liquidity is not just about cancellations; the speed of limit order arrival also matters. A relevant question is why visible depth might suddenly appear or disappear. Opposite changes linked to buy and sell side indicates a book reorganizing itself at a new price level.

We propose a method to measure offered liquidity, then we propose a method to measure the degree of responsiveness — a high response indicates a high propensity for new orders to enter and old orders to leave the book via cancellations. We focus on very short time intervals of 25, 50, and 100 milliseconds. Using limit order book data, we measure OL and ΔOL based on a random selection of timestamps. This provides us with information on expected variation in the two measures. Using limit order book data, we measure OL and ΔOL based on timestamps corresponding to changes in quote segments. We study the tendency for quote segments to appear in sequences that directionally move the market up or down. We explore how OL and ΔOL can be useful for building expectations of future market direction.

We find variation does exist in OL and ΔOL within the sample period — variation explained by volume and volatility, which arises from large directional trading. We find the start of quote segments is linked to higher responses than we would otherwise expect based on a random selection of timestamps. We find measurement of ΔOL changes based on the amount of time considered — larger

responses are linked to shorter time intervals, which indicates reversals in responses (e.g., immediate large negative response on the buy side is followed by a positive response as depth rebuilds). We find there does exist valuable information in the pre-interval — negative response in the pre-interval is linked to a negative response in the post-interval. This indicates that cancellations, and a book in motion with adjusting depth on both sides, starts prior to the event that changes the best quotes.

Chapter 7

Flash Quotes

7.1 Introduction

This chapter focuses on flash events resulting from the arrival and departure of quotes in the limit order book that improve the market. This improvement results when a new limit order arrives and defines a new best bid or ask quote. While these improvements should be viewed as always beneficial to the market, questions arise when we see a sudden departure unrelated to trades.

Should there be a concern about quotes that flash a better bid or ask quote and almost immediately cancel if no other market participants act on the newly available better price? Are these actions suggestive of an attempt to manipulate the limit order book? Do these actions reflect a true intent to trade?

One of the primary concerns should be the question of fairness. In a fully electronic market, participants can be either human or machine. The concern regarding flash events relates to whether one group has the ability to act on such information. If the new best price is only available for five milliseconds, then

it seems unlikely that human traders will be able to notice the better price and react. This then raises a question about who these improved offers are designed for? And if they are designed for algorithms to trade with other algorithms, is such a system fair?

Our prior analysis of the limit order book provides empirical facts that suggest flash quotes exist in the sample period. For example, in Section 3.3, the study of QI and QP events and the bid-ask spread, showed a common pattern of a QI event followed by a QP event on the same side of the market. This indicates a flash event if the duration between these events is short. The duration of a wide bid-ask spread also suggests we can expect the duration to be very short.

Furthermore, based on prior chapters, we know that the majority of orders enter and exit very close to the market; there exists a significant amount of quote cancellation (see Subsection 3.2.2); and a decent share of total volume occurs at the end of a quote (see Subsection 3.2.3). Therefore, a detailed analysis of flickering quotes adds to our general understanding of quote cancellations, the dynamics of market depth, and the informational value of a trade.

Based on visual inspection of the high frequency book update information, we find evidence of such market activity. Figure 7.1 shows six examples from the data. In each example, we show the adjusted best bid (blue) and offer (red) paths and, directly below, the unadjusted best bid and offer quote series. Yellow highlighted boxes are added to mark the activity of interest. In each example, we observe many events move toward the other side of the market by establishing a position at the open quote level between the best quotes. They then move away within the next few book updates. Some result in trades while others do not (see Chapter

3 for more information related to the visualization). The examples demonstrate situations where the events occur on a single market side and situations where there is a mixture of events related to both sides of the market. The examples also demonstrate periods of events followed by periods where we do not observe such events.

This chapter is focused on investigating the limit order book data to study why we observe flash quotes. We claim this activity is fundamentally different from the fleeting nature of liquidity, which was the subject of Chapter 6. We believe it is important to separate the study of fleeting liquidity from flash quoting since the concept of fleeting liquidity concerns a larger market movement as the limit order book repositions visible depth on both sides of the market. Flash quotes, on the other hand, represent small changes in visible depth related to the arrival of limit orders that have a very short life in the limit order book.

Having studied the larger patterns in Chapter 6, we incorporate those measurements of offered liquidity (OL) and the responsiveness of offered liquidity (ΔOL) into the study of flash quotes. We hope to connect these two concepts given their prevalence in the limit order book visualization. We focus on the following questions: How often do we observe flash quotes? Are flash events related to the desire to trade? Is there an intent to trade? How are flash events related to market volume, volatility, OL , and ΔOL .

The remainder of the chapter is organized as follows: Section 7.2 introduces the information source and steps taken to rebuild the limit order book; once the data is processed, we introduce the event of interest. Section 7.3 provides an analysis of flash quotes identified within the sample. Section 7.4 analyzes the

relationship between flash quotes and liquidity provision. Section 7.5 studies the relationship between flash quotes, volume, and volatility. Section 7.6 investigates the relationship between flash quotes and minutes surrounding expected release of information.

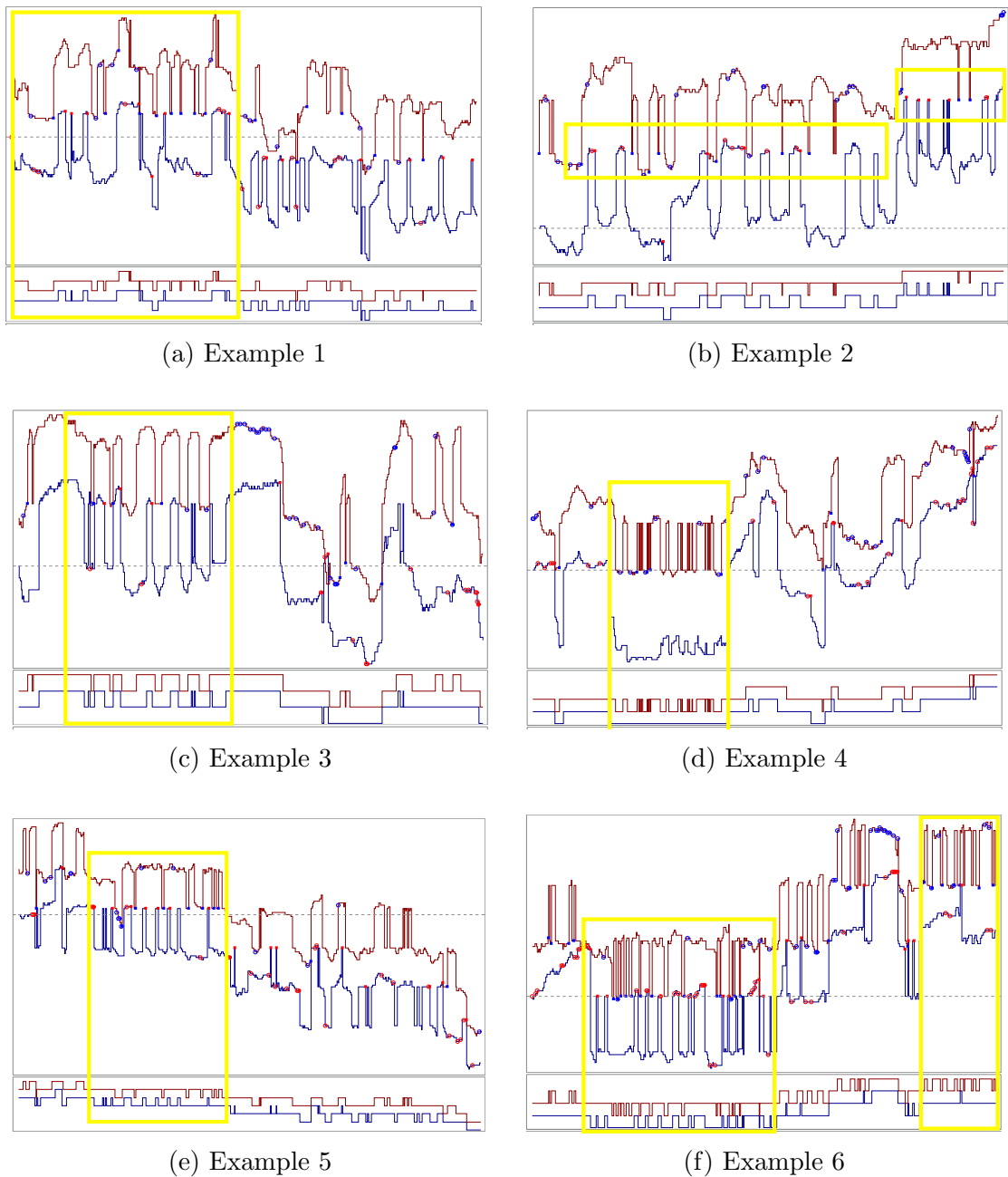


Figure 7.1: Six Examples of Quote Improvements

Notes: Each figure shows a sample of order book data for a fixed interval of 500 book updates (updates not shown spaced out according to time) to the best bid (blue) and ask quote (red). Within each figure, the upper section shows the best bid and ask quotes adjusted to reflect changes in market depth. Ask quote is adjusted to move toward the best bid quote if the arrival of limit orders add depth to the quote. Trades are marked on the quote that provides liquidity — red for seller-initiated trades and blue for buyer-initiated; open circles indicate trades that did not extract all available liquidity, filled in squares indicate all liquidity removed from trades and the best quote expires. The lower section of each figure shows the standard quotes for both sides without incorporating information on market depth. Highlighted sections display the quote behavior of interest.

7.2 Data & Methodology

This section introduces the data used in this chapter, defines the event of interest, and outlines the method to identify events and relevant information related to each event.

7.2.1 Data

Starting Raw Data — CME Message Data:

This study uses individually purchased market depth data from the CME for one year related to crude oil futures. The sample period includes all trade dates between August 1, 2012 to July 31, 2013. This is the same information sold by the exchange to market participants and allows us to rebuild the limit order book for the top ten levels of the book. The data include timestamps precise to the millisecond and also contains the complete set of transactions fully integrated with the observed changes to the limit order book.

The raw information, delivered by the exchange, represents a collection of messages created by the exchange whenever there is a change in the visible limit order book. For example, if a market participant places a new limit order for one contract at the second-best bid, then, if prior visible depth is 4, the action will force the exchange to issue a message that an updated depth at the second-best bid is now 5.

A single message can have many updates. For example, if a market participant places a market order to buy and if the requested quantity is equal to the resting depth at the best ask quote, then the trade event removes the best ask quote. The exchange will issue a message with trade details followed by all ten ask levels with

updated price levels and visible depth for each book level one through ten.

Each message needs to be broken down into individual pieces of information (referred to as data blocks), which adds to the challenge of using such data. Information within the message is defined using TAGs. Within each message there is an ID to correctly sort components of the message.

Rebuilt Limit Order Book for Both Sides:

To rebuild the limit order book, it is necessary to break down each raw message and then sort the set of book updates using both the message ID and the data block ID. We select all information connected to the nearby expiration with the largest observed daily volume.

We rebuild the limit order book for each side of the market separately. Focusing on a single side, we rebuild the limit order book by starting with the first update and place the information in a grid of ten levels. We retain information at each level if updates do not adjust the level. We keep the last update within a related set of updates. For example, if all ten levels change, then there will be ten rows with updates starting with an update at level one only followed by and updated level one with an update to level two, etc. We only need to keep the last row here, which contains the updates to all book levels in a single row.

For a single side of the market, the rebuilt limit order book contains two sort IDs, a timestamp, product information, ten quote levels and ten visible depth numbers corresponding to each book level one to ten.

Best Quote Combined Book Update Table:

After processing the raw message data to rebuild the limit order book, we want to combine information from both sides of the market into a single observation linked to the millisecond the update was recorded.

Focusing on the best quotes from each side of the market, we pull all rows displaying a change in either the price level or visible depth for the best bid and ask quotes. We then build a table with each row corresponding to a unique timestamp that can be linked to some observed change at one or both of the best quotes. Then, for each row, we fill in the observed updates and pull down price and depth over time if one side does not change. This process results in a table with observations linked to a change at the best quotes and allows us to observe changes over time on both sides of the market.

To add information related to quotes above the best ask and below the best bid, we collect snapshot information of visible depth for the next four book levels. Since differences between observed quotes within a single side are almost always the minimum allowed tick size (\$0.01 in crude oil futures), we only keep visible depth information for each book level.

