EMPIRICAL EVIDENCE OF KNIGHTIAN UNCERTAINTY IN EQUITY MARKETS

A Dissertation
submitted to the Faculty of the
Graduate School of Arts and Sciences
of Georgetown University
in partial fulfillment of the requirements for the
degree of
Doctor of Philosophy
in Physics

By

Julie Li Schoening, M.S.

Washington, DC
December 7, 2015
EMPIRICAL EVIDENCE OF KNIGHTIAN UNCERTAINTY IN EQUITY MARKETS

Julie Li Schoening, M.S.

Dissertation Advisors: Jeffrey S. Urbach, Ph.D. and Frank M. Hatheway, Ph.D.

ABSTRACT

Market microstructure theory builds on the assumption that the arrival of new information induces trading, because it alters market participants’ expectations of the true price. The bid-ask spread, which represents microstructure frictions, should straddle the expected true price (Madhavan, 1992). In empirical studies, the bid-ask midpoint is a generally accepted and frequently used proxy for the expected true price of the asset (Hasbrouck, 2007; Huang & Stoll, 1997). In this paradigm, price discovery occurs through trading, with the expected true price unvarying between two adjacent trades.

I consider, on the contrary, that it is not only possible but also common for the expected true price to be outside the bid-ask interval in a modern equity market. I provide empirical evidence of quote movements within time periods of no trade using tick-by-tick equity market data. To explain such previously undocumented evidence, I suggest a Knightian uncertainty model from Easley and O’Hara (2010a), which models liquidity freeze in the fixed-income market. I form hypotheses to test potential causes of the instances where quotation prices clearly deviate from the expected true price of the security. I build a Monte Carlo simulation of the limit order book with order entries, cancellations, and executions following separate Poisson processes, and use the simulation results as a benchmark to test the significance of the empirically observed deviations: they are statistically significant in both frequency and magnitude. Since Knightian uncertainty cannot be quantified by definition, I take a process-of-elimination approach to test competing hypotheses on valuation risk, microstructure risk, and
time of the day effects. I find the empirical evidence of Knightian uncertainty consistent with theoretical prediction, although market risk factors also contribute significantly.

These findings challenge the classical assumption of the relationship between the theoretical construct of true price and the empirically observed quotation prices in a microstructure setting; extend a new approach led by Easley and O’Hara (2010a) and Routledge and Zin (2009) to consider Knightian uncertainty—differentiated from risk—to interpret new market data; and contribute to the empirical literature on Knightian uncertainty.
ACKNOWLEDGEMENTS

The research and writing of this dissertation is dedicated to my dissertation committee, my family, and everyone who helped along the way.

I thank the NASDAQ OMX Group, the NASDAQ OMX Education Foundation, and especially the Economic and Statistical Research Group for providing the data to make this research possible. I also thank the NASDAQ OMX Education Foundation for the Ph.D. Fellowship for financial support.
DISCLAIMER

The views expressed herein are not intended to represent the views of NASDAQ OMX, its employees, or its directors. The author is solely responsible for the content, which is provided for informational and educational purposes only. Nothing contained herein should be construed as investment advice, either on behalf of a particular security or an overall investment strategy. Any mistakes remain the author’s own.

The author finished writing this dissertation while employed with the Commodity Futures Trading Commission (CFTC) of Washington, DC. No data housed at the CFTC or proprietary information from the CFTC were used to produce this dissertation. The research and/or information activities in this dissertation are neither required by nor related to Section 18 of the Commodity Exchange Act (7 U.S.C. § 22). The analyses and conclusions expressed in this dissertation are the author’s own, and do not reflect the views of other Commission staff, or the Commission itself.
TABLE OF CONTENTS

Chapter 1. Introduction .............................................................................................................................................. 1
  1.1. True Price .......................................................................................................................................................... 1
  1.2. No Trade ......................................................................................................................................................... 8
  1.3. Risk versus Uncertainty ............................................................................................................................... 13
  1.4. Overview ...................................................................................................................................................... 15
Chapter 2. Literature Review ........................................................................................................................................ 17
  2.1. Price Theory .................................................................................................................................................. 17
  2.2. Microstructure Theory .................................................................................................................................. 22
  2.3. Knightian Uncertainty ................................................................................................................................... 28
Chapter 3. Theory .......................................................................................................................................................... 38
  3.1. Easley-O’Hara Model .................................................................................................................................... 38
  3.2. Discussion ..................................................................................................................................................... 42
Chapter 4. Data and Methodology .................................................................................................................................. 44
  4.1. Institutional Details ......................................................................................................................................... 44
  4.2. Data ............................................................................................................................................................... 49
     4.2.1. Data Description .................................................................................................................................... 49
     4.2.2. Summary Statistics ............................................................................................................................... 52
  4.3. Constructed Metrics ..................................................................................................................................... 57
Appendix B. Java Code for Monte-Carlo Simulation................................................................. 156

References................................................................................................................................171
LIST OF TABLES

Table 1. SBUX Quotes at 9:30:00. November 5, 2012. Neighbor Cross. ........................................ 115
Table 2. SBUX Quotes at 9:39:46. November 5, 2012. Strict Cross.......................... 116
Table 4. SBUX Quotes at 9:39:47. November 5, 2012. Every Cross............................ 118
Table 5. Descriptive Stock Characteristics. ........................................................................ 119
Table 6. Summary Statistics on Variables of Interest......................................................... 120
Table 7. Relative Cross Frequency. Individual count....................................................... 121
Table 8. Relative Cross Magnitude. Individual count...................................................... 122
Table 9. Relative Cross Frequency. Group count........................................................... 123
Table 10. Relative Cross Magnitude. Group count........................................................ 124
Table 11. Simulation Quality............................................................................................. 125
Table 12. Comparing cross frequency between simulated and real stocks.................. 126
Table 13. Comparing cross magnitude between simulated and real stocks............... 127
Table 14. T-test Results. ................................................................................................. 128
Table 15. Time Series Test: Market Volatility. ................................................................. 129
Table 16. Cross-Sectional Test: Information Quality...................................................... 131
Table 17. Cross-Sectional Test: Insider Holding............................................................. 133
Table 18. Cross-Sectional Test: Analyst Coverage ......................................................... 135
Table 19. Cross-Sectional Test: Microstructure Effect................................................ 139
Table 20. Cross-Sectional Test: Full Model. ...................................................................... 141
LIST OF FIGURES

Figure 1. SBUX Quotes at 9:30:00. November 5, 2012. Neighbor Cross................................. 115
Figure 2. SBUX Quotes at 9:39:46. November 5, 2012. Strict Cross..................................... 116
Figure 3. SBUX Quotes at 9:31:09. November 5, 2012. First Cross...................................... 117
Figure 4. SBUX Quotes at 9:39:47. November 5, 2012. Every Cross.................................... 118
Figure 5. Intraday QBBO price updates per minute................................................................. 146
Figure 6. Intraday share volume per minute.............................................................................. 146
Figure 7. Intraday length of no-trade duration in seconds (median).......................................... 147
Figure 8. Intraday average duration-weighted bid-ask spread.................................................. 147
Figure 9. Intraday standard deviation of the bid-ask spread..................................................... 148
Figure 10. Intraday standard deviation of the bid ask midpoint against market movement...... 148
Figure 11. Intraday relative frequency of strict cross, individual count..................................... 149
Figure 12. Intraday relative magnitude of strict cross, individual count................................... 149
Figure 13. Intraday relative frequency of strict cross, group count.......................................... 150
Figure 14. Intraday relative magnitude of strict cross, group count........................................ 150
Figure 15. Intraday relative frequency of every cross, individual count.................................... 151
Figure 16. Intraday relative magnitude of every cross, individual count................................... 151
Figure 17. Intraday relative frequency of every cross, group count.......................................... 152
Figure 18. Intraday relative magnitude of every cross, group count........................................ 152
Figure 19. Histogram #1 on Number of Analyst Recommendations........................................ 153
Figure 20. Scatter plot on Standard Deviation of Analyst Forecasted EPS vs Number of Analyst Recommendations................................................................. 153
Figure 21. Histogram on Shares Float as percentage of Shares Outstanding.............................. 154
Figure 22. Scatter Plot on the Number of Analyst Recommendations against Average Market Cap. ................................................................................................................................. 154

Figure 23. Scatter Plot on the number of trades per stock against Average Market Cap. ........ 155
Chapter 1. Introduction

1.1. True Price

A financial market is where interested parties gather either physically or electronically to trade financial instruments like stocks, bonds, commodities and financial derivatives. Buyers and sellers, who possess various levels of information and skills, come to the market to search for the best price available. Either through intermediary like brokers, dealers or auctioneers, or on a centralized facility like a limit order book, an equilibrium price is reached when supply meets demand. Transactions occur at this market clearing price, where market participants consider a fair and true representation of the asset value at the given point in time. This price changes over time depending on the varying level of economic conditions, public and private information from all market participants. This process of price discovery is one of the most fundamental functions of a financial market. This dissertation explores issues pertaining to understanding the behaviors of market-determined price, and examines how the market price—through a limit order book mechanism—approximates or deviates from the expected true price.

The research question of this dissertation rests upon the definition of the true price, and how the market price deviates away from the true price. Using a dataset of 1696 NASDAQ-listed securities on November 5-9, 2012, I identify frequently occurring instances with empirical data, when the expected true price must have fallen outside the range of the bid-ask interval regardless of the actual value of the expected true price. I find this occurs frequently across the board, from stocks with very low liquidity to the most active stocks. My findings conflict with not only one of the basic assumptions of theoretical and empirical microstructure research, but also challenge our understanding of the price formation mechanism on a limit order book. Concerns over our proper understanding of the expected true price and the price discovery process pose potential
issues not only in academic research, but also for valuation and risk control in practice. Solvency and margin calculations rely on proper asset valuations, as well as market participants, professional or not, who need to make decisions over investments choice and portfolio selections. Transfers of ownership in the global economy will be troublesome if there is no reliable understanding on the proper price of an asset. In this section, I clarify the definition of true price, and how the true price differs from the market-determined exchange price, and factors contributing to the convergence and divergence between these two constructs in a historical setting of price theory. Not only does such distinction and conceptual understanding vary across economic disciplines, put to different levels of emphasis, but also it is blurred at times with multiple terms used interchangeably. In this dissertation, I view the expected value of the true price as the market clearing price at any given instantaneous moment, which changes constantly. Price formation under full, partial, or asymmetric information with different levels of competition, the function of price as a signal to direct asset allocation through market forces are common themes that connect price theory, asset pricing theory, microstructure theory, and the research question of this particular dissertation. This dissertation does not directly fall under the category of either price theory or general equilibrium theory, and is rather an empirical microstructure examination of how and why the expected true price (fundamental value) strays outside the boundaries of (deviates from) the bid-ask spread (the market-displayed price in absence of transactions).
This dissertation is an investigation of empirically observed quote movements on an electronic limit order book\(^1\). These movements seemingly violate the relationship between the theoretical construct of the true price and the market-displayed quotation prices as generally perceived in a traditional microstructure theory setting.\(^2\) This exercise rests upon three key questions: a) How is the concept of true price (also referred to as fundamental value\(^3\)) of a risky asset defined by economic theorists, and in particular, microstructure theorists? b) How is this value (or a close approximation of this value) discovered by market participants under different trading mechanisms, thus correctly (or incorrectly) reflected and disseminated in a financial market? c) How does transaction price, realized through trading, provide market participants with information on valuation of the security from the marketplace? d) How does this process provoke economists, market operators, and regulators to reexamine, validate, or even question existing models, and inspire more efficient market designs to better facilitate the price discovery process? This iterative process of theorizing, experimenting, observing, modeling, and revising is shared across scientific disciplines—including physics, economics, and finance, along with countless others—both in theory and in practice.\(^4\)

Before embarking on the examination of how and why the expected true price strays outside the boundaries of the bid-ask spread interval, it is worth straightening out the definition of true price and exploring how the relationship between the expected true price and the market-

---

\(^1\) *Trading and Exchanges: Market Microstructure for Practitioners* (Harris, 2002) provides a comprehensive and easy-to-read overview on concepts such as limit and market orders, market makers, exchanges, and limit order books for readers who are not familiar with market microstructure theory.

\(^2\) *Market Microstructure Theory* (O’Hara, 1995) and *Empirical Market Microstructure: The Institutions, Economics, and Econometrics of Securities Trading* (Hasbrouck, 2007) are two standard textbooks at the graduate level and standard references in this area.

\(^3\) The terms true price and fundamental value frequently are used loosely and interchangeably in finance and economics, and are therefore used interchangeably in this dissertation text.

\(^4\) *The Art of Scientific Investigation* (Beveridge, 1957) is a classical text that discusses this process of observation, hypothesis, experimentation, and revision in scientific research.
displayed price (from either transactions or quotations) have been considered in the context of microstructure theory.\(^5\) Chapter 2: Literature Review provides a more comprehensive literature review on this matter from the perspectives of price theory, asset pricing theory, and market microstructure theory.

Madhavan (2000) defines the true price or the fundamental price as the full-information expected present value of future cash flows. In (weakly) efficient markets,\(^6,7\) price reflects all public information. If traders do not possess asymmetric information and frictions can be neglected, prices simply reflect expected values, thus following a martingale process\(^8\). Markets are said to be efficient in the sense that prices at all points in time reflect expected values. Madhavan (2000) also states, however, that asset prices need not equal full information expectations of value because of a variety of frictions—namely the microstructure frictions. Price in this context refers to the market price, which deviates from the sum of all discounted future cash flows (i.e., the fundamental value, or the expected true price) in the presence of investor latent demand\(^9\), market structure, and trading mechanism frictions. A typical

\(^5\) Microstructure theory is defined by O’Hara (1995) as “the study of the process and outcomes of exchanging assets under a specific set of rules. While much of economics abstracts from the mechanics of trading, microstructure theory focuses on how specific trading mechanisms affect the price formation process” (p. 1).

\(^6\) Fama’s (1965) published Ph.D. dissertation, *The Behavior of Stock-Market Prices*, states that stock price follows a random walk. This is often considered the initial formulation of the efficient-market hypothesis.

\(^7\) In addition to the weak-form efficient market, there is also a) semi-strong-form efficiency, which implies trading on publicly available new information cannot generate excessive return, and b) strong-form efficiency, which states that all information—both public and private—is reflected in price. Roberts (1967) defines these efficiencies; Fama (1976) and others formally test them.

\(^8\) In probability theory, martingale is a stochastic process where past events do not affect the future. It is a model of fair game, where past events do not predict future winnings. A related concept is the Markov chain, where probability of the next state only depends on the current state, not the states prior to the current. This process is referred to as memoryless.

\(^9\) Latent demand is also referred to as induced demand, which means the phenomenon when more goods or services are consumed when the supply increases.
microstructure model setup involves one or more market makers\textsuperscript{10} who post limit orders\textsuperscript{11} to buy at the bid and sell at the offer. They make the bid-ask spread as their profit in exchange for serving the functions of facilitating price discovery, providing liquidity and continuity, and stabilizing price during short volatility (Harris, 2002). The dynamics of price setting exhibits itself differently between a dealer-driven market and an order-driven market. For example, Madhavan and Sofianos (1998) examine the magnitude and determinants of dealing strategies by New York Stock Exchange (NYSE) specialists across stocks and over time.\textsuperscript{12} In this section, I discuss the relationship between the expected true price and the market price in two classical setups of microstructure models as examples: inventory-based models and information-based models, while deferring an exhaustive examination to the literature review in Chapter 2: Literature Review.

In inventory-based models such as Garman (1976), Stoll (1978), Amihud and Mendelson (1980), and Ho and Stoll (1981), deviations between the true price and the displayed quotations or the transaction price can temporarily occur due to net inventory accumulated by market makers, which can be induced by temporary supply and demand imbalances. In this case, the bid-ask midpoint becomes a biased proxy for the true price. The dealers prefer trading on either the bid or the offer side in order to exit accumulated long or short positions through trading, instead of being neutral if without a net inventory. Under this setup, the expected true price is by

\begin{itemize}
  \item A market maker is also sometimes referred to as a dealer, although this term within the context of microstructure studies is not to be confused with financial market intermediaries such as banks, which usually hold inventory and are not always risk neutral. The role played by market makers in the price-setting process is the focus of tremendous theoretical and empirical research in microstructure, although it is worth noting that microstructure theory is not limited to that. Even in the microstructure models that focus on market maker behavior and strategy, the setups and assumptions frequently differ.
  
  \item As defined in Madhavan (2000), "limit orders are orders to buy (sell) that specify a maximum (minimum) price at which the trader is willing to transact. A market order is an order to buy (sell) at prevailing prices. A stop order is an order that becomes a market order if and when the market reaches a price pre-specified by the trader" (p. 212).
  
  \item I will provide further review on these distinctions between in Chapter 2.
\end{itemize}
definition straddled by the bid-ask spread interval. In information-based models such as Glosten and Milgrom (1985), Easley and O’Hara (1987), and Kyle (1985), market makers adjust their quotes to anticipate informed order flow and try to stay inventory neutral, while informed traders possess superior information and trade on it. In a sequential trade framework like Glosten and Milgrom (1985), a buy market order signals good news to the market maker; therefore, the stock value conditional on a buy order is greater than the unconditional expected value of the asset, while the same argument goes for the bid price, which is contingent on a sell market order arriving next. In this context, the bid-ask midpoint becomes a reasonable approximation of the unconditional expected value of the underlying asset—the true price. It is worth noting that in either of the two classical setups of microstructure framework, the expected true price should not be a) higher than the best offer, which at any given moment is the lowest price for which someone is willing to sell the stock, or b) lower than the best bid, which is the highest price someone is willing to pay for the stock, as long as the market maker maintains a risk-neutral inventory. Harris (2002) poses the “two brothers and a piece of cake” example\textsuperscript{13} to articulate the reason dealers must set their bids below the expected true price and their offers above them in order to avoid adverse selection from informed traders. This is consistent with the arbitrage pricing theory (APT) under asset pricing theory,\textsuperscript{14} where the market price should equal the expected true price, or at least their differences should not exceed the size of transaction costs in

\begin{footnotes}
\item[13] Harris (2002) discussed a solution to the metaphorical scenario of two selfish brothers attempting to fairly divide a piece of cake. Harris proposed letting one brother divide the cake and the other choose which half he would like (p. 287). This resembles the situation in which dealers quote two-sided markets.
\item[14] Microstructure theory interfaces with asset pricing theory through the impact of liquidity on the cross-sectional variations in stock returns and the differences in liquidity over time. Liquidity explains variations in the risk premium, and therefore may influence stock price levels. For example, Amihud and Mendelson (1986) showed that investors must be compensated for the higher transaction costs in less liquid securities through higher expected returns.
\end{footnotes}
an arbitrage-free world; otherwise, arbitrageurs will buy low and sell high until the two are brought back to close proximity.\textsuperscript{15}

This dissertation pertains to this set of issues in understanding how the market price approximates or deviates from the expected true price. I provide empirical evidence to show instances where the expected true price strays outside the boundaries of the bid-ask spread interval. These quote movements on an electronic limit order book were observed within extremely short time periods—usually on the millisecond scale or even less. This observation cannot be explained by the possible deviations with the inventory-based models where the market makers adjust the bid-ask midpoint when they accumulate undesired long or short inventory, because in that case the purpose of the deviation is to trade and liquidate the inventory.

My observation window is based on time periods of no trades. I compare the relative price levels of bid, ask, and the midpoint between adjacent trades, and identify instances where the expected true price seemingly strays outside the boundaries of the bid-ask spread interval. The occurrences of these instances are not only frequent, but also large in magnitude in stocks across the spectrum of market capitalization (market cap), volume, and price level.\textsuperscript{16} This is not consistent with the market maker adjusting quotes for inventory reasons. It also violates the information-based model where the bid-ask midpoint should be a reasonable proxy for the unconditional expected value of the underlying asset. These empirical findings contradict our

\textsuperscript{15} APT was introduced by Ross in 1976 as an extension to the capital asset pricing model (CAPM). Arrow-Debreu assets/security (Arrow & Debreu, 1954) is a building block to this concept. Becherer and Davis (2010) define the Arrow-Debreu prices as the prices of state contingent claims, which deliver one unit of a specific consumption good if a specific uncertain state realizes at a specific future date. Under the non-arbitrage principle, every asset in the economy is priced with three basic assumptions: a) the market is in equilibrium, b) every agent maximizes her utility, and c) rational agents have preferences that are complete and transitive.

\textsuperscript{16} See further details in Chapter 5 and Tables 7 and 8 in the Appendix.
traditional understanding of the relationship between the theoretical concept of expected true price and the empirically observed quotation prices in a microstructure setting. I suggest using a Knightian uncertainty model to explain these price relationships during periods of no trade, and empirically test the hypothesis.

1.2. No Trade

Cao, Ghysels, and Hatheway (2000) state that it is the fundamental issue of market microstructure theory to study how information is incorporated into price. In turn, price also conveys information about supply and demand on the underlying asset. This price refers to the market clearing transaction price, which maximizes the quantity of goods to change hands. The Financial Accounting Standards Board (FASB) defines "fair value" as "price in an orderly transaction between market participants to sell the asset or transfer the liability in the market in which the reporting entity would transact for the asset or liability, that is, the principal or most advantageous market for the asset or liability" (FASB Statement No. 157, 2010). Even though the weak, semi-strong, and strong forms of market efficiency allow different extents of public and private information to be incorporated into price, the signal nevertheless conveys information to individuals, firms, and other participants in an economy about the scarcity of goods and services. This information then allocates limited resources in an economy through market forces. If this signal does not accurately represent the expected true price, mispricing

17 Known price discovery mechanisms include continuous markets, auction markets, price experimentation, and price signaling. See Chapter 2 for a more detailed review of classical models in each category.

18 One of the classical approaches to address this allocation issue in economic literature is Pareto optimality (also referred to as Pareto efficiency), which describes a state of asset allocation where it is impossible to make any individual better off without causing at least one other individual to be worse off. It is named after Italian economist Vilfredo Pareto (1848–1923). A related concept called weak Pareto optimum refers to a situation where there is no alternative allocation possible to cause every individual to be better off. Greenwald and Stiglitz (1986) consider externalities like tax interventions by whether it is Pareto improving.
will lead to non-optimal allocation (Alti & Tetlock, 2011). The observations of this price signal are made at time intervals according to the needs of different market participants. It can be the daily closing price\(^{19}\) of a stock, settlement price of a derivatives contract,\(^{20}\) or the reference price of a similar or closely related asset. In modern equity and derivatives markets, especially in the age of high-frequency trading, price information is updated in near real-time—at the sub-millisecond time interval for active securities and instruments.\(^{21}\) In the meantime, valuation becomes problematic for less active securities or asset markets that are less liquid. The last trade price becomes an obsolete measure if it had occurred long before economic conditions and market conditions changed.

Easley and O’Hara (2010a) address the need for a new theoretical framework to explain the “market freeze” during the 2007–2009 financial crisis, where little or no trading activity was taking place in the mortgage-backed securities and collateralized debt obligation (CDO) tranches. Traders were unwilling to either buy or sell at a wide range of price levels. When there is no market clearing price, the market has no way of evaluating how much an asset is worth—here lays the crux of the issue. In the absence of a recent transaction price, one might look to the quotation prices as benchmarks for security value—the bid price, the ask price, and sometimes the bid-ask midpoint. This method is troublesome in this case because the bid and the ask prices

\(^{19}\) In the U.S. equity market, the closing price refers to the last trade price when the regular trading hours conclude at 4:00 p.m. on a trading day. In the derivatives market, each exchange has its own specific procedures to determine the settlement price for specific contracts. For example, the Chicago Mercantile Exchange uses a volume-weighted average price (VWAP) of all trades executed in the full-sized futures contract on the trading floor and in the E-mini futures contract executed on CME Globex to calculate the daily settlement price for E-Mini S&P 500 Futures for the lead month from 15:14:30–15:15:00 Central Time (CME Group, E-Mini S&P 500 Futures Daily Settlement Procedure).

\(^{20}\) In the derivatives, there are look-alike contracts that settle off the core reference contract, based on the same commodity, or priced at a fixed differential to a core reference contract.

\(^{21}\) A large set of recent literature addresses the issues involving high-frequency trading; the pertinent research includes but is not limited to Brogaard, Hendershott, and Riordan (2014); Foucault, Hombert, and Roșu (2015); and Hirschey (2013).
are far apart: neither the best case nor the worst case scenario captures the “fair value” in spirit of the FASB definition. Easley and O’Hara (2010a) also illustrates in Proposition 3 that depending on the degree of diversity of opinions of market participants, it is even possible for the fair value—the average of heterogeneous expectations on price—to stand outside the bid-ask spread range. In this dissertation, I document empirical evidence from the equity market that is consistent with this prediction in Proposition 3. I use the Knightian uncertainty framework provided by Easley and O’Hara (2010a) to examine the empirically observed instances where the expected true price seemed to stray outside the bid-ask spread, and I test this prediction empirically.

The no-trade situation modeled in Easley and O’Hara (2010a) is not to be confused with the no-participation scenario in a microstructure setting, where a wide bid-ask spread is also produced. In his seminal essay “Noise”, Black (1986) articulate the importance of noise traders, arguing that the market maker would widen the bid-ask spread as the presence of informed trading increases, and a point of no trade is to be reached if trades are only to be initiated by informed traders. In other words, informed trading can also induce a no-trade situation, but not due to a lack of liquidity or general trading interest. Glosten and Milgrom (1985) discuss this particular market breakdown due to adverse selection, and suggest that the functioning of markets depends on the supply and demand elasticities of uninformed traders. This is considered a market failure that can be repaired by removing information asymmetry or forcing the market maker to trade at a loss (Hasbrouck, 2007).

A related point on trading motivation versus information was made by the Grossman-Stiglitz Paradox (Grossman & Stiglitz, 1980). The paradox states that if the market is strong-form efficient, all public and private information is reflected in the price, and no one has an
incentive to spend money or use resources to gather information to trade on it. In other words, in a strong-formed efficient market, all information has been incorporated into price; no profit can be made through trading. This is a paradox because someone has to do the analysis and trading in order to bring price to an efficient level. This simply proves the market cannot be perfectly informationally efficient; someone has to do the analysis first, which is not instant or free. The nonparticipation scenarios described in both Easley and O’Hara (2010a) and the Grossman-Stiglitz Paradox are similar in appearance to this dissertation in that transactions were absent. However, they both fundamentally differ from the no-trade situation in this dissertation because in both of these two scenarios, the bid and ask prices are unchanging, while the bid-ask spread is wide. To the contrary, the crossing phenomenon I characterize in this dissertation involves constant and drastic movements of the quotation prices between trades.

Another type of nontrading effect—one addressed by Lo and MacKinlay (1990)—is called nonsynchronous trading, which has to do with the time series of asset prices being observed at irregular time intervals. Assuming the returns of two stocks respond to the same news signal, the more active stock will respond to the news first when the less liquid stock will respond in a lag, thus producing a spurious autocorrelation between the two time series of returns. A false sense of predictability is empirically present, even if the true price changes in a statistically independent manner. The Lo and MacKinlay (1990) model differs from the no-trade situation in this dissertation in that a) it focuses on autocorrelation of returns across stocks, while this dissertation captures characteristics of quote movements within each stock b) the empirical

22 An illustration of this model was given in The Econometrics of Financial Markets by Campbell, Lo, and MacKinlay (1997).
examination of the nonsynchronous trading looks at the delayed print of trades instead of quote movements between trades.

Last but not least, Cao, Ghysels, and Hatheway (2000) examine a different type of no-trade setting, where NASDAQ market makers signal pricing information through locked and crossed inside quotes (best bid and offer) during the preopening period. These quotations differ from the quotations that form the best bid and offer I examine in this dissertation not only in that the identities of the market maker providing the quotes are revealed (instead of anonymous), but also that they are noncommittal. The authors find the directions of these nonbinding inside quotes are consistent with price signaling and contribute to price discovery.

In this study, I examine bid and ask prices within time windows between adjacent trades. During these windows, the expected true price presumably stays at the same level. The expected true price (as well as the true price itself) cannot be observed, because it is a theoretical construct. I instead look for violations regardless of where the expected true price had been, and test the reliability and accuracy of the bid-ask midpoint as a proxy for the expected true price. The construction of the test statistics detailed in Chapter 4: Data and Methodology rely on the assumption in market microstructure theory that the process of price discovery—both public and private information—occurs through trading. This assumption is rarely explicitly stated because it is embedded into theoretical models and empirical studies. The informed traders who possess superior information make profits by trading against the uninformed (O’Hara, 1995). Some traders can react to public news announcements faster than others make profits by the speed advantage (Scholtus, Dijk, & Frijns, 2014). Once the valuation of the expected true price of the security changes because new information arrives at the marketplace, trading will occur. An implication of this price discovery framework is that if movements of the expected true price
always induce trades, and trades are indicative of movements in the expected true price, then one can assume the expected true price doesn’t move between trades—particularly over short time horizons—without news occurring. Thus, during periods of no trades (or more accurately speaking, the periods between adjacent trades), the expected true price should stay between the bid price and the ask price, with the bid-ask spread determined by microstructure risks—information asymmetry and inventory cost by market makers—plus order processing costs.\footnote{Decomposition models that estimate the components of the bid-ask spread include but are not limited to George, Kaul, and Nimalendran (1991); Glosten and Harris (1988); Huang and Stoll (1997); and Stoll (1989).} The empirical evidence (which I termed crossing) I present in Chapter 5: Empirical Results violates classical market microstructure theory’s theoretical and empirical assumption that true price always sits within the range of the bid-ask spread while the bid-ask midpoint is its best proxy. The bid-ask midpoint is also not an appropriate representation of the fair value, because it is a constantly changing value, and frequently stands outside the range of another set of bid-ask spread before the next trade. What is shared in this study with the Easley and O’Hara (2010a) no-trade situation is that a) no active market participant is willing to trade at a whole range of price levels, b) it becomes difficult to assign a fair value to the underlying asset, and c) it leaves ambiguous of any economic or financial activity that requires accurate valuation for the security.

1.3. Risk versus Uncertainty

Outside academic research, the concept of uncertainty was popularized by Donald Rumsfeld’s (2002) famous quote regarding the lack of evidence linking the government of Iraq with the supply of weapons of mass destruction to terrorist groups:

Reports that say that something hasn't happened are always interesting to me, because as we know, there are known knowns; there are things we know we know. We also know there are known unknowns; that is to say we know there are some things we do not know.
But there are also unknown unknowns—the ones we don't know we don't know. And if one looks throughout the history of our country and other free countries, it is the latter category that tends to be the difficult ones.

In the formal literature setting, the concept of Knightian uncertainty (more frequently referred to as ambiguity in this set of literature) was introduced by Knight in his seminal work, *Risk, Uncertainty, and Profit* (1921). Knight formally distinguishes risk (which is susceptible to measurement) from uncertainty (which is unmeasurable). In a risk setting, the decision maker possesses known probability distributions, whether that estimation is objective or subjective. Although his concept was revolutionary, it is noteworthy that Knight (1921) does not provide a strict mathematical definition of uncertainty. Various interpretations and treatments emerged subsequently, ultimately formulating the decision theory that provides the framework for how uncertainty affects economic activities. This dissertation conforms to the interpretation of the strand of literature following Ellsberg (1961) that uncertainty is unknown and impossible to quantify. Some econometrics literature takes a different view—one in which risk is viewed as the first order uncertainty over the state space, while ambiguity is the second order uncertainty over the corresponding probability space.

The effects of risk versus Knightian uncertainty have never been distinguished in a microstructure context. Easley and O’Hara (2010a) fill in the literature gap theoretically, but did not test it against real data (no other research has tested the Easley model empirically). One of the difficulties in testing such a model is that most quantities in the concluded theorems are unobservable, as Knightian uncertainty itself cannot be quantified by definition. This is also why empirical research on Knightian uncertainty is extremely rare in general.

This dissertation makes empirical observations from the U.S. equity market, where the relative price levels of the best bid and offer do not conform to the classical microstructure
framework; however, consistent with the prediction made by Proposition 3 in Easley and O’Hara (2010a), the expected true price can stray outside the bid-ask spread in the presence of Knightian uncertainty. I characterize and quantify this phenomenon, then conduct empirical tests with a process-of-elimination approach to investigate if such violations of the classical paradigm are indeed caused by the presence of Knightian uncertainty or risk factors. To model uncertainty, I primarily adhere to Easley and O’Hara’s (2010a) contention that uncertainty turns a statistical distribution from known to unknown. One interpretation of Knightian uncertainty in existing literature—an interpretation I take on in this research—is that the probability distribution is not only unknown, but also impossible to quantify. Since the distribution of a variable is unknown, one cannot model this variable or measure its effects. In an attempt to model the effects of uncertainty on market activities—and, specifically in this study, the effects of uncertainty on quote updates between trades in an active equity market—I turn the known expected mean from a no-uncertainty scenario, which is a specific value, into a range in an under-uncertainty scenario. Within this range, one does not know and has no probability assigned to any given value being more likely than the other. This is the most simplistic case of turning something known into unknown and unknowable. Once we understand the effect of uncertainty under the most simplistic case, we can then turn the known variance into unknown, and ultimately infer the impact of something that is supposedly impossible to measure.

1.4. Overview

The structure of this dissertation is organized as follows: In the next chapter, I first review literature in price theory and asset pricing theory, explaining how seminal work in both areas considers the concept of true price and the relationship between true price and empirically observed price measures. I briefly cover puzzles and empirical failures in asset pricing theory
that led to Knightian uncertainty and its application in asset pricing as a solution. I treat Knightian uncertainty, its decision theories, and their applications in much more detail than in sections of Chapter 1: Introduction. I then review microstructure theory models, including inventory models, information models, and strategic trade models, to compare how these models specifically define the concept of true price and handle the relationship between the true price and the market price differently. I also provide a review of related literature regarding how other studies apply Knightian uncertainty in microstructure. In Chapter 3: Theory, I provide a detailed review of the Knightian uncertainty model in Easley and O’Hara (2010a). I only cover parts of the model that pertains to the conditions in the equity market that I wish to characterize in this dissertation. Chapter 4: Data and Methodology begins with an equity market structure overview and data description, then presents the methodology on test statistics construction to quantify the quote price behavior when true price is clearly outside the range of the bid-ask spread. I also provide details on the hypothesis design, and a Monte Carlo simulation, which is used as a benchmark for significance testing in the following chapter. Because one cannot prove that Knightian uncertainty is or is not the cause of the crossing phenomenon I characterize through hypotheses testing due to its undefinable nature, I design and test alternative competing hypotheses as a process-of-elimination to indirectly provide evidence to that the crossing behaviors of quote prices are consistent with Knightian uncertainty. In Chapter 5: Empirical Results, I present the empirical evidence that are contrary to the predictions from classical microstructure models, but are consistent with the implications of the Easley and O’Hara (2010a) model. Next, I summarize my investigation of the cause for this odd price behavior. Chapter 6: Conclusions presents a summary of the empirical findings, a description of the contributions of this dissertation to existing literature, and a discussion of limitations and future research.
Chapter 2. Literature Review

2.1. Price Theory

Many economic schools of thought have emerged, evolved, and influenced one another throughout history. The conceptual distinction between the expected true price and the market-observed price dates back to the diamond-water paradox\textsuperscript{24} contemplated by philosophers in ancient Greece. Aristotle distinguishes the value in use versus the value in exchange, thus recognizing the influence of demand upon value (Sewall, 1901). Similarly, Petty, in his 17th century economic treatises, considers price the value in exchange, different from the natural value or the intrinsic value, which is based on both land and labor.\textsuperscript{25} The paradox was considered by later economists like Law (1705),\textsuperscript{26} Locke (1692/1969),\textsuperscript{27} and notably Smith in *The Wealth of Nations* (1776),\textsuperscript{28} a text widely considered the beginning of classical economic theory.\textsuperscript{29} Smith (1776) argues that the “natural price” of a commodity is determined by “the rent of the land, the wages of the labour, and the profits of the stock employed in raising, preparing, and bringing it to

\textsuperscript{24} It is also known as the paradox of value, which demonstrates the contradiction that diamond has higher market price even though water is more useful for human's survival. For detailed definition and examples of the diamond-water paradox, see lecture notes by Hill (n.d.).

\textsuperscript{25} Roll (1942), in his *A History of Economic Thought*, contends that Petty had confusion between the two concepts of value versus price, which is inconsistent with Petty’s dealing with exchange value, as evidenced by the fact Petty frequently speaks of labor alone.

\textsuperscript{26} Many texts and articles, including Raghunathan (2010), note that Copernicus also addressed the paradox of value before Smith; however, I could not locate the original essay by these authors that explicitly commented on this paradox.

\textsuperscript{27} Locke’s (1692/1969) consideration of the paradox is similar, but does not involve diamonds. Locke posits: “What more useful and necessary things are there to the being or well-being of men, than air and water? And yet these generally have no price at all, nor yield any money: because their quantity is immensely greater than their rent, in most places of the world. But as soon as ever water comes anywhere to be reduced into any proportion to its consumption, it begins to presently have a price, and is sometimes sold dearer than wine” (pp. 63–64).

\textsuperscript{28} The book was originally published under the title of *An Inquiry into the Nature and Causes of the Wealth of Nations*, but has been generally referred to by its shortened title: *The Wealth of Nations*.