The use of snapshot information related to book levels away from the best quotes means we are not including timestamps linked to an observed change away from the best quotes. For example, if a market participant places an order to buy at the second-best bid, we will not see this update reflected in our table until there is an observed change at the best quotes.

Addition of QI/QP Events, Quote Segments, & Trades:

Once the best quote combined book update table is created from the processed raw message data, we add additional information based on observed changes in the best quotes and transaction information. Specifically, we use changes in the best quotes to define quote segments (see Chapter 4) to divide the day into disjoint intervals. We also incorporate transaction information to identify the reason for observed changes to the limit order book. Trade information is necessary to determine if an observed decrease results from the arrival of market orders or the decision to cancel resting limit orders.

We define a **quote improvement (QI)** event when an update creates an improvement in the inside price. A Bid-QI event occurs when the buy side moves upward, toward the sell side; an Ask-QI event occurs when the sell side moves downward, toward the buy side.

A **quote promotion (QP)** event occurs when the second-best quote is promoted to the first-best quote. A Bid-QP event occurs when the buy side moves downward, away from the sell side; an Ask-QP event occurs when the sell side moves upward, away from the buy side.

We define quote segments as non-overlapping intervals of time which restrict quotes to live within a narrow range. For each day, we start by identifying the first occurrence of a two-tick bid-ask spread and define this at the starting point of the first quote segment; the best ask quote level is defined as the upper quote segment level and the best bid quote level is defined as the lower quote segment level. Since we have a two-tick spread, the open quote level between the best quotes is defined as the inside quote segment level. This quote segment exists

as we encounter subsequent book updates until the market meets conditions for a new quote segment. We mark the start of a new quote segment if the bid-ask spread is two-ticks or less and either the best bid is below the lower quote segment level or the best ask is above the upper quote segment level. After we scan across all book updates, we are left with each observation belonging to a quote segment and we know exactly where the best quotes live within that quote segment.

Besides the updates to the limit order book, the raw message data contains the complete set of trades. For each trade, we can observe the timestamp with millisecond precision, the quantity, and execution price. Since a book update message immediately follows trade information, we are able to correctly sequence observed trades within the evolving limit order book allowing us to determine the status of the book prior and immediately following each trade. This allows us to distinguish between buyer- or seller-initiated trades.

Individual trades are determined based on the market order and characteristics of resting limit orders. If a market order arrives to buy 10 contracts, the number of observed trades will depend on the resting limit orders first in line to trade. If all of these resting limit orders are for one contract, then we will observe ten individual trade records all for one contract — all occurring at the same millisecond and all within the same raw message. We define groups of individual transactions by combining quantity linked to trades uninterrupted by book updates. Since it is unlikely that market orders arrive at the same millisecond, we assume trades sharing the same millisecond are related to a single market order. For each trade group, we place this information at the book update reflecting the limit order book post-trade. Furthermore, we identify if the traded volume is buyer- or seller-

initiated and if this traded volume occurs at the best quotes or away from the best quotes.

7.2.2 QIP Event — Sequence of QI QP Events

Using terminology introduced in Chapter 3, we define the event of interest as a quote improvement (QI) followed by a quote promotion (QP). The initial QI is placed at the inside quote segment level. For example, a Bid-QI followed by a Bid-QP, moves the best bid quote from its initial placement, at the lower quote segment level, to the inside quote segment level, then back down to the lower quote segment level. Given the sequence of a QI followed by a QP, we refer to these combined events as a QIP event and further define a Bid-QIP, with the buy side flashes a better price, or an Ask-QIP, with the sell side flashes a better price.

For QIP events to exist, it is necessary to have the inside quote segment level open. This necessarily implies that the market exists as somewhere between a one and two tick bid-ask spread. Based on evidence displayed in Chapter 3, we know bid-ask spreads are commonly measured close to the minimum tick size. But there exists many observations where the bid-ask spread is larger than the minimum tick size. As we often view the bid-ask spread as a measure of market liquidity this raises the question if flash liquidity provision is optimal under periods where the market is able to expand the bid-ask spread?

Figure 7.2 shows three hypothetical examples of QIP events as we observe them within a single quote segment. For each QI, the blue arrow pointing toward the inside quote represents the moment the QI arrives. QI events start when a limit order arrives that betters the market, meaning it is either a new minimum

ask quote (Ask-QI) or a new maximum bid quote (Bid-QI). The horizontal blue bar at the inside quote segment level, following the directional arrow indicates the duration of the QIP event; during the QIP event, the market is at the minimum bid-ask spread and additional quotes can be added to the QIP event such that visible (or non-visible) depth changes. QIP events end when either an aggressive order hits the quote and removes all resting depth, resulting in a trade, or the resting limit order is cancelled by the trader; we note the occurrence of this by a red directional arrow pointing back toward the upper (lower) quote level for an Ask-QIP (Bid-QIP).

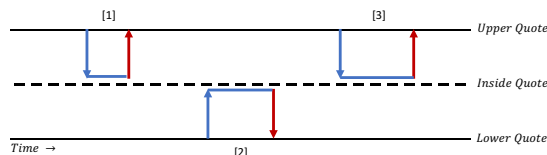


Figure 7.2: Example of three QIPs

Figure shows three examples of QIPs occurring within a single quote segment; we assume there exists visible depth on the ask (bid) side at the upper (lower) quote. QIP [1] and [3] are on the ask side since they are initiated with the arrival of a new minimum quote; conversely, QIP [2] occurs on the bid side.

7.2.3 QIP Identification

We search for QIP events by scanning the best quote combined book update table for each day in the sample. Since QIP events are linked to a specific side of the market, we scan for Bid-QIP events first by only considering movement in the price level associated with the best bid quote. Focusing on price changes within a quote segment, we identify a Bid-QI event followed by a Bid-QP event. Given we want flash quotes that touch the inside quote segment price level, we require the initial Bid-QI event to establish an improvement of the inside price equal to the inside quote segment price level.

For each QIP event, we collect information regarding the side of the market, the quote segment containing the flash event, the duration (defined as the amount of time between the starting QI event and the ending QP event), changes in visible depth at the improved price, transactions that take place at the improved price, and the reason for the ending QP event.

For each QIP event, we track directional movement in depth at the improved price. We define the number of switches in directional movement based on how many times we observe depth switch from increasing to decreasing. Switches capture changing direction of depth associated with the QIP event and QIP events with more switches indicate more activity during the QIP event. We believe this aspect of the QIP event provides information regarding the underlying intent to trade. We use switch counts to trim the sample of QIP events to a subset that we care about, specifically those QIP events with a low number of updates and short durations.

7.3 QIP Summary

Following the identification method introduced above, we identify a total of 3.9 million QIP events during our sample period. There are, on average, 16 thousand QIP events per day. The average QIP event last for 650 milliseconds and contains 5 book updates. We find two-thirds of QIP events have observed trades and the majority of trades occur at the very end of the QIP event associated with the terminating QP event.

7.3.1 Distribution of QIP Characteristics:

Table 7.1 contains a summary of the resulting sample of QIP events during the first month of the sample. We use observable characteristics linked to each event to create distributions related to durations, the number of observed switches, the maximum observed depth during the event, and measured depth at the start of each event across the first two book levels.

The duration for each event is defined as the distance in seconds between the initiating QI event and the ending QP event. This measure provides relevant information regarding the preferences of the market participant submitting the order. We need to make sure and interpret this measure carefully as a short duration can result from either a cancellation or from a trade. Here we show the duration without reference to the ending QP event. The distribution of QIP event duration is concentrated close to zero with a median duration of 36 milliseconds. The observed average of just over 500 milliseconds indicates the presence of some large durations.

As discussed above, for each event, we determine the number of times depth flows switch sign. This was done to distinguish a QIP with a directional or zero build up in depth (zero switches) from a QIP with depth initially increases and then decreases (one switch). We find more than half of the QIP events contain no switches in depth. This finding is a function of the tendency for QIP events to have a very short duration as the probability of observing depth develop and adjust decreases as the duration approaches zero. Similar to durations, we find a large difference between the mean and median, suggesting a few outliers in our sample of QIP events. The recorded number of switches will help us trim the sample to

focus attention on flash events with little fluctuation in available liquidity.

All points of the distribution, related to maximum depth during the QIP event, are below ten contacts and the median and average are just two and three contracts, respectively. This evidence indicates the majority of QIP events add a very small share to total available liquidity. Visible depth at higher points of the distribution is less for QIP events indicating that QIP events tend to not develop a large supply of resting liquidity.

The lower portion of Table 7.1 provides information related to the visible depth at the start of the QIP event. According to how we identify QIP events, the best quote is equivalent to the inside quote segment level and the second-best quote represents either the upper or lower quote segment level depending on if the event is an Ask-QIP or Bid-QIP, respectively. Therefore, we can compare the distribution information with our prior understanding of how we expect depth to be displayed not accounting for flash events. We find, while the averages are very similar, the distribution at the start of QIP events shows lower depths relative during non-QIP events (see Table 4.4 in the chapter on Quote Segments). This indicates that QIP events occur when there is less depth displayed in the book.

Table 7.1: QIP Summary Table

Notes: Table contains information related to the distribution of the main characteristics associated with the QIP events. This is based on QIPs found during the first month of the sample in August 2012.

	Mean	StdDev	P5	P10	Q1	Median	Q3	P90	P95
<u>QIP Characteristics:</u>									
Time Duration (seconds)	0.561	1.579	0.002	0.003	0.006	0.036	0.402	1.544	2.799
Switches	2	10	0	0	0	0	1	4	6
Maximum QIP Depth	3	5	1	1	1	2	4	7	9
<u>Measured Depth at QIP Start:</u>									
Best Quote	2	2	1	1	1	1	1	3	4
Second-Best Quote	13	12	3	5	7	11	15	22	27

Table 7.2 summarizes a subsample of QIP events with a duration less than

500 milliseconds. The upper portion of the table represents QIP events that end without the indication of a trade, which we assume ends with a request to cancel the resting limit order; the lower portion of the table represents QIP events that end with a trade. We find an even split between these two sets with approximately 1.4 million QIP events in each. We find QIP event durations are much shorter if the QIP ends with a cancellation. The number of updates and the maximum depth are also smaller when the QIP ends with a cancellation. We can see a small number of QIP events ending due to a cancellation have trades occurring prior to the QIP that ends the event.

Based on Table 7.2, we can make an important distinction between observed durations of QIP events. Based on the median value, half of all QIP events, of the subset with a duration less than 500 milliseconds, only rest in the limit order book for seven milliseconds or less before they are cancelled.

On the other side, a short duration due to trade represents a case where the opposite side was able to act on the available liquidity. The initial QI event might have been unrelated to a high frequency trader, but this would depend on how soon the initiating QI event arrives relative to the wide bid-ask spread.

7.3.2 QIP Number of Updates & Switches:

Table 7.3 summarizes QIP events conditional on the number of observed book updates (found in the upper part of the table) and number of switches (lower part of the table). For each value of the conditioning variable, we show the total number of QIP events, averages of QIP characteristics, the total traded quantity associated with QIP events, traded quantity occurring at the end of QIP events, and, finally,

Table 7.2: QIP Characteristics Conditional on QP Event

Notes: Table shows a summary of QIP events based on the event that causes the ending quote promotion (QP). Sample is comprised of QIPs with a duration of less than 500 milliseconds, which results in a sample of 2.8 million. This sample is evenly divided across the two groups defined below.

	Mean	StdDev	P5	P10	Q1	Median	Q3	P90	P95
<u>QP -- Cancellation</u>									
Time Duration (Seconds)	0.06	0.10	0.001	0.002	0.003	0.007	0.049	0.205	0.318
Number Book Updates	3.18	2.48	2	2	2	2	4	6	8
Maximum Depth	2.33	2.46	1	1	1	1	3	5	7
Quantity	0.35	1.13	0	0	0	0	0	1	2
<u>QP -- Trade</u>									
Time Duration (Seconds)	0.10	0.13	0.002	0.003	0.007	0.030	0.148	0.318	0.400
Number Book Updates	3.85	3.01	2	2	2	3	5	8	10
Maximum Depth	3.23	3.60	1	1	1	2	4	7	10
Quantity	3.49	6.52	1	1	1	2	4	7	10

the traded quantity occurring within QIP events. We also show percentages to highlight how the total number of QIP events and traded quantity is distributed according to the conditioning variable. We find 11.3 million traded contracts can be linked to QIP events, which is close to 1/3 of total traded quantity. This suggests QIP events account for a decent share of total traded quantity.

The number of updates provides information related to how visible depth changes during the QIP event. QIP events with just two updates are the most common, accounting for 44% of the total, with an average maximum depth of 1.4 contracts and a time duration of 79 milliseconds. We find a positive relationship between the number of updates and the maximum observed depth, time duration, and number of switches.

The lower portion of Table 7.3 summarizes QIP information conditional on the number of switches. As noted in the distribution above, half of the QIP events have no observed switches, which the majority has just two updates. The number of updates, time duration, and maximum observed depth are all increasing in the number of switches.

We find an increased concentration of QIP events with an odd number of switches. It can be the case here that a QIP starts, depth builds as limit orders jump into the new quote, then the decrease happens when trades start to pick away at the visible depth.

More QIP events have zero switches than those with just two book updates. This can happen when depth builds directionally from the start and then is killed off by a single trade or everything is cancelled at the same time. For example, QIP starts with initial depth of 1, new order arrives and now depth is 2, trade arrives that kills the new quote resulting in a QP event.

7.3.3 Intraday Investigation:

Figure 7.3 shows how QIP events are distributed within the day conditional on the number of observed updates. We define five groups according to the number of updates: group 1 (G1) includes all QIP events with two updates; G2 contains QIP events with three to five updates; G3 contains QIP events with six to ten updates; G4 contains QIP events with 11 to 20 updates; and, G5 contains QIP events with 21 to 50 updates. The first subfigure shows counts of QIP events by group and minute between 8 a.m. and 2:30 p.m. The second subfigure shows average durations of QIP events conditional on the group and minute. To highlight durations for QIP events with updates between two and five, the last subfigure again shows average durations, but only for G1 and G2.

We find jumps in the number of QIP events starting with 8:30 a.m. and then for each 30-minute mark till 11:30 a.m. This is a very similar pattern as found in the distribution of volume within the day and indicates a positive relationship

Table 7.3: Summary Statistics of QIPs by Number of Updates & Switches

Notes: Table shows a summary of QIPs conditional on the number of observed updates. Switches are defined as a change in the direction of visible depth movement. Duration is defined as the time difference in seconds between the initial QI and the later QP. Volume is divided between that which happens at the end of the QIP and that which occurs during the QIP.

N Updates	Total QIPs	Avg Switches	Avg Max Depth	Avg Duration	Avg Trade Size	Total Qty	Total Qty End	Total Qty In				
2	1,716,780	44%	0.0	1.4	0.079	1	1,502,104	13%	1,466,604	21%	35,500	1%
3	381,589	10%	0.5	2.8	0.184	3	1,041,843	9%	835,619	12%	206,224	5%
4	392,340	10%	1.0	3.2	0.389	3	1,032,892	9%	718,899	10%	313,993	7%
5	213,368	5%	1.3	4.5	0.502	4	830,577	7%	555,972	8%	274,605	6%
6	217,846	6%	1.7	4.8	0.681	4	769,712	7%	482,210	7%	287,502	7%
7	149,333	4%	2.1	5.8	0.786	4	647,248	6%	388,910	6%	258,338	6%
8	138,145	4%	2.4	6.2	0.968	4	586,067	5%	336,965	5%	249,102	6%
9	103,261	3%	2.8	7.1	1.103	5	505,045	4%	280,846	4%	224,199	5%
10	91,038	2%	3.2	7.5	1.289	5	451,123	4%	240,264	3%	210,859	5%
11	70,922	2%	3.6	8.2	1.437	6	394,460	3%	206,439	3%	188,021	4%
12	61,850	2%	4.0	8.6	1.632	6	353,018	3%	179,162	3%	173,856	4%
13	49,751	1%	4.4	9.2	1.799	6	306,025	3%	150,542	2%	155,483	4%
14	42,759	1%	4.9	9.6	2.007	6	274,931	2%	131,591	2%	143,340	3%
15	35,428	1%	5.3	10.1	2.200	7	241,862	2%	115,018	2%	126,844	3%
16	30,759	1%	5.7	10.4	2.355	7	216,235	2%	99,053	1%	117,182	3%
17	26,123	1%	6.1	11.0	2.548	7	194,810	2%	86,246	1%	108,564	2%
18	22,590	1%	6.6	11.2	2.742	8	175,502	2%	77,696	1%	97,806	2%
19	19,179	0%	7.0	11.6	2.983	8	154,888	1%	66,007	1%	88,881	2%
20	16,769	0%	7.5	11.9	3.160	8	141,558	1%	58,558	1%	83,000	2%
gt 20	131,548	3%	250.6	16.6	9.562	19.9	1,540,847	14%	530,464	8%	1,010,383	23%
Total:	3,911,378						11,360,747		7,007,065		4,353,682	

N Switches	Total QIPs	Avg Updates	Avg Max Depth	Avg Duration	Avg Trade Size	Total Qty	Total Qty End	Total Qty In				
0	1,971,032	50%	2.2	1.6	0.110	1	2,707,776	24%	2,556,063	36%	151,713	3%
1	941,144	24%	5.1	4.3	0.509	3	2,474,452	22%	1,402,009	20%	1,072,443	25%
2	176,682	5%	6.3	6.0	0.647	6	991,112	9%	690,594	10%	300,518	7%
3	340,055	9%	9.0	6.5	1.151	4	1,382,896	12%	676,879	10%	706,017	16%
4	78,292	2%	10.2	8.2	1.249	7	553,286	5%	340,776	5%	212,510	5%
5	146,633	4%	13.1	8.2	1.897	5	798,339	7%	356,784	5%	441,555	10%
6	37,278	1%	14.1	10.0	1.940	9	324,260	3%	180,758	3%	143,502	3%
7	71,660	2%	17.2	9.5	2.706	7	480,455	4%	192,719	3%	287,736	7%
8	19,252	0%	18.1	11.3	2.698	10	197,567	2%	99,762	1%	97,805	2%
9	38,281	1%	21.3	10.8	3.527	8	302,519	3%	113,526	2%	188,993	4%
10	10,530	0%	22.2	12.6	3.503	12	126,017	1%	58,637	1%	67,380	2%
11	21,836	1%	25.4	11.7	4.332	9	200,341	2%	70,440	1%	129,901	3%
12	6,127	0%	26.2	13.4	4.248	13	79,423	1%	34,897	0%	44,526	1%
13	13,234	0%	29.3	12.6	5.167	10	135,674	1%	44,720	1%	90,954	2%
14	3,775	0%	30.2	14.4	4.959	15	55,162	0%	22,674	0%	32,488	1%
15	8,304	0%	33.2	13.3	5.886	11	95,121	1%	31,075	0%	64,046	1%
16	2,442	0%	34.1	15.1	5.876	16	39,296	0%	15,033	0%	24,263	1%
17	5,269	0%	37.2	14.2	6.675	13	66,165	1%	20,366	0%	45,799	1%
18	1,589	0%	38.3	17.3	6.511	19	29,481	0%	10,551	0%	18,930	0%
19	3,496	0%	41.0	14.7	7.282	13	46,532	0%	13,233	0%	33,299	1%
20	1,095	0%	42.0	17.1	7.143	19	20,944	0%	7,358	0%	13,586	0%
gt 20	13,372	0%	305.0	12.0	6.585	14	253,929	2%	68,211	1%	185,718	4%
Total:	3,911,378						11,360,747		7,007,065		4,353,682	

between QIP events and volume. In this market, demand becomes elevated at each half hour mark. We know some of these correspond to release of relevant news (e.g., 8:30 a.m. with the jobs report and 10:30 a.m. with the oil inventory report), while others appear to correspond with market-specific events, such as when the trading pits would open at 9am. The spike in volume just before 2:30 p.m. comes from the 2-minute daily settlement trading period starting at 2:28 p.m.

For the group with just two updates (G1), indicating the minimum required

sequence of a QI followed by a QP book update, we find these jumps to be the most significant at each half hour mark. Between each jump, we observe a steady decrease until the next large jump arrives. Across all updates groups, there is a noticeable interval of low QIP activity between 11:30 a.m. and 2 p.m., with the lowest point occurring around 1 p.m.

The bottom two subfigures of 7.3 shows average duration of QIP events conditional on the number of updates. We find QIP events prior to 8:30 a.m. have the longest observed duration. Similar to the frequency chart, durations significantly decrease on half hour increments. There exists a tendency for these durations to increase between half hour marks — most clearly displayed between 9 a.m. and 9:30 a.m. This indicates differences in how the market is operating just prior to the half hour marks during the morning hours.

Conditional Averages of Market Activity:

Figure 7.4 explores how market information corresponds to the frequency of observed QIP events. For each minute within our sample we determine the number of QIP events and the average duration for these events. For each minute, we also compute the time-weighted bid-ask spread and visible depth at quote segment levels one, two and three.

We find as the number of QIP events per minute increases we can expect shorter durations for these events. Furthermore, minutes with more QIP events tend to also have lower visible depth across the first three quote segment levels and have a higher bid-ask spread. This finding is significant since it connects QIP events to *OL*, the topic of Chapter 6. As the limit order book thins out, there

exists less pressure from both sides of the market. As a result, the bid-ask spread is allowed to widen and there exists an opportunity for QIP events.

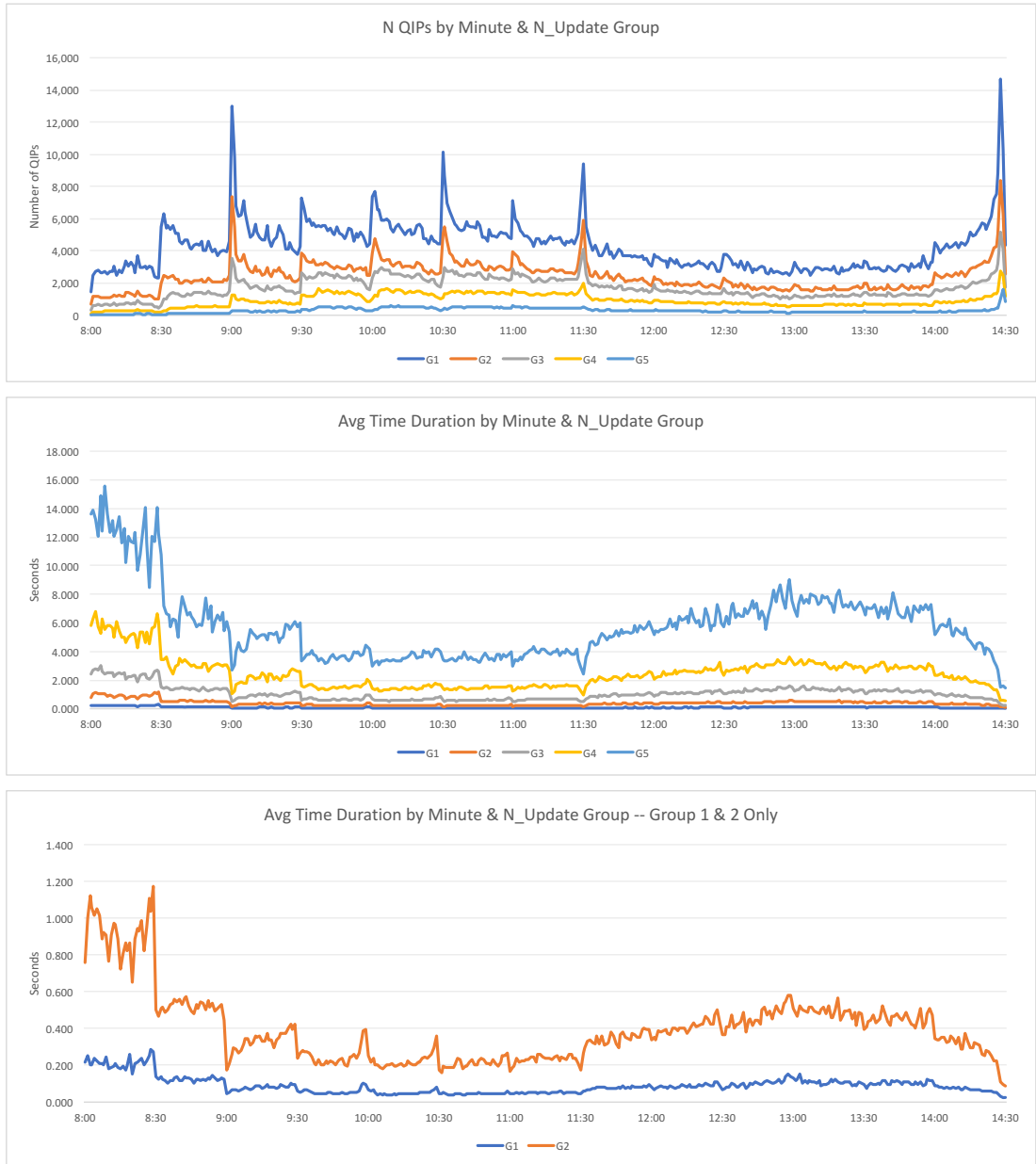


Figure 7.3: QIPs by Minute Intervals

Figure shows QIP counts (top panel) and average duration (lower panel) by one minute intervals. Group 1 (G1) corresponds to QIPs with just two updates; Group 2 (G2) corresponds to QIPs with 3 to 5 updates; Group 3 (G3) corresponds to QIPs with 6 to 10 updates; Group 4 (G4) corresponds to QIPs with 11 to 20 updates; and finally, Group 5 (G5) corresponds to QIPs with 21 to 50 updates.