\textsuperscript{29} Stigler (1976) declares the systematic analysis in *The Wealth of Nations* of individuals pursuing self-interest under conditions of competition to be “the most important substantive proposition in all of economics” and “the foundation of the theory of the allocation of resources” (p. 3).
market, according to their natural rates” (Book I, Chapter VII). These natural prices might temporarily be above, below, or exactly the same as the “market price” due to temporary fluctuations of supply and demand; however, free competition and market forces—the invisible hand—will eventually drive them to converge. In Smith’s own words:

The natural price, therefore, is, as it were, the central price, to which the prices of all commodities are continually gravitating. Different accidents may sometimes keep them suspended a good deal above it, and sometimes force them down even somewhat below it. But whatever may be the obstacles which hinder them from settling in this centre of repose and continuance, they are constantly tending towards it. (Book I: On the Causes of Improvement in the Productive Powers. On Labour, and on the Order According to Which its' Produce is Naturally Distributed Among the Different Ranks of the People. Chapter I: On the Natural and Market Price of)

Additionally, systematic obstructions like monopoly, regulation, and secrets in manufactures can enhance the market price above the natural price for a long time by keeping the supply below the effectual demand. In his On the Principles of Political Economy and Taxation, Ricardo (1817) acknowledges and agrees with these observations by Smith, although he differentiates his views from Smith, positing that rent is price-determined instead of price-determining; he also disagrees with Smith’s contention that a rise in cost of labor would cause a uniform rise in the price of all commodities. Ricardo struggles with the problem of value through the end of his life.

The diamond-water paradox was not solved satisfactorily until the Marginalist Revolution in the 19th century. Marginalism refutes the labor theory of value, and revises the concept of value from the usefulness of the overall quantity to the additional unit of the asset or commodity. Marshall (1890) makes a luminous case for the marginal principles in his Principles

---

30 This “adding-up” approach is sometimes referred to as the cost-of-production theory of value, where the price of a commodity is determined by adding up the costs of producing it, including factors like labor, capital, land, etc.
31 For the history of Marginalism, see Rhoads (2008).
32 Marshall’s (1890) model solves the diamond-water paradox by illustrating that water is cheap because it has both a low marginal value and a low marginal cost of production. Diamonds are expensive because they have both a high marginal value (desirability) and a high marginal cost of production.
of Economics (1890) by illustrating that supply and demand simultaneously determine the equilibrium price using the Marshallian cross\textsuperscript{33} through the function of the marginal utility. Even though the term “value” was employed, Marshall uses it synonymously with “price” as “exchange value” or “money value.” Additionally, he defines consumer surplus as the difference between the price paid for the commodity and the value (gratification or utility) enjoyed by the individual—a subjective element. The Marshallian partial equilibrium model of considering one market at a time later became the basis for establishing the general equilibrium theory pioneered by Walras (1874).

Modern mainstream economic theory primarily builds upon neoclassical economics,\textsuperscript{34} which determines price with three basic assumptions: a) economic agents have rational preferences, b) individuals (consumers) maximize utilities and firms maximize profits, and c) agents act based on full and relevant information (Weintraub, 1993). The rational man assumption was criticized as unrealistic by institutional economist Veblen (1898/1931) in his essay “Why Is Economics Not an Evolutionary Science?” Samuelson’s seminal work Foundations of Economic Analysis in 1947 introduces neoclassical synthesis, combining neoclassical microeconomics and Keynesian macroeconomics, as well as developing the theory into a mathematical structure. The general equilibrium theory following Arrow and Debreu (1954) characterizes the economy by a large number of simultaneous equations, recognizing that

\textsuperscript{33} Microeconomic Theory: Basic Principles and Extensions (2007, 10th edition) by Nicholson and Snyder provides a more detailed history on the development of the economic theory of value, the marginal principles, and illustrations of the Marshallian cross.

\textsuperscript{34} It is debatable whether Marginalism falls under the classical or neoclassical framework.
one cannot discuss any single market in isolation.\textsuperscript{35} Once the price changes in one market, the effects reverberate through other markets. The general equilibrium theory considers equilibrium prices as long-term prices, while actual prices are deviations from equilibrium. Achieving equilibrium is the process of price setting.

Keynesian and post-Keynesian economists criticize the equilibrium assumption in both neoclassical economics and general equilibrium theory for being misleading,\textsuperscript{36} because they apply fundamentally static analysis instead of dynamic analysis, as “in the long run we are all dead.” (\textit{A Tract on Monetary Reform}, Keynes 1923, Chapter 3) Imperfect competition (monopoly and oligopoly) and sticky wages (also known as nominal rigidity) are resistant to changes in nominal price—in other words, hindering the convergence from the nominal value to the real value. Hayek, a leading figure of the competing Austrian school, contends that market prices reflect information—the totality of which is not known to any single individual—that determines the allocation of resources in an economy. In particular, \textit{Economics and Knowledge} (Hayek, 1937) challenges the assumption of full and correct information in equilibrium theory for not being applicable to the real world. His view on price signals—that they communicate information, thus enabling individuals to coordinate their plans—resonates with arguments later made by Friedman on the functions of price. Schumpeter, another leading scholar of the Austrian school, criticizes the Keynesian approach for reasons in terms of abstract models, where they would freeze all but a few variables. Instead, he advocates for more dynamic, change-oriented,

\textsuperscript{35} For treatment of the Arrow-Debreu model, which is the basis of modern price theory, and other elementary models like Robinson Crusoe, the Edgeworth Box, and a two-commodity two-household two-firm model, etc., see \textit{General Equilibrium Theory: An Introduction} (2011) by Starr.
\textsuperscript{36} See \textit{A Tract on Monetary Reform} (1923) by Keynes, and “The Debt-Deflation Theory of Great Depressions” (1933) by Fisher.
and innovation-based economics. Friedman, along with other Chicago school price theorists, refocuses on rational expectations and free markets. In his classical text *Price Theory* (1962), Friedman defines the three functions of price: a) to transmit information, b) to provide an incentive to users of resources to be guided by this information, and c) to provide an incentive to owners of resources to follow this information. These three functions of price directly address five problems laid out in Knight’s 1933 book, *Economic Organization*: a) fixing standards, b) organizing production, c) distributing the product, d) providing for economic maintenance and progress, and e) adjusting consumption to production over short periods.

From the modern portfolio theory of Markowitz (1952) to the capital asset pricing model of Treynor (1961, 1962) and Sharpe (1964), the expected true price of a financial asset is viewed as the sum of discounted future cash flows. Valuations are dependent on both future asset payoffs and stochastic discount factors. This view of the true price has been adopted in many areas of studies in finance and economics. In the context of asset pricing theory, a security is correctly priced when the market price is the same as the true price, otherwise the security would be either undervalued or overvalued. As one of the leading scholars of the Chicago school of economics, Lucas is known to have incorporated rational expectations into a dynamic general equilibrium model. In his seminal work *Expectations and the Neutrality of Money* (1972), Lucas asserts that agents make decisions based on available information; they form expectations about future prices and quantities, and based on these expectations they act to maximize their expected lifetime utility.

---

37 Both Friedman and Samuelson studied at the University of Chicago under Knight, after whom Knightian uncertainty is named.
Fama (1970) defines an efficient market as one in which prices always fully reflect available information instantaneously. He differentiates the weak-form, semi-strong form, and the strong-form of market efficiency, while stating that the efficient market hypothesis is impossible to test through the joint hypothesis; although through their no-trade paradox Grossman and Stiglitz (1980) states that the market cannot be perfectly efficient. Later disputes of the hypothesis include but are not limited to cognitive biases modeled in behavioral economics and finance. Nevertheless, if the market were to be efficient, then price and the expected true price should be identical, assuming no transaction cost, no information cost, and no diversity of opinions.

2.2. Microstructure Theory

Market microstructure theory studies the price formation and discovery mechanism: how exactly the equilibrium price between supply and demand is attained under specific market rules. This area of study is important because the price-setting process itself directly influences the asset price (Madhavan, 2000). Financial markets are designed to aggregate buyers and sellers through different types of price-setting mechanisms, as I will discuss in this section, and to find a fair market clearing price that satisfies utility maximizing conditions, sometimes under constraints like inventory or risk aversion. The price derived from this process also serves as a signal—as noted by Friedman (1962)—of price theory, disseminating valuable information to the marketplace to guide firms and individuals on resource and asset allocation. Microstructure theory does not directly confront equilibrium theory, but rather have a different focus to examine how the equilibrium price is obtained instead of the exact value. It pertains to the temporary equilibrium in Hicks (1939), instead of the long-term equilibrium on which neoclassical economic theory and general equilibrium theory focus. Traditional rational expectations
literature assumes the irrelevance of trading mechanisms in affecting the resulting equilibrium price (O’Hara, 1995). Market microstructure as a field of study addresses this issue.

In a comprehensive survey of microstructure literature, Madhavan (2000) gives the definition of several price concepts in a canonical model: $V_t$ of a risky asset is the expected present value of future cash flows under full information, while $P_t$ is its market price at time $t$. This differentiation corresponds to the expected true price versus exchange price discussed in my review of price theory. $V_t$ can change over time due to variations in expected cash flows or the discount rate, while $\mu_t$ is the conditional expectation of $V_t$ given the set of public information at time $t$. As we shall see in the review of microstructure models, this market price also incorporates private information instantaneously or gradually, depending on the model setup. If traders do not possess asymmetric information (implicit costs), and trading frictions (explicit costs) can be neglected, price simply reflects the expected value: $P_t = \mu_t$. In this model, $\mu_t$ follows a martingale process while applying the Law of Iterated Expectations,\(^{38}\) and returns are serially uncorrelated. Markets are efficient in the sense that prices at all points in time reflect expected values. In Chapter 1: Introduction, I discussed two types of classical microstructure model setup: inventory-based models and information-based models. The bid-ask spread is primarily viewed as a risk spread with information asymmetry and inventory risk factors. The inventory effect on price is transitory, while the effect from information is permanent.

Walras (1874) provides a simple price-setting process he named tâtonnement.\(^{39}\) Each market participant would submit his supply and demand to an auctioneer, who would then

\(^{38}\) Also referred to as Law of Total Expectation, or Adam’s Law.

\(^{39}\) The French word means trial and error, and in this context refers to a series of preliminary auctions the Walrasian auctioneer conducts. It is a trial and error process because every trader is given the opportunity to revise his order before trading actually occurs. The auction will continue until no one needs to revise his order anymore.
aggregate supply and demand to find a market clearing price for equilibrium. The model assumes perfect information and no transaction cost. It is a simple mechanism, without a time dimension, that introduces the bid-ask spread as cost of immediacy. The time dimension was later introduced by Demsetz (1968) in a one-period model. This study formally sets the state of market microstructure as a field of study, because the specific market rules achieving the equilibrium price provides an input to the price itself.

In his pioneering essay *The Only Game in Town* in 1971, Bagehot (pseudonym for Treynor) first distinguishes market gains from trading gains. He is also the first to categorize market participants into three types: market makers, informed traders, and noise traders. Market makers in general lose money to informed traders, and break even or make money from noise traders. This three-trader setup later became a classical framework for information-based models with variations. Unlike inventory-based models, the bid-ask spread would still exist in information-based models even without explicit trading costs. However, Bagehot does not directly attribute this relationship to the size of the bid-ask spread. Copeland and Galai (1983) accomplish this task using a one-trade model. They specify the true price as an exogenous signal, known to the informed trader while unknown to the uninformed trader, and then solves the optimal bid and ask price placements by the dealer to maximize profit. The probabilities of order flow from informed versus uninformed traders determine the size of the bid-ask spread. In this model, the true price is always higher than the bid price and lower than the ask price.

Copeland and Galai (1983) also demonstrate that volatility and price play parts in determining the bid-ask spread through a separate setup, which prices the bid and ask quotations as call and

40 Also referred to as liquidity traders or uninformed traders. Motivations of trades by noise traders are not always specified in microstructure models, and are usually considered as exogenous to the model, although their presence in the market is essential to prevent market failure.
put options. These two factors are adopted by this dissertation as independent variables for hypothesis testing in Chapter 5: Empirical Results. Later models incorporate the effect of competition from multiple dealers by adding a zero-profit constraint.

Sequential trade models like Glosten and Milgrom (1985) further advanced the argument of transactions themselves in the setting of continuous trading, sending signals of information to the marketplace. In models involving multiple rounds of trading, private information is gradually (instead of instantaneously) incorporated into price. The market maker adjusts his quotes by learning from the order flow, as his quotes converge to the true price as expected value of the asset based on all information. Easley and O’Hara (1987) extended Glosten and Milgrom (1985) in that they allowed trade size to be greater than one, and thus studied the effect of trade size and showed that trade size is a determinant of the equilibrium price. Additionally, they did not assume the existence of new information, but rather first considered the existence of new information before the direction. Through the Bayesian learning process of market makers, they studied how fast price adjusts to full information. Transaction prices in both models formed a martingale, thus supporting the efficient market hypothesis by stating that market price is semi-strong-form efficient (Easley & O’Hara, 1987; Glosten & Milgrom, 1985). The martingale property also indicated that the first differences of transaction prices should be serially uncorrelated. This became the basis for empirical testing in later literature, including but not limited to Glosten and Harris (1988) and Hasbrouck (1988).

Differing from strategic trade models, Kyle (1985) introduced a strategy of the informed trader as a new dimension into the price-setting process. He considered the order placement strategies of both the informed trader and the market maker through sequential auctions. The market maker set the market clearing price on condition of the aggregate order quantities from
one single informed trader and multiple uninformed traders. As trading occurs through multiple rounds of auctions, the market maker updates his beliefs using Bayes Rules. Just like sequential trade models, information is incorporated into price gradually as trading occurs. The true price is still received as an exogenous signal to the informed trader. Both sequential trade models and strategic trade models are consistent with rational expectations because the bid and ask quotes set by the market makers are regret-free, as they believe the price is fair as the trades occur. In this class of models, transaction prices always converge to full information by the end of either one or multiple trading sessions.

Price discovery mechanisms are not limited to continuous market or batch auctions. Price signaling and experimentation are others considered by microstructure literature. Spatt and Srivastava (1991) examined one type of price signaling during an initial public offering (IPO). The authors examined the nonbinding fixed-price mechanism and the limited participation features of a typical IPO, and concluded that they are optimal in maximizing seller revenue (Spatt and Srivastava, 1991). This is because there is informational exchange between the underwriter and potential buyers prior to the offering due to the constraint of limited information for a security that has never been publically traded. Cao, Ghysels, and Hatheway (2000) studied a different type of price signaling in an empirical setting, where NASDAQ market makers signal pricing information through locked and crossed inside quotes (best bid and offer) during the time period before daily market open. These quotations differ from the quotations during continuous trading not only because the identities of the market makers behind the quotes are known, but also because the quotes themselves are noncommittal. Through empirical testing, the authors

\[^{41}\text{Later models following Kyle (1985) incorporate competition, multiple informed traders, and the strategic choice of uninformed traders.}\]
found the directions of these nonbinding inside quotes contributed to price discovery (Cao, Ghysels, & Hatheway, 2000). Also regarding the nature and information content of quotes, O’Hara (2010) questioned the definition of a quote in her article “What Is a Quote?” published in Journal of Trading. Specifically, she raised the question of the impact of modern computerized trading following market structure changes in recent decades. On one side, there are quotes posted on a limit order book that are quotes by legal definition, but don’t represent true trading intention; on the other side, there are also indications of interests (IOIs) that are not quotes by definition, but are real trading intentions. I have covered this set of literature to show that transactions are not the only method of price discovery considered in microstructure theory; it largely depends on the market setup and rules that apply.

Leach and Madhavan (1993) studied the issue of price experimentation by market makers, exploring whether they have the incentive to experiment with price in order to induce more informed order flow; they tested the hypothesis with NYSE specialist data. The answers are different between a specialist system, where the cost of experimentation can be easily recovered by the monopolist advantage, and a competing dealer’s market, where a free-rider’s problem exists. This dissertation examined quote movements that are not price experimentation, and in fact is consistent with the views of Leach and Madhavan (1993), because the NASDAQ market is a hybrid between order-driven and quote-driven. It is far from having a monopolistic specialist, and is more like a competing dealers setup in Leach and Madhavan (1993).

The classical market microstructure models I reviewed included price discovery mechanisms such as continuous market, batch auctions, price signaling, and price experimentation. In different forms, all of these models examine the manner by which new information is incorporated into price—in the case of continuous market, through trading. The
bid-ask spread is modeled through the effects of inventory and information, while straddling the expected true price. Within this paradigm, price discovery occurs through trading while the expected true price does not vary between two adjacent trades. Because the true price (or fair price, fundamental value) is a theoretical construct and is unobservable, when it comes to empirical research the bid-ask midpoint is its most commonly used empirical proxy (Hasbrouck, 2007; Huang & Stoll, 1997). I consider, however, that the expected true price frequently stray outside the boundaries of the bid-ask spread. Empirical evidence (as this dissertation shows) not only conflicts with a basic assumption of theoretical and empirical microstructure research, but also challenges our understanding of the price formation mechanism on a limit order book, and raises doubt about many asset and capital allocation models and investment decisions, which all are based on the assumption that true price has been correctly modeled.

2.3. Knightian Uncertainty

Before introducing the distinction between risk and uncertainty, two related and important concepts are objective and subjective probability. I provide a brief historical context in this section:

The word probability has two distinct meanings in economics and statistics literature, dating as far back as the 1650s. Objective probability refers to the relative frequency of a random outcome in repeated trials, while subjective probability is a measure of a decision maker’s degree of belief in the truth of propositions, although the two terms themselves were not explicitly named and distinguished until relatively recently.42

42 The Emergence of Probability: A Philosophical Study of Early Ideas about Probability, Induction and Statistical Inference (1975) by Hacking gives a historical account of probability theory.
Bernoulli (1738) originally introduces the ideas of utility and expected utility maximizing behavior to counter the St. Petersburg Paradox. In his work, the terms probability and utility are based upon and addressing the objective, not subjective, interpretation of probability. This paradox studied the reasons for individuals’ unwillingness to engage in fair games, which Bernoulli solved by introducing a utility function, an expected utility hypothesis, and the presumption of diminishing marginal utility of money.

Ramsey pioneered the concept of subjective probability in Truth and Probability (1926), one of his three famous essays. He suggests that there is a difference between the definition of probability in physics and in logic. He posits that probability is related to each individual's knowledge, which leads to subjective probability—and subjective probabilities can be inferred by observing actions that reflect individuals' personal beliefs. He suggests a way of deriving a consistent theory of choice under uncertainty that could isolate beliefs from preferences while still maintaining subjective probabilities. Decades later, von Neumann and Morgenstern (1944) recognize Ramsey’s work by mentioning that a theory of subjective probability could be provided—even though they used objective probabilities in formulating the expected utility hypothesis regarding risk aversion, supposing that all the agents had the same probability distribution, as a convenience. It was not until Savage (1954) that a complete theory of decision making under risk with a preference structure emerged. In his theory, Savage combined a

43 The paradox involves a fair coin toss that doubles its payoff each time a head appears. The game has infinite expected value, but few are willing to pay to play, thus constitutes a paradox. Bernoulli provided a solution by introducing the concept of utility and expected utility maximizing behavior, although this solution was not completely satisfactory because the lottery can easily be changed in such a way that the paradox reappears. One only needs to modify the game so it provides an even greater payoff for the paradox to reappear.
44 For more information on Ramsey's biography and work, see Sahlin (2001): http://www.nilsericsahlin.se/the-philosophy-of-f-p-ramsey/.
45 Risk here means no more than the unknown. When referencing Savage's principles, some literature uses the terms risk and uncertainty interchangeably.
personal utility function and a personal probability distribution based on Bayesian probability theory, which reflects the decision maker’s personal beliefs. Subjective probability that is choice-based expresses a decision maker’s beliefs. The key here is to interpret probability as subjective instead of objective.

Similar concepts and arguments on uncertainty are also explored in Keynes’s seminal work, *A Treatise on Probability* (1921): that probability does not have to be quantifiable or even comparable, and numerical probabilities are only special cases of probability. In his example of taking an umbrella in case of rain to illustrate the idea of “irreducible uncertainty,” Keynes (1921) states:

> Is our expectation of rain, when we start out for a walk, always more likely than not, or less likely than not, or as likely as not? I am prepared to argue that on some occasions none of these alternatives hold, and that it will be an arbitrary matter to decide for or against the umbrella. If the barometer is high, but the clouds are black, it is not always rational that one should prevail over the other in our minds, or even that we should balance them, though it will be rational to allow caprice to determine us and to waste no time on the debate. (p. 30)

Keynes (1921) extends the deductive logic of conclusive inference to an inductive logic of inconclusive inference, and argued that there is an objective relationship between knowledge and probabilities, as knowledge is disembodied and not personal. Therefore, subjectivity of probabilities does not matter. The paradigm of expected utility maximizing agents under rational expectations has encountered significant difficulties from empirical tests against asset market data. Among many examples is the high-equity premium puzzle illustrated by Mehra and Prescott (1985). They examine the disparity between returns on stocks versus government bonds, which implied an impossibly high level of risk aversion by the investors (Mehra & Prescott,

---

40 Note the term risk here means no more than unknown. Regarding Savage’s principles, some sources use the words risk and uncertainty interchangeably. At this point in history, even after proposition of the concept by Keynes and Knight, mainstream economic research has not adopted the distinction between risk and uncertainty.
Related work by Weil (1989) demonstrates that this implausibly high level of risk aversion leads to a high level of risk-free rates. Weil (1989) names this phenomenon the “risk-free rate puzzle,” and asked why the risk-free rate is so low if agents are so averse to intertemporal substitution. Additionally, according to the mean-variance framework of portfolio theory, all investors should hold risky portfolios that differ only in their relative percentage between the riskless asset and the market portfolio on the efficient frontier. However, investors’ behaviors have been shown to violate this classical two-fund separation theorem. Only a modest level of exposure to risky securities is often observed for a majority of investors. Investors also show a geographical preference for equities from their home country in international portfolio diversification that cannot be explained by the traditional expected utility maximization paradigm.\footnote{The puzzle has led to extensive theoretical and empirical research in macroeconomics and financial economics. Kocherlakota (1996) and Mehra and Prescott (2003) examine some proposed explanations and conclude the puzzle is real and remains unexplained, but so far there has not been one generally accepted solution.}

Confronted with these compelling empirical evidences that conflict with the existing framework, later research has taken several directions. Knightian uncertainty and its decision theory is one of them, along with behavioral finance (Barberis & Thaler, 2003; Hirshleifer, 2001) and Bayesian learning under incomplete information (Lewellen & Shanken, 2002; Veronesi, 1999, 2000).\footnote{For a comprehensive treatment of empirical puzzles that confronted the traditional expected utility framework, see “Ambiguity in Asset Pricing and Portfolio Choice: A Review of the Literature” (2012) by Guidolin and Rinaldi. For a review of empirical puzzles that faced the asset pricing theory, and other-than-behavioral economics and Knightian uncertainty, see Guidolin and Rinaldi (2012). For a comprehensive survey of behavioral finance, see “A Survey of Behavioral Finance” (2003) by Barberis and Thaler, or Inefficient Markets: An Introduction to Behavioral Finance (2000) by Shleifer.} Knight (1921) distinguishes risk (with known probability distribution) from uncertainty (with unknown probability distribution). He suggests economic returns could be earned for bearing uncertainty but not for bearing risk. According to Knight, risk applies to cases
in which one does not know the outcome of a given situation, but can accurately measure the odds. Uncertainty, on the other hand, applies to situations where one does not have access to all the information he needs in order to set accurate odds in the first place. There is a fundamental distinction between the reward for taking a known risk and the reward for assuming a risk whose probability distribution is not known. A known risk is easily converted into an effective certainty, while true uncertainty is not susceptible to measurement. Ellsberg (1961) use a simple thought experiment to illustrate that economic agents as decision makers overwhelmingly prefer risk situations in which they know of a probability distribution and thus can estimate the odds, rather than uncertainty situations in which there are no distributions, or odds are either difficult or impossible to calculate. This illustration showed Savage’s independence axiom is too strong for reality, because it would be impossible to infer meaningful probabilities from the choices made by decision makers in this paradox (Ellsberg, 1961). In other words, uncertainty is nonprobabilistic. The Ellsberg Paradox implies that a decision maker may consider a set of probability distributions instead of a unique single prior when he does not have enough

50 Similar to the Ellsberg Paradox (1961) is the Allais Paradox (1953), which also confronts implications of Savage’s “sure thing principle.”
51 Ellsberg (1961) Paradox illustrates the behavioral aspect of uncertainty aversion. One simple version of the experiment is: an individual is given an opportunity to bet on the draw of a ball from one of two urns. Urn one has 50 red and 50 black balls. Urn two has 100 balls, which are an unspecified mix of red balls and black balls. First, an individual is offered a choice between two gambles: $1 if the ball drawn from urn one is red and nothing if it is black, or $1 if the ball drawn from urn two is red and nothing if it is black. If the individual chooses the first gamble, and has a prior on urn two, then the predicted probability of red in urn two must be less than 0.5. Next, the individual is offered a choice between two new gambles: $1 if the ball drawn from urn one is black and nothing if it is red, or $1 if the ball drawn from urn two is black and nothing if it is red. If the individual again chooses the first gamble, and has a prior on urn two, then the predicted probability of black in urn two must be less than 0.5. But this cannot be, so an individual making these choices does not act as if they have a single prior on urn two.
52 Of Savage’s (1954) seven postulates, the first postulate (“complete ordering of preferences”) and the second postulate (“sure thing principle” or independence axiom) indicate outcomes that occur—regardless of which actions are chosen—should not affect one’s preferences. In other words, the preference between acts depends solely on the consequences in states in which the payoffs of the two acts being compared are distinct. The consequences are assigned utilities that are independent of the underlying state of the world, and events are assigned probabilities that are independent of acts. For a mathematical treatment and summary of Savage’s seven postulates, see Savage’s Subjective Expected Utility Model (2005) by Karni.
information. It also implied that individuals take into account their own confidence in estimates of subjective probability when making decisions. Later works frequently cite the Ellsberg Paradox as evidence of ambiguity aversion.

Decision theory that distinctly distinguished Knightian uncertainty from risk under utility theory began as early as Choquet (1953). The Choquet integral was initially used in statistical mechanics and potential theory, but found its way into decision theory in the 1980s, where it is used as a way of measuring the expected utility of an uncertain event. Using the Choquet integral to denote the expected utility of belief functions measured with capacities is one way toward reconciling the Ellsberg Paradox. Anscombe and Aumann (1963) provide a fully axiomized theory of decision making under subjective probability. Gilboa and Schmeidler (1989) extend the Choquet expected utility framework and provided a set of multiple-prior preferences, where a decision maker is always able to rank-sort decisions.

Schmeidler (1989) and Gilboa and Schmeidler (1989) make the observation that the probability attached to an uncertain event may not reflect the heuristic amount of information that has led to that particular probability assignment. They reflected ambiguity aversion in their model by using a min operator to represent that the decision maker considers the least favorable probability distribution (Gilboa & Schmeidler, 1989; Schmeidler, 1989). Under this decision rule, uncertainty-averse agents act very cautiously, even pessimistically, and choose the worst-case scenario beliefs from a set of probability distributions. They axiomatized Knightian uncertainty by weakening the independence axiom while keeping other axioms unchanged; they also produced a Bernoulli utility function defined over payoffs, which can induce nonparticipation (Gilboa & Schmeidler, 1989; Schmeidler, 1989). The shortcoming of this approach is that it can only explain pessimistic behaviors, but not optimistic or mixed behaviors.
by construction. This approach has been widely followed and frequently treated as the standard decision rule under Knightian uncertainty. For example, Epstein and Schneider (2008) show that if investors observe negative news about a specific stock, they assume the highest probability in their subjective set, but if investors observe positive news instead, they assume the lowest probability. This is consistent with Hansen and Sargent (2008), who state that “a pessimist thinks that good news is temporary but the bad news will endure” (p. 1).

Bewley (2002) provides a different set of decision-making rules, which induce nonparticipation due to inertia rather than worst estimate. Bewley (2002) removes the completeness assumption from the Anscombe-Aumann (1963) formation of Savage’s theory (Savage, 1954), and introduced an inertia principle stipulating that there is such a thing as the status quo and an alternative is accepted only if it is preferred to the status quo. Additionally, Savage’s completeness assumption (Savage, 1954) states that any two choices can be rank ordered; either one is preferred over the other, or they are equivalent in terms of ordering. Instead, Bewley’s (2002) incomplete preference demonstrates that when a decision maker is confronted with a situation where two choices cannot be rank ordered, he finds the choices incomparable, which is different from equivalent. This choice behavior distinguishes indifference from incomparability.

Empirical evidence of Knightian uncertainty has been rare because of its nonquantifiable and nonmeasurable nature. Williams (2015) is one of the few who have produced such evidence, using the CBOE Volatility Index (VIX)53 as a proxy for Knightian uncertainty to empirically

53 According to the VIX profile on the Bloomberg website (http://www.bloomberg.com/quote/VIX:IND): “The Chicago Board Options Exchange (CBOE) Volatility Index reflects a market estimate of future volatility, based on the weighted average of the implied volatilities for a wide range of strikes. 1st & 2nd month expirations are used until 8 days from expiration, then the 2nd and 3rd are used.”
examine investors’ asymmetric response to earnings announcement news, even though VIX traditionally has been accepted as a risk measure because it derives from the sigma term in the Black–Scholes model (Black & Scholes, 1973). Drechsler (2013) provides the theoretical base that VIX contains an important ambiguity-related component through constructing a general equilibrium framework incorporating time-varying Knightian uncertainty. The model suggests a strong correlation between the variance premium and the level of Knightian uncertainty in predicting returns.

Knightian uncertainty and decision theory not only have established a separate realm within asset pricing theory with great success, but also have become a framework, sometimes combined with econometrics, used by economists to solve problems in many areas like public policy (Caballero & Kurlat, 2008), international finance (Lang, Lins, & Maffett, 2012), and market microstructure theory (Easley & O’Hara, 2010a, 2010b; Routledge & Zin, 2009). This dissertation has applied Knightian uncertainty and its decision theory as a toolbox to examine empirical phenomena from microstructure data, in light of the existing research in this specific area.

A new set of literature has emerged using Knightian uncertainty as a framework to solve empirical microstructure problems following in the footsteps of Easley and O’Hara (2010a). The effects of ambiguity aversion in these models usually involve creating or widening the bid-ask spread, inducing nonparticipation, and reducing liquidity. Easley and O’Hara (2010a) develop a theoretical model using Knightian uncertainty to explain the no-trade situations in the debt market during the 2008 financial crisis. In this model, all traders are heterogeneous, and the

54 For a comprehensive survey of the application of Knightian uncertainty in asset pricing, see “Ambiguity in Asset Pricing and Portfolio Choice: A Review of the Literature” (2012) by Guidolin and Rinaldi.
equilibrium price and holdings are determined by supply and demand. When an ambiguous shock occurs, each trader has a range of beliefs instead of a unique expectation of the mean value of the security. Following the incomplete preference over portfolios and inertia assumptions in Bewley’s (2002) decision rules, Easley and O’Hara (2010a) derive the conditions under which a no-trade situation would occur, and thus determined an equilibrium no-trade bid-ask spread. Although it introduces Knightian uncertainty to microstructure theory—a significant contribution—this model has not been tested against real data, as most quantities in their concluded theorems are unobservable. Also aiming to examine the extreme market outcomes in the mortgage-backed securities or credit derivatives—as do Easley and O’Hara (2010a)—Routledge and Zin (2009) build a theoretical model to explore how ambiguity can increase the bid-ask spread and reduce liquidity even though the underlying environment is stationary. In this model, the market maker’s ambiguity aversion limits his ability to hedge a position, and thus reduces his desire to enter a transaction. The authors followed a worst-case scenario approach with the Gilboa-Schmeidler (1989) framework instead of Bewley’s (2002). Also using the Gilboa-Schmeidler decision rules, Easley and O’Hara (2010b) design an uncertainty model to examine microstructure mechanisms and trading rules. This model differentiated two types of traders: sophisticated traders with rational expectations who maximized expected utility, and unsophisticated traders who were rational as well but faced uncertainty about the payoffs to participate in trading. The unsophisticated traders were ambiguity averse, while both types of traders were heterogeneous. This model examined the market rules and trading practices by different exchanges to analyze ways of improving certain microstructure features to reduce perceived ambiguity by unsophisticated market participants, thus not only generating revenue for
the exchanges, but also reducing execution cost by inducing higher volume through increased participation.

More theoretical work followed this strand of literature to incorporate ambiguity and ambiguity aversion into a traditional microstructure framework. Examples include but are not limited to: Gradojevic & Gençay (2013) who examine the uncertainty embedded in technical trading strategies like market timing and order size; and Easley, O’Hara, and Yang (2013), who investigate the effect of ambiguity about hedge fund investment strategies on asset prices and aggregate welfare. Again, however, due to the challenges of observing and testing Knightian uncertainty in real data, empirical work is extraordinarily rare. This dissertation seeks to contribute to this area.
Chapter 3. Theory

3.1. Easley-O’Hara Model

This dissertation investigates the frequent occurrences of significant quotation price deviations away from the expected true price of the security within extremely short time intervals, although it is important to recognize that my focus was not to determine the value of the true price, but rather to examine the relationship between the expected true price and the market price—or, more specifically, violations of the assumptions made by classical microstructure models on this relationship, regardless of where true price actually is at any given point in time.

This chapter presents a simple theoretical scenario that allows such deviations in the presence of Knightian uncertainty, which would be a violation in a classical microstructure model. This study applied Easley and O’Hara’s (2010a) model, not one of my own creation—although I present the conclusions in their paper that directly relates to my empirical tests.\textsuperscript{55} In this theoretical framework, the true price is defined as the market clearing price in the absence of Knightian uncertainty, as I follow the authors’ notation \( \tilde{v} \) to represent this value. The notional price \( p_1^* \) is used as the market price proxy for true price, as the authors state:

\[
[p_1^*] \text{ This price represents the market’s consensus belief about the asset’s true value. It is the average across the population of the future mean value of the asset minus the risk premium necessary to induce the traders to hold the per capita supply of the asset on average. (Easley & O’Hara, 2010a, p. 8)}
\]

\textsuperscript{55} For a complete representation of this model, see Easley & O’Hara (2010a).
The essence of this model I wish to capture is that neither the quotations—the bid and ask price—nor the midpoint is a suitable representation for the true price in the presence of Knightian uncertainty.

**Model Setup**

There are two types of assets in the economy: a risky asset with price $p_t$ at time $t$, and an uncertain future value; and a risk-free asset: cash, which has the value of 1. The uncertain future value will be realized at the end of period 1 at $\bar{v}$. I assume $\bar{v}$ is normally distributed with variance $\sigma^2$.

There are $I$ traders in the economy, indexed by $i = 1, \ldots, I$. Each has heterogeneous beliefs on the expected mean value of $\bar{v}$—trader $i$ believes the expected mean of $\bar{v}$ is $\bar{v}_i$. I assume at least two traders disagree: $\bar{v}_i \neq \bar{v}_j$. Traders do not try to infer information from prices—in other words, there is no strategy or asymmetric information.

I use $\bar{x}_i$ to denote trader $i$'s endowment of the risky asset, thus the per capita endowment is $\bar{x} = \frac{1}{I} \sum_{i=1}^{I} \bar{x}_i$. There is no cash constraint on any of the $I$ traders. I assume constant absolute risk aversion (CARA) utility of wealth $w$ at the end of $t = 1$. In a standard mean-variance framework, each trader seeks to maximize his expected utility.

Trades may occur at $t = 0$ and $t = 1$.

$t = 0$

At $t = 0$, traders choose portfolios to maximize the expected utility of wealth using their initial beliefs. Trader $i$ chooses portfolio $(x_i, m_i)$, his future wealth (after the first period) will be the random variable $\tilde{w}_i = \bar{v}x_i + m_i$. Distribution of his wealth is normal. His expected utility is given by $(\bar{v}_i - p_0)x_i - \frac{1}{2}\sigma^2 (x_i)^2 + \tilde{w}_i$ with demand for the risky asset $x_i^* = \frac{\bar{v}_i - p_0}{\sigma^2}$. Equating per
capita supply and per capita demand, then solving for equilibrium price, I get $p_0^* = \hat{\theta} - \sigma^2 \bar{x}$, thus the risky asset holding becomes $x_{i0}^* = \frac{\bar{v}_i - \hat{\theta}}{\sigma^2} + \bar{x}$. Price for the risky asset at $t = 0$ is the average mean value reduced by a factor that compensates the traders for holding the market risk.

The purpose of the $t = 0$ period is to generate endogenously heterogeneous portfolios of the risky asset and cash. At the next time period $t = 1$, I consider two scenarios: where price moves due to the presence of risk and where price moves due to the presence of uncertainty. Asset payoffs are realized after $t = 1$.

**$t = 1$: Change in the market price of risk**

At $t = 1$, trader $i$’s expectation of the mean value of the risky asset declines by $(1 - \alpha)$ and becomes $\bar{v}_{i1} = \alpha \bar{v}_i$ where $\alpha < 1$. In the presence of risk, each trader’s coefficient of variation of the risky asset’s value increases as the price of the risky asset falls, because each trader considers the asset to be more risky now. Under the same mean variance framework, the new equilibrium price and holdings are $p_1^* = \alpha \hat{\theta} - \sigma^2 \bar{x}$ and $x_{i1}^* = \alpha \left( \frac{\bar{v}_i - \hat{\theta}}{\sigma^2} \right) + \bar{x}$.

**$t = 1$: Introduce the presence of uncertainty**

At $t = 1$ in the alternative scenario, traders are presented with an uncertainty shock instead of a risk shock. Instead of considering that the mean expected value of the risky asset declines by $(1 - \alpha)$, now $\alpha$ can be a set of values within the range of $\alpha \in [\underline{\alpha}, \bar{\alpha}]$ without a known distribution. Traders now make decisions over trades rather than over final portfolios. Following Bewley’s (2002) incomplete preferences and inertia assumption over portfolios, a trader only moves away from the status quo if—and only if—his move (a trade) is expected utility improving for every belief in the set of beliefs that represent the trader’s preferences; one
portfolio is preferred to another if—and only if—it yields greater expected utility for every belief in the set of beliefs that represent the trader’s preferences.