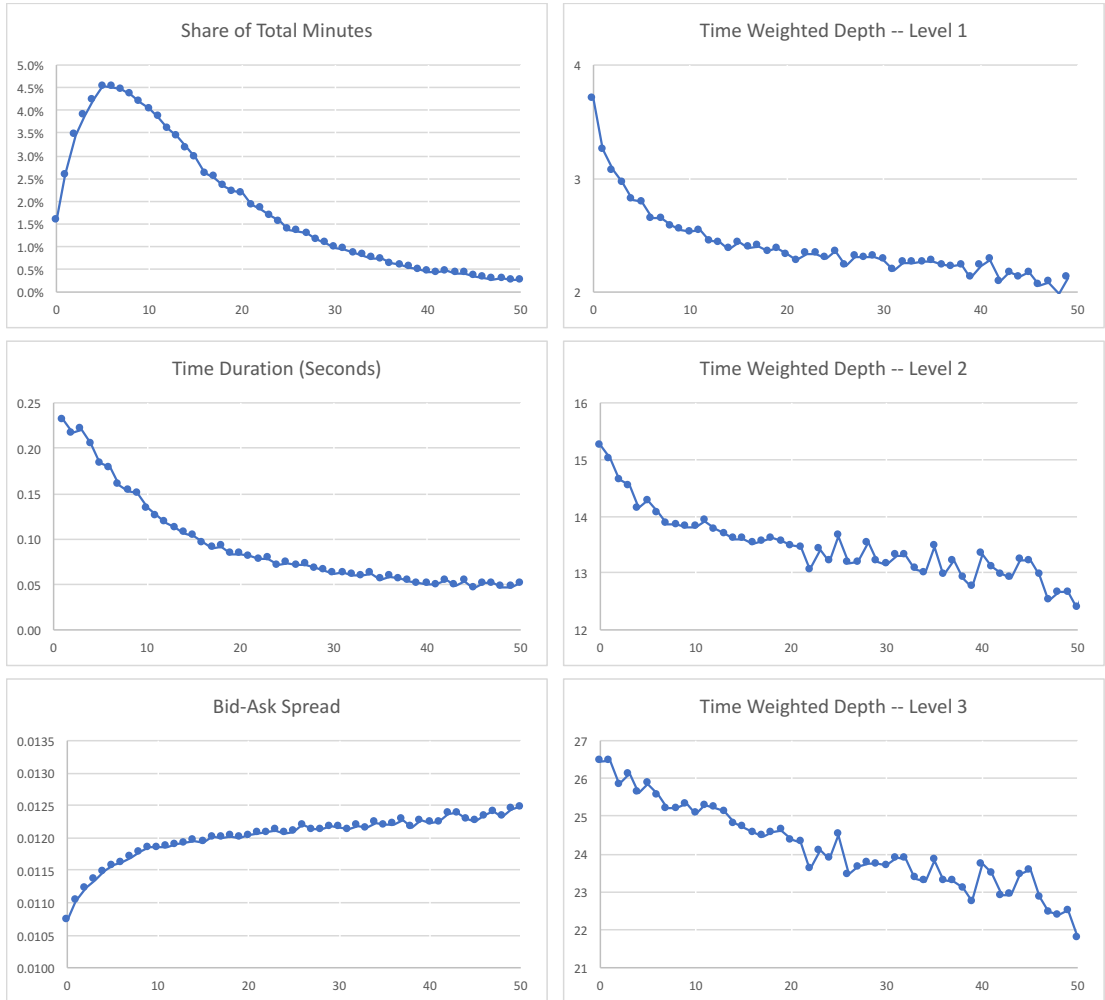


Figure 7.4: Time Weighted Information Conditional on Number of QIPs per Minute

Figure shows how book information varies conditional on the number of QIP events per minute; for each minute, we count the number of QIP events, and then find averages of market information for minutes with the same number of QIP events. The top left shows the share of total minutes for a conditional on the number of QIP events. The middle left shows the conditional average time duration of QIP events. The bottom left shows the conditional average bid-ask spread. The right column shows conditional averages of visible depth during the minute at the inside quote (upper right), the lower quote (middle right), and the below lower quote segment level (bottom right) for the bid side conditional on the number of QIPs observed per minute.

7.4 QIP Events and Liquidity Provision

This section presents an analysis of QIP events and liquidity provision. We often think of liquidity provision resulting from resting limit orders, but if a quote placed in the limit order book is immediately matched in a trade, should we think of this initial limit order as also demanding liquidity?

We focus on the division between liquidity provision and demand since this is often the main distinction between a market participant who makes the market and one who is trading to express a directional position.

For this analysis, we build a table summarizing QIP events for each minute in the sample. Along with the total number of observed QIP events per minute, we add the number of QIP events conditional on the side which initiates the event (e.g., a Bid-QIP event is a sequence of a Bid-QI followed by a Bid-QP). Furthermore, we include a separate count, by market side, of QIP events linked to liquidity provision.

We are curious if there exists a tendency for QIP events to be concentrated on a single market side. If, for a single minute, QIPs are observed to have a higher frequency from buyers (e.g., Bid-QIPs), then this might indicate that there is a directional preference. On the other hand, if QIP events are more evenly distributed, then that indicates a different situation with market participants on both sides agreeing on the price and use QIP events to initiate trades.

Figure 7.5 shows four subfigures based on 244 observations of one-minute intervals observed between 9:00 a.m. and 9:01 a.m. We focus on this single minute due to the increase in observed volume and QIP events as seen in Figure 7.3. For each minute, we include the number of QIP events, the number of QIP events by

market side, and the number conditional on a trade. The upper left figure shows daily counts of QIP events (in blue) and QIP events associated with trades (in orange), sorted according to the total number of QIP events per minute. We find a close relationship between the total number of QIP events and those linked to trades. The upper right figure, showing a scatter plot of daily QIP event count against QIP event count linked to trades, confirms this positive relationship.

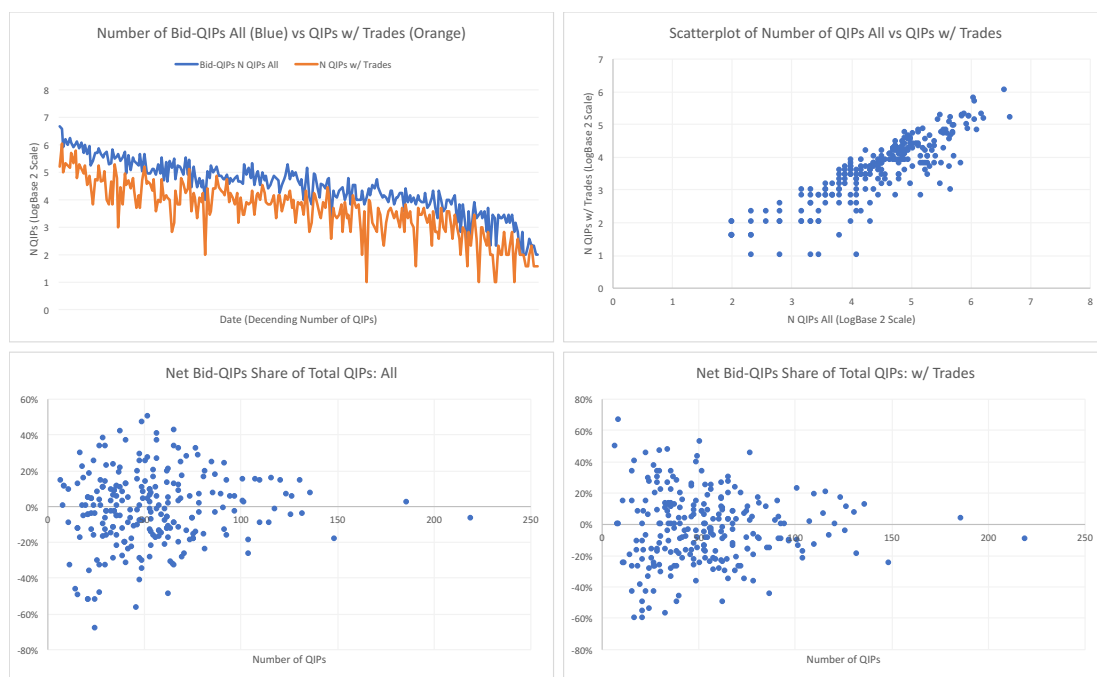


Figure 7.5: Balance between QIPs Arrival on Both Sides of the Market

Figure shows the frequency of QIPs during the 9am minute over each day of the sample. The table of daily values are sorted according to the number of observed QIPs from high to low. The first figure shows the frequency of QIPs (blue) along with those QIPs that result in trades (orange). The scatter plot showing the relationship between total QIPs to those with trades is to the right. Lower two figures show the scatter of our imbalance measure given the frequency of observed QIPs (left) and QIPs with trades (right).

Next, we look at the imbalance between market sides. QIP events can be initiated by either side of the market. Is there a difference between observing a balance in QIP events from both sides of the market compared to an imbalance? If a balance is observed, this would indicate preferences from both sides to improve the market. To explore the balance of QIP events, we take the difference between the number of Bid-QIP events and the number of Ask-QIP events and divide this

difference by the total number of QIP events within the minute. The lower two subfigures plot this measure of imbalance computed with all QIP events (lower left) and QIP events associated with trades (lower right). We find, for both measures, that this imbalance moves toward zero as the number of QIP events increase. This is evidence that the number of QIP events is more likely to come from both sides of the market during minutes where there are many QIP events.

Table 7.4 shows correlations between observed counts of QIP events for each minute of the sample. We show correlations between the number of QIP events conditional on the side of the market and the number of QIP events linked to liquidity provision. We find statistically significant positive correlations for all pairs. This evidence suggests, at the minute level, the number of QIP events are related to the demand for trades and QIP events are linked to both sides of the market; the QIP events tend to display a non-directional demand for liquidity.

Table 7.4: Correlations — QIP Counts to QIPs with Trades

Notes: Table shows correlations between counts of QIP events conditional on market side and conditional on if the QIP events can be linked to liquidity provision. Counts are found for each minute of the sample.

Pearson Correlation Coefficients, N = 94,442				
Prob > r under H0: Rho=0				
	N QIP Bid	N QIP Ask	N QIP Bid T	N QIP Ask T
N QIP Bid		0.85288 <.0001	0.9113 <.0001	0.85392 <.0001
N QIP Ask			0.85529 <.0001	0.91186 <.0001
N QIP Bid T				0.90775 <.0001
N QIP Ask T				

Figure 7.6 shows ten grouped bar charts with imbalance information conditional on the number of observed QIP events. Specifically, we group minutes into ten buckets from the lowest number of QIP events per minute (group 1) to the largest number of QIP events per minute (group 10). We find the absolute imbalance measure decreases toward zero as we increase the number of QIPs per

minute. We also find, within each decile group, the difference between averages of all QIP events and those with trades becomes less meaningful.

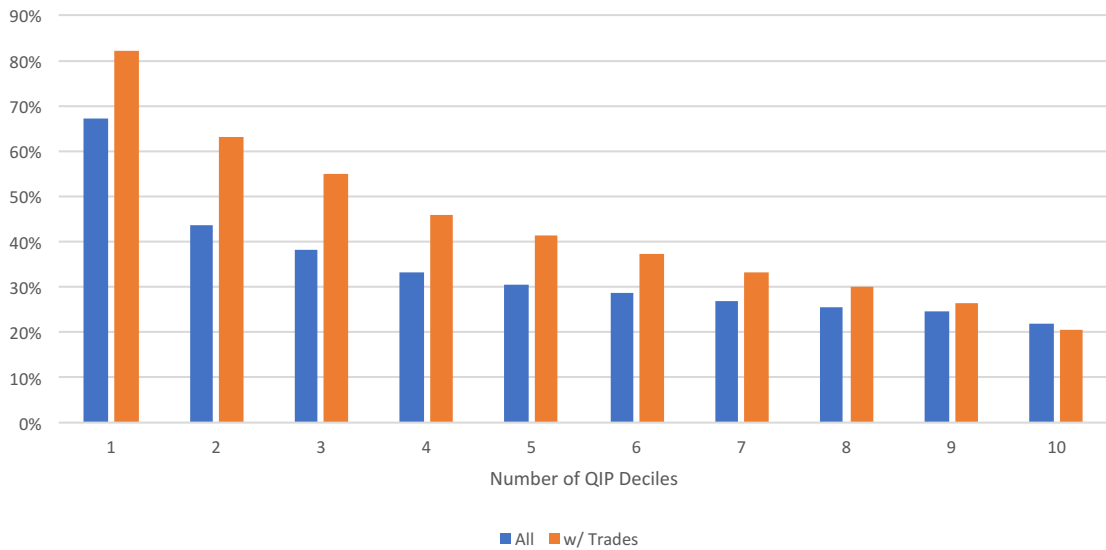


Figure 7.6: Average Absolute Imbalance by Number of QIP Deciles

Figure groups minutes based on the number of observed QIPs. We then show the average absolute imbalance for counts of all QIPs (blue) and those QIPs with trades (orange).

7.5 QIP Events, Volume & Volatility

This section explores the relationship between observed QIP events and visible depth from the limit order book. The research is focused on addressing the question: How can visible depth provide information or why QIP events occur?

By definition, QIP events can only occur when there exists an open inside quote segment level. Following the prior section, during minutes with above average observed QIP events, we can expect both sides of the market to participate in initiating these events. Therefore, traded volume during a sequence of QIPs will often be some combination of both buyer- and seller-initiated trades, where buyer-initiated trades are associated with Ask-QIP events and seller-initiated trades are associated with Bid-QIP events. There are two possibilities for the market to

exhibit a larger than expected bid-ask spread: during periods of elevated volatility, or during calm periods where there is a lack of directional pressure from either side of the market.

We build a table that collects information related to the market for one-minute intervals. For each minute observation, we add information related to volume, the share of volume linked to QIP events, the count of QIP events, offered liquidity (OL), the responsiveness of offered liquidity (ΔOL), measured volatility (following a 2SRV procedure), and a time-weighted measure of the bid-ask spread.¹

7.5.1 Conditional Expectations: ΔOL

For each minute, we incorporate information from Chapter 6 related to OL and ΔOL . At this point we add in this information to explore if there is a relationship between these and QIP events. Why might we expect a relationship? If QIP events are designed to signal the desire to trade without information, this might become more valuable during periods where offered liquidity is more responsive.

Figure 7.7 shows average market activity as a function of the responsiveness of offered liquidity (ΔOL). We find a positive relationship between the number of QIP events and ΔOL . We also find the share of per minute volume linked to QIP events increases as the market becomes more responsive. Interestingly, this also appears directly related to volume and volatility and occur during periods of higher offered liquidity. This provides evidence that QIP events are liquidity

¹To measure volatility, we use a two-scale realized volatility (2SRV) estimator using the returns from an evenly spaced measurement of the inside quote segment level. We select a distance of 100 milliseconds between measurements and then within each hour define 20 sub-grids. For each sub-grid, we compute sum of squared log returns. We then estimate volatility within each minute by averaging these returns within each minute of the sample. See Chapter 6 for details related to OL and ΔOL .

motivated and take place during minutes when there is an increased responsiveness of offered liquidity.

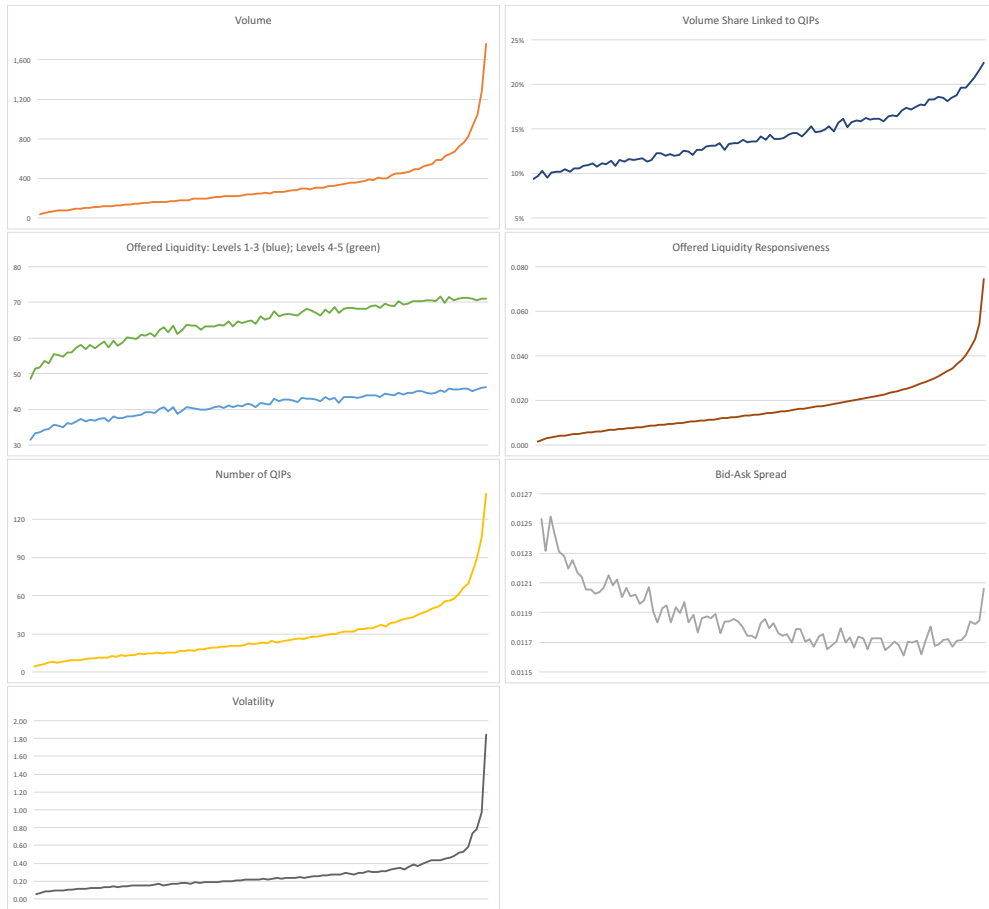


Figure 7.7: Analysis of One Minute Intervals — Ranked by ΔOL

Figure shows market activity (volume; bid-ask spread; volatility), offered liquidity and measured responsiveness, and QIP count and volume share. One minute intervals are sorted according to measured responsiveness of offered liquidity (levels one to three) into 100 buckets, then averages computed for each bucket.

7.5.2 Conditional Expectations: Bid-Ask Spread

Figure 7.8 shows average market activity as a function of the bid-ask spread. We define 100 groups of minutes sorted from low BAS to high BAS and then compute averages of minutes within each group. We expect more QIP events as the bid-ask spread increases. This, however, shows there exists a point when the average number of QIPs actually decreases for the largest bid-ask spread values. This

appears related to minutes where the volatility is the greatest and the offered liquidity is the lowest. We expect these minutes to be related to major news events and therefore, it makes sense to see a reduction in the number of QIP events. This decrease further supports the relationship with the responsiveness of offered liquidity as we can see there is also a clear decrease in the responsiveness for minutes with the largest bid-ask spread.

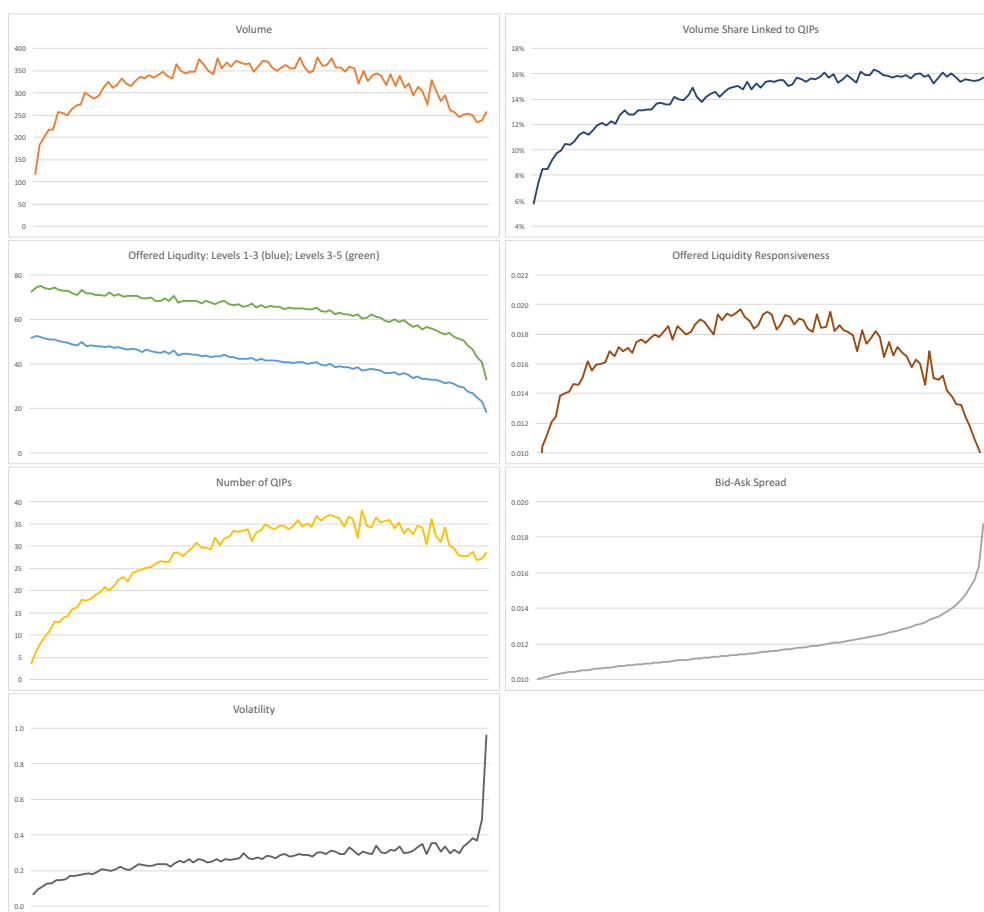


Figure 7.8: Analysis of One Minute Intervals — Ranked by Bid-Ask Spread

Figure shows market activity (volume; bid-ask spread; volatility), offered liquidity and measured responsiveness, and QIP count and volume share. One minute intervals are sorted according to bid-ask spread into 100 buckets, then averages computed for each bucket.

7.5.3 Information Conditional on Number of QIP Events

Figure 7.9 shows how market characteristics change with the number of observed QIP events. For this analysis, we sort minute observations during the sample period into ten equally sized groups with group 1 containing the minutes with the lowest number of QIP events and group 10 containing the minutes with the largest number of QIP events. For each of the ten groups, we show distribution information (25th, 50th, 75th percentile) related to the bid-ask spread, the signed volume imbalance, OL , ΔOL , volume, and volatility.

This analysis extends the work above focusing on the one minute interval starting at 9 a.m., which focused on exploring the relationship between QIP events linked to trades and the imbalance between buy and sell QIP events. This figure shows that minutes with the largest number of QIP events are associated with minutes that have the largest range of signed volume imbalance. We find a significant increase in volume, volatility, the bid-ask spread, and ΔOL in the group of minutes with the largest number of QIP events.

7.5.4 Regression Analysis of QIP Volume Share

Here we test if the volume associated with QIP events is a constant share of total volume. We define p_t as the share of volume, over a one minute interval, which corresponds to QIP events. If QIP events are simply a function of volume, then we would expect other observable characteristics of the market, such as the bid-ask spread or volatility, to not influence p_t holding constant volume. Along with standard measures of the market, we also include ΔOL and OL as measures during the minute interval.