A given trader chooses a non-zero trade if—and only if:

1) Trader is better than status quo:
   \[ Utility_{\text{trade}} \geq Utility_{\text{no trade}} \]

2) This trade is not dominated by another trade (no alternative is better):
   \[ Utility_{\text{trade}} \geq Utility_{\text{trade'}} \]

Under this set of assumptions, Easley and O’Hara (2010a) derive the following items: Theorem 1, the condition under which a no-trade equilibrium occurs: the lowest possible price at which any trader is willing to sell is higher than the highest possible price at which any trader is willing to buy; Proposition 2 derives the bid price, ask price, and the bid-ask spread under the no-trade equilibrium: the ask price is set by the least optimistic trader’s best estimate, while the bid price is set by the most optimistic trader’s worst estimate; finally, Proposition 3, which is most pertinent to this dissertation: the relative relationship between the notional price and the bid-ask quotes. Proposition 3 states that the bid price is less than the notional price of the asset \( p_1^* \geq \text{bid} \) if—and only if—\( p_o^* - (1 - \alpha) \hat{\theta} \geq p_o^* - (1 - \alpha) \min_i \{ \hat{v}_{i0} \} \). Similarly, the ask price is above the notional price of the asset \( \text{ask} \geq p_i^* \), if—and only if—\( p_o^* - (1 - \alpha) \hat{\theta} \geq p_o^* - (1 - \alpha) \max_i \{ \hat{v}_{i0} \} \). During discussion of Proposition 3, Easley and O’Hara (2010a) discuss the scenario where the notional price can lie outside the bid-ask interval by either being below the bid or above the ask. If below the bid, the condition can be satisfied if prior beliefs are very dispersed and the notional decline in future value is large. This is due to price-setting shifts from the beliefs of the average trader to the beliefs of the most optimistic trader, thus posting an
upward bias to the bid price. Vice versa, if above the ask, the condition can be satisfied if the effect of diversity of opinions is greater than the notional decline in future value.

3.2. Discussion

This model falls under the category of microstructure models because it discusses the price-setting process and, more specifically, the impact of Knightian uncertainty on the bid and ask price in the absence of trading. The price-setting mechanism indeed influences asset value. However, this model does not treat true price in the manner of classical microstructure model I reviewed in Chapter 2: Literature Review. Kyle (1985), Glosten and Milgrom (1985), and many others consider the expected true price a signal exogenous to the model. Additionally, in this model, the bid-ask spread arises for a totally different reason. As it will become clear in the Chapter 4: Data and Methodology, this is the reason I selected this model to characterize the empirical evidence presented in this dissertation: it is consistent with the predictions made in this model but not that of microstructure models. In Easley and O’Hara (2010a), the true price was defined as the market clearing price before the uncertainty shock. Once uncertainty is present, it becomes unclear (or impossible to identify) where the expected true price actually lies, because it is a set of distributions instead of a specific distribution, which has a range of expected mean values instead of one expectation. The authors instead used the notional price as a proxy for the true price, which they define as “the average across the population of the future mean value of the asset minus the risk premium necessary to induce the traders to hold the per capita supply of the asset on average,” because “this price represents the market’s consensus belief about the asset’s true value” (Easley & O’Hara, 2010a, p. 8).
The central argument of Easley & O’Hara (2010a) is that fair value accounting and true price assessment become ambiguous in the absence of Knightian uncertainty. None among the bid price, the ask price, or the bid-ask midpoint are appropriate metrics. My view is that the notional price from scenario 1 without uncertainty was used as a proxy or reference to assess the degree of bias imposed on the market price due to uncertainty, but should not be considered the true price definition, nor is it an appropriate representation of the expected true price. The purpose of having this benchmark was so Easley & O’Hara (2010a) could compare two pricing scenarios of the risky asset (or two equivalent securities): one only impacted by risk, the other impacted by both risk and uncertainty. This dissertation provides empirical evidence of the expected true price sits outside the bid-ask spread. Or more accurately speaking, this dissertation provides evidence of cases in which—if we follow the definition of true price from classical microstructure models—true price would have been sitting outside the bid-ask spread. The identification of this case is regardless of where the true price is. This is because, as this model has demonstrated, the true price could be anywhere within a range, and it is impossible to locate exactly. For the purposes of this study I have not attempted to determine how much true price is, but rather to provide empirical evidence of Knightian uncertainty (or, more accurately speaking, to provide empirical evidence that is consistent with the existence of Knightian uncertainty, because its existence cannot be proved).
Chapter 4. Data and Methodology

4.1. Institutional Details

Different types of market structure have been established throughout the history of financial markets to facilitate the price discovery process. A quote-driven market operates through market makers (also frequently referred to as dealers or specialists), who buy from sellers and sell to buyers. By providing the intermediary service of connecting other market participants through time and/or space, market makers attempt to make the bid-ask spread as their profit, which at the same time is transaction cost to other market participants who use market orders for immediacy. As trading interest and liquidity grow, many active financial products have moved away from this type of market structure, although more customized instruments with low levels of liquidity and/or large notional/face value (e.g., fixed-income securities and complex derivatives) are still primarily bought and sold through dealers. More standardized securities and instruments like stocks, futures, and options are traded on order-driven platforms, which usually operate by maintaining an electronic limit order book. Modern order-driven markets may or may not have designated market makers. Any trader can post, buy, or sell limit orders to the book, or transact through marketable orders against the prevailing best bid or offer. Modern markets like the NASDAQ Stock Market are of a hybrid nature, which offers an electronic limit order book while at the same time having one or more registered market makers for each security as part of the listing service.

56 Marketable orders in this context include both true market orders and marketable limit orders, which are marketable until the limit price is reached. The NASDAQ INET matching engine only accepts/processes buy/sell orders that come with a price. So technically all market orders are marketable limits.
Technological advancements, regulatory changes, and competition among exchanges and among market participants have led to tremendous changes in the U.S. equity market structure in the past few decades. The market share of the NYSE on its own listed stocks dropped from 78.9% in February 2005 to 20.1% in February 2014 (U.S. Securities and Exchange Commission [SEC], 2015). As of April 2015, trading in U.S. stocks is dispersed among 11 registered exchanges dominated by three corporate groups (Intercontinental Exchange, Inc. [ICE], NASDAQ OMX Group, and BATS) and many dark venues that do not contribute to pre-trade transparency on the National Market System. Transaction volume and quote message traffic have grown substantially, average trade size is substantially smaller, and competition among exchanges has intensified in the past 25 years (Angel, Harris, & Spatt, 2011). In this section, I discuss a few market structure issues most relevant to the empirical analysis of this dissertation: exchange fee structure, market maker obligations, and high-frequency trading (HFT) market making strategies. The goal is not to ascribe causal relationships between any or a combination of these factors and the empirical observations I call crossing later in this dissertation, but rather to put this empirical anomaly (one that seems to be in violation of classical theories) in the context of modern markets. I consider that evolutions in market structure, not limited to the ones discussed in this dissertation, have contributed to significant changes to the price-setting

---

57 The Securities and Exchange Commission put out a concept release on equity market structure in 2010 that provides an overview of the current market structure of the U.S. equity market and solicited public opinions on market structure issues including market quality, high-frequency trading, and undisplayed liquidity.
60 The Chicago Stock Exchange (CHX) is the only registered stock exchange not operated by one of these three corporate groups.
61 Dark venues here refer to dark pool alternative trading systems (ATSs) and internalization by broker dealers.
mechanism and price discovery functions in the modern era. As such, classical models require revisiting, questioning, and revising.

Many exchanges in the United States provide a rebate to each passive order executed on their platform in order to attract liquidity providers, while at the same time charging an access fee for traders who remove liquidity by executing through market orders. This fee structure is frequently referred to as the maker-taker model.\(^2\) The exact charge and rebate amount vary across exchange platforms, but is capped at 30 cents per 100 shares traded by the SEC. Regulators, academics, and market practitioners have criticized this fee structure for distorting the price discovery process.\(^3\) In fact, NASDAQ OMX Group tested the effect of this fee structure in a pilot program by reducing the maker/taker fees from 29 cents/30 cents to 4 cents/5 cents per 100 shares on 14 active stocks for 18 trading days in February, 2015. While a comprehensive report to assess the impact of the pilot study is still in progress, some early analysis by Pearson (2015) shows that NASDAQ lost market share on the pilot stocks. According to Pearson (2015), one group that responded to this fee structure change includes the HFT firms who favor higher rebates due to their strategy of posting limit orders. This is not to say market-making strategies solely rely on exchange rebates to stay profitable, as there is the presence of market makers on exchange platforms with inverted pricing models,\(^4\) as well as the futures market like the CME GLOBEX where the fee structure is very different. Nevertheless,

\(^2\) The fee structure was first invented in 1997 by Joshua Levine, creator of the Island ECN (electronic communications network), and later became popular on other platforms. The platform was later acquired by Instinet, then subsequently acquired by NASDAQ in 2005. This fee structure is not to be confused with payment for order flow, which is the exchange or off-exchange market centers pay brokers to route their client’s order flow to their platform.


\(^4\) Inverted pricing platforms include NASDAQ BX, EDGA, and BATS BYX (Y).
the maker-taker model in the rather dominant platforms favors executions with limit orders over market orders. Additionally, exchanges charge participants by volume executed, not number of messages. These two features of exchange pricing models, combined with competition between market makers, encourage trading strategies that experiment with limit orders in the presence of uncertainty, consistent with the empirical observations made by this dissertation.

Another market structure issue that relates to this study is market maker obligations. Firms could engage in market-making strategies by primarily executing through limit orders and profiting from the bid-ask spread without registering as market makers with the exchanges. However, there are benefits to registering with an exchange, such as exemption to naked short selling rules while engaging in bona fide market-making activities. Once registered, market makers are to maintain continuous two sides of the quote at all times in the securities for which they are registered market makers (“Equity Regulatory Alert,” 2011). Some market makers used to fulfill their responsibility for providing active quotations at all times by placing stub quotes. After the Flash Crash on May 6, 2010, this practice was heavily criticized. Subsequently, the SEC approved proposals by the exchanges and the Financial Industry Regulatory Authority (FINRA) to ban stub quotes (SEC, 2010b). Such responsibility for registered market makers to maintain two-sided continuous quotes within 8% of the national best bid and offer (NBBO) ensures there are available quotes at all times; however, it does not necessarily encourage quote

66 FINRA is a non-government private regulatory body that is a registered self-regulatory organization (SRO), which regulates its member exchanges and broker-dealers.
67 According to the SEC (2010b), “a stub quote is an offer to buy or sell a stock at a price so far away from the prevailing market that it is not intended to be executed, such as an order to buy at a penny or an offer to sell at $100,000. A market maker may enter stub quotes to nominally comply with its obligation to maintain a two-sided quotation at those times when it does not wish to actively provide liquidity. Executions against stub quotes represented a significant proportion of the trades that were executed at extreme prices on May 6, and subsequently broken.”
movements examined in this study. For most securities traded on the National Market System, 8% of the NBBO is still a relatively wide range compared to the tick size.68

The final (but not less important) market structure issue I consider in this section is trading strategies by modern market makers on an electronic limit order book. As has been discussed in this section and in Chapter 2: Literature Review, market makers post limit orders on the limit order book, buy at the bid, sell at the ask, and make the bid-ask spread as their profit. This used to be the strategy of the specialists on the NYSE floor. With the increasing use of computerized trading and information processing technology in the financial industry, this activity has shifted to be conducted by algorithms instead of human traders. The specialists and floor traders used to be able to observe each other on a trading floor without necessarily clearly spoken instructions for order placements. This type of observation and information inference is not possible on a fully anonymized electronic platform. Instead, market participants use software technology like machine learning and artificial intelligence to infer information or intent of others by observing the order flows. Understanding that their own orders are also being observed, market participants can potentially post orders to sniff out further information, or post pretentious orders they have no intent to execute, which qualifies as spoofing, a type of market manipulation. The Dodd-Frank Act of 2010 specifically outlawed this tactic, although there is still a large middle ground where firms can post limit orders as part of their strategy to induce further information because the intent to cancel while posting a limit order is difficult to prove. Even a bona fide limit order placed on the limit order book and subsequently canceled without

68 Tick size is the smallest possible movement allowed for a stock’s price. In the United States, the tick size has been reducing over the past decades. Currently, the tick size is 1 cent for stocks priced above $1, and 1/100 of 1 cent for stocks priced below $1. In many European equity markets like the London Stock Exchange, the tick size follows a dynamic schedule that is dependent on the stock price.
executing is still information for the firm posting the order and all others observing that no one was willing to execute that quantity at that price level at that particular point in time.

The purpose of the above discussion is not to attribute the occurrence of the crossing phenomenon characterized in this dissertation to HFT or a specific exchange fee structure, but rather to point out that the U.S. equity market is constantly evolving in pace with new technology and strategies. Issues like HFT, off-exchange trading, and high-frequency market-making strategies are being debated in a growing set of new literature, but are far from having definitive conclusions. Any one or a combination of these issues could be contributing factors (even if not the deciding factor) to empirical observations made using equity market data. The intent of this dissertation is to offer new ideas and pose questions regarding existing models, and ultimately inspire further studies on this subject.

4.2. Data

4.2.1. Data Description

The sample dataset consists of quotation prices from the top of the NASDAQ limit order book—the NASDAQ best bids and offers (QBBOs) on 1,671 NASDAQ-listed common stocks, and trades from all market centers—including both NASDAQ and other market centers that report to the UTP Securities Information Processor (SIP). The sample period is between

69 The original data collection from NASDAQ was made on May 16, 2013, and contains all NASDAQ-listed securities at that time, which has more than 1,671 ticker symbols. As will be explained in this section, in selecting the sample stocks for this study, I exclude non-common stocks; stocks that traded at least once below $1 during the five-day sample period; stocks that never traded during the sample period; stocks that did not have quotations posted on at least one of the five-day sample period; test symbols; and ticker symbols that did not have descriptive statistics available on Bloomberg and/or IBES database, which are necessary for regression analysis later.

70 UTP stands for Unlisted Trading Privileges—an outdated term from a period when NASDAQ primarily traded over-the-counter (OTC) securities that are not listed on a national exchange. Nowadays, the term UTP represents an electronic network that collects and disseminates quote and trade information on NASDAQ-listed securities, which are also frequently referred to as Tape C stocks.
Monday, November 5, 2012, and Friday, November 9, 2012.\textsuperscript{71} I use quotes only from NASDAQ to avoid synchronization issues introduced by potentially different time delays from different market centers reporting to the national consolidated tape. According to the quarterly metrics reported by the UTP Plan, the average quote/trade latency is less than two milliseconds as of the fourth quarter of 2012 (NASDAQ OMX Group, 2013). A delay differential at the millisecond level could pose a problem for this study, because more than one transaction could occur within a millisecond for active stocks. Using quotes from all market centers to construct the NBBOs could erroneously count more crossing occurrences during certain no-trade periods, while less for other periods introduced by the latency in data transmission. Additionally, since QBBOs take part in forming the consolidated data feed composing the NBBOs, I take a conservative approach in designing this study in the sense that quote deviations I observe from QBBOs definitely occur on NBBOs, because the NBBO spread is always tighter than the QBBO spread.\textsuperscript{72} This inference is not true the other way around; therefore, positive conclusions still hold if quotes from all market centers are used.\textsuperscript{73} In contrary to the quote records, I use trade observations from all market centers instead of only NASDAQ in order to control for the no-trade time windows to be strictly no trading, thus no public news or private information has been incorporated into price through trading.

\textsuperscript{71} I have chosen one week in early November as sample period to avoid the summer and the year-end holidays, to avoid a particularly turbulent market period with significant events, and to limit the observation window to a relatively short time period to avoid seasonality changes and longer-term market structure shifts.

\textsuperscript{72} In theory, QBBOs or BBOs of any particular market center should always be wider than the NBBO. In the case of NASDAQ listed securities, the QBBO and the NBBO are usually identical.

\textsuperscript{73} Market operators, including NASDAQ, occasionally drop trade and quote records due to buffering, which could possibly introduce a bias into this study. However, as this buffering occurs much more often to quote records than to trade records, the bias is in favor of finding positive conclusions in this study.
In addition to the NASDAQ dataset, I also collect additional stock-specific information, including market cap, analyst coverage, and shares outstanding from Bloomberg and the Institutional Brokers’ Estimate System (IBES) from the Wharton Research Data Services (WRDS) database. I further exclude stocks for which Bloomberg or IBES do not have information on the variables necessary for regression analysis later. The selection procedure eliminates many stocks with few or no activities during the sample period. While this procedure potentially presents bias against less active stocks, the selection procedure is appropriate because the constructed metrics in this study are not possible or meaningful if the quotes barely moved during the trading day. I also stratify the full sample by market capitalization to control for structural shifts: the mechanisms for quote movements are likely to be different between large-cap stocks, mid-cap stocks, and small-cap stocks. Market cap is collected from Bloomberg for each day of the sample period, and then averaged across the five trading days. The large-cap stratum includes 456 stocks with market cap greater than one billion dollars. The mid-cap stratum includes 963 stocks with market cap between 100 million dollars up to one billion dollars. The small-cap stratum includes 252 stocks with market cap less than 100 million dollars. It is also common in empirical studies to divide the full sample into equal-sized strata by market cap or other stratification measures. I have chosen to stratify in this manner because the quotation and trading mechanism might differ once the market cap passes a certain threshold, such as one billion dollars, because of the type of traders actively participating and their trading strategies. Additionally, certain index funds, mutual funds, and hedge funds select stocks to invest in once they pass a certain market cap threshold. Due to these reasons, this stratification

74 Volume and trade counts are other commonly used measures for stratification in empirical microstructure studies. Because the observation windows in this study are restricted to periods of no trade, any trade measures like volume are not exogenous. Therefore, I have chosen market cap as the stratification measure.
method is more likely to capture structural shifts. However, this method introduces the issue of different normalization scales for metrics that are cumulative in nature and are proportional to the number of stocks in the stratum in time series metrics. I address this issue by dividing these metrics by the number of stocks within each stratum to make them comparable across categories.

Records during pre-open periods and after-hour trading are excluded from the sample observation window, except for the stocks for which two-sided quotes are not immediately available on the limit order book. In other words, these stocks do not have a quote update at the exact moment when the market opens at 9:30. In this case, the best bid, or the best offer, or both, from pre-open periods carry through. I retain the price value of those quotes before dropping the quote and trade records before 9:30 a.m. There is no locked or crossed quote in the dataset.\(^75\) Additionally, each trading day is divided into three segments: 9:30 a.m. to 11:00 a.m., 11:00 a.m. to 3:00 p.m., and 3:00 p.m. to 4:00 p.m. Similar to structural shifts across strata by market cap, significant differences also exist between the three segments of the trading day. For all calculated metrics to prepare for regression analysis I use quote and trade observations only during the midday section (11:00 a.m. to 3:00 p.m.) in order to exclude effects from the time periods close to market open and close where price is more volatile, as demonstrated in Figures 5 through 10.

4.2.2. Summary Statistics

Table 5 provides summary statistics for the full sample and its each stratum for this study. Market capitalization (Avg Mkt Cap) is collected from Bloomberg based on the closing price of each trading day, then averaged across the sample period of five days. Market capitalization in

\(^75\) Locked quotes refer to situations when the best bid price is the same as the best ask price at a given moment. Crossed quotes refer to situations when the best bid price is higher than the best ask price. Both locked quotes and crossed quotes occur in the consolidated tape across all market centers, sometimes due to latency, sometimes due to non-routable limit orders. They usually resolve themselves very quickly, especially locked quotes due to arbitrage. They should not occur on a single limit order book like NASDAQ.
millions of dollars ranges from $10.70 million to $528,722.06 million, with the median of large-, mid-, and small-cap stratum being $2,420.06 million, $307.96 million, and $57.37 million, respectively. Among the three strata, the larger the market cap, the higher the stock price. The median large-cap stock price is $32.49, compared to $11.42 for mid-cap stocks and $3.39 for small-cap stocks. Price, volume, trade count, turnover, quoted spread, and relative quoted spread are calculated from the trade and quote dataset from 11:00 a.m. to 3:00 p.m. during each trading day, and averaged across five sample days. Price (Avg Price) is the average transaction price in U.S. dollars. Volume is the average number of shares traded. Trade count (Num Trades) is the average number of trades for the day—by number of transactions, not number of shares. Turnover is the average dollar volume traded in U.S. dollars. These three volume metrics are lower compared to the equivalent metrics collected from Bloomberg for the entire trading day because only observations between 11:00 a.m. and 3:00 p.m. are taken into consideration. They are used later for regression analysis on crossing metrics (frequency and magnitude) for the midday period. The same goes for other metrics labeled as calculated in all tables. Metrics on average volume per day range from 27,923,628 in number of shares (NASDAQ:SIRI), 66,765 in number of trades (NASDAQ:MSFT), and $6,201,902,525 in dollar amount (NASDAQ:AAPL) for the most active stocks, to 350 in number of shares (NASDAQ:TORM), three in number of trades (NASDAQ:CASH), and $1,953 in dollar amount (NASDAQ:RGDX) for the least active stocks. The most active stocks by each of the three volume measures are all in the large-cap category; the least active stocks by number of shares and dollar amount are in the small-cap category, but by number of trades are in the mid-cap category. Median share volume is 363,084 for large-cap stocks, 50,276 for mid-cap stocks, and 13,302 for small-cap stocks; median number of trade count is 2,522 for large-cap stocks, 307 for mid-cap stocks, and 46 for small-cap stocks;
median dollar volume is $10,757,748 for large-cap stocks, $568,234 for mid-cap stocks, and $52,373 for small-cap stocks. Quoted spread is the duration-weighted average difference between the QBBO measured in U.S. dollars. Individual observations of the bid-ask spread above $2 are eliminated in order to control for significant outliers. I also eliminate one-sided quote records where either the bid price or the ask price is missing. Relative quoted spread is calculated on a quote-by-quote basis using the quoted spread as percentage of the bid-ask midpoint before taking the average for the trading day. The bid-ask midpoint is rounded to the nearest one-tenth of a cent to avoid floating point issues. Quoted spread ranges from $0.01 to $1.35 for the full sample, with standard deviation of $0.14. Median quoted spreads are larger for small-cap stocks ($0.11) than mid-cap stocks ($0.06), which is larger than large-cap stocks ($0.03). This inverse relationship exists in the median price in each of the three strata; thus, median in relative quoted spread shows an even stronger difference of 0.12% in large-cap stocks, 0.52% in mid-cap stocks, and 2.81% in small-cap stocks.

Table 6 provides summary statistics on variables of interest. All metrics are calculated or collected on a daily basis, then averaged across days from November 5, 2012, to November 9, 2012. As with Table 5, the calculated metrics in Table 6 are based on trade and quote records between 11:00 a.m. and 3:00 p.m. during each trading day in order to keep consistency with regression analysis later. QBBO Updates is the cumulative number of updates on either the bid side or the ask side of the QBBO for each stock. Quantities are not considered in this calculation because they are not part of later analysis: the study concerns the relative price levels; thus, subsequent quantity updates are consolidated into the first observation with the same bid and ask prices. The number of QBBO updates ranges from two to 61,218 from 11:00 a.m. to 3:00 p.m. of the trading day, with a standard deviation of 2,266. The median number of QBBO updates is
2,342 for large-cap stocks, 682 for mid-cap stocks, and 158 for small-cap stocks. As expected, the larger the market cap, the more active the stock, and the more often quote updates occur. No-trade duration is calculated using the trade observations, as the average length of time (in number of seconds) for which no transactions occurred—in other words, the passage of time from one trade to the next adjacent trade. If multiple trades occur within the same millisecond, I consolidate these records into one trade of a larger trade size so the zero duration would not bias this calculation. It ranges from one millisecond (shown as zero seconds in Table 6 to only display the integer part) for each of the market cap strata, to 7,993 seconds (approximately 2.22 hours), with a standard deviation of 770 seconds (approximately 12.8 minutes). The median no-trade duration is three seconds for large-cap stocks, 27 seconds for mid-cap stocks, and 528 seconds for small-cap stocks. Compared to the number of trade count metrics, these durations are within a reasonable order of magnitude. The less active the stock, the longer it takes for the next trade to occur. No-trade duration and trade count are not the exact inverse of each other because I set the lower bound of one millisecond for no-trade duration and consolidated all trades occurring within the same millisecond. I gather the number of analyst recommendations (Num Analyst Rec) from the IBES database on a stock-by-stock basis. Each market cap category has stocks with zero or only one analyst coverage. The maximum number of coverage of 62 is Apple Inc. (NASDAQ:AAPL) in the large-cap group. Median number of analyst recommendations is 16 for large-cap stocks, six for mid-cap stocks, and two for small-cap stocks. As expected, more active stocks with higher market capitalization receive more analyst coverage and utilitarian trading interest from the general public. Alongside number of analyst recommendations, I also

76 Summary Recommendations contains one record for each forecast period for each Thomson Reuters statistical period. The forecast period is the fiscal year for which the recommendations were made, while the Thomson Reuters statistical (stat) period is the date when the set of summary statistics was calculated.
collect the standard deviation between the forecast earnings per share (EPS) by different analyst (Analyst StdDev) reports from the IBES database. This represents the diversity of opinions among security analysts on the value of the stock. This number ranges from zero to 2.12, with the median being 0.83 for large-cap stocks, 0.72 for mid-cap stocks, and 0.45 for small-cap stocks. This disagreement measure is the largest for the most active stock and trends less for less active stocks because a sufficient number of analyst coverage is necessary for disagreement to exist. For the 109 stocks within the 1,671 full stock sample that have zero or only one analyst covering, the standard deviation would clearly be zero. As will be discussed in the next chapter, this diversity of opinion is not Knightian uncertainty or its proxy. Float/outstanding is the number of common stock shares float divided by the number of stock shares outstanding for the stock; I collect both numbers from Bloomberg and calculate the ratio by dividing the two. The number of shares float comprises the stock shares available for trade by the general public. The rest of the shares outstanding are closely held by insiders, institutions, major stakeholders or employees, and the restricted shares within the lock-in period. I use this metric as a proxy for the likelihood that an execution was initiated by an insider. All market cap strata have stocks with less than 20% out floating, with the minimum of a mid-cap stock Crescent Financials (NASDAQ:CRFN) with only 7.74% of outstanding shares floating. Median float/outstanding ratio is 95.47% for large-cap stocks, 87.23% for mid-cap stocks, and 80.02% for small-cap

---

77 There are nine anomalies in this field—BIOC, CTIB, DUSA, EDGW, PENX, PXLW, RMCF, SYNL, and TACT—where there is zero or one analyst covering the stock but the standard deviation between EPS forecasts is not zero, but rather takes on the value of 0.5, 0.71, and 1.41. I suspect it is because during the sample period, the number of analysts covering the stock had changed, and this standard deviation reflects that change but the coverage number only reflects the status quo at the end of each quarter when the database updates. I have chosen not to make changes to a third-party source so that the treatment for this field is consistent across stocks.

78 Crescent Financials (NASDAQ:CRFN) merged with Vantage South BancShares (NYSE:VSB), then merged with Yadkin Financial (NYSE:YDKN) in 2014. The ticker symbols CRFN and VSB are no longer traded.
stocks. This indicates more active stocks have less insider holding as percentage of shares outstanding, and trading interest are more likely to be from the general public.

4.3. Constructed Metrics

4.3.1. Examples and Definitions

Chapter 3: Theory presents a scenario under which in presence of Knightian uncertainty, it is possible for the expected true price to stray outside the market-wide best bid and offer interval during periods of no trade. Market participants are unwilling to either buy or sell at a wide range of prices under this condition. In this chapter, I use equity market data provided by NASDAQ, Inc. to illustrate a process I have designed to identify characteristic events when the expected true price is outside the bid-ask interval. I construct and define new metrics to quantify these events both in frequency and in magnitude. I form testable hypotheses to investigate the cause of these events, and consider whether empirical evidence is consistent with the theoretical prediction of the presence of Knightian uncertainty when the expected true price is clearly outside the bid-ask interval during these periods of no trade.

Since the research interest of this dissertation focuses on the movements of quotation prices during periods of no trade, I interleave trade observations with quote observations throughout the trading day. The combined dataset uses trade observations to separate quote observations into groups. Within each group, I compare relative price levels of the best bid and offer. This step also builds in an initial level of control to limit the observation spectrum to periods of “relative quietness” where there is no trading, and thus no significant news arrival to the marketplace; in other words, the expected true price should not move much—if at all—during the short time interval between adjacent trades. In the four groups of tables and figures in
the Appendix A (Table 1, Table 2, Table 3, Table 4, and Figure 1, Figure 2, Figure 3 and Figure 4), I provide concrete illustrations on the type of quote movements examined in this study. As an example, I use the common stock for Starbucks Corporation (NASDAQ:SBUX)\textsuperscript{79} soon after market opens on trading day November 5, 2012. Because SBUX is a large-cap stock that is relatively active, the observations are made in extremely short time windows. Shown in Table 1 are the first three quotes at 9:30 a.m., immediately following market open before the first transaction of the day occurs. It is worth noting that the following examples are chosen only to illustrate the concept and methodology of metric design. In empirical calculation and analysis in Chapter 5: Empirical Results, only quote and trade records between 11:00 a.m. and 3:00 p.m. are used to avoid structural differences during time periods soon after the market opens and before the market closes. As the readers will see, the quote prices move more dramatically immediately following market open.

In the events illustrated in Table 1 and Figure 1,\textsuperscript{80} no trade occurs and the expected true price should—in theory—stay the same during the 180 millisecond interval. Market makers can still revise bid and ask prices for accumulated inventory in related stocks or other reasons; however, the best bid and ask should still straddle the expected true price despite inventory adjustments, otherwise arbitrage should occur. In this example, the expected true price—in theory—should be between the best bid $50.80 and ask $51.20 at 9:30:00.000 at the first observation, between the best bid $50.72 and ask $51.20 at 9:30:00.180 at the second observation, and between the best bid $50.77 and ask $50.80 at 9:30:00.180 at the third

\textsuperscript{79} I have selected SBUX as the example because it is a relatively well known and active stock with which readers are likely familiar. The stock price was around $50 to $55 in November 2012, with a market capitalization of $38 billion.

\textsuperscript{80} The horizontal axes in Figure 1 through Figure 4 reflect event space not time space in order to illustrate quote movements, and do not correctly reflect the time lapse between the data points.
observation. However, less than a millisecond later in the third observation, posted quotation prices shift, with both best bid and best ask prices below $50.80. If either of the previous midpoint prices $51.00 or $50.96 were an accurate proxy for the expected true price less than a fraction of a second later, the best bid and ask would not straddle the true price anymore. This is a clear example among many where the true price moves (or stays) outside the range of at least one—but more likely many—sets of bid-ask intervals. This is a mechanism I name *neighbor cross*, which will be defined mathematically later in this section.

To give another example in a different type of price movement mechanism (as will be named *strict cross* later in this section), I provide Table 2 for SBUX’s quotes on the same day less than 10 minutes later. In this example, I do not use the bid-ask midpoint as a proxy for the expected true price to demonstrate the violation of microstructure assumption in a more strict sense—the best bid and ask were at $51.15/$51.18 at 9:39:46.710, and at $51.15/$51.17 at 9:39:47.100, but at $51.13/$51.14 at 9:39:47.775, which was below the lower bound of the bid-ask interval of $51.15 at 9:39:46.710 and 9:39:47.100. Regardless of where true price had been during this period of time, it must have been outside the spread boundaries at one of these three instances as long as it had not moved.

Searching through the full sample of 1,671 NASDAQ-listed securities across five trading days in a week between every given two sets of quotes \([bid_t, ask_t]\) and \([bid_{t'}, ask_{t'}]\) when no trade occurs between \(t\) and \(t'\), I find an overwhelming number of instances where either \(bid_{t'} > ask_t\) or \(ask_{t'} < bid_t\) within a brief time window with no trade occurring in between. When I use the bid-ask midpoint as a proxy for the expected true price (\(\text{midpoint}_{t,t'} = \frac{(bid_{t'}+ask_{t'})}{2}\)) as in the first example, occurrences of \(\text{midpoint}_{t'} > ask_t\) or \(\text{midpoint}_{t'} < bid_t\)
are violations as well. In this case, either the midpoint is not an accurate proxy, or the expected true price has been sitting outside the bid-ask interval. I refer to both versions of violations as crossing to note that the expected true price at one time crossed outside the range of the bid-ask spread of another time, while no news or trades occurred in between. I name the first type of violation (namely, without using a proxy) the strict version; while the second type (using the midpoint as a proxy for the expected true price) is the relaxed version. This crossing I refer to here is not to be confused with the “lock and cross” in equity market quotation data, which can happen for a combination of several reasons, including non-routable orders, access fees, and data feed delays between different market centers and the Securities Information Processor (SIP). In the calculations of this dissertation, I have also chosen not to include the equal signs in these comparisons in order to exclude the possibility that the true price is sitting exactly at \( bid_t \) or \( ask_t \). (Namely, \( \text{midpoint}_{t'} > ask_t \) or \( \text{midpoint}_{t'} < bid_t \) instead of \( \text{midpoint}_{t'} \geq ask_t \) or \( \text{midpoint}_{t'} \leq bid_t \)). The counterargument against this design is that if the bid at one point in time is the same as the ask at another point in time without trading in between, there should have been a violation because these two quotes should have traded against each other. I have chosen the former approach because it is more conservative in criteria for what qualifies as a crossing. Choosing the more conservative methodology makes potential positive conclusions (if any) stronger.

In order to characterize and accurately define these type of crossing events for a given stock \( i \) within a given no-trade period that expands across time window \( \Delta t \), I first denote \( freq_{i,t} \) to quantify the frequency of crossing occurrences for stock \( i \)’s quote record at time \( t \), which starts from zero, and increases by one each time this particular quote record crosses with another quote record (or the midpoint of another quote record if the proxy is used) within the same no-
trade period. There is one unique frequency number kept for every quote record. I also define $mag_{i,t}$ to quantify the magnitude or distance in dollar amount of crossing each time it occurs. The value initiates with null and is kept at null if no other quote record crosses with this one. Once crossing occurs for the first and subsequent times, this magnitude represents the distance marking how far one side of the quote or a midpoint of one set of quotes crosses outside the bid or ask of the other set of quotes. By this design, for any given quote record for stock $i$ at time $t$, where crossing does not occur, $freq_{i,t} = 0$ and $mag_{i,t} = null$ instead of $mag_{i,t} = 0$. Thus if $mag_{i,t} > 0$, it takes on the meaning of the cross magnitude conditional that crossing has occurred. Still using SBUX as an example, I illustrate below how $freq_{i,t}$ and $mag_{i,t}$ are constructed using four different counting methods. While defining these variables, I drop the subscript $i$ for simplicity, because $i$ represents a specific stock, and the definition below is stock specific. Once the metrics are defined, I explain how aggregation $freq_{i,t}$ and $mag_{i,t}$ are done across stocks within time buckets to get $freq_{i,\Delta t}$ and $mag_{i,\Delta t}$ for time series, and across time within each stock to get $freq_{i,t}$ and $mag_{i,t}$ for cross-sectional regressions.

I define the crossing measures for a given set of quote $[bid_t, ask_t]$, while $t \in [1,T]$, with $t = 1$ being the first quote record immediately following a trade, and $t = T$ is the last quote record in this group before the next trade. Here $t$ is defined in the event space. In other words, $t + 1$ represents the next QBBO update instead of the next millisecond or second. $\Delta t$ is used to represent the length of this no-trade duration. In the strict form of crossing without using midpoint as a proxy for the expected true price, for a given $t' < t$ if $bid_{t'} > ask_t$ then $freq_t = 1$ and $mag_t = |bid_{t'} - ask_t|$. (If the crossing occurs in the other direction, $ask_{t'} < bid_t$, instead $mag_t = |bid_t - ask_{t'}|$.) I search through all $t' < t$ for $t'$ from 1 to $(t - 1)$, I
accumulate the frequency count and the magnitude $freq_t = freq_t + 1$ and $mag_t = mag_t + |bid_t' - ask_t| \text{ (or } mag_t = mag_t + |bid_t - ask_t'| \text{)}, and then I divide the cumulative $mag_t$ for the no-trade duration by $freq_t$ to achieve the average cross magnitude per occurrence for this particular quote within this group. The magnitude measure $mag_t$ can take on the value of missing but not zero by design. A missing value represents absence of crossing for this quote record within the observation window. (Zero distance would be possible if I chose to include the equal sign in comparison, in which case it would be possible for the true price to lay on the edge of the bid-ask interval. I have chosen not to include these cases, as explained earlier.)

Additionally, if within this group crossing occurred at all, regardless of how many times, I define $gfreq_{\Delta t} = 1$ instead of zero. For all four counting methods I define below, $gfreq_{\Delta t}$ only takes on the value of either zero or one to signify whether crossing occurred at all within this no-trade duration. $gmag_{\Delta t}$ takes on the value of the largest value of $mag_t$ within the group $gmag_{\Delta t} = \max (mag_t)_{t \in [1, T]}$, or the value of null if no crossing occurred.\textsuperscript{81}

In Table 2, the calculation of frequency and magnitude for $freq_t$, $mag_t$, $gfreq_{\Delta t}$, $gmag_{\Delta t}$ is given as an example using SBUX. I have named this counting method strict cross. Within this no-trade duration, the best bid and offer of $51.13/51.14$ at time 9:39:47.775 crossed with two sets of previous quotes: $51.15/51.18$ for $0.01$ at 9:39:46.710 and $51.15/51.17$ also for $0.01$ at 9:39:47.; therefore, the frequency $freq_t = 2$, and the magnitude

\textsuperscript{81} $gmag_{\Delta t}$ could also be defined as the weighted average of all non-zero magnitude $mag_t$ within the group using $gfreq_{\Delta t}$ as weight—it signifies the average cross magnitude for each individual occurrence of crossing. Instead, I assign $gmag_{\Delta t}$ to take on the value of the largest value of $mag_t$ within the group, or the value of null if no crossing occurred. This is in order to have the average measure in aggregation (defined later within this section) $mag_{\Delta T}$ and $gmag_{\Delta T}$ to represent different aspects of the phenomenon I characterize instead as being somewhat repetitive, especially with $freq_{\Delta T}$ and $gfreq_{\Delta T}$ taking on different meanings.
$mag_t = 0.01$ as the average of the two occurrences. For the entire group, $gfreq_\Delta t$ takes on the value of one because strict cross did occur. $gmag_\Delta t = 0.01$, which is the same with $mag_t$ because there is only one value from which to take the maximum. Next, I define three variations of crossing metrics in the relaxed form using the bid-ask midpoint as a proxy for the expected true price, in order to see how often this proxy stands outside other sets of quotes within a no-trade duration. Possible counting methods are not limited to the four kinds I propose in this dissertation. These methods are designed to capture variations of benchmarks in the intuitive sense. The conclusions are found to be robust if they hold, regardless of counting methods.

Among the three relaxed forms of crossing, I name the first form neighbor cross as illustrated in Table 1 and Figure 1: for each given set of quotes $[\text{bid}_t, \text{ask}_t]$, if the midpoint from the immediate previous set of quote $t' = t - 1$ is outside the boundaries of $\text{bid}_t$ and $\text{ask}_t$, then $freq_t = 1$ and $mag_t = mid_{t-1} - \text{ask}_t$ when $mid_{t-1} > \text{ask}_t$, $mag_t = \text{bid}_t - mid_{t-1}$ when $mid_{t-1} < \text{bid}_{t-1}$. In Table 1, I illustrate the calculation of neighbor cross using SBUX as an example. Comparing the best bid $50.77$ and ask $50.80$ at $9:30:00.180$ at the third observation against the midpoint of the immediate previous quote $50.96$, the previous midpoint was above the current ask price $50.80$ for $0.16$, thus $freq_t = 1$ and $mag_t = 0.16$ for this individual quote. For the entire group, $gfreq_\Delta t = 1$ because neighbor cross did occur. $gmag_\Delta t = 0.16$, which is the same with $mag_t$ because this is the only non-zero value in the group. If there are more than one non-zero $mag_t$, then $gmag_\Delta t$ would have been the largest of all $mag_t$ within this no-trade period.

Once the midpoint is introduced as a proxy, various other counting methods can be used. I discuss two more here: one I name first cross, and the other every cross. The first cross method is similar to neighbor cross, except that I use the first midpoint of the group ($t = 1$) as a
benchmark instead of the immediate previous neighbor set of quotes. The rationale is that if true price has not moved after the previous trade, this midpoint should be an accurate proxy, although comparatively it is less strict than neighbor cross because a longer time period elapses between the two quote records in comparison. For the same given quote \([bid_t, ask_t]\) like above, if the midpoint at \(t = 1\) is outside the boundaries of \(bid_t\) and \(ask_t\), then \(freq_t = 1\) and \(mag_t = mid_1 - ask_t\) when \(mid_1 > ask_t\), \(mag_t = bid_t - mid_1\) when \(mid_1 < bid_t\). See Table 3 for an example using SBUX at 9:31:09 on the same trading day as the examples in Tables 1 and 2. The first midpoint of the group is taken by averaging the best bid $50.73 and ask $50.81 to be $50.77 at 9:31:09.189. Every set of quotations within the group from \((t + 1)\) to \(T\) is compared against $50.77 at 9:31:09.189. There are three occurrences of first cross within the same millisecond in subsequent quotes $50.80/$50.81 of the second observation, $50.80/$50.81 of the fourth observation, and $50.79/$50.81 of the last observation. For each of these observations \(freq_t = 1\) to record that first cross, and the magnitude is the distance of $50.77 being below the best bid: $0.03, $0.03, and $0.02. For the entire group, \(gfreq_{\Delta t}\) takes on the value of one because first cross did occur. \(gmag_{\Delta t} = $0.03\), which is the maximum all three non-zero values of \(mag_t\).

The last counting method I introduce is named every cross, where every midpoint up to \(t\) is compared and counted for. From \(t' = 1\) to \(t' = t - 1\), if midpoint at \(t'\) is outside the boundaries of \(bid_t\) and \(ask_t\), then \(freq_t = 1\) and \(mag_t = |mid_{t'} - ask_t|\) when \(mid_{t'} > ask_t\), \(mag_t = |bid_t - mid_{t'}|\) when \(mid_{t'} < bid_t\). Similar to strict cross, I conduct more than one comparison for each quote; as I go through all \(t' < t\) from 1 to \(t - 1\), I keep adding to the frequency count and the magnitude, \(freq_t = freq_t + 1\) and \(mag_t = mag_t + |mid_{t'} - ask_t|\) or \(mag_t = mag_t + |bid_t - mid_{t'}|\) then divide the cumulative \(mag_t\) for the time period with
the cumulative $f_{req_t}$ in order to get the average cross magnitude for this particular quote within this group of quotes. Still, if there are any occurrences of crossing in this group, regardless of how many, I record $gf_{req\Delta t} = 1$. And $gma_{gat}$ was the largest of all $mag_t$. In Table 4, I again present the example of SBUX with the fifth observation of the best bid and ask at $51.14/\$51.15$. There are three previous midpoints outside this range: $51.165$ of the first observation, $51.16$ of the second observation, and $51.155$ of the third observation. Therefore $freq_t = 3$ for this particular quote of the fifth observation, while the magnitude is $0.01$ as the average magnitude of ask price below previous midpoints in all three occurrences. For the entire group, $gf_{req\Delta t} = 1$ because every cross occurs in this group, and $gma_{gat} = 0.015$. Other variations of counting methods can be used, but the general ideas are similar. In Chapter 5: Empirical Results, I examine whether the conclusions are robust compared across the four counting methods.

4.3.2. Normalization and Aggregation

All four counting methods I introduce for crossing characterize instances where the expected true price is outside the boundaries of the bid-ask interval. The strict cross is the most rigid of all because no proxy is used. The other three counting methods each have their own characteristics. The neighbor cross is still relatively rigid because the bid-ask spread at $t + 1$ must have narrowed for at least two ticks compared to the bid-ask spread at $t$ with no other activity between $t$ and $t + 1$, either quote update or trade. This counting method captures instances where either the midpoint (as a proxy for the expected true price) or the bid-ask spread itself has shifted. This shift is comparatively more dramatic for more active stocks when the actual clock time lapse in between $t$ and $t + 1$ is much shorter, and the at-least-two-tick-change in its bid-ask spread is relatively much larger compared to its average quoted spread. The first
cross counting method instead captures price movement trends comparing the beginning of the no-trade duration to the later sets of quotes. First cross frequency is more likely to be one toward the later side of the no-trade duration than the earlier side if true price indeed has drifted. Every cross instead has no particular bias toward either kind, but rather takes each set of quote and midpoint into consideration. The same all-combinations approach is also taken by the strict cross method.

Comparatively speaking, every quote record is used only once in neighbor cross counting methods to be compared against other quote records; this is also the case in the first cross counting method except the first quote within each no-trade duration that is used as a reference; however in the strict cross and every cross counting methods, the quote record at \( t \) is used to make \((t - 1)\) number of comparisons. This gives an upward bias for counting frequency for the strict cross method and the every cross method. In order to adjust for this bias, I normalize the frequency in these two methods by the number of comparisons made, which is \((t - 1)\). For example, in Table 4 the normalized frequency for the fifth observation would be \(\frac{3}{4} = 0.75\) after the normalization. In Table 2, the normalized frequency for the seventh observation would be \(\frac{2}{6} = 0.3333\). (I did not make this adjustment for Table 1 through Table 4 because these tables are set up for illustration purposes.) After this adjustment, I aggregate frequency and magnitude measures across individual stocks or across a specified time period, like a minute or a day (I use \(\Delta T\) to represent the aggregation time period) prospectively in the case of time series and cross-sectional analysis, then take the average across the five days in my sample. For the group measures of all four counting methods, the numeric sum of each group frequency \(gfreq_{\Delta t}\) is taken to form \(gfreq_{\Delta T}\) for the aggregation time period \(\Delta T\): \(gfreq_{\Delta T} = \sum_{\Delta T} gfreq_{\Delta t}\). The
average group magnitude is taken using the numeric sum of each $g mag_{\Delta t}$ divided by the sum of group frequency for that period, then normalized by the average duration weighted quoted spread: $g mag_{\Delta T} = \frac{\sum_{\Delta T} g mag_{\Delta t}}{(g freq_{\Delta t})(qs\text{spread}_{\Delta t})}$. The purpose of the magnitude measure is not to get a cumulative measure, but rather to determine the approximate distance for each crossing occurrence. The aggregation of individual measures is treated differently than the group measures, because the frequency count $f req_{\Delta t}$ is proportional to the number of QBBO updates within this group, and thus should be normalized by this number in order to reflect the severity of crossing frequencies $f req_{\Delta T} = \sum_{\Delta T} \frac{f req_{\Delta t}}{(QBBO_{\Delta t})}$. This normalization takes into consideration the size of the sample within an event space across stocks. Without this normalization, the frequency count would have a bias for stocks with longer periods of no trades in event space, thus more quote updates, and more opportunities for crossings to occur, but against stocks that have shorter periods of no trade. The average individual count magnitude measure in aggregation is taken by the numeric average of all non-zero $m ag_{\Delta t}$, and then also normalized by the average duration weighted quoted spread: $m ag_{\Delta T} = \frac{m ag_{\Delta t}}{qs\text{spread}_{\Delta t}}$. This treatment is consistent with the group magnitude aggregation by representing the distance for each crossing occurrence. The rationale for normalizing the magnitude measure by the bid-ask spread is to capture the relative effect of how far the cross magnitude is, given the average width of its quoted spread.

By design of these metrics, the actual value of true price is irrelevant. The goal is to identify clear violations of the assumption that true price sits between the quotes, regardless of how much the true price actually is. One additional caveat worth pointing out is that the group frequency and magnitude are arbitrarily assigned along the first quote observation of the group, regardless of whether $g freq_{\Delta t}$ takes on the value of zero or one, or how large the group
magnitude is, while these two numbers in fact represent the crossing characteristics of the entire no-trade duration. From the second quote observation to the last, I assign these two measures to be null (missing). I deem this choice fair since it is consistently assigned to the first observation. Few differences between individual count and group count occur when a particular no-trade duration crosses over the 11:00 a.m. mark and 3:00 p.m. mark I designate to carve out the midday section, when the group frequency and magnitude belong to a different segment of the day than part of the group. For these cases, individual measures are still kept intact, while group measures could have a lower bias at the 11:00 a.m. mark and a higher bias at the 3:00 p.m. mark. The two biases are in the opposite direction and presumably balance out for the midday period between 11:00 a.m. and 3:00 p.m.

4.3.3. Summary Statistics

Table 7 through Table 10 present an overview on the crossing metrics defined earlier for the four different counting methods, both in frequency and in magnitude, and both in individual count and in group count. All measures are calculated or collected on a daily basis and then averaged across the sample period between November 5, 2012, and November 9, 2012. Again, only quote and trade records during market open hours between 11:00 a.m. and 3:00 p.m. EST are used. Table 7 provides summary statistics on the relative frequency of the crossing episodes (relative frequency has been defined earlier as the frequency count as percentage of the number of QBBO price updates) with the individual counting method, while Table 9 does the same except for the group counting method.

Examining Table 7, as expected the relative cross frequency is higher in first cross compared to neighbor cross because the criteria are less stringent. This pattern is consistent in the mean, median, standard deviation, and min and max between these two counting methods, while
the statistics for every cross are always lower than first cross while higher than neighbor cross, because every cross fairly considers both methodologies. Strict cross happens much less often than all of the other three counting methods because it does not use the bid-ask midpoint as a proxy. The highest relative frequency for strict cross is 2.17%, compared to neighbor cross as high as 34.15%, first cross as high as 43.53%, and every cross as high as 25.87%. The same pattern exists in measures of mean, median, and standard deviation, with strict cross having the lowest median at 0.03%, neighbor cross at the second-lowest median at 3.82%, every cross next at 6.15%, and first cross the highest median at 8.36%. Examining the three strata of neighbor cross breakdown by market cap, the relative cross frequency is the highest for small-cap with a median of 6.27%, while the median is 4.28% for mid-cap and 2.43% for large-cap. This is expected because in order for neighbor cross to occur, the bid-ask spread needs to change for at least two ticks. This is much less likely for active large-cap stocks that are more constrained by the one cent tick size, while being much more likely for less active small-cap stocks with an average quoted spread of $0.15 (as shown in Table 5), it is a much smaller percentage move for the bid-ask spread to move for at least two ticks. Comparing the three strata for first cross, the relative cross frequency has a median of 7.78% for large-cap, while the median rises to 8.89% for mid-cap and dips slightly to 8.39% for small-cap. This slightly different pattern is partially because the average price levels differ significantly across the three strata: $43.29 for large-cap, $14.39 for mid-cap, and $4.73 for small-cap. The higher the absolute stock price, the wider a price range is allowed for the midpoint to be drifting away from the first set of quotes within a no-trade duration. Comparing the three strata for every cross, the same trend as neighbor cross is shown: the median is the highest for small-cap at 6.81%, decreases to 6.61% for mid-cap, and is the lowest for small-cap at 5.23%. The difference across strata is less dramatic than neighbor
cross but more obvious than first cross because every cross takes into consideration both counting methods. Interestingly, this trend is reversed for strict cross, where the highest relative frequency both in mean and in median is with large-cap, with a median of 0.07%, but 0.03% for mid-cap, and less than 0.01% for small-cap. This is primarily because of the more rigid criteria for strict cross, where one side of the quotes has to be entirely above or below the bounds of another set. Frequent quote updates are necessary for this condition to be met. This is much more likely for large-cap stocks with significantly higher activities and competitions among market participants.

Table 9 provides summary statistics for the same four counting methods for the relative cross frequency measure, except that group count is used instead of individual count. Overall, all corresponding cells have lower values in group count compared to individual count because the frequency count is capped at one for each no-trade duration while it can be much higher than one for individual count. Besides this observation, the trends comparing among the four counting methods and three market cap strata are consistent with individual count, with a few exceptions. The median for the full sample is 3.02% for first cross, higher than neighbor cross with a median of 2.42%, which is the same trend with individual count. However, every cross median is higher in mean, median, standard deviation, and max than both neighbor cross and first cross. This is because any occurrence of crossing within a no-trade duration by any methodology puts every cross frequency at minimum of one (before taking the percentage of the number of QBBO updates). This is inclusive of neighbor cross and first cross; thus, the frequency is higher than these two methods, unlike the individual count, where the every cross is a combination and is in between neighbor cross and first cross. Nevertheless, of all four counting methods, strict cross has the lowest median of 0.07%, as well as the lowest mean of 0.11%, lowest standard deviation
of 0.12%, and the lowest maximum value of 1.12%. Examining the three strata of neighbor cross breakdown by market cap, the relative cross frequency is the highest for small-cap with a median of 3.05%, while the median is 2.66% for mid-cap and 1.89% for large-cap. This is expected and the same trend with individual cross. Comparing the three strata for first cross, the opposite trend is shown. The relative cross frequency is the highest for large-cap with a median of 3.12%, while the median is 3.07% for mid-cap and 2.66% for small-cap. This is a clear and dramatic demonstration of the average price level difference across the three strata discussed earlier: the higher the absolute stock price, the wider a price range is allowed for the midpoint to be drifting away from the first set of quotes within a no-trade duration. Comparing the three strata for every cross, the median holds steady from large-cap to mid-cap, both at 4.49%, then dropping to 3.87% for small-cap. Strict cross where the highest relative frequency both in mean and in median is with large-cap, with a median of 0.14%, but 0.07% for mid-cap, and less than 0.01% for small-cap, due to the same reasons discussed before.

Turning to relative cross magnitude, I provide summary statistics for individual count in Table 8, and for group count in Table 10. As discussed earlier in this section, magnitude—regardless of counting methods—indicates the distance of how far the crossing violation occurred in dollar amount, then taken as percentage of the duration-weighted average quoted spread for this stock. The magnitude is null, not zero, when crossing does not occur; thus, the value is meaningful in condition of a non-zero frequency. Looking at the full sample in Table 8, the median is highest for neighbor cross at 26.18%, which is only slightly higher than the median of 26.08% for first cross, which is then higher than the median of 24.85% for every cross, and 23% for strict cross. In theory, neighbor cross should have the highest magnitude because it requires the bid-ask spread to change for at least two ticks. Strict cross should have the lowest
magnitude because the condition is the most difficult to meet. Additionally, the calculation for
strict cross magnitude should be approximately a half a spread lower than the other two methods
because the absolute boundaries of the best bid and offer are used instead of the bid-ask
midpoint. This pattern observed in the full sample is mostly consistent across the three strata
breakdown by market cap except in large-cap, in which the trend is reversed for neighbor cross
and strict cross. The median relative cross magnitude in the large-cap stratum is 21.66% for
neighbor cross, 23.94% for first cross, 22.17% for every cross, and 26.69% for strict cross. For
similar reasons discussed for frequency, large-cap stocks are more active with on average much
lower quoted spread; thus, the tick size constraint is much more apparently observed in the lower
magnitude as well as frequency in neighbor cross. Comparing across the three strata for strict
cross, the median decreases from 26.69% for large-cap, to 22.03% for mid-cap, to 14.65% for
small-cap. This trend is consistent with relative frequency observed in Table 7 and Table 9. This
trend clearly indicates that more active quote updates and general activities encourage both the
occurrence of strict cross and the severity of its occurrence. The other three counting methods in
this table exhibit a pattern that is not observed in relative frequency: the relative cross magnitude
for mid-cap is higher than small-cap, which is then higher than large-cap. The relative cross
magnitude for neighbor cross has the median of 21.66% for large-cap, 28.27% for mid-cap, and
23.62% for small-cap. First cross has the median of 23.94% for large-cap, 27.09% for mid-cap,
and 25.89% for small-cap. Every cross has the median of 22.17% for large-cap, 26.04% for mid-
cap, and 22.43% for small-cap. Average by mean follows the exact same pattern.

The exact same pattern across the three market cap strata is also depicted in Table 10,
which provides summary statistics for relative cross magnitude in group count. Ranging from
large-cap to mid-cap to small-cap, the relative cross magnitude steadily declines in strict cross,
but rises first and then decline for the other three counting methods: neighbor cross, first cross, and every cross. Comparing across all four counting methods, neighbor cross still has the highest median relative cross magnitude at 28.15% for the full sample, while every cross has the second-highest median at 26.65%, first cross comes next with the median at 25.94%, and finally strict cross with the smallest median at 25.28%. Unlike the relative cross frequency comparison between Table 7 and Table 9—where individual count is greater than group count because group count considers multiple occurrences of individual count within the same no-trade duration as one increment for frequency—the relative cross magnitude in individual count in Table 8 is smaller than group count in Table 10 because the group magnitude as defined in this section is the maximum magnitude of all individual magnitudes within one no-trade duration. For this reason, all measures in Table 10 are larger than their corresponding counterparts in Table 8.

Even though all measures are calculated during the midday hours between 11:00 a.m. and 3:00 p.m., it is still interesting to see how cross frequency and magnitude changes throughout the course of the day. As discussed above, every cross takes into consideration both neighbor cross and first cross, while strict cross is unique in its own way; I have chosen to use every cross and strict cross as examples to chart the intraday changes of relative cross frequency and magnitude, both in individual count and group count, as shown from Figure 11 to Figure 18. All measures are calculated in a minute-by-minute bucket for each trading day, then averaged across the five-day sample period. Unlike other market activity metrics like volume or the bid-ask spread, none of the crossing metrics (in any of the eight figures shown) clearly demonstrate a pattern of spikes

---

82 At the earlier stage of this research, I decided to use the midday section between 11:00 a.m. and 3:00 p.m. due to this intraday observation on trade count, QBBO updates, length of no-trade duration, etc., observed in Figure 5 through Figure 9. As it is shown, if cutting off the morning section before 11:00 a.m. and 3:00 p.m., the line graph indicates a relatively steady and flat trend during the middle of the day.
or dips close to market open or market close. Since this is the case, it is even more necessary to
choose the relative flat midday period for the other market activity metrics as they will be used
as independent variables and control variables in the regression analysis. Similar to what has
been shown in the cross-sectional comparisons in Table 7 through Table 10, large-cap has the
highest frequency in strict cross by both individual count and group count, followed by mid-cap,
then small-cap. Relative cross magnitude in Figure 12 and Figure 14 shows the same pattern but
in a less distinct way. It is interesting to note that even for strict cross, there are many large
outliers for relative cross magnitude from the small-cap stratum, in both individual count and
group count. Turning to every cross, I observe the relative cross frequency competing between
mid-cap and small-cap, while both higher than large-cap in individual count, although distinctly
highest for large cap, followed by mid-cap, then by small-cap. The relative cross magnitude is
highest for mid-cap in both individual count and group count, which is higher than small-cap,
and then higher than large-cap. All of these patterns observed in the intraday by-minute graphs
are consistent with what is observed in the cross-sectional tables.

4.4. Monte Carlo Simulation

4.4.1. Methodology and Assumptions

Even though the occurrences of crossing by any of the four counting methods discussed
in the previous section violate assumptions in microstructure theory as discussed in Chapter 2:
Literature Review, it is important to test whether the occurrences observed empirically are
statistically significant. Such a test can be performed to compare the mean of the crossing
metrics against zero as a benchmark, although the newly documented empirical evidence will be
much stronger if tested against a benchmark that simulates non-informational noise in the
marketplace. I construct a Monte Carlo simulation of order arrivals on the limit order book in order to provide such a benchmark for the first null hypothesis, presented in Chapter 5: Empirical Results, to determine whether the crossing episodes observed and documented in this dissertation are statistically significant. This simulation also provides valuable information for understanding the mechanism by which different market factors contribute to the occurrences of crossing, either in frequency or in magnitude. Regarding constructing a model for such a simulation, the existing literature follows one of two paths: one is to construct an agent-based simulation with strategies and behavioral game-theoretic assumptions of modern markets; the other is to have a zero-intelligence model following a few simple of assumptions that captures the key empirical properties of a continuous double auction on a limit order book. In order to build an agent-based strategic trader model following game theory, it is necessary to make certain assumptions on the behaviors and the interactions between market makers, informed traders, and noise traders. Some frameworks become analytically intractable or time consuming to estimate from empirical data. One can follow the assumptions on these issues from the classical microstructure framework, but such assumptions would potentially prevent crossings from occurring, as not having crossings to occur is one of the features of a classical microstructure model. I build a simulation according to the second approach, following the simplest form of the limit order book simulation models, which is to have three separate Poisson processes for order arrivals: limit orders, market orders (includes marketable limit), and limit order cancellations. The purpose of this part of the dissertation is not to have a realistic model on all or most aspects of the market mechanism, but rather to capture the basic properties of the limit order book and produce noise order flows to provide a basis for empirical testing.
Fortunately, there is a small set of econophysics and quantitative finance literature that follows the second approach to study the empirical features of limit order books and then construct simulations of the limit order book to study these features. These literatures provide guidance, building blocks, and assumptions for the simulation I build for this dissertation. Due to the complex nature of the limit order book dynamics, it is difficult to have a model that is statistically realistic in every aspect—or even on most aspects. Most of the proposed models (including both the simple zero-intelligence models and complex strategy-based models) focus on features of interest to the author, while ignoring the fidelity of other features. For example Tóth, Kertész, and Farmer (2009) assume that new limit order placements follow a uniform distribution established around the bid-ask midpoint: new buy orders are placed below the midpoint and new sell orders are placed above the midpoint. Following this approach, any given limit order placed on the book could never cross the current bid-ask midpoint. It is an unrealistic assumption for the real market, although it still produces informative results for the purpose of the authors. The goal of their simulation is to study the mechanism of how the width of the bid-ask spread is restored after large price movements. The same set of assumptions is clearly not suitable for this dissertation. Instead, I present the methodology I have chosen to construct the simulation, while comparing and contrasting the different approaches taken by three papers that have the most relevance: Bouchaud, Mézard, and Potters (2002); Zovko and Farmer (2002); and Cont, Stoikov, and Talreja (2010). The first two of these three papers—Bouchaud, Mézard, and Potters (2002) and Zovko and Farmer (2002)—are empirical studies that observe the limit order arrivals outside the current best bid and offer, which follows a power-law distribution. The third paper—Cont, Stoikov, and Talreja (2010)—builds upon the power-law distribution to construct an actual simulation, and then evaluates a simple trading strategy in the simulation. Cont,
Stoikov, and Talreja (2010) also calibrate against empirical data to show that the model can effectively capture the short-term dynamics of the limit order book. This is the primary reference for my simulation. Cont, Stoikov, and Talreja (2010) attempt to construct a relatively generic form of the zero-intelligence models; however, some features are simplified in ways that affect features critical to crossings. Therefore, I make necessary modifications to extend this model.

I follow the general approach from Cont, Stoikov, and Talreja (2010), which models the limit order book as a continuous-time Markov process by tracking the number of limit orders—at each price level on the limit order book—that are in queue to be executed against market orders or cancelled. The arrivals of limit orders, market orders, and cancellations follow separate Poisson processes. I select three sample stocks, one from each market cap stratum for calibration and testing: Starbucks Corporation (NASDAQ: SBUX) for large-cap, IGATE Corp (NASDAQ:IGTE) for mid-cap, and American Pacific Corporation (NASDAQ:APFC) for small-cap. Choosing SBUX as a representation of the large-cap group is to have a stock that is relatively actively traded but not too high in price (approximately $50 in November 2012), and with which the readers are likely to be familiar. Sorting all stocks by market cap, SBUX ranks 18th out of 456 all large-cap stocks, which is approximately 39th out of 100. I have chosen proportionally from 963 mid-caps to arrive at IGTE, then from 252 small-caps to arrive at APFC. To stay consistent with empirical analysis, all metrics for the simulated stocks are also calculated using trade and quote records between 11:00 a.m. and 3:00 p.m. I achieve the Poisson distribution by taking the negative natural log of a uniform distribution, divided by the arrival rate of the particular order type in count per second. Market order is the easiest among the three order types to simulate, as the arrival rate comes directly from the data as the number of trades
per second between 11:00 a.m. and 3:00 p.m. I call this parameter $\mu$, which takes on the value of 1.424 for SBUX, 0.0279 for IGTE, and 0.00168 for APFC.

All three of the reference papers—Bouchaud, Mézard, and Potters (2002); Zovko and Farmer (2002); and Cont, Stoikov, and Talreja (2010)—analyze empirical data and observe that limit order placement from the best bid and offer into further depth of the book follows a power-law distribution. All three models assume order quantity to be one, and have no discussion on modifications or more complex order types. Bouchaud, Mézard, and Potters (2002) obtain order book level data from the Paris Bourse\textsuperscript{83} on the three most liquid stocks during February 2001: France-Telecom (F.T.), Vivendi, and Total. The authors present the power-law distribution of the incoming limit order prices in the following form:

$$P(\Delta) \propto \frac{\Delta_0^\mu}{(\Delta_1+\Delta)^{(1+\mu)}}, \quad \Delta \geq 1 \quad (1)$$

The power-law exponent $\mu$ takes on the value of 0.6 for all three stocks the authors analyze, extending from the best bid and offer to 100 ticks away. They also observe the probabilities for a new limit order to be placed at the best quote (either the bid side or the ask side, which are symmetrical), versus one tick wider, versus one tick narrower, are approximately the same—that is, $P(\Delta= -1) \sim P(\Delta= 0) \sim P(\Delta= 1)$.

Cont, Stoikov, and Talreja (2010) disregard the differences between these three probabilities, instead placing the new limit orders at distance $i$ away from the opposite best quote, following the power-law distribution of:

$$\lambda(i) = \frac{k}{i^\alpha} \quad (2)$$

\textsuperscript{83} Now part of NYSE Euronext.
This setup imposes a strong bias toward a one-tick spread, as new limit order placement has the highest probability of being directly adjacent to the opposite side best price. The authors obtain order book data on Sky Perfect Communications from the Tokyo Stock Exchange from the best bid and offer to five deep over a period of 126 days (August to December 2006). They provide the parameter estimate to be 0.52 for $\alpha$ and 1.92 for $k$.

Zovko and Farmer (2002) obtain a dataset of roughly two million orders from the London Stock Exchange from August 1, 1998, to April 31, 2000. The authors choose a sample of 50 different names with a balance between high- and low-volume stocks. Their version of the power-law distribution follows:

$$P(\delta) = \frac{A}{(x_0 + \delta)^\beta}$$

Fitting the available data across ticker symbols and time period to the sample, the authors find the parameters to be $\beta = 1.49 \pm 0.0001$, $x_0 = 7.01 \pm 0.05$ and the $A$ value depends on the stock. $\delta$ here is equivalent to $\Delta$ in Bouchaud, Mézard, and Potters (2002) and $i$ in Cont, Stoikov, and Talreja (2010), which represents the price level for limit order placement. Zovko and Farmer (2002) discard partial fill orders and only focus on limit orders that enter the book. Using Vodafone as an example, Zovko and Farmer (2002) find that 74% of the orders are submitted at the best quotes ($\delta = 0$), while only 1% of the orders improve the current bid-ask spread ($\delta < 0$), and the remaining 25% of the orders are out of the prevailing quotes ($\delta > 0$). This observation

---

84 SKY Perfect Communications is an 100% owned subsidiary of SKY Perfect JSAT Corporation (TYO: 9412), which is the only provider of multichannel pay TV broadcasting and satellite communications in Japan, and the largest in Asia and Oceania. As of October 2015, the estimated market cap is 217.79 billion JPY. Source: http://www.bloomberg.com/quote/9412:JP.

85 According to Zovko and Farmer (2002), the ticker symbols in their sample include the following: AIR, AL, ANL, AZN, BAA, BARC, BAY, BLT, BOC, BOOST, BPB, BSCT, BSY, BT.A, CCH, CCM, CS, CW., GLXO, HAS, HG., ICI, III, ISYS, LAND, LLOY, LMI, MKS, MNI, NPR, NU., PO, PRU, PSON, RB., RBOS, REED, RIO, RR., RTK, RTO, SB, SBRY, SHEL, SLP, TSCO, UNWS, UU., VOD, and WWH.
conflicts with Bouchaud, Mézard, and Potters (2002), who find the three probabilities to be roughly the same. Therefore, Zovko and Farmer (2002) only model limit orders from the current best bid and offer to further depths in the book using this power-law distribution.

I implement all three models for limit order placement with a range of parameters, and find that all produce realistic results for SBUX, but none for IGTE in mid-cap and APFC in small-cap, although the Zovko and Farmer (2002) fair better than the other two. Cont, Stoikov, and Talreja (2010) produce the bid-ask spread for all stocks to be consistently at one tick wide. I find the observation from Zovko and Farmer (2002) more compelling than Bouchaud, Mézard, and Potters (2002) regarding the probabilities between order placements at the best quotes, wider, or narrower. This approach is only reasonable for highly liquid stocks where the bid-ask spread consistently stays at one tick wide. The lack of proper models for the distribution of limit orders improving the best quotes could be because there is less empirical data available to allow for clear analysis. I follow the method from Zovko and Farmer (2002) on limit order placement outside the current best quotes, but extend an additional Poisson process for limit orders placed inside the best bid and offer, where I allow the spread to narrow at one tick at a time. I calibrate these two separate processes for limit order placements to produce a consistent average bid-ask spread that aligns with empirical observation. The one-tick-at-a-time assumption produces reasonable results for both SBUX and IGTE, but not as well as expected for APFC in small-cap. This is because as market-cap decreases, the average bid-ask spread increases, and the bid-ask spread changes at a higher range. However, this improvement still takes a step toward the direction of a better model for less active securities. Summary statistics comparing the real stocks and the simulated stocks will be discussed in the next section.
Cont, Stoikov, and Talreja (2010) observe empirical data from depths one to five and provide estimates for the cancellation rate of limit orders to be 0.71, 0.81, 0.68, 0.56, 0.47, then stay constant at 0.47 beyond depth five. I take this form, but make the observation that the cancellation rate from depths two to five follows an approximate power-law distribution. Instead of setting the cancellation rate at constant beyond depth five, I make the rate trail off at this power-law. This added trailing-off effect makes little difference for very active stocks with narrow spreads, but fits much better with less active stocks with wider spreads. A global scale factor similar to $k$ in equation (2) for limit order placements is applied to the cancellation rate. The overall model is tuned using the global scale factors of limit order entry and cancellation to match the number of QBBO updates and average quoted spread in the empirical data.

4.4.2. Simulation Quality

Table 11 examines the quality of the simulation by comparing summary statistics between the simulation and observations from empirical data on the three sample stocks: Starbucks Corporation (NASDAQ: SBUX) for large-cap, IGATE Corp (NASDAQ:IGTE) for mid-cap, and American Pacific Corporation (NASDAQ:APFC) for small-cap. Again, all metrics for the simulated stocks are also calculated using trade and quote records between 11:00 a.m. and 3:00 p.m. The summary statistics are averaged across five trading days for the real stocks, and 10,000 simulated trading days for the simulated stocks. Looking at SBUX, average price in dollar amount ranges from $51.04 to $51.72 for the real stock, spanning across $0.68, while the average price ranges from $49.82 to $50.17 for the simulated stock, spanning across $0.35. The median average price for the real stock is $51.36 with a standard deviation of $0.17, while the median average price for the simulated stock is $49.99 with a standard deviation of $0.08. The no-trade duration for the real stock ranges from 0.001 second to 40 seconds, with a median of
one second and a standard deviation of four seconds, while the simulated stock ranges from 0.001 second to seven seconds, with a median of one second and a standard deviation of one second. This distinct difference is because the rate of order arrival on the real stock does not follow the exact Poisson process and is more skewed. The real stock has a quoted spread that ranges from $0.01 to $0.042, which has a mean of $0.014, median of $0.012, and standard deviation of $0.006. The simulated stock has a quoted spread that ranges from $0.01 to $0.062, and a mean of $0.014, median of $0.011, and standard deviation of $0.008. The real stock totals 20,493 trades and 7,190 QBBO updates between 11:00 a.m. and 3:00 p.m., while the simulated stock scores 20,506 trades and 7,199 QBBO updates.

Turning to IGTE in the mid-cap category, average price in dollar amount ranges from $15.57 to $15.80 for the real stock, spanning across $0.23, while average price ranges from $15.34 to $15.64 for the simulated stock, spanning across $0.30. The median average price for the real stock is $15.69 with a standard deviation of $0.05, while the median average price for the simulated stock is $15.49 with a standard deviation of $0.07. The no-trade duration for the real stock ranges from 0.001 second to 905 seconds, with a median of nine seconds, and a standard deviation of 134 seconds, while the simulated stock ranges from 0.197 second to 235 seconds, with a median of 31 seconds and a standard deviation of 38 seconds. As with SBUX, the simulation for IGTE also has a much less skewed distribution of the length of no-trade duration than the real stock. The real stock for IGTE has a quoted spread that ranges from $0.01 to $0.222, which has a mean of $0.049, median of $0.044, and standard deviation of $0.037. The simulated stock has a quoted spread that ranges from $0.01 to $0.195, and a mean of $0.049, median of $0.037, and standard deviation of $0.041. The real stock totals 402 trades and 915
QBBO updates between 11:00 a.m. and 3:00 p.m., while the simulated stock also scores also 402 trades and, similarly, 910 QBBO updates.

Finally, I examine the simulation quality for APFC in the small-cap category, for which average price in dollar amount ranges from $12.17 to $12.88 for the real stock, spanning across $0.70, while average price ranges from $11.99 to $12.98 for the simulated stock, spanning across $0.99. The median average price for the real stock is $12.44 with a standard deviation of $0.23, while the median average price for the simulated stock is $12.49 with a standard deviation of $0.99. The no-trade duration for the real stock ranges from 176 seconds to 3,705 seconds, with a median of 1,190 seconds and a standard deviation of 1,475 seconds, while the simulated stock ranges from 39 seconds to 2,321 seconds, with a median of 495 seconds and a standard deviation of 610 seconds. The real stock for IGTE has a quoted spread that ranges from $0.254 to $1.35, which has a mean of $0.848, median of $0.778, and standard deviation of $0.325. The simulated stock has a quoted spread that ranges from $0.714 to $1.001, and a mean of $0.086, median of $0.861, and standard deviation of $0.079. The real stock totals 24 trades and 81 QBBO updates between 11:00 a.m. and 3:00 p.m., while the simulated stock also scores 24 trades and, similarly, 87 QBBO updates.

Comparing across the three stocks in each market cap stratum between the real stocks and the simulated stocks, the number of trades and QBBO updates range from being identical to less than 1% different. The absolute value of the average price is less relevant because it is an offsetting variable in the simulation that is input as a constant in the simulation. However, comparing the standard deviation and the variation range between the real stocks and the simulated stocks, the simulated SBUX is at roughly 50% variation of the real SBUX, the simulated IGTE has roughly 30%-40% more variation than the real IGTE, and the simulated
APFC has roughly 40%–80% more variation than the real APFC. The quoted spread measures are very close in comparison between the simulated stock and the real stock, with the mean from being identical to 1% different, and the median from 8% different to 11% different. The no-trade duration has the highest skewness in the real stock, and is thus the least well resembled by the simulated stock. Looking across all three strata, the mean and the median are farther apart, from 66% different in APFC to eight times as large in IGTE; however, the simulated stocks have similar to identical means and medians generated from the Poisson process.