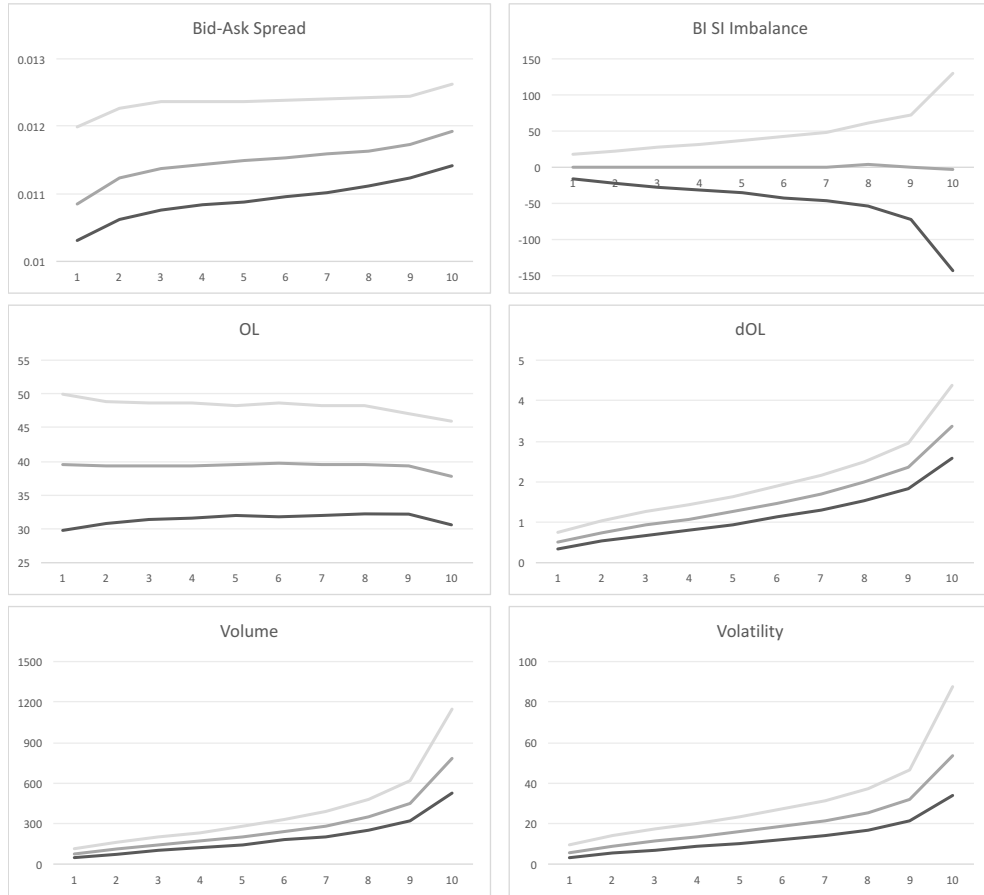


Figure 7.9: Distributions of Market Activity Conditional on Observed QIP Events

Figure shows distribution information (Q2 (dark grey), median (grey), Q3 (light grey)) for six types of market activity for one-minute intervals conditional on the observed number of QIP events. We sort minutes into ten equally sized groups according to the number of QIP events. The horizontal axis sorts these groups from minutes with the lowest number of QIP events per minute ($X = 1$) to the highest number of QIP events per minute ($X = 10$). For each group, we find distributions for the bid-ask spread, signed volume imbalance (buyer-initiated quantity less seller-initiated quantity), OL , ΔOL , volume, and volatility.

We define two models with p_t as the dependent variable. While this variable is determined within the minute interval, we argue that volatility can potentially influence the number of QIP events, but we do not expect the QIP events to influence measured volatility. Similarly, since flash events are often for a very small quantity, we do not expect the occurrence of flash events to influence ΔOL or OL .

The first model is defined according to the following equation:

$$p_t = \beta_0 + \beta_1 * p_{t-1} + \beta_2 * BAS_t + \beta_3 * VOL_t + \beta_4 * \Delta OL_t + \beta_5 * OL_t + \beta_6 * OL_{(4,5)t} \quad (7.1)$$

The specified model shows the current minute volume share linked to QIP events (p_t) to be a function of the prior minute volume share linked to QIP events (p_{t-1}), the current minute time-weighted bid-ask spread (BAS_t), the current minute volatility (VOL_t), the current minute average responsiveness of offered liquidity (ΔOL_t), the current minute time-weighted offered liquidity across the first three quote segment levels (OL_t), and the current minute time-weighted offered liquidity across the fourth and fifth quote segment levels ($OL_{(4,5)t}$).

According to the regression results, found in Table 7.5, we find evidence that the p_t is more than just a function of volume due to the statistically significant estimated parameter values. The results indicate there is a positive autocorrelation across minutes. The positive coefficient on the bid-ask spread follows from the need for a larger bid-ask spread to allow for QIP events. We find the share of volume linked to QIP events decreases as volatility increases once we control for all other variables.

We find the share of QIP events increases as ΔOL increases. This provides support for the claim that QIP events become optimal during periods when the market is responding to changes in the book. For example, during minutes with an elevated ΔOL , it becomes easy to start a bidding war by placing limit orders in the book and allowing them to rest at quote levels close to the market. Algorithmic traders who are able to monitor the books willingness to react, and who do not wish to contribute to large movement in the book, can use QIP events to flash an

intent and engage in trades without letting the intent contribute to ΔOL .

The opposite signs linked to the two OL variables also are interesting since they help explain why we observe QIP events and how they contribute to the total volume during one-minute intervals. The negative sign on OL supports the claim that more QIP events occur when there is less pressure in the limit order book. As OL increases, we expect more directional movement, and therefore, less opportunities for QIP events to be an option. The positive sign related to OL at the fourth and fifth quote segment levels indicates support for the current market level at deeper levels of the book.

Table 7.5: Regression 1 — QIP Volume Shares Per Minute

Notes: Table shows regression results with the share of per-minute volume linked to QIP events as our dependent variable. Our set of independent variables includes the prior share of per-minute volume linked to QIP events, the bid-ask spread, volatility, ΔOL , and two measures of OL based on visible depth at quote segment levels 1, 2, and 3 and the second based on levels 4 and 5.

Variable	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	0.0093	0.0031	2.98	0.0028
Lag QIP VM %	0.1215	0.0032	37.59	<.0001
Bid-Ask Spread	0.0083	0.0002	39.21	<.0001
Volatility	-0.0022	0.0004	-5.23	<.0001
dOL	0.0252	0.0002	102.09	<.0001
OL	-0.0074	0.0004	-18.91	<.0001
OL Levels 4, 5	0.0022	0.0003	8.61	<.0001
Observations	85,037			
F Value	3,159			
Pr > F	<.0001			
R-Square	0.1823			

The second model is defined according to the following equation:

$$p_t = \beta_0 + \beta_1 * p_{t-1} + \beta_2 * VOL_t/VM_t + \beta_3 * \Delta OL_t/VM_t + \beta_4 * OL_t/VM_t \quad (7.2)$$

The second model now shows the current minute volume share linked to QIP events (p_t) to be a function of the prior minute volume share linked to QIP

events (p_{t-1}), the current minute volatility per unit of volume (VOL_t/VM_t), the current minute average responsiveness of offered liquidity per unit of volume ($\Delta OL_t/VM_t$), and the current minute time-weighted offered liquidity across the first three quote segment levels per unit of volume (OL_t/VM_t).

Table 7.6 shows the results from the regression and further supports the findings from the first model. These results again show a negative relationship between volatility and the share of volume linked to QIP events once we control for the other variables. As above, we find a positive relationship between the share of QIP events and measured responsiveness and a negative relationship related to offered liquidity.

Table 7.6: Regression 2 — QIP Volume Shares Per Minute

Notes: Table shows regression results with the share of per-minute volume linked to QIP events as our dependent variable. Our set of independent variables includes the prior share of per-minute volume linked to QIP events, volatility per unit of volume, ΔOL per unit of volume, and OL based on visible depth at quote segment levels 1, 2, and 3 per unit of volume.

Variable	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	0.1386	0.0008	175.47	<.0001
Lag QIP VM %	0.1770	0.0033	53.28	<.0001
Volatility / VM	-0.9895	0.3038	-3.26	0.0011
dOL / VM	0.2338	0.0166	14.09	<.0001
OL / VM	-0.0831	0.0013	-66.01	<.0001
Observations	85,037			
F Value	2,395			
Pr > F	<.0001			
R-Square	0.1013			

7.5.5 Minutes Before and After Unexpected Market Events

Above we have shown how we can expect a larger per-minute imbalance between buyer- and seller-initiated volume is linked to a larger number of QIP events. The goal of this section is to explore how QIP events relate before, during, and after minutes associated with unusually large directional trading.

To accomplish this, we first exclude minutes surrounding expected news events. We exclude ten minutes before and after 8:30 a.m. and 10:30 a.m. For the remaining minutes, we compute a measure of directional trading to identify when the volume is made up of a majority of buyer-initiated volume (a positive imbalance), or sell-initiated volume (a negative imbalance). We then define 20 equally sized groups of minutes and select those with the largest positive and negative trade imbalances.

For each minute, corresponding to a large imbalance in traded volume, we identify ten minutes before and after. Furthermore, we keep only non-overlapping intervals, which ensures we do not have two minutes with a large imbalance within ten minutes from each other.

Figure 7.10 shows median (orange) along with Q2 and Q3 from distributions of market activity conditional on proximity to an identified unexpected event. We show one-minute summaries of the bid-ask spread, number of QIP events, OL , ΔOL , volume, and volatility. In the figures, the information corresponding to 0 indicates the minute with large directional trading. To allow for comparison to expected levels, we include the median value across all minutes in the sample, displayed as the horizontal line in each figure.

Regarding the minutes prior to the event, we find an increase in volume, ΔOL , and number of QIPs relative to the expected median value. This evidence suggests the market is becoming more responsive and allowing for a larger number of QIPs per minute than we otherwise would expect. There is no observable difference in the bid-ask spread, which is surprising given the increase in the number of QIPs.

As expected, the minute containing the unusually large imbalance in traded

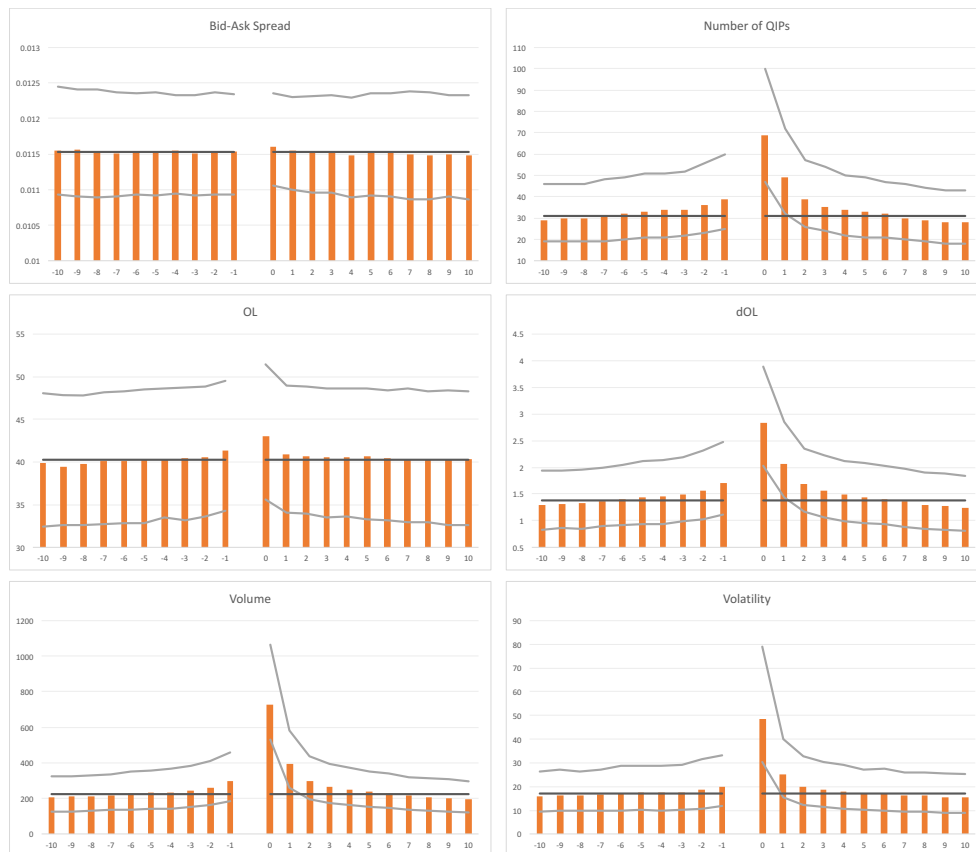


Figure 7.10: Activity Conditional on Proximity to Unexpected Event

Figure shows median (orange) along with Q2 and Q3 from distributions of market activity conditional on proximity to an identified unexpected event. We show one-minute summaries of the bid-ask spread, number of QIP events, OL , ΔOL , volume, and volatility. We define unexpected events as minutes with unusually large directional trading; then for each minute, we identify the ten minutes before and after. In the figures, the information corresponding to 0 indicates the minute with large directional trading. To allow for comparison to expected levels, we include the median value across all minutes in the sample, displayed as the horizontal line in each figure.

quantity shows a jump in volume, volatility, number of QIPs and ΔOL . The evidence also shows an increase in OL during this minute, which is likely related to large values of OL that are uncovered during the period of elevated volatility.