Table 12 and Table 13 provide crossing summary statistics both in frequency and in magnitude comparing between the empirically observed stocks and the simulated stocks. In Table 12, the four columns compare the relative cross frequency, comparing between the four different counting methods: neighbor, first, every, and strict. I compare the real stock and the simulated stock in rows right next to each other, organized into the mean average of individual count, standard deviation of the individual count, mean average of group count, and standard deviation of the group count. As discussed earlier in this section, due to the methodology design of this simulation, neighbor cross does not occur because it requires the bid-ask spread to change for at least two ticks comparing to the immediately previous set of quotes. Therefore, the cells for mean and standard deviations are given “-” to indicate missing values in both Table 12 and Table 13. The relative cross frequencies for first cross, every cross, and strict cross exhibit a clear pattern with the simulated value significantly smaller than is observed in empirical data. Take SBUX first cross individual count as an example: the relative cross frequency for the real stock has a mean of 4.15% with a standard deviation of 1.46%, while the simulated stock has a mean of 0.4% and a standard deviation of 0.14%. The same pattern is shown using group count first cross, where the real SBUX has the relative cross frequency with a mean of 3.13% and a
standard deviation of 1.07%, while the simulated SBUX has the relative cross frequency with a mean of 0.37% and a standard deviation of 0.13%. Take IGTE every cross as another example: the relative cross frequency for the real stock has a mean of 6.95% with a standard deviation of 2.15%, while the simulated stock has a mean of 2.25% and a standard deviation of 0.07%. The same pattern is shown using group count every cross, where the real IGTE has the relative cross frequency with a mean of 5.82% and a standard deviation of 1.77%, while the simulated IGTE has the relative cross frequency with a mean of 2.83% and a standard deviation of 0.14%. As one last example: in APFC strict cross, which in theory should be rare, the relative cross frequency for the real stock has a mean of 0.01% with a standard deviation of 0.02%, while the simulated stock does not have strict cross occurrences at all. The same pattern is shown using group count strict cross, where the real APFC has the relative cross frequency with a mean of 0.08% and missing standard deviation because this crossing only occurs once, while the simulated APFC has zero occurrences in strict cross. I perform a t-test to examine the statistical significance in the difference between the real stock and the simulated stock at the beginning of the next chapter.

Table 13 provides the same set of comparisons in magnitude instead of frequency. It is very interesting to observe the relative cross magnitude for SBUX is the same 35.96% with the standard deviation of 4.38% for neighbor cross, first cross, and every cross, both in individual count and group count. The same characteristic is observed both in the real stock and in the simulated stock. This happens because SBUX is an active stock with a relatively narrow bid-ask spread (average $0.014), and is constrained by the tick size $0.01 frequently. When one side of the quote moves, the other side often follows to maintain the quoted spread at approximately $0.01 wide. This limits the possible value for crossing magnitude, which represents the distance of how far crossing occurred. The crossing magnitude is in fact 35.96% of the average quoted
spread, approximately $0.005 in dollar amount, which is half the tick size. This is consistent with the methodology for neighbor cross, first cross, and every cross, which considers the cross magnitude using the bid-ask midpoint from the previous set of quotes. Unsurprisingly, the relative cross magnitude for strict cross, which does not use the midpoint proxy, is twice as much at 71.93%, which is approximately $0.01. I use the median instead of mean for averaging across all possible magnitude amounts, which is the reason why the percentages and dollar amounts are quite exact. Turning to IGTE, which is less active than SBUX and less constrained by tick size: Looking at every cross as an example, the relative cross magnitude for the real stock has a mean of 25.73% with a standard deviation of 5.39%, while the simulated stock has a mean of 11.04% with a standard deviation of 7.93%. The same pattern is shown using group count first cross, where the real SBUX has the relative cross magnitude with a mean of 22.85% and a standard deviation of 6.60%, while the simulated SBUX has the relative cross frequency with a mean of 11.13% and a standard deviation of 7.93%. Both in the real stock and the simulated stock, crossing occurs more often with smaller-cap stocks, but at a lesser relative magnitude. This effect is due to the normalization using the average quoted spread. As the market cap value increases, the quoted spread gets significantly larger and at a faster pace than the magnitude of cross by any counting method. I expect this effect from the normalization as well as the tick-size constraint to be reflected in the regression analysis when price and quoted spread are used as independent variables for relative magnitude. Due to the design of the simulation, the crossing results for APFC are less realistic compared to SBUX and IGTE. Neither neighbor cross nor strict cross occurs at all for the simulated APFC, although the mean for first cross and every cross are not too far apart between the real stock and the simulated stock. Using first cross as example, by individual count, the simulated stock has a mean of 17.53% with a standard deviation of 63.77%,
compared to the real stock with a mean of 12.65% with a standard deviation of 0.59%. By group count, the simulated stock has a mean of 21.91% with a standard deviation of 62.74%, compared to the real stock with a mean of 11.75% with a standard deviation of 9.82%. The statistical significance in the difference in relative cross magnitude between the real stock and the simulated stock is also provided at the beginning of the next chapter.
Chapter 5. Empirical Results

5.1. T-test Results

The null of the first hypothesis states that true price $V_{i,t}$ always stays within the range of the bid $B_{i,t}$ and ask $A_{i,t}$ price range (in other words, it never strays outside); therefore, crossings should not ever occur on the limit order book in the equity market. In what I call the strict version, without using the bid-ask midpoint as a proxy for true price, I use the strict cross frequency and magnitude defined in the previous section for testing. Between $t$ and $t'$ if no new information arrives in the market on a given stock $i$, the relationship $V_{i,t}' = V_{i,t}$ holds. In the relaxed version, where midpoint is used as a proxy for true price, I use one of the other three counting methods defined: neighbor cross, first cross, and every cross, where $M_{i,t}' = V_{i,t}'$ and therefore $V_{i,t}' = V_{i,t}$. To reject this hypothesis, I use a series of t-tests to examine if the relative frequency and relative magnitude are statistically significantly different from results produced from the simulation as a benchmark. In the cases where the frequency or magnitude is zero from the simulation (like neighbor cross), a benchmark of zero is used. I take the difference in the means between the real stock and the benchmark, and divide by the standard error of the corresponding variables from the real stock.

Details regarding the results of the t-test are provided in Table 14, in both relative frequency and relative magnitude, in individual count and in group count. ***, **, and * in this table indicate significance levels of 0.1%, 1%, and 5% respectively based on a two-tailed t-test. Looking through a total of 48 t-tests performed, the null hypothesis is rejected 25 times at 95% confidence, 32 times at 99% confidence, and three times at 99.9% confidence. For SBUX alone, of 16 t-tests, the null hypothesis is rejected eight times at 95% confidence, six times at 99%
confidence, and twice at 99.9% confidence. Among these rejections, the two with the highest level of significance are neighbor cross magnitude, once in individual count and once in group count. These two rejections are made because the benchmark is zero, instead of a minimum of half a tick, which will turn these two high-significance rejections to statistically insignificant instead, just like the other three counting methods: first cross, every cross, and strict cross. However, in considering the relative frequency t-tests, crossing is observed more often than simulated noise in a statistically significant manner in six out of eight tests for both individual count and group count. For the 16 t-tests performed on IGTE, the null hypothesis is rejected for neighbor cross, first cross, and every cross at a minimum of 95% confidence for all measures including relative frequency and relative magnitude, individual count and group count, even though strict cross has mixed results that are not statistically significant. The simulation setup is the least realistic for APFC in the small-cap stratum as discussed in the previous chapter, although the high levels of standard deviation from the simulation are not used in these t-tests. The standard error is calculated from the standard deviation of the real stocks instead. Results for APFC are more mixed and inconclusive. The first cross t-test reflects statistical significance, although in the opposite direction than anticipated. The relative magnitude for neighbor cross are rejected both in individual count and group count, and is also rejected in every cross group count.

Once the null hypothesis is rejected at the first stage—even if more reliably for large-cap and mid-cap stocks and less so for the small-cap stock, I conclude that unlike classical microstructure literature has assumed, the true price does occasionally—or even often—fall outside the range of the bid-ask spread. This provides evidence that quote price behavior in the equity market is consistent with the predictions from the uncertainty model in Easley and O’Hara.
This consistency does not, however necessarily indicate the crossing episodes are caused by or even correlated with Knightian uncertainty until further testing.

5.2. Hypothesis Testing

The null of my second hypothesis states that these crossing episodes are not induced by Knightian uncertainty. As discussed earlier, Knightian uncertainty is by definition unquantifiable; I cannot reject this null hypothesis directly. Instead, I use a process-of-elimination method and determine whether I can reject a list of competing hypotheses on different types of factors that can reasonably induce movements in quotation prices. If I fail to reject any of the competing hypotheses, and the price behavior is consistent with a Knightian uncertainty model that would be a step closer in providing empirical evidence on the existence of Knightian uncertainty in the equity market. Two control variables are inherently built into the test. One is stratification by market cap that is controlling for structural shifts to create an all-else-equal statistical sample. The other is interleaving trades into quotes and thus separate quotes into groups, which limits my scope to periods of no trade. This is to control for significant news arrival to the market. If I fail to reject the null for each of my competing hypotheses, I can conclude under a common alternative to these competing hypotheses that crossings are indeed caused by Knightian uncertainty. In the series of five competing hypotheses, the null of the first competing hypothesis states that the crossing episodes are not caused by time series or cross-sectional effects from the specific market factor in question. The five competing alternative hypotheses include one time series regression, which examines general market volatility, and four cross-sectional regressions, which test the impact on relative cross frequency and magnitude from information quality, insider holding, analyst coverage, and microstructure effect one by
one. In the end, I conduct a full model cross-sectional regression that combines the independent variables from the four smaller regressions together to compare the effects from the different factors. I use every cross by group count for all regression analysis since it is a balanced representation of neighbor cross and first cross. Strict cross would be a better choice since its rigidness by design indicates the expected true price must be outside the boundaries of the bid-ask interval. However, because of this rigidness, strict cross does not ever occur for many stocks, and the high number of missing values make the regressions less informative. (For group count, there are 506 stocks out of the 1,671-stock sample for which strict cross never occurred, compared to only six out of 1,671 for every cross, 23 out of 1,671 for first cross, and 127 out of 1,671 for neighbor cross.) As with previous tables on summary statistics, all measures are calculated on a daily basis and then averaged across the sample period between November 5, 2012, and November 9, 2012. Only quote and trade records during market open hours between 11:00 a.m. and 3:00 p.m. are used. All non-dummy variables are log-transformed; thus, the coefficients may be interpreted as percentage changes. ***, **, and * indicate significance levels of 0.1%, 1% and 5% respectively.

5.2.1. Time Series: Market Volatility

In the first of the five competing alternative hypotheses, I test the impact of market volatility on the relative cross frequency and magnitude on a minute-to-minute basis. When the market as a whole or the industry sector is experiencing a high-level of trading or volatility, the cross-asset effect usually spills over to other securities that have no news on the particular stock. I run a linear time series test to examine this effect using the QQQ Volatility as the proxy for market movements of NASDAQ listed securities:
\[ \ln(\text{Rel. } Freq_t) \text{ or } \ln(\text{Rel. Mag}_t) \]
\[ = \beta_1 \ln(QQQ \text{ volatility}_t) + \beta_2 \ln(\text{NumTrades}_t) + \beta_3 \ln(\text{AvgPrice}_t) \\
+ \beta_4 \ln(\text{RelQspread}_t) + \beta_5 \ln(\text{QspreadStdDev}_t) + \beta_0 + \epsilon_t \]

\[ H_0: \beta_1 = 0 \]
\[ H_1: \beta_1 \neq 0 \]

QQQ volatility is calculated using the transaction price of the PowerShare NASDAQ-100 Index Tracking Stock ETF (NASDAQ:QQQ) as the difference between the highest price within the one-minute window and the lowest price of the one-minute window, as a percentage of the low: \( \frac{\text{high} - \text{low}}{\text{low}} \). Num Trades is the total number of transaction counts, used here as a control variable for potential differences across minute windows in trading activities. Share volume and Turnover in dollar amount are also calculated in the full sample, but are not as appropriate to be used here because crossing episodes are characterized with price level changes while discarding quantity updates either in number of shares or dollar amount. Avg Price is the mean of all transaction prices, used here as a control variable for tick size constraint. Because subpenny stocks have been eliminated from the sample, all 1,671 stocks in the sample have the tick size of $0.01, thus \( \frac{1}{\text{Avg Price}} \) represents the relative ticks size. The higher the average stock price, the lower its inverse amount, and the less constrained it is by tick size—although tick size constraint also has other contributing factors, like trading activity indicated by Num Trade, as well as being reflected in the quoted spread. The smaller the average quoted spread, the more constrained the stock price is by tick size. Two more control variables are put in this regression: Rel Qspread is the duration-weighted average quoted spread as percentage of the bid-ask midpoint; Qspread
StdDev is the standard deviation of the duration-weighted average quoted spread. They are also control variables for microstructure effect.

I expect crossing to occur more often and in greater magnitude if the market is more volatile overall during the one-minute window, which is represented by a greater value in QQQ Volatility; therefore, I expect the coefficient and t-value to be positive. I reject the null of the first competing alternative hypothesis if $\beta_1 \neq 0$ at a minimum of 95% confidence. The same goes for Num Trades. If the market is experiencing a higher level of activities overall during the one-minute window, it will be reflected in a higher Num Trades amount, and I expect crossing to occur more often and in greater magnitude. Therefore, I expect $\beta_2 \neq 0$ in a statistically significant manner, and I expect the coefficient and t-value for Num Trades to be positive. Since Avg Price serves as the proxy as tick size constraint, I expect it to have a mixed effect across the three market cap strata, because on one hand, the higher the stock price, the lower the relative tick size as it is $\frac{0.01}{Avg\ Price}$, and the less the price movements are limited by the tick size; on the other hand, however, the higher-priced stocks are more concentrated in the large-cap category, which has higher level of activities and smaller quoted spread. For this reason, Rel Qspread and Qspread StdDev are also in this time series regression as control variables, in addition to their effect to control for minute-to-minute changes of microstructure risk, which includes asymmetric information and inventory risk. I use the relative quoted spread in this regression instead of the quoted spread in dollar amount because the average price itself is another control variable. I expect $\beta_4 \neq 0$ and $\beta_5 \neq 0$ both in a statistically significant manner.

Table 15 provides regression results in separate panels for the full sample, the large-cap, the mid-cap, and the small-cap. The left-hand side of the table provides the coefficient of
estimate, standard error, t-value, and p-value for the relative cross frequency, while the right-hand side of the table provides the same set of statistics for relative cross magnitude. The control variables—Num Trades, Avg Price, Rel Qspread, and Qspread StdDev—show consistent statistical significance across the three strata for relative cross frequency, with the two exceptions of Rel Qspread and Qspread StdDev for small-cap. These two exceptions are not statistically significant, although still hold the same direction of the signs. This could be because small-cap stocks experience fewer minute-to-minute variations in microstructure effects, as well as minute-to-minute changes in their restrictions by tick size. The effect from average price is negative for large-cap at t-value of -2.93, but a positive effect for both mid-cap and small-cap, with t-values of 2.51 and 2.09, respectively. This shows that within the large-cap category—in which most stocks experience more quoting and trading activities, and thus the quoted spread is closer to one tick—the higher the price, the lower the tick size constraint, and the more likely it is for crossing to occur. For mid-cap and small-cap, however, where the quoted spread is wider and is less restricted by tick size, the positive relationship with statistical significance is associated with higher levels of activities and utilitarian trading interest. It is interesting to observe consistent negative relationships between relative cross frequency against Rel Qspread and Qspread StdDev across the three strata, which control for variations in market microstructure risk. I do not anticipate average relative quoted spread or the standard deviation of this variable to change significantly on a minute-to-minute basis throughout the course of a day (in this case, 11:00 a.m. to 3:00 p.m.), especially when averaged across all stocks within each strata. Nevertheless, these two relationships indicate the lower the bid-ask spread, and the less the bid-ask spread varies within the one-minute window, the more likely it is for crossing to occur. This is worth further testing and consideration in a cross-sectional context. Turning to the right-hand side of Table 15
for the relative cross magnitude, the relationships are much simpler. None of the control
variables are statistically significant except the consistent negative relationship against relative
quoted spread. This indicates that the smaller the relative quoted spread, the greater the cross
magnitude. This inverse relationship is expected because the relative cross magnitude is the
absolute cross magnitude in dollar amount as defined in Chapter 4: Data and Methodology as a
percentage of the relative quoted spread.

After all the control variables are taken into consideration, Table 15 shows statistical
significance from \textit{QQQ Volatility} only in the large-cap category for relative cross frequency, but
no statistical significance for other strata or the full sample for relative cross frequency, and none
for relative cross magnitude. The only statistical significance is shown with a t-value of -2.78,
which is the opposite sign from what I expect if news and general market volatility have an
impact on the frequency and magnitude of crossing. I fail to reject the null of the first competing
alternative hypothesis on market volatility.

5.2.2. Cross-Sectional: Information Quality

From the second to the fifth competing alternative hypotheses, I test various market risk
factors in their cross-sectional effect, and then wrap up with the last competing alternative
hypothesis that combines all these cross-sectional factors into a full model. In this particular test,
I examine the impact from the quality of information on the cross-sectional differences in
relative cross frequency and relative cross magnitude. The variables of interest are two dummy
variables: \textit{If US} and \textit{If Primary}. \textit{If US} indicates whether the company is based in the United
States, collected from Bloomberg. \textit{If Primary} indicates whether NASDAQ is the primary listing
venue, and this is the primary ticker symbol for the stock. For both dummy variables, 1 indicates
true, 0 indicates false. They are used as proxies for the quality and reliability of the company's
reported accounting information. This risk factor is also related to the industry, in which the
company conducts business, and whether the company is domestic or international—
international stocks are considered more risky because they are subject to country-specific risks
and different accounting standards. Lack of reliable information leads to a higher level of risk
and higher level of volatility in the stock price.

\[
\ln(\text{Rel. Freq}_i) \text{ or } \ln(\text{Rel. Mag}_i) = \beta_1 \text{IfUS}_i + \beta_2 \text{IfPrimary}_i + \beta_3 \ln(\text{AvgMktCap}_i) + \beta_4 \ln(\text{NumTrades}_i) \\
+ \beta_5 \ln(\text{AvgPrice}_i) + \beta_0 + \epsilon_i
\]

\[H_0: \beta_1 = 0, \beta_2 = 0\]

\[H_1: \beta_1 \neq 0, \beta_2 \neq 0\]

I expect negative coefficients and t-values between both dummies and the relative cross
frequency and relative cross magnitude, if information quality and reliability have a statistically
significant impact over relative cross frequency and relative cross magnitude. \textit{Avg Mkt Cap}, \textit{Num
Trades}, and \textit{Avg Price} are control variables to manage cross-sectional differences between
stocks. \textit{Avg Mkt Cap} is market capitalization in U.S. dollar amount based on daily closing price
collected from Bloomberg, and then taken on average over the sample period. As demonstrated
in all summary statistics tables presented so far, almost all variables show a certain level of
correlation with \textit{Avg Mkt Cap}. \textit{Num Trades} is the total number of transaction counts to control
for cross-sectional differences in utilitarian trading interest. \textit{Avg Price} is the mean of all
transaction prices in U.S. dollars, as control for tick size constraint. The higher the average stock
price, the lower its inverse amount, and the less constrained it is by tick size. As discussed in the
previous test, tick size constraint is also reflected in other contributing factors like \textit{Num Trade}
and \textit{Avg Mkt Cap}. Similar to the time series test, I expect the control variables to be statistically
significant; however, the signs may vary among the large-cap, mid-cap, and small-cap strata due to the differences, or even structural shifts in combination of utilitarian trading interest and tick size constraint.

Table 16 provides regression results in separate panels for the full sample, the large-cap, the mid-cap, and the small-cap. The left-hand side of the table provides the coefficient of estimate, standard error, t-value, and p-value for the relative cross frequency, while the right-hand side of the table provides the same set of statistics for relative cross magnitude. As discussed earlier, I use every cross by group count for all regression analysis because it is a balanced representation of neighbor cross and first cross, and it drops the least number of data points out of the sample because there are only six stocks (in both individual count and group count) to which every cross never occurred. This is also why the full sample in the regression tables shows the number of observations to be 1,665 instead of 1,671. The control variables—Num Trades and Avg Price—are statistically significant for both relative cross frequency and relative cross magnitude for large-cap stocks at 99% and 99.9% confidence. The mid-cap and small-cap strata show similar patterns on the control variables, although at lower confidence levels. Avg Mkt Cap has a negative correlation with relative cross frequency for large-cap stocks, but positive for mid-cap stocks, and not statistically significant for small-cap stocks. Instead, it is consistently significant and shows a positive relationship with relative cross magnitude. This indicates that the higher the market cap amount, the farther (or in the relative term, more number of ticks) the crossing is likely to occur, although the likelihood of occurrence is more mixed. This is not because stocks with higher market cap have higher prices, as the control variable Avg Price also shows consistent statistical significance; instead, with opposite signs for relative cross magnitude, the lower the stock price, the farther the crossing magnitude is likely to be. The
impact from \textit{Avg Price} on relative cross frequency is mixed, showing a positive relationship at 99.9\% confidence in the large-cap category, but negative signs and no statistical significance in the mid-cap and small-cap categories.

Once the effects from the control variables are taken into consideration, Table 16 shows statistical significance from \textit{If US} only once in the large-cap category for relative cross magnitude at 99\% confidence, but no statistical significance for other strata for relative cross magnitude, and none for relative cross frequency. This statistical significance is shown with a t-value of 2.67, which is the opposite sign with what I expect if being a domestic U.S. company with higher quality of information has an impact on the frequency and magnitude of crossing. The \textit{If Primary} dummy variable does not for once show statistical significance across the spectrum. I fail to reject the null of the second competing alternative hypothesis on information quality.

5.2.3. Cross-Sectional: Insider Holding

In this section, I test the cross-sectional impact from insider holding on the relative cross frequency and magnitude. I collect both the number of common stock shares outstanding and common stock shares float from Bloomberg, and calculate the ratio between them, \textit{Float/Outstanding}, and average the ratio across the five days of the data sample. The number of shares float are the shares available for trade by the general public. The rest of the shares outstanding are closely held by insiders, institutions, major stakeholders or employees, and the restricted shares within the lock-in period. I use this metric as a proxy for the likelihood that an execution was initiated by an insider. The higher this ratio, the lower the likelihood that a trade is initiated by an insider. A histogram on the distribution of this variable is provided in Figure 21. If the level of market risk as the probability of trades initiated by insiders contributes to the
occurrences of the crossing episodes, I expect $\beta_1 \neq 0$ with statistical significance and both negative coefficient and negative t-value. The choice of control variables $\text{Avg Mkt Cap}, \text{Num Trades}$, and $\text{Avg Price}$ and their corresponding reasoning is the same as the previous test on information quality.

$$\ln(\text{Rel. Freq}_i) \text{ or } \ln(\text{Rel. Mag}_i)$$

$$= \beta_1 \ln\left(\frac{\text{Float}_{i}}{\text{Outstanding}_{i}}\right) + \beta_2 \ln(\text{AvgMktCap}_i) + \beta_3 \ln(\text{NumTrades}_i) + \beta_4 \ln(\text{AvgPrice}_i) + \beta_0 + \epsilon_i$$

$H_0: \beta_1 = 0$

$H_1: \beta_1 \neq 0$

Table 17 provides results on this cross-sectional linear regression on the relative cross frequency and magnitude across stocks within the full sample, and the large-cap, mid-cap, and small-cap strata. The control variables show mixed significance and signs, similar to the previous test, and mostly consistent with the previous test. Once the effects from the control variables are taken into consideration, Table 17 shows statistical significance from $\text{Float/Outstanding}$ only once in the mid-cap category for relative cross magnitude at 99% confidence, but no statistical significance for other strata for relative cross magnitude, and none for relative cross frequency. This statistical significance is shown with a t-value of 2.59, which is the opposite sign from what I expect if the probability of trades initiated by insiders contributes to the occurrences of the crossing episodes. I fail to reject the null of the third competing alternative hypothesis on insider holding.
5.2.4. Cross-Sectional: Analyst Coverage

In this section, I test the cross-sectional impact from valuation risk (i.e., analyst coverage on the relative cross frequency and magnitude). I collect the number of analyst recommendations \textit{Num Analyst} on earnings per share (EPS) forecasts on each sample stock from the IBES database from Wharton Research Data Services (WRDS). This number ranges from zero to 62. A histogram on the distribution of this variable is provided in Figure 19. Additionally, the \textit{Num Analyst} shows a strong positive correlation with \textit{Avg Mkt Cap}, which is shown in Figure 22. As expected, more active stocks with higher market capitalization receive more analyst coverage and utilitarian trading interest from the general public. Additionally, I also collect the standard deviation between the forecast EPS by different analyst \textit{EPS StdDev} reports from the same IBES database. This represents the diversity of opinions among security analysts on the value of the stock. This number ranges from zero to 2.12. This disagreement measure is the largest for the most active stock and trends less for less active stocks because a sufficient number of analyst coverage is necessary for disagreement to exist. This relationship is explored in Figure 20 in Appendix A. Furthermore, I break \textit{Num Analyst} into categorical dummies, at the ranges of \(=0\), \(=1\), \([2,5]\), \([6,10]\), \([11,20]\), \([20,40]\), \([41,62]\). For all seven categorical variables, 1 indicates true, 0 indicates false. Again, \textit{Avg Mkt Cap}, \textit{Num Trades}, and \textit{Avg Price} are control variables calculated for each stock.
\[
\ln(\text{Rel. Freq}_i) \text{ or } \ln(\text{Rel. Mag}_i) = \beta_1 \text{If Num Analyst}_0 + \beta_2 \text{If Num Analyst}_1 + \beta_3 \text{If Num Analyst}_5 \\
+ \beta_4 \text{If Num Analyst}_0 + \beta_5 \text{If Num Analyst}_20 + \beta_6 \text{If Num Analyst}_40 \\
+ \beta_7 \text{If Num Analyst}_\text{More}_i + \beta_9 \text{EPS StdDev}_i + \beta_9 \ln(\text{Avg Mkt Cap}_i) \\
+ \beta_{10} \ln(\text{Num Trades}_i) + \beta_{11} \ln(\text{Avg Price}_i) + \beta_0 + \epsilon_i
\]

\[H_0: \beta_1 = 0, \beta_2 = 0, \beta_3 = 0, \beta_4 = 0, \beta_5 = 0, \beta_6 = 0, \beta_7 = 0, \beta_9 = 0\]

\[H_1: \beta_1 \neq 0, \beta_2 \neq 0, \beta_3 \neq 0, \beta_4 \neq 0, \beta_5 \neq 0, \beta_6 \neq 0, \beta_7 \neq 0, \beta_9 \neq 0\]

I expect the control variables to continue to show a mixed level of significance like they have in the previous three tests. I expect some—if not all—of the categorical variables on If Num Analyst to have statistical significance. The more analysts cover the stock, and the less they disagree, the lower the valuation risk. This is expressed in higher Num Analyst, or the higher categories of If Num Analyst, and lower EPS StdDev. If valuation risk contributes to the relative cross frequency and magnitude, I expect the If Num Analyst to show statistical significance in the lower-level categories like when If Num Analyst=0 and when If Num Analyst=1, as well as EPS StdDev to test statistically significant but positive signs.

Table 18 provides the results on this cross-sectional regression within the full sample, and the large-cap, mid-cap, and small-cap strata. Due to the construction of the categorical variables, not every market cap stratum has at least one stock present in each of the categories. For example, there is no stock in the large-cap category that receives zero analyst coverage, and no stock receives analyst coverage between the numbers of 41 and 62. This all-zero-value independent variable can lead to the number of least-squares solutions for the parameters not being unique, some of the regression statistics being misleading, and the estimates being biased.
In these situations, I dropped the corresponding variables in the stratified regression samples, as they are indicated with zero coefficient and missing standard error, t-value, and p-value. Once the effects from the control variables are taken into consideration, both large-cap and mid-cap stocks show the pattern of statistical significance in the positive sign for relative cross frequency as the number of analysts increases. This statistical significance for Num Analyst is the opposite sign from what I expect if analyst coverage—thus valuation risk—has an impact on the frequency and magnitude of crossing. None of the categorical variables show statistical significance for relative cross magnitude in these two strata. Nor do they show statistical significance for small-cap stocks in either relative frequency or relative magnitude. EPS StdDev shows statistical significance only once at 95% confidence for relative cross frequency in the mid-cap stratum, and none for other market cap strata or the full sample, and never for relative cross magnitude. Based on this observation, I fail to reject the null of the fourth competing alternative hypothesis on analyst coverage.

5.2.5. Cross-Sectional: Microstructure Effect

In this section, I provide the last standalone cross-sectional test as a competing alternative hypothesis. I examine the impact from microstructure effect on the relative cross frequency and cross magnitude across stocks within the full sample, and the large-cap, mid-cap, and small-cap strata. As the readers have seen in the summary statistics in previous tables and the simulation exercise, the quoted spread and liquidity level vary tremendously across stocks. To provide liquidity, the market maker faces the information asymmetry risk from trading against informed traders, and the inventory risk from the security price moving away while they are holding net inventory. Average duration-weighted quoted spread has been established by microstructure
literature—much of it summarized in Chapter 2: Literature Review—to represent both these types of risks. I use the relative quoted spread \( Rel \ Qspread \) and the standard deviation of the quoted spread \( Qspread \ StdDev \) as proxies for information asymmetry and microstructure noise to capture cross-stock differences.

\[
\ln(\text{Rel. Freq}_i) \text{ or } \ln(\text{Rel. Mag}_i) = \beta_1 \ln(\text{RelQspread}_i) + \beta_2 \ln(\text{QspreadStdDev}_i) + \beta_3 \ln(\text{AvgMktCap}_i)
+ \beta_4 \ln(\text{Num Trades}_i) + \beta_5 \ln(\text{Avg Price}_i) + \beta_0 + \epsilon_i
\]

\( H_0: \beta_1 = 0, \beta_2 = 0 \)

\( H_1: \beta_1 \neq 0, \beta_2 \neq 0 \)

I expect statistical significance from both \( Rel \ Qspread \) and \( Qspread \ StdDev \), and the signs to be positive: the higher the relative quoted spread, the larger the microstructure risk from information asymmetry and inventory risk, and the more this quantity changes, the more often the crossing episodes are expected to occur, in the case where information asymmetry and microstructure noise contribute to the occurrences of the crossing episodes. Again, \( \text{Avg Mkt Cap} \), \( \text{Num Trades} \), and \( \text{Avg Price} \) are control variables calculated for each stock. Turning to Table 19 for results of this cross-sectional regression, I see that indeed both \( Rel \ Qspread \) and \( Qspread \ StdDev \) are consistently statistically significant, but with the coefficient and t-value negative instead of positive for \( Rel \ Qspread \), even though the signs for \( Qspread \ StdDev \) are positive as expected. This result indicates that even after controlling for cross-sectional differences by \( \text{Avg Mkt Cap} \), \( \text{Num Trades} \), and \( \text{Avg Price} \), the smaller the relative quoted spread, not only is it more likely that the crossing episodes will occur, they also occur in a larger magnitude. It is worth noting that \( Rel \ Qspread \) and \( Qspread \ StdDev \) also appear in the first competing alternative hypothesis, where the sign for \( Rel \ Qspread \) is negative (the same applies in this hypothesis), but
the sign for \( Qspread \) \( StdDev \) is also negative (the opposite in this hypothesis). With the mixed observation, I fail to reject the null of the fifth competing alternative hypothesis on microstructure effect. However, the consistent statistical significance for quoted spread with a negative sign is worth further consideration.

### 5.2.6. Cross-Sectional: Full Model

In the last cross-sectional test, I combine all previous market risk factors into one large regression that I call the full model. The purpose of this regression is to examine the relative strength and impact between information quality, insider holding, analyst coverage, and microstructure effect with their impact on the relative cross frequency and relative cross magnitude. So far I have not been able to reject any of these market risk factors.

\[
\ln(\text{Rel. Freq}_i) \text{ or } \ln(\text{Rel. Mag}_i) = \beta_1 \text{IfUS}_i + \beta_2 \text{IfPrimary}_i + \beta_3 \ln\left(\frac{\text{Float}}{\text{Outstanding}}\right)_i + \beta_4 \text{IfNumAnalyst}_0i \\
+ \beta_5 \text{IfNumAnalyst}_1i + \beta_6 \text{IfNumAnalyst}_5i + \beta_7 \text{IfNumAnalyst}_10i \\
+ \beta_8 \text{IfNumAnalyst}_20i + \beta_9 \text{IfNumAnalyst}_40i + \beta_{10} \text{IfNumAnalystMore}_i \\
+ \beta_{11} \text{EPSStdDev}_i + \beta_{12} \ln(\text{RelQspread}_i) + \beta_{13} \ln(\text{QspreadStdDev}_i) \\
+ \beta_{14} \ln(\text{AvgMktCap}_i) + \beta_{15} \ln(\text{NumTrades}_i) + \beta_{16} \ln(\text{AvgPrice}_i) + \beta_0 + \epsilon_i
\]

\( H_0: \beta_1 = 0, \beta_2 = 0, \beta_3 = 0, \beta_4 = 0, \beta_5 = 0, \beta_6 = 0, \beta_7 = 0, \beta_8 = 0, \beta_9 = 0, \beta_{10} = 0, \beta_{11} = 0, \beta_{12} = 0, \beta_{13} = 0 \)

\( H_1: \beta_1 \neq 0, \beta_2 \neq 0, \beta_3 \neq 0, \beta_4 \neq 0, \beta_5 \neq 0, \beta_6 \neq 0, \beta_7 \neq 0, \beta_8 \neq 0, \beta_9 \neq 0, \beta_{10} \neq 0, \beta_{11} \neq 0, \beta_{12} \neq 0, \beta_{13} \neq 0 \)
Table 20 provides the results on this cross-sectional linear regression on the relative cross frequency and magnitude across stocks within the full sample, and the large-cap, mid-cap, and small-cap strata. Comparing the results between the full model and the four previous standalone cross-sectional tests, the significance level has decreased for some variables. For example, If Num Analyst $\in [21,40]$ for the large-cap category has the t-value of 4.20 on relative cross frequency in the standalone model, but only 2.7 in the full model. The confidence level for this test dropped from 99.9% to 99%; however, the sign has stayed consistent. Similarly, while comparing the second standalone test on insider holding, the only statistical significance among all tests is found in mid-cap relative cross magnitude with a t-value of 2.59 with 99% confidence. This variable in the full model test drops to a t-value of 2.2 with 95% confidence. This effect is expected when more independent variables that have positive correlation with each other join in the regression. It is also interesting to note that some variables are not statistically significant in the standalone model, but are subsequently exposed as being significant in the full model when potentially more controls have been put in place. Reviewing the tests on Float/Outstanding on relative cross frequency, none of the four tests show up as significant. However, in the full model, this factor is shown as statistically significant with a t-value of -2.81 for the full sample with 99% confidence, a t-value of -2.22 for the large-cap stratum with 95% confidence, and a t-value of -4.56 for the mid-cap stratum with 99.9% confidence. This negative relationship is in the direction of my earlier expectation, which indicates that the possibility of orders from insiders could be a contributing factor for the occurrences of crossing episodes for large-cap and mid-cap stocks, even though it is still not statistically significant for small-cap stocks. The control variables show a mixed significance and signs throughout the cross-sectional tests within each test, but have shown consistency comparing across tests as I use the same control variables.
Comparing the strength of the four market risk factors in question—information quality, insider holding, analyst coverage, and market microstructure effect on the context of the full model—the last factor (microstructure) is by far the strongest and dominates the others, even though none of the four factors tested ever show statistical significance. Across all models and all tests performed on the impact of relative cross frequency, the adjusted $R^2$ is the highest for the large-cap stratum in the full model at 34.64%, while the adjusted $R^2$ is below 10% for many categories, including full sample and mid-cap for information quality, full sample and mid-cap for insider holding, and full sample and mid-cap for analyst coverage. Microstructure effects have the highest consistent level of adjusted $R^2$ at 23.59% for the full sample, 31.73% for large-cap, 29.55% for mid-cap, and 30.35% for small-cap. Overall, the cross-sectional test variables and control variables collectively have better explanatory power in the large-cap and small-cap categories than in mid-cap. Turning to relative cross magnitude, the adjusted $R^2$ is consistently the highest for the large-cap category at 53.95% for the full model. The lowest adjusted $R^2$ for relative cross magnitude is 6.05% at the small-cap stratum in the analyst coverage test. Overall, the various market risk factors each have some level of explanatory power over the occurrences of the crossing episodes, both in frequency and in magnitude, but leave a large part of the significance unexplained. This is consistent with the second null hypothesis at the beginning of this chapter, which indicates that the presence of Knightian uncertainty contributes to the occurrences of the crossing episodes, which is consistent with the implications from the theoretical model in Easley and O’Hara (2010b).

As observed while comparing the full model and each stand-alone regression, the independent variables from one test to the next are correlated and not entirely exogenous. A classic example is that volume (or trade count) is frequently used as an independent variable.
while studying bid-ask spread. So we are assuming the causal relationship goes one way but not the other. Is it, however, only that higher volume leads to lower transaction cost, or that stocks with lower transaction cost attract more volume? I make my best effort to resolve this conflict by keeping highly correlated independent variables in separate tests and keeping each test simple; this is why I use smaller separate regressions before integrating all factors into a full model.
Chapter 6. Conclusions

6.1. Summary and Significance

Classical market microstructure models assume that the expected value of the true price lies within the range of the bid-ask interval. Because the expected true price (or fair price) is a theoretical construct and is unobservable, when it comes to empirical research, the bid-ask midpoint is its most commonly used empirical proxy. I observe from empirical data of frequently occurring instances that the expected true price is clearly outside the range of the bid-ask interval, regardless of where the expected true price is. I find this occurs frequently across the board, from stocks with very low liquidity to the most active stocks. My findings not only conflict with one of the basic assumptions of theoretical and empirical microstructure research, but also challenge our understanding of the price formation mechanism on a limit order book. The implication of this research raises doubts about many asset and capital allocation models and investment decisions, all of which are based on the assumption that true price has been correctly modeled, and that the price discovery mechanism is relatively accurate and well understood.

Knightian uncertainty by definition cannot be quantified and measured, which makes empirical research on this topic extremely rare and difficult. In applying the uncertainty concept in an equity market microstructure setting and testing a group of hypotheses with high-frequency market data, this dissertation contributes to the empirical literature that documents the evidence of Knightian uncertainty in the equity market. Easley and O’Hara (2010b) construct a heterogeneous trader model attempting to explain a scenario where traders are not willing to execute at a wide range of price levels in the presence of quotations. This model implies a scenario where the expected true price could stand outside the range of the bid-ask interval in the
presence of Knightian uncertainty. I take this empirical scenario to empirical testing to see if the crossing episodes I observe are consistent with the theoretical prediction of the existence of Knightian uncertainty.