7.6 QIP Events & Expected News

We have shown evidence that there exists a relationship between QIP events during periods with elevated volume and volatility. We also have shown that QIP events are related to the desire to trade. But what about during periods of extreme

market stress commonly associated with the expected release of news?

7.6.1 Sample of Events

Here we focus on a subset of 62 expected news events. This sample of events is a mix of U.S. jobs reports released at 8:30 a.m. and oil inventory reports released at 10:30 a.m. We focus on one-minute intervals and select ten minutes prior and ten minutes immediately following each event. To compare news with non-news minutes, we select the same minutes during dates without expected events. We compare minutes related to news dates with minutes related to non-news dates with the same distance to the event.

7.6.2 Analysis

Figure 7.11 shows a comparison between minutes corresponding to either news events or non-news events. For each minute, we show the median value and minutes are arranged such that the minute immediately following the news release corresponds to group 1.

Regarding activity prior to expected release, we find bid-ask spread becomes elevated during the two minutes prior to the release. The widening of the spread reflects the uncertainty in the market just prior to an expected news event. We also observe a significant drop in OL , which starts six minutes prior to the expected event. The observed increase in the bid-ask spread along with the decrease in OL indicate a market preparing for a potentially significant flow of information. Relative to non-news days, we would expect a widening of the spread to be corrected with the arrival of new limit orders to build depth and increase OL .

During the few minutes prior to the expected release, we also find a decrease in ΔOL , which likely is related to the lower observed OL and elevated bid-ask spread. Prior to news, there is less interest in providing liquidity to the limit order book since the market expects a jump soon. Less interest in providing resting liquidity also translates into less observed response over short intervals of time. We find very little difference between news and non-news days, prior to the event, in volume, volatility, or the number of QIP events.

During the minute immediately following the news release, there is a significant jump in volume, volatility, the number of QIPs, and ΔOL . Our responsiveness measure here combines the willingness to add and remove liquidity from both sides of the market. While there is a decrease over time, all these per-minute measures remain elevated during the full ten minute window relative to the expected values during non-news days.

Figure 7.11 highlights how the market fundamentally changes following an expected news event and highlights how this change differs across our six measures of market activity. The bid-ask spread and OL are different from the rest since they return to levels expected during non-news days within five minutes following the release. We find the bid-ask spread is the fastest to return back to expected levels.

Figure 7.12 extends the number of minutes following an expected news event from 10 to 20. Comparing median values within the same minutes location to the event, we find higher observed number of QIPs, ΔOL , volume, and volatility extending 20 minutes after the news event.

This analysis clearly shows how the bid-ask spread and OL are unreliable

measures of market liquidity. If we only focused on these two measures of market quality, then we would assume the impact of the news event only extends for a few minutes following the release. The bid-ask spread recovers to a level very close to the minimum tick size and OL also returns to normal levels indicating a willingness of market participants to add limit orders which build depth on both sides of the market. Based on these two measures, the market appears to have fully recovered.

The appearance of a fully recovered market falls apart once we consider ΔOL and the number of observed QIP events. The elevated nature of ΔOL indicates whatever liquidity is resting in the book can quickly disappear, resulting in directional movement in quote segments as reflected by the elevated volatility.

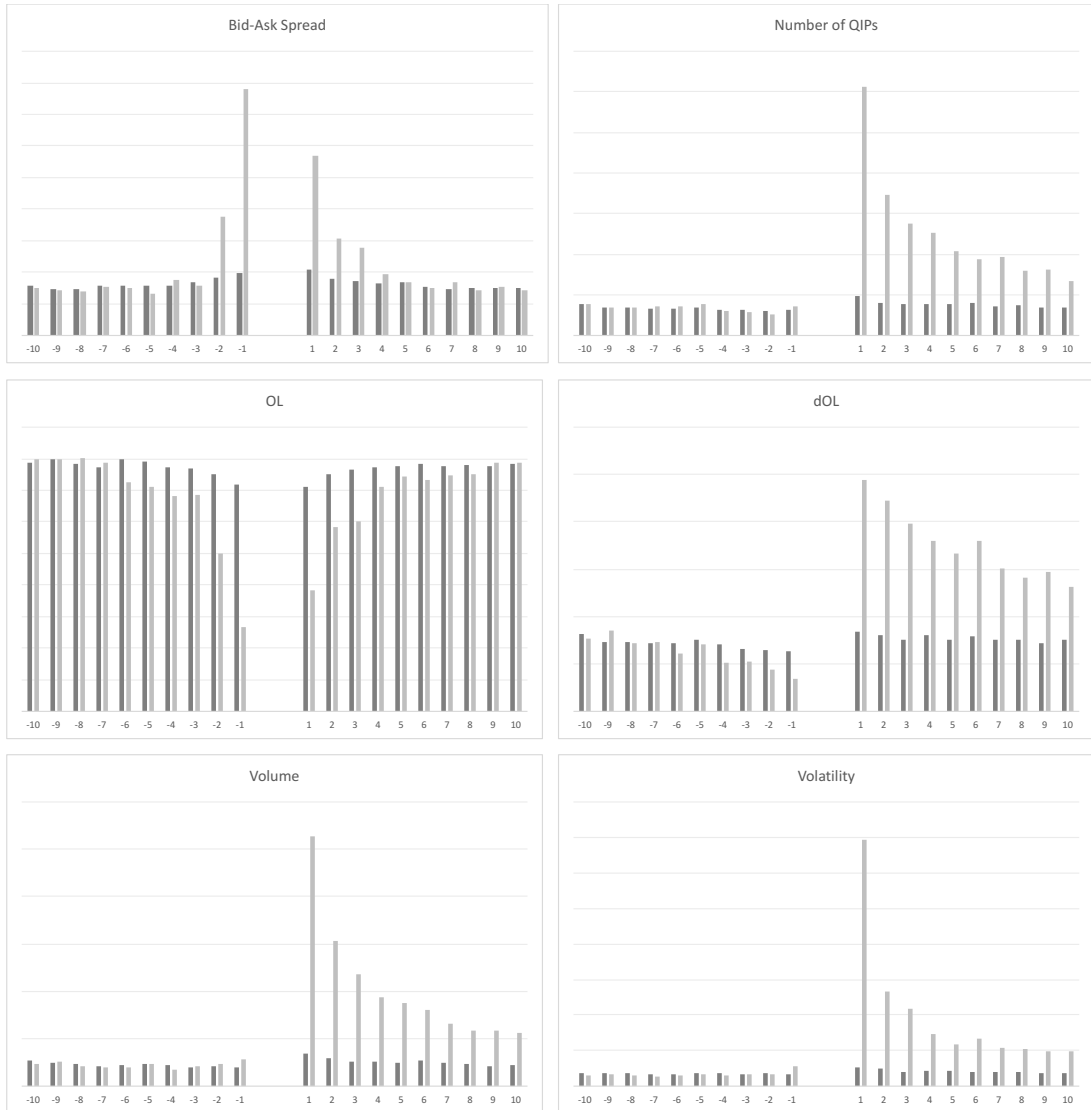


Figure 7.11: Analysis Before & After Expected News — 10 Minutes After

Figure shows median values of observed market activity for both non-news days (dark grey) and news days (light grey) conditional on the proximity to expected news events. There are 62 days with expected news related to either U.S. Jobs or crude oil inventory reports, occurring at 8:30am and 10:30am, respectively. Non-news information controls for the time of day differences. The figure shows median values for time-weighted bid-ask spread, number of QIP events per minute, OL , ΔOL , volume and volatility.

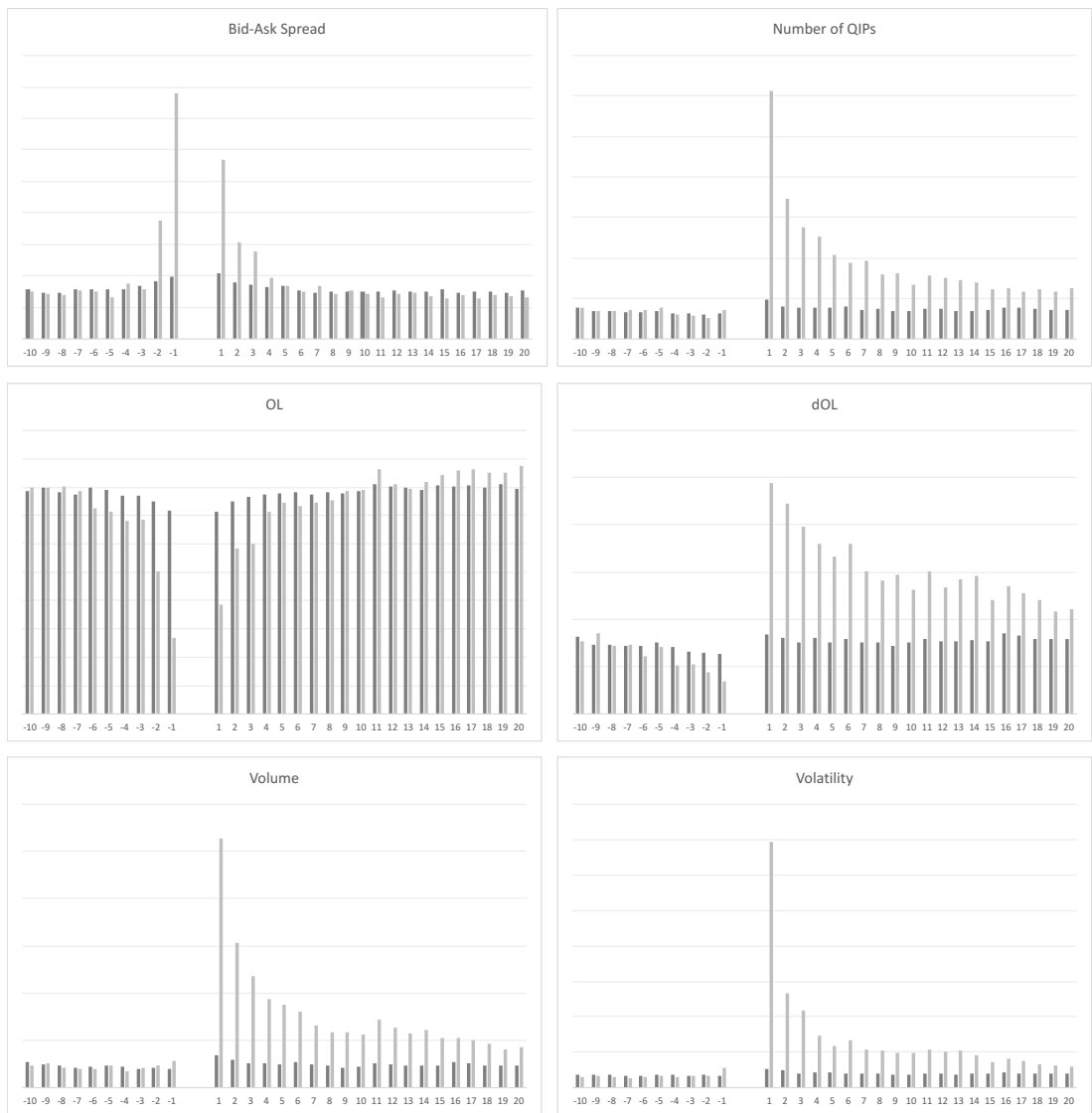


Figure 7.12: Analysis Before & After Expected News — 20 Minutes After

Figure shows median values of observed market activity for both non-news days (dark grey) and news days (light grey) conditional on the proximity to expected news events. There are 62 days with expected news related to either U.S. Jobs or crude oil inventory reports, occurring at 8:30am and 10:30am, respectively. Non-news information controls for the time of day differences. The figure shows median values for time-weighted bid-ask spread, number of QIP events per minute, OL , ΔOL , volume and volatility.