I first construct empirical measures to quantify the quote price behavior when true price is clearly outside the range of the bid-ask interval, both in frequency and in magnitude. T-tests against the benchmark of market noise show that these occurrences are statistically significant, which is contrary to the prediction from classical microstructure models, but consistent with the implications of a Knightian uncertainty scenario. To provide such a benchmark, I build a Monte Carlo simulation of the limit order book based on methodology suggested by Bouchaud, Mézard, and Potters (2002); Zovko and Farmer (2002); and Cont, Stoikov, and Talreja (2010). Next, I investigate the cause for this price behavior. Because Knightian uncertainty by definition cannot be observed and quantified, I design and test a series of alternative hypotheses as a process-of-elimination to indirectly provide evidence to the hypothesis that the crossing behavior of quote prices is consistent with a Knightian uncertainty model. I test the frequency and magnitude of the crossing occurrences in one time series regression, four cross-sectional regressions, and one full sample to examine market risk factors including market volatility, information quality, insider holding, analyst coverage, and market structure effect. I fail to reject the competing alternative hypotheses consistently for other market factors, even though certain market risk factors show up positive for either frequency or magnitude in a subset of market cap strata.

Market microstructure as a subfield in finance plays a more important role, as security trading is occurring at increasingly short time intervals. This research identifies a phenomenon—which is constantly occurring across all stocks at all hours—that violates one of the basic assumptions of microstructure theory. Identifying and investigating the cause of the incidents
will provide insight into one of the most fundamental questions in price modeling in general. The empirical evidence I provide in this study shows that crossings are common and prevalent among stocks from all market cap categories. After controlling information and risk factors, they would still occur to a large degree. The empirical evidence I provide suggests it is possible for this type of quote price behavior to occur in the presence of uncertainty—but that does not exclude other possible reasons besides the market factor reasons I test. Nevertheless, the crossing of quote prices leads to us to reconsider the price discovery mechanism on a limit order book from existing theory, to further advance studies in this field, and to eventually incorporate Knightian uncertainty to become part of a standard microstructure framework in order to better understand the impact of uncertainty on price discovery.

As technology advances and information availability improves, it becomes increasingly important to be able to correctly price a security at an increasingly short time interval. Deviations between the market price and the expected true price also occur in shorter time intervals; for instance, the Flash Crash of May 6, 2010, is but one of many examples where security prices crash and quickly rebound, accompanied by a shortage in liquidity, high trading volume, and extraordinary volatility. In even shorter time intervals (within minutes or seconds), the market price can encounter a volatile ride due to microstructure noise and momentary supply and demand imbalances. In theory, the equilibrium price is determined by supply and demand. A stock or any other financial asset is worth the price for which it can be bought or sold at any

\[\text{stock} \times \text{price} = \text{worth}\]

\(^{86}\) The cause of the Flash Crash has drawn significant attention from the public, researchers, and policy makers. For more detailed discussions regarding the event, see the SEC publication “Recommendations Regarding Regulatory Responses to the Market Events of May 6, 2010” (2011), and “The Microstructure of the ‘Flash Crash’: Flow Toxicity, Liquidity Crashes and the Probability of Informed Trading” (2010) by Easley, de Prado, and O’Hara; see also “The Flash Crash: The Impact of High Frequency Trading on an Electronic Market” (2014) by Kirilenko, Samadi, Kyle, and Tuzun.
given time. As instances like the 2010 Flash Crash become common, both our conceptual understanding of the expected true price of the security and the way we model it are being challenged. In this dissertation, I review how the expected true price has been considered historically in finance and economics, and I document empirical evidence that the expected true price—if modeled under the classical microstructure framework—would not work, regardless of where it actually is. This dissertation raises the question of where the true price actually is, or whether there is a better way to define it, because in presence of uncertainty the traditional definition of true price falls apart. The ways of market making, accessing liquidity, and price discovery have changed rapidly in recent decades. New empirical evidence like this call attention to new theoretical settings, modeling tools, and empirical tests to be brought into microstructure research to design better market structure and protect investors from manipulation strategies.

6.2. Future Research

This study is exploratory in nature, and many interesting future research directions can be spun forward from its foundation. In this section I present a few ideas: In the Chapter 3: Theory, I use the static model from Easley and O’Hara (2010b) because it keeps the model simple and captures a scenario where it is possible for true price to be sitting outside the bid-ask interval. However, an important aspect of the crossing episodes I observe is the actual mechanism of price updates. A Bayesian learning dynamic model can perhaps better explain why quotes move between trades, and make sense of the manner in which they move. This model also introduces the impact of uncertainty on price in the most simplistic sense to turn the expected mean from a single, specific value into an unknowable range. It is worthwhile to look at price impact if I do the same for variance or other aspects of the distribution.
The empirical section of the dissertation processes tick-by-tick market data, which has millions of observations. However, with the day-by-day aggregation, an extraordinarily small statistical sample size of five is produced. One possible solution I consider is to chop up each day’s data into five minute- or hour-long chunks, and treat each five minutes for each stock as an individual data point. Following this approach, I conduct fixed-effect panel regressions without much success. The primary concern with that is controlling the length of the time window. In order to capture transient effects within short time intervals, I need to keep the time window relatively short; however, that results in a majority of zero values for most individual stocks in most time windows for the test statistics. Lengthening the time window, however, eliminates the possibility of capturing short time effects. A fixed-effect regression would also limit the scope to a time-based instead of an event-based framework, even though the crossing metrics are—by construction—event-based.

Further research can potentially shed light on whether the market maker truly intends to trade while posting the limit order: access to order-book-level data, additional proxies like quote-to-trade ratio, quote cancellation rates, the total number of active market makers providing liquidity, and the sequence of events on limit order posting versus cancelling. With order level data, I can also look at whether the conflicting/crossing quote records are posted consistently by a single market marker or a small group of competing market makers. This would be worthwhile because the according to the current fee structure, exchanges may post and cancel for free, and only pay to trade. If the market makers do not trade, they cannot profit. Some might specialize in posting and then canceling if they detect an informed trader coming along or observe the price moving away from their quotes. This may incentivize a very high quote-to-trade ratio. Additionally, it would be interesting to test and see if the quote movements are in the general
direction of returns, and thus part of the price discovery process. This may prove positive for some less active stocks with less trading interest from the general public—and therefore the market makers are adjusting the quote price levels as part of the price discovery process. The longer the time periods of no trades, the more likely it is for this scenario to occur. The quote updates between trades could also be a type of learning behavior. Another interesting experiment would be to conduct the same empirical tests on over-the-counter securities, where presumably higher levels of both risk and uncertainty exist.

The Monte Carlo simulation that precedes empirical testing is for the sole purpose of providing a more robust benchmark of market noise compared to zero. However, the simulation itself presents interesting features. Without access to order-book-level data, I cannot properly estimate the parameters for some of the power-law distributions. Upon examining all three approaches by Bouchaud, Mézard, and Potters (2002), Zovko and Farmer (2002), and Cont, Stoikov, and Talreja (2010), I choose the approach by Zovko and Farmer (2002) as a starting point from which to expand, because the features produced by this model most closely resemble the three sample stocks I calibrate to. It is not surprising because the data sample Zovko and Farmer (2002) acquire from the London Stock Exchange has a balance between high-and low-volume stocks, while Bouchaud, Mézard, and Potters (2002) use the three most active stocks from Paris Bourse, and Cont, Stoikov, and Talreja (2010) use a large-cap stock from the Tokyo Stock Exchange. However, Zovko and Farmer (2002) use the time period of August 1, 1998, to April 31, 2000. Not only could potential differences exist between the U.S. markets and the London Stock Exchange, but the nature of electronic trading and the types of market participants are also likely to have changed in the past 15 years. It would be interesting and worthwhile to acquire recent order-book-level equity market data in the United States to examine whether such
power-law distribution or the parameters I borrow from Zovko and Farmer (2002) are appropriate.
Appendix A. Tables and Figures


<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>9:30:00.0000</td>
<td>$50.80</td>
<td>$51.20</td>
<td>$0.40</td>
<td>$51.000</td>
<td></td>
<td>0</td>
<td>.</td>
<td>1</td>
<td>$0.160</td>
</tr>
<tr>
<td>9:30:00.180</td>
<td>$50.72</td>
<td>$51.20</td>
<td>$0.48</td>
<td>$50.960</td>
<td>$51.000</td>
<td>0</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>9:30:00.180</td>
<td>$50.77</td>
<td>$50.80</td>
<td>$0.03</td>
<td>$50.785</td>
<td>$50.960</td>
<td>1</td>
<td>$0.160</td>
<td>.</td>
<td>.</td>
</tr>
</tbody>
</table>

*Note.* Freq. = Frequency; Mag. = Magnitude.

Figure 1. *SBUX* Quotes at 9:30:00. November 5, 2012. Neighbor Cross.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>9:39:46.710</td>
<td>$51.15</td>
<td>$51.18</td>
<td>$0.03</td>
<td>0</td>
<td></td>
<td>1</td>
<td>$0.010</td>
</tr>
<tr>
<td>9:39:47.100</td>
<td>$51.15</td>
<td>$51.17</td>
<td>$0.02</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9:39:47.101</td>
<td>$51.14</td>
<td>$51.17</td>
<td>$0.03</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9:39:47.160</td>
<td>$51.14</td>
<td>$51.16</td>
<td>$0.02</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9:39:47.773</td>
<td>$51.14</td>
<td>$51.15</td>
<td>$0.01</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9:39:47.773</td>
<td>$51.13</td>
<td>$51.15</td>
<td>$0.02</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9:39:47.775</td>
<td>$51.13</td>
<td>$51.14</td>
<td>$0.01</td>
<td>2</td>
<td>$0.10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9:39:47.776</td>
<td>$51.13</td>
<td>$51.15</td>
<td>$0.02</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. Freq. = Frequency; Mag. = Magnitude.

Figure 2. SBUX Quotes at 9:39:46. November 5, 2012. Strict Cross.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>9:31:09.189</td>
<td>$50.73</td>
<td>$50.81</td>
<td>$0.08</td>
<td>$50.770</td>
<td>0</td>
<td>.</td>
<td>1</td>
<td>$0.030</td>
</tr>
<tr>
<td>9:31:09.189</td>
<td>$50.80</td>
<td>$50.81</td>
<td>$0.01</td>
<td>$50.770</td>
<td>1</td>
<td>$0.030</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>9:31:09.189</td>
<td>$50.73</td>
<td>$50.81</td>
<td>$0.08</td>
<td>$50.770</td>
<td>0</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>9:31:09.189</td>
<td>$50.80</td>
<td>$50.81</td>
<td>$0.01</td>
<td>$50.770</td>
<td>1</td>
<td>$0.030</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>9:31:09.189</td>
<td>$50.73</td>
<td>$50.81</td>
<td>$0.08</td>
<td>$50.770</td>
<td>0</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>9:31:09.189</td>
<td>$50.79</td>
<td>$50.81</td>
<td>$0.02</td>
<td>$50.770</td>
<td>1</td>
<td>$0.020</td>
<td>.</td>
<td>.</td>
</tr>
</tbody>
</table>

Note. Freq. = Frequency; Mag. = Magnitude.

Figure 3. SBUX Quotes at 9:31:09. November 5, 2012. First Cross.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>9:39:46.710</td>
<td>$51.15</td>
<td>$51.18</td>
<td>$0.03</td>
<td>$51.165</td>
<td>0</td>
<td></td>
<td>1</td>
<td>$0.015</td>
</tr>
<tr>
<td>9:39:47.100</td>
<td>$51.15</td>
<td>$51.17</td>
<td>$0.02</td>
<td>$51.160</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9:39:47.101</td>
<td>$51.14</td>
<td>$51.17</td>
<td>$0.03</td>
<td>$51.155</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9:39:47.160</td>
<td>$51.14</td>
<td>$51.16</td>
<td>$0.02</td>
<td>$51.150</td>
<td>1</td>
<td>$0.005</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9:39:47.773</td>
<td>$51.14</td>
<td>$51.15</td>
<td>$0.01</td>
<td>$51.145</td>
<td>3</td>
<td>$0.010</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9:39:47.773</td>
<td>$51.13</td>
<td>$51.15</td>
<td>$0.02</td>
<td>$51.140</td>
<td>3</td>
<td>$0.010</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9:39:47.775</td>
<td>$51.13</td>
<td>$51.14</td>
<td>$0.01</td>
<td>$51.135</td>
<td>5</td>
<td>$0.015</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9:39:47.776</td>
<td>$51.13</td>
<td>$51.15</td>
<td>$0.02</td>
<td>$51.140</td>
<td>3</td>
<td>$0.010</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. Freq. = Frequency; Mag. = Magnitude.

Figure 4. SBUX Quotes at 9:39:47. November 5, 2012. Every Cross.
Table 5. Descriptive Stock Characteristics.

This table provides descriptive statistics of sample stocks. The sample dataset is consisted of: UTP reported trades from all markets on 1671 NASDAQ-listed securities; compiled NASDAQ best bid and offer (QBBO) based on UTP reported NASDAQ quotes on these stocks. The sample period is from 11/05/2012 to 11/09/2012. *Avg Mkt cap* is market capitalization in US dollar amount based on daily closing price collected from Bloomberg, and then average over the sample period. The large-cap stratum contains stocks with market cap greater than $1 billion; the mid-cap stratum contains stocks with market cap between $100 million and $1 billion; the small-cap stratum contains stocks with market cap less than $100 million. *Avg Price* is the average transaction price in US dollars. *Volume* is the average number of shares traded for the day. *Num Trades* is the total number of transactions for the day. *Turnover* is the average dollar amount traded in US dollars. *Quoted spread* is the duration-weighted average difference between the best bid and offer in US dollars. All measures except market cap are calculated with quote and trade records during market open hours between 11am and 3pm EST.

<table>
<thead>
<tr>
<th></th>
<th>Avg Mkt Cap ($ mil)</th>
<th>Avg Price ($)</th>
<th>Volume (Count)</th>
<th>Num Trades (Count)</th>
<th>Turnover ($)</th>
<th>Quoted Spread ($)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Full Sample (n=1671)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>$2,670.98</td>
<td>$20.82</td>
<td>403,059</td>
<td>1,846</td>
<td>$12,782,943</td>
<td>$0.11</td>
</tr>
<tr>
<td>Median</td>
<td>$391.60</td>
<td>$12.91</td>
<td>69,031</td>
<td>424</td>
<td>$840,542</td>
<td>$0.05</td>
</tr>
<tr>
<td>StdDev</td>
<td>$17,297.13</td>
<td>$35.24</td>
<td>1,505,809</td>
<td>4,705</td>
<td>$155,759,475</td>
<td>$0.14</td>
</tr>
<tr>
<td>Min</td>
<td>$10.70</td>
<td>$1.07</td>
<td>350</td>
<td>3</td>
<td>$1,953</td>
<td>$0.01</td>
</tr>
<tr>
<td>Max</td>
<td>$528,722.06</td>
<td>$670.67</td>
<td>27,923,628</td>
<td>66,765</td>
<td>$6,201,902,525</td>
<td>$1.35</td>
</tr>
<tr>
<td><strong>Large-Cap (n=456)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>$8,936.73</td>
<td>$43.29</td>
<td>1,158,298</td>
<td>5,192</td>
<td>$43,891,753</td>
<td>$0.07</td>
</tr>
<tr>
<td>Median</td>
<td>$2,420.06</td>
<td>$32.49</td>
<td>363,084</td>
<td>2,522</td>
<td>$10,757,784</td>
<td>$0.03</td>
</tr>
<tr>
<td>StdDev</td>
<td>$32,308.57</td>
<td>$59.39</td>
<td>2,714,401</td>
<td>7,945</td>
<td>$296,137,809</td>
<td>$0.12</td>
</tr>
<tr>
<td>Min</td>
<td>$1,002.12</td>
<td>$2.20</td>
<td>2,011</td>
<td>10</td>
<td>$69,468</td>
<td>$0.01</td>
</tr>
<tr>
<td>Max</td>
<td>$528,722.06</td>
<td>$670.67</td>
<td>27,923,628</td>
<td>66,765</td>
<td>$6,201,902,525</td>
<td>$1.02</td>
</tr>
<tr>
<td><strong>Mid-Cap (n=963)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>$387.85</td>
<td>$14.39</td>
<td>141,133</td>
<td>717</td>
<td>$1,369,736</td>
<td>$0.11</td>
</tr>
<tr>
<td>Median</td>
<td>$307.96</td>
<td>$11.42</td>
<td>50,276</td>
<td>307</td>
<td>$568,234</td>
<td>$0.06</td>
</tr>
<tr>
<td>StdDev</td>
<td>$242.13</td>
<td>$11.60</td>
<td>274,886</td>
<td>1,093</td>
<td>$2,546,704</td>
<td>$0.14</td>
</tr>
<tr>
<td>Min</td>
<td>$101.74</td>
<td>$1.32</td>
<td>374</td>
<td>3</td>
<td>$7,324</td>
<td>$0.01</td>
</tr>
<tr>
<td>Max</td>
<td>$999.54</td>
<td>$103.36</td>
<td>3,840,311</td>
<td>9,642</td>
<td>$34,712,886</td>
<td>$1.35</td>
</tr>
<tr>
<td><strong>Small-Cap (n=252)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>$57.77</td>
<td>$4.73</td>
<td>37,370</td>
<td>107</td>
<td>$105,565</td>
<td>$0.15</td>
</tr>
<tr>
<td>Median</td>
<td>$57.37</td>
<td>$3.39</td>
<td>13,302</td>
<td>46</td>
<td>$52,373</td>
<td>$0.11</td>
</tr>
<tr>
<td>StdDev</td>
<td>$23.24</td>
<td>$3.71</td>
<td>81,798</td>
<td>207</td>
<td>$153,899</td>
<td>$0.16</td>
</tr>
<tr>
<td>Min</td>
<td>$10.70</td>
<td>$1.07</td>
<td>350</td>
<td>4</td>
<td>$1,953</td>
<td>$0.01</td>
</tr>
<tr>
<td>Max</td>
<td>$99.91</td>
<td>$24.10</td>
<td>699,681</td>
<td>1,784</td>
<td>$975,167</td>
<td>$1.14</td>
</tr>
</tbody>
</table>
This table provides summary statistics on variables of interest. All measures are calculated or collected on a daily basis and then averaged across the sample period between 11/05/2012 and 11/09/2012. *QBBO Updates* is the cumulative number of price updates on either the bid side or the ask side of the NASDAQ best bid and offer for each stock. *No-Trade Duration* is the median average length of time period during which no trading occurs. *Num Analyst* is the number of analyst recommendations on Earnings per share (EPS) forecasts on this stock, collected from the IBES database. *EPS StdDev* is the standard deviation between the forecasted EPS by different analysts, also collected from the IBES database. *Float/Outstanding* is the number of stock shares float divided by the number of shares outstanding for the stock, collected from Bloomberg. *QBBO Updates* and *No-Trade Duration* are calculated with quote and trade records during market open hours between 11am and 3pm EST.

<table>
<thead>
<tr>
<th>Calculated</th>
<th>Collected</th>
</tr>
</thead>
<tbody>
<tr>
<td>QBBO Updates</td>
<td>Num Analyst</td>
</tr>
<tr>
<td>(Count)</td>
<td>(Count)</td>
</tr>
<tr>
<td>No Trade Duration</td>
<td>EPS StdDev</td>
</tr>
<tr>
<td>(Second)</td>
<td>(Count)</td>
</tr>
</tbody>
</table>

**Full Sample (n=1671)**

<table>
<thead>
<tr>
<th></th>
<th>Calculated</th>
<th>Collected</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>1,416</td>
<td>9</td>
</tr>
<tr>
<td>Median</td>
<td>794</td>
<td>7</td>
</tr>
<tr>
<td>StdDev</td>
<td>2,266</td>
<td>8</td>
</tr>
<tr>
<td>Min</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Max</td>
<td>61,218</td>
<td>62</td>
</tr>
</tbody>
</table>

**Large-Cap (n=456)**

<table>
<thead>
<tr>
<th></th>
<th>Calculated</th>
<th>Collected</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>3,008</td>
<td>17</td>
</tr>
<tr>
<td>Median</td>
<td>2,342</td>
<td>16</td>
</tr>
<tr>
<td>StdDev</td>
<td>3,615</td>
<td>10</td>
</tr>
<tr>
<td>Min</td>
<td>31</td>
<td>1</td>
</tr>
<tr>
<td>Max</td>
<td>61,218</td>
<td>62</td>
</tr>
</tbody>
</table>

**Mid-Cap (n=963)**

<table>
<thead>
<tr>
<th></th>
<th>Calculated</th>
<th>Collected</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>939</td>
<td>7</td>
</tr>
<tr>
<td>Median</td>
<td>682</td>
<td>6</td>
</tr>
<tr>
<td>StdDev</td>
<td>953</td>
<td>4</td>
</tr>
<tr>
<td>Min</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Max</td>
<td>12,031</td>
<td>28</td>
</tr>
</tbody>
</table>

**Small-Cap (n=252)**

<table>
<thead>
<tr>
<th></th>
<th>Calculated</th>
<th>Collected</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>356</td>
<td>3</td>
</tr>
<tr>
<td>Median</td>
<td>158</td>
<td>2</td>
</tr>
<tr>
<td>StdDev</td>
<td>620</td>
<td>2</td>
</tr>
<tr>
<td>Min</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>Max</td>
<td>4,489</td>
<td>12</td>
</tr>
</tbody>
</table>
Table 7. Relative Cross Frequency. Individual count.

This table provides summary statistics on the constructed metrics: the relative frequency of the crossing episodes, as defined in Chapter 4.3. The relative frequency is the frequency count as percentage of the number of QBBO price updates. Individual count takes into consideration of each QBBO price update. All measures are calculated on a daily basis and then averaged across the sample period between 11/05/2012 and 11/09/2012. Only quote and trade records during market open hours between 11am and 3pm EST are used.

<table>
<thead>
<tr>
<th></th>
<th>Relative Cross Frequency</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Neighbor Cross (Percent)</td>
<td>First Cross (Percent)</td>
<td>Every Cross (Percent)</td>
<td>Strict Cross (Percent)</td>
</tr>
<tr>
<td><strong>Full Sample (n=1671)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>4.49%</td>
<td>8.85%</td>
<td>6.38%</td>
<td>0.06%</td>
</tr>
<tr>
<td>Median</td>
<td>3.82%</td>
<td>8.36%</td>
<td>6.15%</td>
<td>0.03%</td>
</tr>
<tr>
<td>StdDev</td>
<td>3.71%</td>
<td>5.52%</td>
<td>3.50%</td>
<td>0.12%</td>
</tr>
<tr>
<td>Min</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
<tr>
<td>Max</td>
<td>34.15%</td>
<td>43.53%</td>
<td>25.87%</td>
<td>2.17%</td>
</tr>
<tr>
<td><strong>Large-Cap (n=456)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>2.39%</td>
<td>7.13%</td>
<td>4.91%</td>
<td>0.08%</td>
</tr>
<tr>
<td>Median</td>
<td>2.43%</td>
<td>7.78%</td>
<td>5.23%</td>
<td>0.07%</td>
</tr>
<tr>
<td>StdDev</td>
<td>1.85%</td>
<td>4.13%</td>
<td>2.58%</td>
<td>0.08%</td>
</tr>
<tr>
<td>Min</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.08%</td>
<td>0.00%</td>
</tr>
<tr>
<td>Max</td>
<td>11.30%</td>
<td>22.95%</td>
<td>14.91%</td>
<td>0.67%</td>
</tr>
<tr>
<td><strong>Mid-Cap (n=963)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>4.69%</td>
<td>9.35%</td>
<td>6.72%</td>
<td>0.06%</td>
</tr>
<tr>
<td>Median</td>
<td>4.28%</td>
<td>8.89%</td>
<td>6.61%</td>
<td>0.03%</td>
</tr>
<tr>
<td>StdDev</td>
<td>3.12%</td>
<td>5.23%</td>
<td>3.29%</td>
<td>0.12%</td>
</tr>
<tr>
<td>Min</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
<tr>
<td>Max</td>
<td>20.52%</td>
<td>34.35%</td>
<td>20.32%</td>
<td>2.17%</td>
</tr>
<tr>
<td><strong>Small-Cap (n=252)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>7.50%</td>
<td>10.04%</td>
<td>7.73%</td>
<td>0.03%</td>
</tr>
<tr>
<td>Median</td>
<td>6.27%</td>
<td>8.39%</td>
<td>6.81%</td>
<td>0.00%</td>
</tr>
<tr>
<td>StdDev</td>
<td>5.57%</td>
<td>7.70%</td>
<td>4.67%</td>
<td>0.14%</td>
</tr>
<tr>
<td>Min</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
<tr>
<td>Max</td>
<td>34.15%</td>
<td>43.53%</td>
<td>25.87%</td>
<td>1.75%</td>
</tr>
</tbody>
</table>
This table provides summary statistics on the constructed metrics: the relative magnitude of the crossing episodes, as defined in Chapter 4.3. The relative magnitude is the cross magnitude measured in dollar amount as percentage of the quoted spread. Individual count takes into consideration of each QBBO price update. All measures are calculated on a daily basis and then averaged across the sample period between 11/05/2012 and 11/09/2012. Only quote and trade records during market open hours between 11am and 3pm EST are used.

<table>
<thead>
<tr>
<th>Relative Cross Magnitude</th>
<th>Neighbor Cross</th>
<th>First Cross</th>
<th>Every Cross</th>
<th>Strict Cross</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(Percent)</td>
<td>(Percent)</td>
<td>(Percent)</td>
<td>(Percent)</td>
</tr>
<tr>
<td><strong>Full Sample (n=1671)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>28.46%</td>
<td>30.41%</td>
<td>28.21%</td>
<td>30.49%</td>
</tr>
<tr>
<td>Median</td>
<td>26.18%</td>
<td>26.08%</td>
<td>24.85%</td>
<td>23.00%</td>
</tr>
<tr>
<td>StdDev</td>
<td>13.28%</td>
<td>16.85%</td>
<td>14.77%</td>
<td>22.62%</td>
</tr>
<tr>
<td>Min</td>
<td>3.62%</td>
<td>1.25%</td>
<td>4.71%</td>
<td>1.19%</td>
</tr>
<tr>
<td>Max</td>
<td>126.70%</td>
<td>298.25%</td>
<td>298.25%</td>
<td>99.99%</td>
</tr>
<tr>
<td><strong>Large-Cap (n=456)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>24.84%</td>
<td>29.45%</td>
<td>27.29%</td>
<td>36.00%</td>
</tr>
<tr>
<td>Median</td>
<td>21.66%</td>
<td>23.94%</td>
<td>22.17%</td>
<td>26.69%</td>
</tr>
<tr>
<td>StdDev</td>
<td>10.91%</td>
<td>14.50%</td>
<td>13.09%</td>
<td>25.27%</td>
</tr>
<tr>
<td>Min</td>
<td>9.51%</td>
<td>10.39%</td>
<td>8.06%</td>
<td>3.11%</td>
</tr>
<tr>
<td>Max</td>
<td>86.92%</td>
<td>96.84%</td>
<td>83.19%</td>
<td>99.12%</td>
</tr>
<tr>
<td><strong>Mid-Cap (n=963)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>30.49%</td>
<td>30.70%</td>
<td>28.81%</td>
<td>28.31%</td>
</tr>
<tr>
<td>Median</td>
<td>28.27%</td>
<td>27.09%</td>
<td>26.04%</td>
<td>22.03%</td>
</tr>
<tr>
<td>StdDev</td>
<td>13.41%</td>
<td>14.41%</td>
<td>12.37%</td>
<td>20.81%</td>
</tr>
<tr>
<td>Min</td>
<td>6.07%</td>
<td>5.12%</td>
<td>6.38%</td>
<td>1.57%</td>
</tr>
<tr>
<td>Max</td>
<td>126.70%</td>
<td>167.17%</td>
<td>97.51%</td>
<td>99.99%</td>
</tr>
<tr>
<td><strong>Small-Cap (n=252)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>26.73%</td>
<td>31.06%</td>
<td>27.61%</td>
<td>20.57%</td>
</tr>
<tr>
<td>Median</td>
<td>23.62%</td>
<td>25.89%</td>
<td>22.43%</td>
<td>14.65%</td>
</tr>
<tr>
<td>StdDev</td>
<td>14.77%</td>
<td>26.80%</td>
<td>23.50%</td>
<td>17.06%</td>
</tr>
<tr>
<td>Min</td>
<td>3.62%</td>
<td>1.25%</td>
<td>4.71%</td>
<td>1.19%</td>
</tr>
<tr>
<td>Max</td>
<td>100.94%</td>
<td>298.25%</td>
<td>298.25%</td>
<td>81.86%</td>
</tr>
</tbody>
</table>
This table provides summary statistics on the constructed metrics: the relative frequency of the crossing episodes, as defined in Chapter 4.3. The relative frequency is the frequency count as percentage of the number of QBBO price updates. Group count considers all QBBO price update within each no-trade duration as one observation. All measures are calculated on a daily basis and then averaged across the sample period between 11/05/2012 and 11/09/2012. Only quote and trade records during market open hours between 11am and 3pm EST are used.

<table>
<thead>
<tr>
<th>Relative Cross Frequency</th>
<th>Neighbor Cross</th>
<th>First Cross</th>
<th>Every Cross</th>
<th>Strict Cross</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(Percent)</td>
<td>(Percent)</td>
<td>(Percent)</td>
<td>(Percent)</td>
</tr>
<tr>
<td><strong>Full Sample (n=1671)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>2.53%</td>
<td>3.06%</td>
<td>4.32%</td>
<td>0.11%</td>
</tr>
<tr>
<td>Median</td>
<td>2.42%</td>
<td>3.02%</td>
<td>4.41%</td>
<td>0.07%</td>
</tr>
<tr>
<td>StdDev</td>
<td>1.81%</td>
<td>1.81%</td>
<td>2.18%</td>
<td>0.12%</td>
</tr>
<tr>
<td>Min</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
<tr>
<td>Max</td>
<td>18.52%</td>
<td>16.27%</td>
<td>19.44%</td>
<td>1.12%</td>
</tr>
<tr>
<td><strong>Large-Cap (n=456)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>1.78%</td>
<td>3.14%</td>
<td>4.42%</td>
<td>0.16%</td>
</tr>
<tr>
<td>Median</td>
<td>1.89%</td>
<td>3.12%</td>
<td>4.49%</td>
<td>0.14%</td>
</tr>
<tr>
<td>StdDev</td>
<td>1.23%</td>
<td>1.79%</td>
<td>2.00%</td>
<td>0.13%</td>
</tr>
<tr>
<td>Min</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.16%</td>
<td>0.00%</td>
</tr>
<tr>
<td>Max</td>
<td>5.52%</td>
<td>7.27%</td>
<td>8.62%</td>
<td>0.67%</td>
</tr>
<tr>
<td><strong>Mid-Cap (n=963)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>2.70%</td>
<td>3.04%</td>
<td>4.33%</td>
<td>0.10%</td>
</tr>
<tr>
<td>Median</td>
<td>2.66%</td>
<td>3.07%</td>
<td>4.49%</td>
<td>0.07%</td>
</tr>
<tr>
<td>StdDev</td>
<td>1.64%</td>
<td>1.64%</td>
<td>2.04%</td>
<td>0.10%</td>
</tr>
<tr>
<td>Min</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
<tr>
<td>Max</td>
<td>9.08%</td>
<td>9.73%</td>
<td>11.82%</td>
<td>0.86%</td>
</tr>
<tr>
<td><strong>Small-Cap (n=252)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>3.29%</td>
<td>2.98%</td>
<td>4.12%</td>
<td>0.04%</td>
</tr>
<tr>
<td>Median</td>
<td>3.05%</td>
<td>2.66%</td>
<td>3.87%</td>
<td>0.00%</td>
</tr>
<tr>
<td>StdDev</td>
<td>2.66%</td>
<td>2.40%</td>
<td>2.89%</td>
<td>0.13%</td>
</tr>
<tr>
<td>Min</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
<tr>
<td>Max</td>
<td>18.52%</td>
<td>16.27%</td>
<td>19.44%</td>
<td>1.12%</td>
</tr>
</tbody>
</table>
Table 10. Relative Cross Magnitude. Group count.

This table provides summary statistics on the constructed metrics: the relative magnitude of the crossing episodes, as defined in Chapter 4.3. The relative magnitude is the cross magnitude measured in dollar amount as percentage of the quoted spread. Group count considers all QBBO price update within each no trade duration as one observation. All measures are on a daily basis and then averaged across the sample period between 11/05/2012 and 11/09/2012. All measures are calculated with quote and trade records during market open hours between 11am and 3pm EST.

<table>
<thead>
<tr>
<th>Relative Cross Magnitude</th>
<th>Neighbor Cross (Percent)</th>
<th>First Cross (Percent)</th>
<th>Every Cross (Percent)</th>
<th>Strict Cross (Percent)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Full Sample (n=1671)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>30.51%</td>
<td>29.90%</td>
<td>29.79%</td>
<td>32.23%</td>
</tr>
<tr>
<td>Median</td>
<td>28.15%</td>
<td>25.94%</td>
<td>26.65%</td>
<td>25.28%</td>
</tr>
<tr>
<td>StdDev</td>
<td>13.71%</td>
<td>13.96%</td>
<td>12.50%</td>
<td>22.73%</td>
</tr>
<tr>
<td>Min</td>
<td>1.67%</td>
<td>1.67%</td>
<td>1.67%</td>
<td>1.19%</td>
</tr>
<tr>
<td>Max</td>
<td>130.01%</td>
<td>149.99%</td>
<td>123.43%</td>
<td>119.34%</td>
</tr>
<tr>
<td><strong>Large-Cap (n=456)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>25.49%</td>
<td>28.45%</td>
<td>27.86%</td>
<td>36.70%</td>
</tr>
<tr>
<td>Median</td>
<td>22.09%</td>
<td>23.27%</td>
<td>23.04%</td>
<td>26.91%</td>
</tr>
<tr>
<td>StdDev</td>
<td>11.00%</td>
<td>13.84%</td>
<td>12.57%</td>
<td>24.89%</td>
</tr>
<tr>
<td>Min</td>
<td>10.70%</td>
<td>11.84%</td>
<td>10.09%</td>
<td>5.04%</td>
</tr>
<tr>
<td>Max</td>
<td>86.70%</td>
<td>96.84%</td>
<td>68.40%</td>
<td>99.12%</td>
</tr>
<tr>
<td><strong>Mid-Cap (n=963)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>32.77%</td>
<td>30.71%</td>
<td>31.11%</td>
<td>30.47%</td>
</tr>
<tr>
<td>Median</td>
<td>30.22%</td>
<td>27.05%</td>
<td>28.36%</td>
<td>24.89%</td>
</tr>
<tr>
<td>StdDev</td>
<td>13.38%</td>
<td>13.60%</td>
<td>11.79%</td>
<td>21.20%</td>
</tr>
<tr>
<td>Min</td>
<td>1.67%</td>
<td>1.67%</td>
<td>1.67%</td>
<td>1.57%</td>
</tr>
<tr>
<td>Max</td>
<td>130.01%</td>
<td>149.99%</td>
<td>123.43%</td>
<td>119.34%</td>
</tr>
<tr>
<td><strong>Small-Cap (n=252)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>30.22%</td>
<td>29.44%</td>
<td>28.27%</td>
<td>23.25%</td>
</tr>
<tr>
<td>Median</td>
<td>26.47%</td>
<td>25.60%</td>
<td>25.23%</td>
<td>15.84%</td>
</tr>
<tr>
<td>StdDev</td>
<td>16.46%</td>
<td>15.32%</td>
<td>14.33%</td>
<td>19.90%</td>
</tr>
<tr>
<td>Min</td>
<td>5.93%</td>
<td>2.96%</td>
<td>7.85%</td>
<td>1.19%</td>
</tr>
<tr>
<td>Max</td>
<td>125.02%</td>
<td>129.77%</td>
<td>117.76%</td>
<td>92.56%</td>
</tr>
</tbody>
</table>
Table 11. Simulation Quality.
This table provides summary statistics comparing the simulated and the actual stocks. The simulation consists of 10,000 iterations. All measures are first calculated within each trading day between 11am and 3pm, and averaged across all sample days.

<table>
<thead>
<tr>
<th></th>
<th>Observed Mean</th>
<th>Observed Median</th>
<th>Observed StdDev</th>
<th>Simulated Mean</th>
<th>Simulated Median</th>
<th>Simulated StdDev</th>
<th>Num Trades Sum</th>
<th>QBBO Updates Count</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SBUX (Large-Cap)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Avg Price ($</td>
<td>$51.38</td>
<td>$51.36</td>
<td>$0.17</td>
<td>$49.99</td>
<td>$49.99</td>
<td>$0.08</td>
<td>20,493</td>
<td>7,190</td>
</tr>
<tr>
<td>Price Range ($)</td>
<td>$0.68</td>
<td>$1</td>
<td>$0.014</td>
<td>$0.35</td>
<td>$1</td>
<td>$0.003</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No Trade Duration (Second)</td>
<td>2</td>
<td>1</td>
<td>$0.012</td>
<td>1</td>
<td>$0.011</td>
<td>$0.008</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quoted Spread ($)</td>
<td>$0.014</td>
<td>$0.012</td>
<td>$0.006</td>
<td>$0.014</td>
<td>$0.011</td>
<td>$0.008</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Num Trades</td>
<td>20,493</td>
<td>7,190</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>QBBO Updates Count</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>IGTE (Mid-Cap)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Avg Price ($</td>
<td>$15.69</td>
<td>$15.69</td>
<td>$0.05</td>
<td>$15.49</td>
<td>$15.49</td>
<td>$0.07</td>
<td>402</td>
<td>915</td>
</tr>
<tr>
<td>Price Range ($)</td>
<td>$0.23</td>
<td>$0.23</td>
<td>$0.30</td>
<td>$0.30</td>
<td>$0.30</td>
<td>$0.30</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No Trade Duration (Second)</td>
<td>71</td>
<td>9</td>
<td>$0.049</td>
<td>41</td>
<td>$0.049</td>
<td>$0.049</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quoted Spread ($)</td>
<td>$0.010</td>
<td>$0.010</td>
<td>$0.037</td>
<td>$0.010</td>
<td>$0.037</td>
<td>$0.041</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Num Trades</td>
<td>402</td>
<td>915</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>QBBO Updates Count</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>APFC (Small-Cap)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Avg Price ($</td>
<td>$12.48</td>
<td>$12.44</td>
<td>$0.23</td>
<td>$12.49</td>
<td>$12.49</td>
<td>$0.43</td>
<td>24</td>
<td>81</td>
</tr>
<tr>
<td>Price Range ($)</td>
<td>$0.70</td>
<td>$0.70</td>
<td>$0.325</td>
<td>$0.860</td>
<td>$0.861</td>
<td>$0.861</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No Trade Duration (Second)</td>
<td>1,558</td>
<td>1,190</td>
<td>$0.778</td>
<td>24</td>
<td>87</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quoted Spread ($)</td>
<td>$0.848</td>
<td>$0.778</td>
<td>$0.325</td>
<td>$0.860</td>
<td>$0.861</td>
<td>$0.861</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Num Trades</td>
<td>24</td>
<td>81</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>QBBO Updates Count</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
This table compares the crossing relative frequency comparing the simulated and the actual stocks. The simulation consists of 10,000 iterations. All measures are first calculated within each trading day between 11am and 3pm, and averaged across all sample days.

<table>
<thead>
<tr>
<th>Mean Individual</th>
<th>Neighbor Cross</th>
<th>First Cross</th>
<th>Every Cross</th>
<th>Strict Cross</th>
</tr>
</thead>
<tbody>
<tr>
<td>SBUX - obs.</td>
<td>0.26%</td>
<td>4.15%</td>
<td>2.48%</td>
<td>0.03%</td>
</tr>
<tr>
<td>SBUX - sim.</td>
<td>-</td>
<td>0.40%</td>
<td>0.21%</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>IGTE - obs.</td>
<td>5.30%</td>
<td>8.06%</td>
<td>6.95%</td>
<td>0.06%</td>
</tr>
<tr>
<td>IGTE - sim.</td>
<td>-</td>
<td>4.06%</td>
<td>2.25%</td>
<td>0.08%</td>
</tr>
<tr>
<td>APFC - obs.</td>
<td>3.97%</td>
<td>2.45%</td>
<td>4.23%</td>
<td>0.01%</td>
</tr>
<tr>
<td>APFC - sim.</td>
<td>-</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>-</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>StdDev Individual</th>
<th>Neighbor Cross</th>
<th>First Cross</th>
<th>Every Cross</th>
<th>Strict Cross</th>
</tr>
</thead>
<tbody>
<tr>
<td>SBUX - obs.</td>
<td>0.18%</td>
<td>1.46%</td>
<td>0.77%</td>
<td>0.03%</td>
</tr>
<tr>
<td>SBUX - sim.</td>
<td>-</td>
<td>0.14%</td>
<td>0.07%</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>IGTE - obs.</td>
<td>1.71%</td>
<td>1.27%</td>
<td>2.15%</td>
<td>0.06%</td>
</tr>
<tr>
<td>IGTE - sim.</td>
<td>-</td>
<td>0.14%</td>
<td>0.07%</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>APFC - obs.</td>
<td>6.58%</td>
<td>3.40%</td>
<td>5.67%</td>
<td>0.02%</td>
</tr>
<tr>
<td>APFC - sim.</td>
<td>-</td>
<td>3.16%</td>
<td>2.56%</td>
<td>-</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Mean Group</th>
<th>Neighbor Cross</th>
<th>First Cross</th>
<th>Every Cross</th>
<th>Strict Cross</th>
</tr>
</thead>
<tbody>
<tr>
<td>SBUX - obs.</td>
<td>0.24%</td>
<td>3.13%</td>
<td>4.09%</td>
<td>0.08%</td>
</tr>
<tr>
<td>SBUX - sim.</td>
<td>-</td>
<td>0.37%</td>
<td>0.40%</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>IGTE - obs.</td>
<td>3.77%</td>
<td>3.93%</td>
<td>5.82%</td>
<td>0.27%</td>
</tr>
<tr>
<td>IGTE - sim.</td>
<td>-</td>
<td>2.17%</td>
<td>2.83%</td>
<td>0.16%</td>
</tr>
<tr>
<td>APFC - obs.</td>
<td>2.42%</td>
<td>1.86%</td>
<td>2.77%</td>
<td>0.08%</td>
</tr>
<tr>
<td>APFC - sim.</td>
<td>-</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>-</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>StdDev Group</th>
<th>Neighbor Cross</th>
<th>First Cross</th>
<th>Every Cross</th>
<th>Strict Cross</th>
</tr>
</thead>
<tbody>
<tr>
<td>SBUX - obs.</td>
<td>0.18%</td>
<td>1.07%</td>
<td>1.10%</td>
<td>0.04%</td>
</tr>
<tr>
<td>SBUX - sim.</td>
<td>-</td>
<td>0.13%</td>
<td>0.14%</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>IGTE - obs.</td>
<td>1.01%</td>
<td>1.06%</td>
<td>1.77%</td>
<td>0.21%</td>
</tr>
<tr>
<td>IGTE - sim.</td>
<td>-</td>
<td>0.13%</td>
<td>0.14%</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>APFC - obs.</td>
<td>2.61%</td>
<td>1.74%</td>
<td>2.62%</td>
<td>-</td>
</tr>
<tr>
<td>APFC - sim.</td>
<td>-</td>
<td>2.24%</td>
<td>3.16%</td>
<td>-</td>
</tr>
</tbody>
</table>

Note. obs. = observed. sim. = simulated.
Note. '-' means cell value is zero, while <0.0001 represents small non-zero values.
This table compares the crossing relative magnitude comparing the simulated and the actual stocks. The simulation consists of 10,000 iterations. All measures are first calculated within each trading day between 11am and 3pm, and averaged across all sample days.

<table>
<thead>
<tr>
<th>Relative Cross Magnitude</th>
<th>Neighbor Cross (Percent)</th>
<th>First Cross (Percent)</th>
<th>Every Cross (Percent)</th>
<th>Strict Cross (Percent)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mean Individual</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SBUX - obs.</td>
<td>35.96%</td>
<td>35.96%</td>
<td>35.96%</td>
<td>71.93%</td>
</tr>
<tr>
<td>SBUX - sim.</td>
<td>-</td>
<td>37.24%</td>
<td>37.24%</td>
<td>81.41%</td>
</tr>
<tr>
<td>IGTE - obs.</td>
<td>30.88%</td>
<td>23.79%</td>
<td>25.73%</td>
<td>21.38%</td>
</tr>
<tr>
<td>IGTE - sim.</td>
<td>-</td>
<td>12.14%</td>
<td>11.04%</td>
<td>22.93%</td>
</tr>
<tr>
<td>APFC - obs.</td>
<td>16.29%</td>
<td>12.65%</td>
<td>14.09%</td>
<td>1.19%</td>
</tr>
<tr>
<td>APFC - sim.</td>
<td>-</td>
<td>17.53%</td>
<td>8.77%</td>
<td></td>
</tr>
<tr>
<td><strong>StdDev Individual</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SBUX - obs.</td>
<td>4.38%</td>
<td>4.38%</td>
<td>4.38%</td>
<td>8.77%</td>
</tr>
<tr>
<td>SBUX - sim.</td>
<td>-</td>
<td>7.93%</td>
<td>7.93%</td>
<td>13.76%</td>
</tr>
<tr>
<td>IGTE - obs.</td>
<td>8.59%</td>
<td>6.05%</td>
<td>5.39%</td>
<td>3.70%</td>
</tr>
<tr>
<td>IGTE - sim.</td>
<td>-</td>
<td>7.93%</td>
<td>7.93%</td>
<td>13.76%</td>
</tr>
<tr>
<td>APFC - obs.</td>
<td>8.78%</td>
<td>0.59%</td>
<td>3.30%</td>
<td></td>
</tr>
<tr>
<td>APFC - sim.</td>
<td>-</td>
<td>63.77%</td>
<td>69.97%</td>
<td></td>
</tr>
<tr>
<td><strong>Mean Group</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SBUX - obs.</td>
<td>35.96%</td>
<td>35.96%</td>
<td>35.96%</td>
<td>71.93%</td>
</tr>
<tr>
<td>SBUX - sim.</td>
<td>-</td>
<td>37.24%</td>
<td>37.24%</td>
<td>81.41%</td>
</tr>
<tr>
<td>IGTE - obs.</td>
<td>28.96%</td>
<td>22.89%</td>
<td>22.85%</td>
<td>26.72%</td>
</tr>
<tr>
<td>IGTE - sim.</td>
<td>-</td>
<td>11.92%</td>
<td>11.13%</td>
<td>24.83%</td>
</tr>
<tr>
<td>APFC - obs.</td>
<td>15.89%</td>
<td>11.75%</td>
<td>14.44%</td>
<td>1.19%</td>
</tr>
<tr>
<td>APFC - sim.</td>
<td>-</td>
<td>21.91%</td>
<td>8.77%</td>
<td></td>
</tr>
<tr>
<td><strong>StdDev Group</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SBUX - obs.</td>
<td>4.38%</td>
<td>4.38%</td>
<td>4.38%</td>
<td>8.77%</td>
</tr>
<tr>
<td>SBUX - sim.</td>
<td>-</td>
<td>7.93%</td>
<td>7.93%</td>
<td>13.76%</td>
</tr>
<tr>
<td>IGTE - obs.</td>
<td>7.39%</td>
<td>7.33%</td>
<td>6.60%</td>
<td>15.10%</td>
</tr>
<tr>
<td>IGTE - sim.</td>
<td>-</td>
<td>7.93%</td>
<td>7.93%</td>
<td>13.76%</td>
</tr>
<tr>
<td>APFC - obs.</td>
<td>5.72%</td>
<td>9.82%</td>
<td>3.71%</td>
<td></td>
</tr>
<tr>
<td>APFC - sim.</td>
<td>-</td>
<td>62.74%</td>
<td>68.94%</td>
<td></td>
</tr>
</tbody>
</table>

*Note. obs. = observed. sim. = simulated.*

*Note. '-' means cell value is zero, while <0.0001 represents small non-zero values.*
Table 14. T-test Results.

This table provides results on t-test performed to examine the statistical significance of crossing episodes in both frequency and magnitude, against the simulated stocks as benchmarks. ***, ** and * indicate significance levels of 0.1%, 1% and 5% respectively based on a two-tailed t-test. The simulation consists of 10,000 iterations.

<table>
<thead>
<tr>
<th></th>
<th>SBUX</th>
<th></th>
<th>IGTE</th>
<th></th>
<th>APFC</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>t-value</td>
<td>Pr &gt;</td>
<td>t-value</td>
<td>Pr &gt;</td>
<td>t-value</td>
<td>Pr &gt;</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(df=4)</td>
<td></td>
<td></td>
<td>(df=4)</td>
</tr>
<tr>
<td><strong>Frequency. Individual.</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neighbor Cross</td>
<td>2.81</td>
<td>0.0482</td>
<td>*</td>
<td>6.20</td>
<td>0.0034</td>
<td>**</td>
</tr>
<tr>
<td>First Cross</td>
<td>5.14</td>
<td>0.0068</td>
<td>**</td>
<td>6.31</td>
<td>0.0032</td>
<td>**</td>
</tr>
<tr>
<td>Every Cross</td>
<td>5.95</td>
<td>0.0040</td>
<td>**</td>
<td>4.36</td>
<td>0.0120</td>
<td>*</td>
</tr>
<tr>
<td>Strict Cross</td>
<td>2.21</td>
<td>0.0914</td>
<td>-</td>
<td>0.81</td>
<td>0.4648</td>
<td></td>
</tr>
<tr>
<td><strong>Magnitude. Individual.</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neighbor Cross</td>
<td>16.40</td>
<td>&lt;0.0001</td>
<td>***</td>
<td>7.19</td>
<td>0.0020</td>
<td>**</td>
</tr>
<tr>
<td>First Cross</td>
<td>-0.58</td>
<td>0.5903</td>
<td></td>
<td>3.85</td>
<td>0.0183</td>
<td>*</td>
</tr>
<tr>
<td>Every Cross</td>
<td>-0.58</td>
<td>0.5903</td>
<td></td>
<td>5.45</td>
<td>0.0055</td>
<td>**</td>
</tr>
<tr>
<td>Strict Cross</td>
<td>-2.16</td>
<td>0.0965</td>
<td></td>
<td>-0.84</td>
<td>0.4496</td>
<td></td>
</tr>
<tr>
<td><strong>Frequency. Group.</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neighbor Cross</td>
<td>2.73</td>
<td>0.0526</td>
<td></td>
<td>7.49</td>
<td>0.0017</td>
<td>**</td>
</tr>
<tr>
<td>First Cross</td>
<td>5.16</td>
<td>0.0067</td>
<td>**</td>
<td>3.32</td>
<td>0.0295</td>
<td>*</td>
</tr>
<tr>
<td>Every Cross</td>
<td>6.73</td>
<td>0.0025</td>
<td>**</td>
<td>3.38</td>
<td>0.0279</td>
<td>*</td>
</tr>
<tr>
<td>Strict Cross</td>
<td>3.88</td>
<td>0.0178</td>
<td></td>
<td>1.05</td>
<td>0.3526</td>
<td></td>
</tr>
<tr>
<td><strong>Magnitude. Group.</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neighbor Cross</td>
<td>16.40</td>
<td>&lt;0.0001</td>
<td>***</td>
<td>7.83</td>
<td>0.0014</td>
<td>**</td>
</tr>
<tr>
<td>First Cross</td>
<td>-0.58</td>
<td>0.5903</td>
<td></td>
<td>2.99</td>
<td>0.0402</td>
<td>*</td>
</tr>
<tr>
<td>Every Cross</td>
<td>-0.58</td>
<td>0.5903</td>
<td></td>
<td>3.55</td>
<td>0.0238</td>
<td>*</td>
</tr>
<tr>
<td>Strict Cross</td>
<td>-2.16</td>
<td>0.0965</td>
<td></td>
<td>0.25</td>
<td>0.8144</td>
<td></td>
</tr>
</tbody>
</table>
Table 15. Time Series Test: Market Volatility.

This table provides a time series test on the impact of market volatility on the relative cross frequency and magnitude on a minute to minute basis. QQQ Volatility is the calculated using the transaction price of the PowerShare Nasdaq-100 Index Tracking Stock ETF (NASDAQ:QQQ), using (high-low)/low within a one minute window, and is used as a proxy for market movements generated by news at a given minute. Num Trades, Avg Price, Rel Qspread, Qspread StdDev are control variables, each calculated within a one minute window averaging across all stocks within the full sample, or the large-cap, mid-cap, and small-cap strata. Num Trades is the total number of transaction counts. Avg Price is the mean of all transaction prices. Rel Qspread is the duration-weighted average quoted spread as percentage of the bid ask midpoint. Qspread StdDev is the standard deviation of the duration-weighted average quoted spread. I use every cross by group count for all regression analysis.

All measures are calculated on a daily basis and then averaged across the sample period between 11/05/2012 and 11/09/2012. Only quote and trade records during market open hours between 11am and 3pm EST are used. All non-dummy variables are log-transformed, thus the coefficients may be interpreted as percentage changes. ***, ** and * indicate significance levels of 0.1%, 1% and 5% respectively.

\[
\ln(\text{Rel. Freq}_t) \text{ or } \ln(\text{Rel. Mag}_t) = \beta_1 \ln(\text{QQQVolatility}_t) + \beta_2 \ln(\text{NumTrades}_t) + \beta_3 \ln(\text{AvgPrice}_t) \\
+ \beta_4 \ln(\text{RelQspread}_t) + \beta_5 \ln(\text{QspreadStdDev}_t) + \beta_0 + \epsilon_t
\]

<table>
<thead>
<tr>
<th>Relative Cross Frequency</th>
<th>Relative Cross Magnitude</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coefficient</td>
<td>Std Error</td>
</tr>
<tr>
<td><strong>Full Sample (obs=240 minutes, averaging over 1696 stocks)</strong></td>
<td></td>
</tr>
<tr>
<td>QQQ Volatility</td>
<td>-0.0518</td>
</tr>
<tr>
<td>Num Trades</td>
<td>0.1922</td>
</tr>
<tr>
<td>Avg Price</td>
<td>-0.1163</td>
</tr>
<tr>
<td>Rel Qspread</td>
<td>-0.2459</td>
</tr>
<tr>
<td>Qspread StdDev</td>
<td>-0.3704</td>
</tr>
<tr>
<td>Intercept</td>
<td>-6.8571</td>
</tr>
</tbody>
</table>

Adj-Rsquare | 0.2208 | 0.1127 |
Table 15 – continued.

<table>
<thead>
<tr>
<th>Relative Cross Frequency</th>
<th>Relative Cross Magnitude</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coefficient</td>
<td>Std Error</td>
</tr>
<tr>
<td><strong>Large-Cap (obs=240 minutes, averaging over 453 stocks)</strong></td>
<td></td>
</tr>
<tr>
<td>QQQ Volatility</td>
<td>-0.0725</td>
</tr>
<tr>
<td>Num Trades</td>
<td>0.1414</td>
</tr>
<tr>
<td>Avg Price</td>
<td>-0.1745</td>
</tr>
<tr>
<td>Rel Qspread</td>
<td>-0.2462</td>
</tr>
<tr>
<td>Qspread StdDev</td>
<td>-0.1059</td>
</tr>
<tr>
<td>Intercept</td>
<td>-5.9598</td>
</tr>
<tr>
<td><strong>Adj-Rsquare</strong></td>
<td>0.2886</td>
</tr>
<tr>
<td><strong>Mid-Cap (obs=240 minutes, averaging over 968 stocks)</strong></td>
<td></td>
</tr>
<tr>
<td>QQQ Volatility</td>
<td>-0.0189</td>
</tr>
<tr>
<td>Num Trades</td>
<td>0.2053</td>
</tr>
<tr>
<td>Avg Price</td>
<td>0.5969</td>
</tr>
<tr>
<td>Rel Qspread</td>
<td>-0.4339</td>
</tr>
<tr>
<td>Qspread StdDev</td>
<td>-0.2218</td>
</tr>
<tr>
<td>Intercept</td>
<td>-8.9925</td>
</tr>
<tr>
<td><strong>Adj-Rsquare</strong></td>
<td>0.2554</td>
</tr>
<tr>
<td><strong>Small-Cap (obs=240 minutes, averaging over 272 stocks)</strong></td>
<td></td>
</tr>
<tr>
<td>QQQ Volatility</td>
<td>0.1135</td>
</tr>
<tr>
<td>Num Trades</td>
<td>0.3955</td>
</tr>
<tr>
<td>Avg Price</td>
<td>0.3842</td>
</tr>
<tr>
<td>Rel Qspread</td>
<td>-0.1890</td>
</tr>
<tr>
<td>Qspread StdDev</td>
<td>-0.0025</td>
</tr>
<tr>
<td>Intercept</td>
<td>-6.0256</td>
</tr>
<tr>
<td><strong>Adj-Rsquare</strong></td>
<td>0.1019</td>
</tr>
</tbody>
</table>
**Table 16. Cross-Sectional Test: Information Quality.**

This table provides a cross-sectional test on the impact of information quality on the relative cross frequency and magnitude across stocks within the full sample, and the large-cap, mid-cap, and small-cap strata. If US is a dummy variable that indicates whether the company is based in the United States, collected from Bloomberg. If Primary is a dummy variable that indicates whether NASDAQ is the primary listing venue and this is the primary ticker symbol for the stock. For both dummy variables, 1 indicates true, 0 indicates false. They are used as proxies for the quality and reliability of the company’s reported accounting information. Avg Mkt Cap, Num Trades, Avg Price are control variables calculated for each stock stocks. Avg Mkt Cap is market capitalization in US dollar amount based on daily closing price collected from Bloomberg, and then taken average over the sample period. Num Trades is the total number of transaction counts. Avg Price is the mean of all transaction prices in US dollars. I use every cross by group count for all regression analysis. All measures are calculated on a daily basis and then averaged across the sample period between 11/05/2012 and 11/09/2012. Only quote and trade records during market open hours between 11am and 3pm EST are used. All non-dummy variables are log-transformed, thus the coefficients may be interpreted as percentage changes. ***, ** and * indicate significance levels of 0.1%, 1% and 5% respectively.

\[
\ln(\text{Rel. Freq}_i) \text{ or } \ln(\text{Rel. Mag}_i) = \beta_1 \text{IfUS}_i + \beta_2 \text{IfPrimary}_i + \beta_3 \ln(\text{AvgMktCap}_i) + \beta_4 \ln(\text{NumTrades}_i) + \beta_5 \ln(\text{AvgPrice}_i) + \beta_0 + \epsilon_i
\]

<p>| Relative Cross Frequency | Coefficient | Std Error | t-value | Pr &gt; |t| | Relative Cross Magnitude | Coefficient | Std Error | t-value | Pr &gt; |t| |
|-------------------------|-------------|-----------|---------|-------|-----------|--------------------------|-------------|-----------|---------|-------|-----------|
| <strong>Full Sample (obs=1665)</strong> |             |           |         |       |           |                          |             |           |         |       |           |
| If US                   | -0.1058     | 0.0676    | -1.57   | 0.1176|           | 0.1021                   | 0.0374      | 2.73      | 0.0064  | **    |           |
| If Primary              | 0.0678      | 0.0995    | 0.68    | 0.4960|           | 0.0663                   | 0.0551      | 1.20      | 0.2289  |       |           |
| Avg Mkt Cap             | -0.1678     | 0.0278    | -6.04   | &lt;.0001| ***       | 0.0914                   | 0.0154      | 5.95      | &lt;.0001  | ***   |           |
| Num Trades              | 0.1655      | 0.0174    | 9.52    | &lt;.0001| ***       | 0.0083                   | 0.0096      | 0.86      | 0.3895  |       |           |
| Avg Price               | 0.1069      | 0.0278    | 3.84    | 0.0001| ***       | -0.2288                  | 0.0154      | -14.86    | &lt;.0001  | ***   |           |
| Intercept               | -3.5208     | 0.1196    | -29.43  | &lt;.0001| ***       | -1.4904                  | 0.0662      | -22.53    | &lt;.0001  | ***   |           |
| Adj-Rsquare             | 0.0577      |           |         |       |           | 0.1722                   |             |           |         |       |           |</p>
<table>
<thead>
<tr>
<th></th>
<th>Large-Cap (obs=456)</th>
<th></th>
<th>Relative Cross Frequency</th>
<th>Relative Cross Magnitude</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coefficient</td>
<td>Std Error</td>
<td>t-value</td>
<td>Pr &gt;</td>
<td>t</td>
<td>Coefficient</td>
</tr>
<tr>
<td>If US</td>
<td>-0.0042</td>
<td>0.1115</td>
<td>-0.04</td>
<td>0.9697</td>
<td>0.1566</td>
<td>0.0586</td>
</tr>
<tr>
<td>If Primary</td>
<td>0.1423</td>
<td>0.2090</td>
<td>0.68</td>
<td>0.4962</td>
<td>0.1667</td>
<td>0.1098</td>
</tr>
<tr>
<td>Avg Mkt Cap</td>
<td>-0.2655</td>
<td>0.0455</td>
<td>-5.84</td>
<td>&lt;.0001 ***</td>
<td>0.1911</td>
<td>0.0239</td>
</tr>
<tr>
<td>Num Trades</td>
<td>0.1977</td>
<td>0.0385</td>
<td>5.14</td>
<td>&lt;.0001 ***</td>
<td>-0.0595</td>
<td>0.0202</td>
</tr>
<tr>
<td>Avg Price</td>
<td>0.4172</td>
<td>0.0510</td>
<td>8.18</td>
<td>&lt;.0001 ***</td>
<td>-0.4592</td>
<td>0.0268</td>
</tr>
<tr>
<td>Intercept</td>
<td>-4.2379</td>
<td>0.3160</td>
<td>-13.41</td>
<td>&lt;.0001 ***</td>
<td>-1.1899</td>
<td>0.1660</td>
</tr>
<tr>
<td>Adj-Rsquare</td>
<td>0.1260</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mid-Cap (obs=959)</td>
<td></td>
<td></td>
<td>Relative Cross Frequency</td>
<td>Relative Cross Magnitude</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Coefficient</td>
<td>Std Error</td>
<td>t-value</td>
<td>Pr &gt;</td>
<td>t</td>
<td>Coefficient</td>
</tr>
<tr>
<td>If US</td>
<td>-0.0925</td>
<td>0.0846</td>
<td>-1.09</td>
<td>0.2745</td>
<td>0.0160</td>
<td>0.0481</td>
</tr>
<tr>
<td>If Primary</td>
<td>0.0501</td>
<td>0.1217</td>
<td>0.41</td>
<td>0.6810</td>
<td>0.0919</td>
<td>0.0692</td>
</tr>
<tr>
<td>Avg Mkt Cap</td>
<td>-0.0923</td>
<td>0.0499</td>
<td>-1.85</td>
<td>0.0648</td>
<td>0.0586</td>
<td>0.0284</td>
</tr>
<tr>
<td>Num Trades</td>
<td>0.1174</td>
<td>0.0210</td>
<td>5.60</td>
<td>&lt;.0001 ***</td>
<td>0.0064</td>
<td>0.0119</td>
</tr>
<tr>
<td>Avg Price</td>
<td>-0.0246</td>
<td>0.0367</td>
<td>-0.67</td>
<td>0.5019</td>
<td>-0.1695</td>
<td>0.0208</td>
</tr>
<tr>
<td>Intercept</td>
<td>-3.3323</td>
<td>0.2241</td>
<td>-14.87</td>
<td>&lt;.0001 ***</td>
<td>-1.3090</td>
<td>0.1275</td>
</tr>
<tr>
<td>Adj-Rsquare</td>
<td>0.0563</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Small-Cap (obs=250)</td>
<td></td>
<td></td>
<td>Relative Cross Frequency</td>
<td>Relative Cross Magnitude</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Coefficient</td>
<td>Std Error</td>
<td>t-value</td>
<td>Pr &gt;</td>
<td>t</td>
<td>Coefficient</td>
</tr>
<tr>
<td>If US</td>
<td>-0.2061</td>
<td>0.2098</td>
<td>-0.98</td>
<td>0.3269</td>
<td>0.1770</td>
<td>0.1060</td>
</tr>
<tr>
<td>If Primary</td>
<td>0.2164</td>
<td>0.2576</td>
<td>0.84</td>
<td>0.4018</td>
<td>-0.0425</td>
<td>0.1301</td>
</tr>
<tr>
<td>Avg Mkt Cap</td>
<td>0.1009</td>
<td>0.1327</td>
<td>0.76</td>
<td>0.4475</td>
<td>0.1639</td>
<td>0.0670</td>
</tr>
<tr>
<td>Num Trades</td>
<td>0.2748</td>
<td>0.0559</td>
<td>4.92</td>
<td>&lt;.0001 ***</td>
<td>0.0408</td>
<td>0.0282</td>
</tr>
<tr>
<td>Avg Price</td>
<td>-0.1357</td>
<td>0.0998</td>
<td>-1.36</td>
<td>0.1751</td>
<td>-0.1525</td>
<td>0.0504</td>
</tr>
<tr>
<td>Intercept</td>
<td>-4.8183</td>
<td>0.5277</td>
<td>-9.13</td>
<td>&lt;.0001 ***</td>
<td>-2.1046</td>
<td>0.2665</td>
</tr>
<tr>
<td>Adj-Rsquare</td>
<td>0.1450</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 17. Cross-Sectional Test: Insider Holding.

This table provides a cross-sectional test on the impact of the percentage of shares outstanding held by insiders on the relative cross frequency and magnitude across stocks within the full sample, and the large-cap, mid-cap, and small-cap strata. *Float/Outstanding* is the number of stock shares float divided by the number of shares outstanding for the stock, collected from Bloomberg. It is used as a proxy for the percentage of shares outstanding held by corporate owners and large institutions instead of floating. *Avg Mkt Cap, Num Trades, Avg Price* are control variables calculated for each stock. *Avg Mkt Cap* is market capitalization in US dollar amount based on daily closing price collected from Bloomberg, and then taken average over the sample period. *Num Trades* is the total number of transaction counts. *Avg Price* is the mean of all transaction prices in US dollars. I use every cross by group count for all regression analysis. All measures are calculated on a daily basis and then averaged across the sample period between 11/05/2012 and 11/09/2012. Only quote and trade records during market open hours between 11am and 3pm EST are used. All non-dummy variables are log-transformed, thus the coefficients may be interpreted as percentage changes. ***, ** and * indicate significance levels of 0.1%, 1% and 5% respectively.

\[
\ln(\text{Rel. Freq}_i) \text{ or } \ln(\text{Rel. Mag}_i) = \beta_1 \ln\left(\frac{\text{Float}}{\text{Outstanding}}\right)_i + \beta_2 \ln(\text{Avg Mkt Cap}_i) + \beta_3 \ln(\text{Num Trades}_i) + \beta_4 \ln(\text{Avg Price}_i) + \beta_0 + \epsilon_i
\]

<table>
<thead>
<tr>
<th></th>
<th>Relative Cross Frequency</th>
<th></th>
<th>Relative Cross Magnitude</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coefficient</td>
<td>Std Error</td>
<td>t-value</td>
<td>Pr &gt;</td>
</tr>
<tr>
<td><strong>Full Sample (obs=1665)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Float/Outstanding</td>
<td>-0.0744</td>
<td>0.0536</td>
<td>-1.39</td>
<td>0.1651</td>
</tr>
<tr>
<td>Avg Mkt Cap</td>
<td>-0.1715</td>
<td>0.0280</td>
<td>-6.14</td>
<td>&lt;.0001 ***</td>
</tr>
<tr>
<td>Num Trades</td>
<td>0.1711</td>
<td>0.0179</td>
<td>9.54</td>
<td>&lt;.0001 ***</td>
</tr>
<tr>
<td>Avg Price</td>
<td>0.1110</td>
<td>0.0280</td>
<td>3.96</td>
<td>&lt;.0001 ***</td>
</tr>
<tr>
<td>Intercept</td>
<td>-3.5931</td>
<td>0.0723</td>
<td>-49.70</td>
<td>&lt;.0001 ***</td>
</tr>
<tr>
<td><strong>Adj-Rsquare</strong></td>
<td>0.0579</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[
\ln(\text{Rel. Freq}_i) \text{ or } \ln(\text{Rel. Mag}_i) = \beta_1 \ln\left(\frac{\text{Float}}{\text{Outstanding}}\right)_i + \beta_2 \ln(\text{Avg Mkt Cap}_i) + \beta_3 \ln(\text{Num Trades}_i) + \beta_4 \ln(\text{Avg Price}_i) + \beta_0 + \epsilon_i
\]
<table>
<thead>
<tr>
<th></th>
<th>Relative Cross Frequency</th>
<th></th>
<th>Relative Cross Magnitude</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coefficient</td>
<td>Std Error</td>
<td>t-value</td>
<td>Pr &gt;</td>
</tr>
<tr>
<td><strong>Large-Cap (obs=456)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Float/Outstanding</td>
<td>0.0025</td>
<td>0.1173</td>
<td>0.02</td>
<td>0.9832</td>
</tr>
<tr>
<td>Avg Mkt Cap</td>
<td>-0.2693</td>
<td>0.0459</td>
<td>-5.86</td>
<td>&lt;.0001 ***</td>
</tr>
<tr>
<td>Num Trades</td>
<td>0.2012</td>
<td>0.0405</td>
<td>4.97</td>
<td>&lt;.0001 ***</td>
</tr>
<tr>
<td>Avg Price</td>
<td>0.4208</td>
<td>0.0520</td>
<td>8.10</td>
<td>&lt;.0001 ***</td>
</tr>
<tr>
<td>Intercept</td>
<td>-4.1114</td>
<td>0.2684</td>
<td>-15.32</td>
<td>&lt;.0001 ***</td>
</tr>
<tr>
<td><strong>Adj-Rsquare</strong></td>
<td>0.1271</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Mid-Cap (obs=959)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Float/Outstanding</td>
<td>-0.0998</td>
<td>0.0585</td>
<td>-1.71</td>
<td>0.0882</td>
</tr>
<tr>
<td>Avg Mkt Cap</td>
<td>-0.1049</td>
<td>0.0501</td>
<td>-2.09</td>
<td>0.0364 *</td>
</tr>
<tr>
<td>Num Trades</td>
<td>0.1279</td>
<td>0.0217</td>
<td>5.90</td>
<td>&lt;.0001 ***</td>
</tr>
<tr>
<td>Avg Price</td>
<td>-0.0120</td>
<td>0.0370</td>
<td>-0.32</td>
<td>0.7459</td>
</tr>
<tr>
<td>Intercept</td>
<td>-3.4159</td>
<td>0.1838</td>
<td>-18.59</td>
<td>&lt;.0001 ***</td>
</tr>
<tr>
<td><strong>Adj-Rsquare</strong></td>
<td>0.0590</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Small-Cap (obs=250)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Float/Outstanding</td>
<td>0.2457</td>
<td>0.1935</td>
<td>1.27</td>
<td>0.2054</td>
</tr>
<tr>
<td>Avg Mkt Cap</td>
<td>0.1318</td>
<td>0.1322</td>
<td>1.00</td>
<td>0.3199</td>
</tr>
<tr>
<td>Num Trades</td>
<td>0.2543</td>
<td>0.0559</td>
<td>4.55</td>
<td>&lt;.0001 ***</td>
</tr>
<tr>
<td>Avg Price</td>
<td>-0.1726</td>
<td>0.0979</td>
<td>-1.76</td>
<td>0.0791</td>
</tr>
<tr>
<td>Intercept</td>
<td>-4.7212</td>
<td>0.4604</td>
<td>-10.25</td>
<td>&lt;.0001 ***</td>
</tr>
<tr>
<td><strong>Adj-Rsquare</strong></td>
<td>0.1492</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
This table provides a cross-sectional test on the impact of analyst coverage on the relative cross frequency and magnitude across stocks within the full sample, and the large-cap, mid-cap, and small-cap strata. I collect the number of analyst recommendations on Earnings per share (EPS) forecasts on this stock from the IBES database, which ranges from zero to 62. I break down this variable into categorical dummies, at the ranges of =0, =1, [2,5], [6,10], [11,20], [20,40], [41,62]. Analyst StdDev is the standard deviation between the forecasted EPS by different analysts, also collected from the IBES database. Avg Mkt Cap, Num Trades, Avg Price are control variables calculated for each stock stocks. Avg Mkt Cap is market capitalization in US dollar amount based on daily closing price collected from Bloomberg, and then taken average over the sample period. Num Trades is the total number of transaction counts. Avg Price is the mean of all transaction prices in US dollars. I use every cross by group count for all regression analysis. All measures are calculated on a daily basis and then averaged across the sample period between 11/05/2012 and 11/09/2012. Only quote and trade records during market open hours between 11am and 3pm EST are used. All non-dummy variables are log-transformed, thus the coefficients may be interpreted as percentage changes. ***, ** and * indicate significance levels of 0.1%, 1% and 5% respectively.

\[
\ln(\text{Rel. Freq}) \text{ or } \ln(\text{Rel. Mag}) = \beta_1 I_{\text{Num Analyst}=0} + \beta_2 I_{\text{Num Analyst}=1} + \beta_3 I_{\text{Num Analyst} \in [2,5]} + \beta_4 I_{\text{Num Analyst} \in [6,10]} + \beta_5 I_{\text{Num Analyst} \in [11,20]} + \beta_6 I_{\text{Num Analyst} \in [21,40]} + \beta_7 I_{\text{Num Analyst} \in [41,62]} + \beta_8 \text{EPS StdDev} + \beta_9 \ln(\text{Avg Mkt Cap}) + \beta_{10} \ln(\text{Num Trades}) + \beta_{11} \ln(\text{Avg Price}) + \beta_0 + \epsilon_i
\]