7.7 Conclusions

This chapter has documented evidence regarding flash events — referred to as QIP events due to the sequence of a QI followed by a QP event — with very short durations in the limit order book. While order entry and exit can occur at any level of the limit order book, we focus on the subset of events that produce a change in the best quotes and therefore represent an improvement in the best quotes.

Evidence suggests flash events are related to an intent to trade; while we do observe many flash events that do not result in trades, these appear to exist due to liquidity demand that gets satisfied from flash events which do result in trades. This evidence suggests that flash quotes have become a new way to trade and participate in the limit order book.

Evidence suggests a close relationship between the occurrence of QIP events, volume, volatility, and ΔOL . We have showed how the share of volume linked to QIP events is not a constant share of total volume. The positive relationship between QIP events and ΔOL suggest flash quotes exist as a way for participants to hide their intention to trade. This evidence suggests these two observations of market activity are closely related and likely result from market participants with a technological capability to monitor how the limit order book is reacting to order flow.

Our study of QIP events surrounding major expected news shows how QIP events remain elevated for 20 minutes prior to an expected news release. This analysis raises a number of questions related to the bid-ask spread and OL observed after a news event while ΔOL and QIP events remain elevated.

Future research will explore how these results compare and contrast across different markets. We are particularly interested in investigating how the occurrence of flash quotes has changed over time. With access to regulatory data, we would be able to explore who is doing these flash quotes: Is it a human or non-human automated trading system?

Chapter 8

Conclusion

This research has focused on analyzing high frequency updates to the electronic limit order book. This is a timely research topic given all the unanswered questions about how modern electronic markets function, especially during periods of elevated volatility. By utilizing a sample of book updates for the crude oil futures market, we have investigated how there exist a number of expected dynamic relationships on both sides of the market.

This research contributes to our understanding of these markets by analyzing the most detailed source of information that currently exists. The market depth data, purchased from the CME, is the source of information that is used by trading firms to study market dynamics and is the source of information used by algorithms to spot trading opportunities.

As this information becomes more readily available, we expect a great deal of research to make use of this information. Many questions related to market liquidity, price discovery, and transparency can benefit from the analysis of such information.

This dissertation has provided a detailed look at this new source of high frequency information. There are events that commonly occur, including the widening of the bid-ask spread; we have demonstrated how this information can be used to develop predictions for how the market will return to a one-tick bid-ask spread. We have also documented patterns related to book updates and trades that allowed us to define quote segments. We show how this allows us to keep better track of how the market evolves over time.

The visualizations presented in this dissertation allow for a much richer view of the market and how visible depth evolves over time. The use of such a visualization captures many of the important components of the market including the relationship between the best quote movement and the visible depth. The visual shows clear patterns that emerge from both sides of the market moving and taking turns moving toward and then away from the market.

Using the market depth data, we have documented a method for measuring the responsiveness of the offered liquidity on both sides of the market. We find this responsiveness changes during the day and displays differences based on the occurrence of market events. This responsiveness of offered liquidity is related to the occurrence of fleeting liquidity, which has become a concern and commonly carries a negative connotation.

We find that responsiveness changes as a response to the arrival of information. This information can take the form of trades or related visible depth. If the buy side is aggressively building depth, the sell side should be allowed to respond by cancelling quotes — a response that is directly related to the activity on the buy side prior to the arrival of trades.

Another aspect of modern electronic markets is the tendency for orders to flash intent. We identify and investigate why such events occur in our sample. We find these events are the outcome of market participants desiring to trade during periods when the market has a tendency to respond to new information.

Future work will extend this analysis across multiple products whose market activity varies. For example, we might expect differences in response patterns or QIPs to exist between a very active futures market and a less active futures market. Similarly, we might expect different response patterns to exist across different futures markets that vary by size and activity.

As the data is collected over a longer time range, there exist many opportunities for future work focused on how the market changes over time. We expect differences to arise as trading firms adjust algorithms and continue to improve their execution methods.

This research has focused on the important role visible market depth can play in helping to make sense of market activity. A large part of this comes from the ability of some market participants to learn from this information in close to real time. Recently, the CME has made changes to the way it provides this information that gives more detail. In particular, the new data allows for tracking a single order and its placement in line. This should have significant implications for how traders use this information and how we observe order submissions. Future work will study how the market has changed given this expanded information.

Bibliography

Anat R Admanti and Paul Pflleiderer. Sunshine trading and financial market equilibrium. *The Review of Financial Studies*, 4(3):443–481, 1991.

Hee-Joon Ahn, Kee-Hong Bae, and Kalok Chan. Limit orders, depth, and volatility: Evidence from the stock exchange of hong kong. *The Journal of Finance*, 56(2):767–788, 2001.

Yacine Ait-Sahalia and Mehmet Saglam. High frequency market making: Optimal quoting. Technical report, Working Paper, <http://ssrn.com/abstract=2331613>, 2016.

Yacine Ait-Sahalia and Mehmet Sađlam. High frequency market making: Implications for liquidity. 2017.

Yang An and Ngai Hang Chan. Short-term stock price prediction based on limit order book dynamics. *Journal of Forecasting*, 36(5):541–556, 2017.

Shmuel Baruch and Lawrence R Glosten. Fleeting orders. 2013.

Bruno Biais, Pierre Hillion, and Chester Spatt. An empirical analysis of the limit order book and the order flow in the paris bourse. *the Journal of Finance*, 50(5):1655–1689, 1995.

- Bruno Biais, Christophe Bisiere, and Chester Spatt. Imperfect competition in financial markets: An empirical study of island and nasdaq. *Management Science*, 56(12):2237–2250, 2010.
- Robert Bloomfield and Maureen O’Hara. Market transparency: Who wins and who loses? *The Review of Financial Studies*, 12(1):5–35, 1999.
- Ekkehart Boehmer, Gideon Saar, and Lei Yu. Lifting the veil: An analysis of pre-trade transparency at the nyse. *The Journal of Finance*, 60(2):783–815, 2005.
- Jonathan Brogaard, Terrence Hendershott, and Ryan Riordan. High-frequency trading and price discovery. *Review of Financial Studies*, 27(8):2267–2306, 2014.
- Allen Carrion. Very fast money: High-frequency trading on the nasdaq. *Journal of Financial Markets*, 16(4):680–711, 2013.
- Bidisha Chakrabarty, Konstantin Tyurin, Zhaohui Han, and Xiaoyong Zheng. A competing risk analysis of executions and cancellations in a limit order market. 2006.
- Sugato Chakravarty and Craig W Holden. An integrated model of market and limit orders. *Journal of Financial Intermediation*, 4(3):213–241, 1995.
- Jin-Wan Cho and Edward Nelling. The probability of limit-order execution. *Financial Analysts Journal*, 56(5):28–33, 2000.
- Rama Cont and Adrien De Larrard. Price dynamics in a markovian limit order market. *SIAM Journal on Financial Mathematics*, 4(1):1–25, 2013.

- Rama Cont, Sasha Stoikov, and Rishi Talreja. A stochastic model for order book dynamics. *Operations research*, 58(3):549–563, 2010.
- Jérôme Dugast and Thierry Foucault. Data abundance and asset price informativeness. 2016.
- David Easley and Maureen O’hara. Time and the process of security price adjustment. *The Journal of finance*, 47(2):577–605, 1992.
- Andrew Ellul, Craig W Holden, Pankaj Jain, and Robert Jennings. Order dynamics: Recent evidence from the nyse. *Journal of Empirical Finance*, 14(5):636–661, 2007.
- Mark D Flood, Ronald Huisman, Kees G Koedijk, and Ronald J Mahieu. Quote disclosure and price discovery in multiple-dealer financial markets. *The Review of Financial Studies*, 12(1):37–59, 1999.
- Thierry Foucault. Order flow composition and trading costs in a dynamic limit order market. *Journal of Financial markets*, 2(2):99–134, 1999.
- Thierry Foucault, Ohad Kadan, and Eugene Kandel. Limit order book as a market for liquidity. *Review of Financial Studies*, 18(4):1171–1217, 2005.
- Thierry Foucault, Sophie Moinas, and Erik Theissen. Does anonymity matter in electronic limit order markets? *Review of financial studies*, 20(5):1707–1747, 2007.
- Ronald L Goettler, Christine A Parlour, and Uday Rajan. Equilibrium in a dynamic limit order market. *The Journal of Finance*, 60(5):2149–2192, 2005.

- Martin D Gould, Mason A Porter, Stacy Williams, Mark McDonald, Daniel J Fenn, and Sam D Howison. Limit order books. *Quantitative Finance*, 13(11): 1709–1742, 2013.
- Mark D Griffiths, Brian F Smith, D Alasdair S Turnbull, and Robert W White. The costs and determinants of order aggressiveness. *Journal of Financial Economics*, 56(1):65–88, 2000.
- Yasushi Hamao and Joel Hasbrouck. Securities trading in the absence of dealers: Trades and quotes on the tokyo stock exchange. *The Review of Financial Studies*, 8(3):849–878, 1995.
- Puneet Handa and Robert A Schwartz. Limit order trading. *The Journal of Finance*, 51(5):1835–1861, 1996.
- Oliver Hansch. Island tides: exploring ecn liquidity. 2003.
- Lawrence Harris. *Does a large minimum price variation encourage order exposure?*, volume 96. New York Stock Exchange, 1996.
- Lawrence Harris. Order exposure and parasitic traders. *University of Southern California working paper*, 1997.
- Lawrence Harris and Ethan Namvar. The economics of flash orders and trading. 2011.
- Joel Hasbrouck and Gideon Saar. Limit orders and volatility in a hybrid market: The island ecn. 2001.
- Joel Hasbrouck and Gideon Saar. Technology and liquidity provision: The blurring of traditional definitions. *Journal of financial Markets*, 12(2):143–172, 2009.

- Nikolaus Hautsch and Ruihong Huang. The market impact of a limit order. *Journal of Economic Dynamics and Control*, 36(4):501–522, 2012.
- Terrence Hendershott and Charles M Jones. Island goes dark: Transparency, fragmentation, and regulation. *The Review of Financial Studies*, 18(3):743–793, 2005.
- Nicholas Hirschey. Do high-frequency traders anticipate buying and selling pressure? *Available at SSRN 2238516*, 2013.
- Charles M Jones. What do we know about high-frequency trading? 2013.
- Charles M Jones, Gautam Kaul, and Marc L Lipson. Transactions, volume, and volatility. *Review of Financial Studies*, 7(4):631–651, 1994.
- Ron Kaniel and Hong Liu. So what orders do informed traders use? *The Journal of Business*, 79(4):1867–1913, 2006.
- Albert S Kyle. Continuous auctions and insider trading. *Econometrica: Journal of the Econometric Society*, pages 1315–1335, 1985.
- Albert S Kyle, Anna A Obizhaeva, and Yajun Wang. Smooth trading with overconfidence and market power. *Robert H. Smith School Research Paper No. RHS*, 2423207, 2014.
- Andrew W Lo, A Craig MacKinlay, and June Zhang. Econometric models of limit-order executions. *Journal of Financial Economics*, 65(1):31–71, 2002.
- Ananth Madhavan, David Porter, and Daniel Weaver. Should securities markets be transparent? *Journal of Financial Markets*, 8(3):265–287, 2005.

- Anna A Obizhaeva and Jiang Wang. Optimal trading strategy and supply/demand dynamics. *Journal of Financial Markets*, 16(1):1–32, 2013.
- Christine A Parlour. Price dynamics in limit order markets. *Review of Financial Studies*, 11(4):789–816, 1998.
- Christine A Parlour and Duane J Seppi. Limit order markets: A survey. *Handbook of financial intermediation and banking*, 5:63–95, 2008.
- Angelo Ranaldo. Order aggressiveness in limit order book markets. *Journal of Financial Markets*, 7(1):53–74, 2004.
- Ioanid Roşu. A dynamic model of the limit order book. *Review of Financial Studies*, page hhp011, 2009.
- Ioanid Rosu. Liquidity and information in order driven markets. 2016.
- Johannes A Skjeltorp, Elvira Sojli, and Wing Wah Tham. Flashes of trading intent at nasdaq. *Journal of Financial and Quantitative Analysis*, 51(1):165–196, 2016.
- Jeffrey Smith. Market vs. limit order submission behavior at a nasdaq market maker. 2000.
- Brian M Weller. Efficient prices at any cost: does algorithmic trading deter information acquisition? 2016.
- Liyan Yang and Haoxiang Zhu. Back-running: Seeking and hiding fundamental information in order flows. 2016.