<p>| Relative Cross Frequency | Coefficient | Std Error | t-value | Pr &gt; |t|  | Relative Cross Magnitude | Coefficient | Std Error | t-value | Pr &gt; |t|  |
|--------------------------|-------------|-----------|---------|-------|-----|--------------------------|-------------|-----------|---------|-------|-----|--------------------------|-------------|-----------|---------|-------|-----|--------------------------|-------------|-----------|---------|-------|-----|
| <strong>Full Sample (obs=1665)</strong> |             |           |         |       |     |                           |             |           |         |       |     |                           |             |           |         |       |     |                           |             |           |         |       |     |
| If Num Analyst = 0       | 0.8500      | 0.6960    | 1.22    | 0.2222| 0.5608| 0.3829                   | 1.46        | 0.1432    |         |       |     |                           |             |           |         |       |     |                           |             |           |         |       |     |
| If Num Analyst = 1       | 0.2585      | 0.2117    | 1.22    | 0.2221| 0.0419| 0.1164                   | 0.36        | 0.7191    |         |       |     |                           |             |           |         |       |     |                           |             |           |         |       |     |
| If Num Analyst ∈ [2,5]   | 0.4917      | 0.2050    | 2.40    | 0.0166*| 0.2351| 0.1128                   | 2.08        | 0.0372*   |         |       |     |                           |             |           |         |       |     |                           |             |           |         |       |     |
| If Num Analyst ∈ [6,10]  | 0.5539      | 0.2096    | 2.64    | 0.0083**| 0.1871| 0.1153                   | 1.62        | 0.1050    |         |       |     |                           |             |           |         |       |     |                           |             |           |         |       |     |
| If Num Analyst ∈ [11,20] | 0.5198      | 0.2168    | 2.40    | 0.0166*| 0.0416| 0.1193                   | 0.35        | 0.7271    |         |       |     |                           |             |           |         |       |     |                           |             |           |         |       |     |
| If Num Analyst ∈ [21,40] | 0.5597      | 0.2310    | 2.42    | 0.0155*| 0.0509| 0.1271                   | 0.40        | 0.6886    |         |       |     |                           |             |           |         |       |     |                           |             |           |         |       |     |
| If Num Analyst ∈ [41,62] | -1.3150     | 0.3061    | -0.44   | 0.6592| 0.0358| 0.1684                   | 0.21        | 0.8317    |         |       |     |                           |             |           |         |       |     |                           |             |           |         |       |     |
| EPS StdDev               | 0.0615      | 0.0491    | 1.25    | 0.2106| -0.0235| 0.0270                   | -0.87       | 0.3852    |         |       |     |                           |             |           |         |       |     |                           |             |           |         |       |     |
| Avg Mkt Cap              | -0.1426     | 0.0299    | -4.77   | &lt;.0001***| 0.1008| 0.0165                   | 6.13        | &lt;.0001***|         |       |     |                           |             |           |         |       |     |                           |             |           |         |       |     |
| Num Trades               | 0.1352      | 0.0188    | 7.21    | &lt;.0001***| 0.0242| 0.0103                   | 2.35        | 0.0189*  |         |       |     |                           |             |           |         |       |     |                           |             |           |         |       |     |
| Avg Price                | 0.0798      | 0.0280    | 2.85    | 0.0044**| -0.2230| 0.0154                   | -14.49      | &lt;.0001***|         |       |     |                           |             |           |         |       |     |                           |             |           |         |       |     |
| Intercept                | -3.9995     | 0.2109    | -18.97  | &lt;.0001***| -1.6366| 0.1160                   | -14.11      | &lt;.0001***|         |       |     |                           |             |           |         |       |     |                           |             |           |         |       |     |
| Adj-Rsquare              | 0.0747      |           |         |       |     |                           | 0.1958      |           |         |       |     |                           |             |           |         |       |     |                           |             |           |         |       |     |</p>
<table>
<thead>
<tr>
<th>Relative Cross Frequency</th>
<th>Relative Cross Magnitude</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Coefficient</strong></td>
<td><strong>Std Error</strong></td>
</tr>
<tr>
<td><strong>Large-Cap (obs=456)</strong></td>
<td></td>
</tr>
<tr>
<td>If Num Analyst = 0</td>
<td>0</td>
</tr>
<tr>
<td>If Num Analyst = 1</td>
<td>-0.0737</td>
</tr>
<tr>
<td>If Num Analyst ∈[2,5]</td>
<td>0.3838</td>
</tr>
<tr>
<td>If Num Analyst ∈[6,10]</td>
<td>0.6112</td>
</tr>
<tr>
<td>If Num Analyst ∈[11,20]</td>
<td>0.7208</td>
</tr>
<tr>
<td>If Num Analyst ∈[21,40]</td>
<td>0.8036</td>
</tr>
<tr>
<td>If Num Analyst ∈[41,62]</td>
<td>0</td>
</tr>
<tr>
<td>EPS StdDev</td>
<td>0.1324</td>
</tr>
<tr>
<td>Avg Mkt Cap</td>
<td>-0.1848</td>
</tr>
<tr>
<td>Num Trades</td>
<td>0.1253</td>
</tr>
<tr>
<td>Avg Price</td>
<td>0.3799</td>
</tr>
<tr>
<td>Intercept</td>
<td>-4.8469</td>
</tr>
</tbody>
</table>

**Adj-Rsquare** | 0.1790 | 0.5024
<table>
<thead>
<tr>
<th></th>
<th>Relative Cross Frequency</th>
<th></th>
<th>Relative Cross Magnitude</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coefficient</td>
<td>Std Error</td>
<td>t-value</td>
<td>Pr &gt;</td>
</tr>
<tr>
<td>Mid-Cap (obs=959)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>If Num Analyst = 0</td>
<td>1.1010</td>
<td>0.6403</td>
<td>1.72</td>
<td>0.0858</td>
</tr>
<tr>
<td>If Num Analyst = 1</td>
<td>0.5997</td>
<td>0.2800</td>
<td>2.14</td>
<td>0.0325</td>
</tr>
<tr>
<td>If Num Analyst ∈[2,5]</td>
<td>0.7554</td>
<td>0.2346</td>
<td>3.22</td>
<td>0.0013</td>
</tr>
<tr>
<td>If Num Analyst ∈[6,10]</td>
<td>0.7695</td>
<td>0.2304</td>
<td>3.34</td>
<td>0.0009</td>
</tr>
<tr>
<td>If Num Analyst ∈[11,20]</td>
<td>0.7634</td>
<td>0.2327</td>
<td>3.28</td>
<td>0.0011</td>
</tr>
<tr>
<td>If Num Analyst ∈[21,40]</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>If Num Analyst ∈[41,62]</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>EPS StdDev</td>
<td>0.1288</td>
<td>0.0548</td>
<td>2.35</td>
<td>0.0190</td>
</tr>
<tr>
<td>Avg Mkt Cap</td>
<td>-0.0933</td>
<td>0.0501</td>
<td>-1.86</td>
<td>0.0632</td>
</tr>
<tr>
<td>Num Trades</td>
<td>0.1102</td>
<td>0.0216</td>
<td>5.10</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Avg Price</td>
<td>-0.0283</td>
<td>0.0358</td>
<td>-0.79</td>
<td>0.4299</td>
</tr>
<tr>
<td>Intercept</td>
<td>-4.1547</td>
<td>0.3305</td>
<td>-12.57</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

Adj-Rsquare 0.0712

Pr > |t| 0.1201
<table>
<thead>
<tr>
<th>Small-Cap (obs=250)</th>
<th>Relative Cross Frequency</th>
<th>Relative Cross Magnitude</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coefficient</td>
<td>Std Error</td>
</tr>
<tr>
<td>If Num Analyst = 0</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>If Num Analyst = 1</td>
<td>0.1351</td>
<td>0.2874</td>
</tr>
<tr>
<td>If Num Analyst ∈ [2,5]</td>
<td>0.3137</td>
<td>0.2926</td>
</tr>
<tr>
<td>If Num Analyst ∈ [6,10]</td>
<td>0.3999</td>
<td>0.3452</td>
</tr>
<tr>
<td>If Num Analyst ∈ [11,20]</td>
<td>-0.0358</td>
<td>0.5936</td>
</tr>
<tr>
<td>If Num Analyst ∈ [21,40]</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>If Num Analyst ∈ [41,62]</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>EPS StdDev</td>
<td>-0.1596</td>
<td>0.1427</td>
</tr>
<tr>
<td>Avg Mkt Cap</td>
<td>0.1027</td>
<td>0.1368</td>
</tr>
<tr>
<td>Num Trades</td>
<td>0.2519</td>
<td>0.0606</td>
</tr>
<tr>
<td>Avg Price</td>
<td>-0.1305</td>
<td>0.0993</td>
</tr>
<tr>
<td>Intercept</td>
<td>-4.9068</td>
<td>0.5343</td>
</tr>
<tr>
<td>Adj-Rsquare</td>
<td>0.1398</td>
<td>0.0605</td>
</tr>
</tbody>
</table>
Table 19. Cross-Sectional Test: Microstructure Effect.

This table provides a cross-sectional test on the impact of microstructure effect on the relative cross frequency and magnitude across stocks within the full sample, and the large-cap, mid-cap, and small-cap strata. Rel Qspread is the duration-weighted average quoted spread as percentage of the bid ask midpoint. Qspread StdDev is the standard deviation of the duration-weighted average quoted spread. They are used as proxies for information assymetry and microstructure noise. Avg Mkt Cap, Num Trades, Avg Price are control variables calculated for each stock stocks. Avg Mkt Cap is market capitalization in US dollar amount based on daily closing price collected from Bloomberg, and then taken average over the sample period. Num Trades is the total number of transaction counts. Avg Price is the mean of all transaction prices in US dollars. I use every cross by group count for all regression analysis. All measures are calculated on a daily basis and then averaged across the sample period between 11/05/2012 and 11/09/2012. Only quote and trade records during market open hours between 11am and 3pm EST are used. All non-dummy variables are log-transformed, thus the coefficients may be interpreted as percentage changes. ***, ** and * indicate significance levels of 0.1%, 1% and 5% respectively.

\[
\ln(\text{Rel. Freq}_i) \text{ or } \ln(\text{Rel. Mag}_i) = \beta_1 \ln(\text{RelQspread}_i) + \beta_2 \ln(\text{QspreadStdDev}_i) + \beta_3 \ln(\text{AvgMktCap}_i) + \beta_4 \ln(\text{NumTrades}_i) + \beta_5 \ln(\text{AvgPrice}_i) + \beta_0 + \epsilon_i
\]

<table>
<thead>
<tr>
<th></th>
<th>Relative Cross Frequency</th>
<th>Relative Cross Magnitude</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coefficient</td>
<td>Std Error</td>
</tr>
<tr>
<td><strong>Full Sample (obs=1665)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rel Qspread</td>
<td>-1.1484</td>
<td>0.0582</td>
</tr>
<tr>
<td>Qspread StdDev</td>
<td>0.8405</td>
<td>0.0530</td>
</tr>
<tr>
<td>Avg Mkt Cap</td>
<td>-0.1349</td>
<td>0.0267</td>
</tr>
<tr>
<td>Num Trades</td>
<td>-0.0437</td>
<td>0.0214</td>
</tr>
<tr>
<td>Avg Price</td>
<td>0.2658</td>
<td>0.0383</td>
</tr>
<tr>
<td><strong>Intercept</strong></td>
<td>-3.3054</td>
<td>0.0801</td>
</tr>
<tr>
<td><strong>Adj-Rsquare</strong></td>
<td>0.2359</td>
<td></td>
</tr>
</tbody>
</table>
Table 19 – continued.

<table>
<thead>
<tr>
<th></th>
<th>Relative Cross Frequency</th>
<th></th>
<th>Relative Cross Magnitude</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coefficient</td>
<td>Std Error</td>
<td>t-value</td>
<td>Pr &gt;</td>
</tr>
<tr>
<td>Large-Cap (obs=456)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rel Qspread</td>
<td>-1.1606</td>
<td>0.1539</td>
<td>-7.54</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Qspread StdDev</td>
<td>0.5212</td>
<td>0.1308</td>
<td>3.98</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Avg Mkt Cap</td>
<td>-0.3356</td>
<td>0.0413</td>
<td>-8.13</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Num Trades</td>
<td>-0.0289</td>
<td>0.0397</td>
<td>-0.73</td>
<td>0.4675</td>
</tr>
<tr>
<td>Avg Price</td>
<td>0.9571</td>
<td>0.0666</td>
<td>14.37</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Intercept</td>
<td>-5.4032</td>
<td>0.2510</td>
<td>-21.53</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Adj-Rsquared</td>
<td>0.3173</td>
<td></td>
<td>0.5134</td>
<td></td>
</tr>
<tr>
<td>Mid-Cap (obs=959)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rel Qspread</td>
<td>-1.3237</td>
<td>0.0737</td>
<td>-17.95</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Qspread StdDev</td>
<td>1.0562</td>
<td>0.0665</td>
<td>15.88</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Avg Mkt Cap</td>
<td>-0.1143</td>
<td>0.0446</td>
<td>-2.56</td>
<td>0.0105</td>
</tr>
<tr>
<td>Num Trades</td>
<td>-0.0684</td>
<td>0.0251</td>
<td>-2.72</td>
<td>0.0066</td>
</tr>
<tr>
<td>Avg Price</td>
<td>0.1379</td>
<td>0.0496</td>
<td>2.78</td>
<td>0.0055</td>
</tr>
<tr>
<td>Intercept</td>
<td>-2.8180</td>
<td>0.1693</td>
<td>-16.64</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Adj-Rsquared</td>
<td>0.2955</td>
<td></td>
<td>0.2117</td>
<td></td>
</tr>
<tr>
<td>Small-Cap (obs=250)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rel Qspread</td>
<td>-1.2659</td>
<td>0.1725</td>
<td>-7.34</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Qspread StdDev</td>
<td>1.2511</td>
<td>0.1786</td>
<td>7.00</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Avg Mkt Cap</td>
<td>-0.0352</td>
<td>0.1240</td>
<td>-0.28</td>
<td>0.7772</td>
</tr>
<tr>
<td>Num Trades</td>
<td>0.0755</td>
<td>0.0626</td>
<td>1.21</td>
<td>0.2289</td>
</tr>
<tr>
<td>Avg Price</td>
<td>-0.1016</td>
<td>0.1258</td>
<td>-0.81</td>
<td>0.4198</td>
</tr>
<tr>
<td>Intercept</td>
<td>-2.5393</td>
<td>0.5496</td>
<td>-4.62</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Adj-Rsquared</td>
<td>0.3035</td>
<td></td>
<td>0.4057</td>
<td></td>
</tr>
</tbody>
</table>
This table provides a cross-sectional test on all the factors examined by Table 16 through 19, on the relative cross frequency and magnitude across stocks within the full sample, and the large-cap, mid-cap, and small-cap strata. If US is a dummy variable that indicates whether the company is based in the United states, collected from Bloomberg. If Primary is a dummy variable that indicates whether NASDAQ is the primary listing venue and this is the primary ticker symbol for the stock. For both dummy variables, 1 indicates true, 0 indicates false. They are used as proxies for the quality and reliability of the company’s reported accounting information. Float/Outstanding is the number of stock shares float divided by the number of shares outstanding for the stock, collected from Bloomberg. It is used as a proxy for the percentage of shares outstanding held by corporate owners and large institutions instead of floating. I collect the number of analyst recommendations on Earnings per share (EPS) forecasts on this stock from the IBES database, which ranges from zero to 62. I break down this variable into categorical dummies, at the ranges of =0, =1, [2,5], [6,10], [11,20], [20,40], [41,62]. Analyst StdDev is the standard deviation between the forecasted EPS by different analysts, also collected from the IBES database. Rel Qspread is the duration-weighted average quoted spread as percentage of the bid ask midpoint. Qspread StdDev is the standard deviation of the duration-weighted average quoted spread. They are used as proxies for information assymetry and microstructure noise. Avg Mkt Cap, Num Trades, Avg Price are control variables calculated for each stock stocks. Avg Mkt Cap is market capitalization in US dollar amount based on daily closing price collected from Bloomberg, and then taken average over the sample period. Num Trades is the total number of transaction counts. Avg Price is the mean of all transaction prices in US dollars. I use every cross by group count for all regression analysis. All measures are calculated on a daily basis and then averaged across the sample period between 11/05/2012 and 11/09/2012. Only quote and trade records during market open hours between 11am and 3pm EST are used. All non-dummy variables are log-transformed, thus the coefficients may be interpreted as percentage changes. ***, ** and * indicate significance levels of 0.1%, 1% and 5% respectively.

\[
\ln (\text{Rel. Freq}_i) \text{ or } \ln (\text{Rel. Mag}_i) = \beta_1 \text{IfUS}_i + \beta_2 \text{IfPrimary}_i + \beta_3 \ln \left( \frac{\text{Float}}{\text{Outstanding}} \right)_i + \beta_4 \text{IfNumAnalyst0}_i + \beta_5 \text{IfNumAnalyst1}_i + \beta_6 \text{IfNumAnalyst5}_i + \beta_7 \text{IfNumAnalyst10}_i + \beta_8 \text{IfNumAnalyst20}_i + \beta_9 \text{IfNumAnalyst40}_i + \beta_{10} \text{IfNumAnalystMore}_i + \beta_{11} \text{EPSStdDev}_i \\
+ \beta_{12} \ln (\text{RelQspread}_i) + \beta_{13} \ln (\text{QspreadStdDev}_i) + \beta_{14} \ln (\text{AvgMktCap}_i) + \beta_{15} \ln (\text{NumTrades}_i) + \beta_{16} \ln (\text{AvgPrice}_i) + \beta_0 + \epsilon_i
\]
Table 20 – continued.

<table>
<thead>
<tr>
<th></th>
<th>Full Sample (obs=1665)</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coefficient</td>
<td>Std Error</td>
<td>t-value</td>
<td>Pr &gt;</td>
<td>Coefficient</td>
<td>Std Error</td>
<td>t-value</td>
</tr>
<tr>
<td>Relative Cross Frequency</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Relative Cross Magnitude</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coefficient</td>
<td>Std Error</td>
<td>t-value</td>
<td>Pr &gt;</td>
<td>Coefficient</td>
<td>Std Error</td>
<td>t-value</td>
<td>Pr &gt;</td>
</tr>
<tr>
<td>If US</td>
<td>-0.0971</td>
<td>0.0606</td>
<td>-1.60</td>
<td>0.1091</td>
<td>0.0840</td>
<td>0.0340</td>
<td>2.47</td>
</tr>
<tr>
<td>If Primary</td>
<td>-0.0465</td>
<td>0.0894</td>
<td>-0.52</td>
<td>0.6028</td>
<td>0.0120</td>
<td>0.0502</td>
<td>0.24</td>
</tr>
<tr>
<td>Float/Outstanding</td>
<td>-0.1426</td>
<td>0.0507</td>
<td>-2.81</td>
<td>0.0050</td>
<td>**</td>
<td>0.0500</td>
<td>0.0285</td>
</tr>
<tr>
<td>If Num Analyst = 0</td>
<td>0.4985</td>
<td>0.6263</td>
<td>0.80</td>
<td>0.4262</td>
<td>0.3994</td>
<td>0.3514</td>
<td>1.14</td>
</tr>
<tr>
<td>If Num Analyst = 1</td>
<td>0.0510</td>
<td>0.1908</td>
<td>0.27</td>
<td>0.7892</td>
<td>-0.0505</td>
<td>0.1071</td>
<td>-0.47</td>
</tr>
<tr>
<td>If Num Analyst ∈[2,5]</td>
<td>0.0375</td>
<td>0.1864</td>
<td>0.20</td>
<td>0.8407</td>
<td>0.0231</td>
<td>0.1046</td>
<td>0.22</td>
</tr>
<tr>
<td>If Num Analyst ∈[6,10]</td>
<td>0.0253</td>
<td>0.1908</td>
<td>0.13</td>
<td>0.8944</td>
<td>-0.0583</td>
<td>0.1071</td>
<td>-0.54</td>
</tr>
<tr>
<td>If Num Analyst ∈[11,20]</td>
<td>0.1403</td>
<td>0.1963</td>
<td>0.71</td>
<td>0.4750</td>
<td>-0.1181</td>
<td>0.1102</td>
<td>-1.07</td>
</tr>
<tr>
<td>If Num Analyst ∈[21,40]</td>
<td>0.4203</td>
<td>0.2081</td>
<td>2.02</td>
<td>0.0436</td>
<td>*</td>
<td>-0.0044</td>
<td>0.1168</td>
</tr>
<tr>
<td>If Num Analyst ∈[41,62]</td>
<td>0.0813</td>
<td>0.2771</td>
<td>0.29</td>
<td>0.7692</td>
<td>0.1021</td>
<td>0.1555</td>
<td>0.66</td>
</tr>
<tr>
<td>EPS StdDev</td>
<td>0.0895</td>
<td>0.0442</td>
<td>2.02</td>
<td>0.0433</td>
<td>*</td>
<td>-0.0158</td>
<td>0.0248</td>
</tr>
<tr>
<td>Rel Qspread</td>
<td>-1.2335</td>
<td>0.0625</td>
<td>-19.74</td>
<td>&lt;.0001</td>
<td>***</td>
<td>-0.6033</td>
<td>0.0351</td>
</tr>
<tr>
<td>Qspread StdDev</td>
<td>0.8937</td>
<td>0.0569</td>
<td>15.71</td>
<td>&lt;.0001</td>
<td>***</td>
<td>0.4730</td>
<td>0.0319</td>
</tr>
<tr>
<td>Avg Mkt Cap</td>
<td>-0.1897</td>
<td>0.0302</td>
<td>-6.28</td>
<td>&lt;.0001</td>
<td>***</td>
<td>0.1097</td>
<td>0.0170</td>
</tr>
<tr>
<td>Num Trades</td>
<td>-0.0636</td>
<td>0.0222</td>
<td>-2.87</td>
<td>0.0041</td>
<td>**</td>
<td>-0.0769</td>
<td>0.0124</td>
</tr>
<tr>
<td>Avg Price</td>
<td>0.3151</td>
<td>0.0427</td>
<td>7.39</td>
<td>&lt;.0001</td>
<td>***</td>
<td>-0.1640</td>
<td>0.0239</td>
</tr>
<tr>
<td>Intercept</td>
<td>-3.0786</td>
<td>0.2284</td>
<td>-13.48</td>
<td>&lt;.0001</td>
<td>***</td>
<td>-1.2317</td>
<td>0.1282</td>
</tr>
<tr>
<td>Adj-Rsquare</td>
<td>0.2515</td>
<td>0.3233</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 20 – continued.

<table>
<thead>
<tr>
<th></th>
<th>Relative Cross Frequency</th>
<th></th>
<th>Relative Cross Magnitude</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coefficient</td>
<td>Std Error</td>
<td>t-value</td>
<td>Pr &gt;</td>
</tr>
<tr>
<td><strong>Large-Cap (obs=456)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>If US</td>
<td>-0.0684</td>
<td>0.0998</td>
<td>-0.69</td>
<td>0.4931</td>
</tr>
<tr>
<td>If Primary</td>
<td>-0.0727</td>
<td>0.1858</td>
<td>-0.39</td>
<td>0.6957</td>
</tr>
<tr>
<td>Float/Outstanding</td>
<td>-0.2399</td>
<td>0.1080</td>
<td>-2.22</td>
<td>0.0269</td>
</tr>
<tr>
<td>If Num Analyst = 0</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>If Num Analyst = 1</td>
<td>-0.0091</td>
<td>0.4377</td>
<td>-0.02</td>
<td>0.9834</td>
</tr>
<tr>
<td>If Num Analyst ε[2,5]</td>
<td>0.0583</td>
<td>0.2280</td>
<td>0.26</td>
<td>0.7983</td>
</tr>
<tr>
<td>If Num Analyst ε[6,10]</td>
<td>0.2295</td>
<td>0.2045</td>
<td>1.12</td>
<td>0.2625</td>
</tr>
<tr>
<td>If Num Analyst ε[11,20]</td>
<td>0.3639</td>
<td>0.1900</td>
<td>1.92</td>
<td>0.0561</td>
</tr>
<tr>
<td>If Num Analyst ε[21,40]</td>
<td>0.4698</td>
<td>0.1743</td>
<td>2.70</td>
<td>0.0073</td>
</tr>
<tr>
<td>If Num Analyst ε[41,62]</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>EPS StdDev</td>
<td>0.0928</td>
<td>0.1137</td>
<td>0.82</td>
<td>0.4148</td>
</tr>
<tr>
<td>Rel Qspread</td>
<td>-1.2171</td>
<td>0.1571</td>
<td>-7.75</td>
<td>&lt;.0001***</td>
</tr>
<tr>
<td>Qspread StdDev</td>
<td>0.5821</td>
<td>0.1337</td>
<td>4.35</td>
<td>&lt;.0001***</td>
</tr>
<tr>
<td>Avg Mkt Cap</td>
<td>-0.3098</td>
<td>0.0472</td>
<td>-6.57</td>
<td>&lt;.0001***</td>
</tr>
<tr>
<td>Num Trades</td>
<td>-0.0619</td>
<td>0.0439</td>
<td>-1.41</td>
<td>0.1590</td>
</tr>
<tr>
<td>Avg Price</td>
<td>0.9468</td>
<td>0.0723</td>
<td>13.10</td>
<td>&lt;.0001***</td>
</tr>
<tr>
<td>Intercept</td>
<td>-5.5811</td>
<td>0.4882</td>
<td>-11.43</td>
<td>&lt;.0001***</td>
</tr>
<tr>
<td><strong>Adj-Rsquare</strong></td>
<td>0.3464</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 20 – continued.

<table>
<thead>
<tr>
<th></th>
<th>Relative Cross Frequency</th>
<th>Relative Cross Magnitude</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coefficient</td>
<td>Std Error</td>
</tr>
<tr>
<td>Mid-Cap (obs=959)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>If US</td>
<td>-0.0238</td>
<td>0.0723</td>
</tr>
<tr>
<td>If Primary</td>
<td>-0.1595</td>
<td>0.1047</td>
</tr>
<tr>
<td>Float/Outstanding</td>
<td>-0.2507</td>
<td>0.0549</td>
</tr>
<tr>
<td>If Num Analyst = 0</td>
<td>0.9009</td>
<td>0.5500</td>
</tr>
<tr>
<td>If Num Analyst = 1</td>
<td>0.4637</td>
<td>0.2410</td>
</tr>
<tr>
<td>If Num Analyst ∈[2,5]</td>
<td>0.3991</td>
<td>0.2029</td>
</tr>
<tr>
<td>If Num Analyst ∈[6,10]</td>
<td>0.3867</td>
<td>0.1993</td>
</tr>
<tr>
<td>If Num Analyst ∈[11,20]</td>
<td>0.4356</td>
<td>0.2009</td>
</tr>
<tr>
<td>If Num Analyst ∈[21,40]</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>If Num Analyst ∈[41,62]</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>EPS StdDev</td>
<td>0.1295</td>
<td>0.0471</td>
</tr>
<tr>
<td>Rel Qspread</td>
<td>-1.3746</td>
<td>0.0750</td>
</tr>
<tr>
<td>Qspread StdDev</td>
<td>1.0219</td>
<td>0.0671</td>
</tr>
<tr>
<td>Avg Mkt Cap</td>
<td>-0.1975</td>
<td>0.0481</td>
</tr>
<tr>
<td>Num Trades</td>
<td>-0.0766</td>
<td>0.0256</td>
</tr>
<tr>
<td>Avg Price</td>
<td>0.2649</td>
<td>0.0562</td>
</tr>
<tr>
<td>Intercept</td>
<td>-3.2248</td>
<td>0.3189</td>
</tr>
<tr>
<td>Adj-Rsquare</td>
<td>0.3152</td>
<td></td>
</tr>
</tbody>
</table>
Table 20 – continued.

<table>
<thead>
<tr>
<th>Relative Cross Frequency</th>
<th>Relative Cross Magnitude</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coefficient</td>
<td>Std Error</td>
</tr>
<tr>
<td>Small-Cap (obs=250)</td>
<td></td>
</tr>
<tr>
<td>If US</td>
<td>-0.3433</td>
</tr>
<tr>
<td>If Primary</td>
<td>0.2349</td>
</tr>
<tr>
<td>Float/Outstanding</td>
<td>0.2631</td>
</tr>
<tr>
<td>If Num Analyst = 0</td>
<td>0</td>
</tr>
<tr>
<td>If Num Analyst = 1</td>
<td>-0.0507</td>
</tr>
<tr>
<td>If Num Analyst ∈[2,5]</td>
<td>0.0798</td>
</tr>
<tr>
<td>If Num Analyst ∈[6,10]</td>
<td>0.1049</td>
</tr>
<tr>
<td>If Num Analyst ∈[11,20]</td>
<td>-0.0667</td>
</tr>
<tr>
<td>If Num Analyst ∈[21,40]</td>
<td>0</td>
</tr>
<tr>
<td>If Num Analyst ∈[41,62]</td>
<td>0</td>
</tr>
<tr>
<td>EPS StdDev</td>
<td>-0.0715</td>
</tr>
<tr>
<td>Rel Qspread</td>
<td>-1.2397</td>
</tr>
<tr>
<td>Qspread StdDev</td>
<td>1.2506</td>
</tr>
<tr>
<td>Avg Mkt Cap</td>
<td>-0.0400</td>
</tr>
<tr>
<td>Num Trades</td>
<td>0.0758</td>
</tr>
<tr>
<td>Avg Price</td>
<td>-0.0942</td>
</tr>
<tr>
<td>Intercept</td>
<td>-2.3048</td>
</tr>
<tr>
<td>Adj-Rsquare</td>
<td>0.2962</td>
</tr>
</tbody>
</table>
Figure 5. Intraday QBBO price updates per minute.

Figure 6. Intraday share volume per minute.
Figure 7. Intraday length of no-trade duration in seconds (median).

Figure 8. Intraday average duration-weighted bid-ask spread.
Figure 9. Intraday standard deviation of the bid-ask spread.

Figure 10. Intraday standard deviation of the bid ask midpoint against market movement.
Figure 11. Intraday relative frequency of strict cross, individual count.

Figure 12. Intraday relative magnitude of strict cross, individual count.
Figure 13. Intraday relative frequency of strict cross, group count.

Figure 14. Intraday relative magnitude of strict cross, group count.
Figure 15. Intraday relative frequency of every cross, individual count.

Figure 16. Intraday relative magnitude of every cross, individual count.
Figure 17. Intraday relative frequency of every cross, group count.

Figure 18. Intraday relative magnitude of every cross, group count.
Figure 19. Histogram #1 on Number of Analyst Recommendations.

Figure 20. Scatter plot on Standard Deviation of Analyst Forecasted EPS vs Number of Analyst Recommendations.
Figure 21. Histogram on Shares Float as percentage of Shares Outstanding.

Figure 22. Scatter Plot on the Number of Analyst Recommendations against Average Market Cap.
Figure 23. Scatter Plot on the number of trades per stock against Average Market Cap.
Appendix B. Java Code for Monte-Carlo Simulation

```java
import java.text.SimpleDateFormat;
import java.util.Calendar;
import java.util.Random;
import java.io.FileNotFoundException;
import java.io.PrintWriter;
import java.io.UnsupportedEncodingException;
import java.lang.Math;

public class StochasticDynSim {

    public static final int EVENT_LIMIT = 1; // Limit Order
    public static final int EVENT_MARKET = 2; // Market Order
    public static final int EVENT_CANCEL = 3; // Cancel Limit Order
    public static final int EVENT_INVASION = 4; // Invasive Limit Order

    static Random r = new Random(); // r.nextDouble() returns a random number
    0.0 to 1.0 r.nextBoolean() returns randomly true or false

    static public String displayDate;
    // date for trades csv file
    static public Calendar calendar = Calendar.getInstance();
    // used to generate displayDate
    static SimpleDateFormat sDFormat = new SimpleDateFormat("yyyy-MM-dd HH:mm:ss"); // used to generate displayDate

    public double currentTime; // current time in day in seconds
    public double startTime; // 'market open' time in day in seconds
    public double lastTime; // 'market close' time in day in seconds

    public double lambdaK; // k component of lambda (limit order entry) factor
    public double lambdaA; // alpha component of lambda (limit order entry) factor
    public double lambdaAC; // alpha component constant factor of lambda (limit order entry) factor
    public double mu; // mu (market order) factor
    public double[] theta; // theta factors ((limit order cancellation), filled in at runtime)

    public double lambdaKInvade; // k component of lambda factor (Spread Invading limit order entry)
    public double lambdaAInvade; // alpha component of lambda factor (Spread Invading limit order entry)

    public int bestBid; // current best bid
    public int bestAsk; // current best ask
    public int ridingBestBid; // current best bid (this was previously used for 'sticky' prices -- preserved in code for future use)
    public int ridingBestAsk; // current best ask (this was previously used for 'sticky' prices -- preserved in code for future use)

    public int lastSpecialBestBid; // used to know when to add to qbo csv if only simulating partial book
```
public int lastSpecialBestAsk; // used to know when to add to qbbo csv if only simulating partial book

public int minPrice = 0; // the lowest possible price offset in the book
public int priceOffset; // the actual lowest price in the book
public int maxPrice; // the highest price offset in the book

public int tickBins; // the total number of prices in the book
public int[] currentLOB; // current quantity for each tick bin
public int[] specialLOB; // a subset of the currentLOB for simulating a partial book

public double specialLOBSIZE; // the size of the specialLOB subset of currentLOB: 0.0 to 1.0

public double[] nextLimit; // time in day in seconds of the next limit orders
public double[] nextInvasion; // time in day in seconds of the next spread invading limit orders
public double[] nextCancel; // time in day in seconds of the set of next cancellations
public double nextMarket; // time in day in seconds of next market order

public double invadeSuccess; // self calibration for spread invasion success
public double invadeTotal; // self calibration for spread invasion attempts
public boolean invadeCal; // start using self calibration (once an invasion attempt fails)
public double[] invadeThresh; // time in day in seconds of the set of next cancellations
public double invadeThreshMax; // maximum invasion value

public boolean invadeCalUse; // use spread invasion self calibration (set in initialize)

public int sequenceNumber; // sequence in trades csv and qbbo csv
public String symbol; // stock symbol for trades csv and qbbo csv

PrintWriter writerQBBO; // writes the qbbo csv
PrintWriter writerTrades; // writes the trades csv

/* Debug Statistics Globals */
int middayTrades; // Trades during midday of the current simulation
int middayQBBO; // QBBO updates during midday of the current simulation
static double totalQBBO = 0; // total number of QBBO entries across all simulation runs (to determine the average)
double qspread; // cumulative duration weighted spread during midday (seems broken)
double lastTimeQspread; // time of last spread measurement
public int middayBestBid; // worst best bid during midday
public int middayBestAsk; // worst best ask during midday
public double middayInvadeSuccess; // self calibration invasion successes during midday + 1
public double middayInvadeTotal;   // self calibration invasion attempts during midday + 1
/* End Statistics Globals */

/* Debug Statistics Quotes Globals */
double qInvade;   // number of times a limit order is placed inside the spread
double qBest;     // number of times a limit order is placed at best bid or best ask
double qOutside;  // number of times a limit order is placed outside of best bid and best ask
/* End Statistics Globals */

public static void main(String[] args) {
  StochasticDynSim sim = new StochasticDynSim();
  int startIndex = 1;   // index of first iteration for csv file
  int iterations = 100; // total number of simulations to perform
  sim.initialize();    // set up common data for all simulations
  for(int i = startIndex; i < (startIndex + (iterations)); i = i + 1){
    sim.generate(i);  // run a simulation
  }
  System.out.println("Average QBBO updates: " + totalQBBO / iterations);
}

/**
* Run once for a stock; sets startup values;
*/
public void initialize() {
  double thetaAmp;
  startTime = 25100.0;   // ~7:00 (to give time to settle)
  lastTime = 56700.0;    // 3:45
  tickBins = 400;        // $4.00 range
  specialLOBSize = 1.00; // size of the specialLOB relative to the currentLOB (0.0 to 1.0)
  calendar.set(2012, 10, 20, 0, 0, 0); // 'month' is 0 to 11, not 1 to 12

  /* SBUX *****/
  symbol = "SBUX "; // stock ticker
  priceOffset = 5000 - (tickBins/2); // ~$50.00 SBUX
  mu = 1.424;        // mu factor for SBUX
  (20493/14400)
  lambdaK = 9.2;     // k component of lambda factor for SBUX
  lambdaKInvade = 0.2621354167; // k component of invasion lambda factor
  lambdaA = 1.49;    // alpha component of lambda factor (Zovko 2002)
  lambdaAC = 7.0;    // alpha constant factor (Zovko 2002)
lambdaAInvade = lambdaA;               // UNUSED: lambda invasion
alpha component rate
thetaAmp = 0.05;                      // theta amplification
factor
invadeCalUse = true;                  // use invasion self calibration

 /**************************************************************************
  * Stuff in this method below here shouldn't be changed */
  nextLimit = new double[tickBins];   //initialize the structure for
  next invasion limit orders
  nextInvasion = new double[tickBins]; //initialize the structure for
  next invasive limit orders
  nextCancel = new double[tickBins];  //initialize the structure for
  next sell cancels

  minPrice = 0;  // do not change
  maxPrice = minPrice + tickBins - 1;  // do not change

  }  

  /**************************************************************************
  * Run once for a sim (outputs a trades csv and a qbbo csv) 
  * @param index Simulation number (used to tag csv outputs)
  */
  public void generate(int index) {
    displayDate = sDFormat.format(calendar.getTime()); // get the date
    string for this iteration for trades csv
    sequenceNumber = 1;  // restart the
    sequence number for the csvs
    currentTime = startTime;  // initialize the
time in day

    /**************************************************************************
    * Initialize best bid and ask (and related values) for this
    simulation run */
    bestBid = minPrice;
    bestAsk = maxPrice;
    ridingBestBid = bestBid;
    ridingBestAsk = bestAsk;
    lastSpecialBestBid = bestBid;
    lastSpecialBestAsk = bestAsk;
    invadeSuccess = 1.0;   // 1 to prevent divide by zero
    invadeTotal = 1.0;     // 1 to prevent divide by zero
    invadeCal = false;

    /**************************************************************************
    * Debug Statistics Initialization */
    middayBestBid = maxPrice; //reversed for 'worst'
    middayBestAsk = minPrice; //reversed for 'worst'
middayQBBO = 0;
middayTrades = 0;
middayInvadeSuccess = invadeSuccess;
middayInvadeTotal = invadeTotal;
lastTimeQspread = 39600.0;
qspread = 0.0;

/* End Debug Statistics Initialization */

/* Debug Statistics Quotes Initialization */
qInvade = 0;
qBest = 0;
qOutside = 0;
/* End Debug Statistics Quotes Initialization */

currentLOB = new int[tickBins]; // initialize the limit order book
for this simulation
specialLOB = new int[tickBins]; // initialize the special limit order
book for this simulation

/* determine times for all the next possible events */
updateNextMarket();
for(int i = minPrice; i <= maxPrice; i++){
    updateNextLimit(i);
    updateNextInvasion(i);
    updateNextCancel(i);
}

/* set up our qbbo and trade csv output files */
try {
    writerQBBO = new PrintWriter("qbbo_" + index + ".csv", "UTF-8");
} catch (FileNotFoundException | UnsupportedEncodingException e) {
    e.printStackTrace();
}
try {
    writerTrades = new PrintWriter("trades_" + index + ".csv", "UTF-8");
} catch (FileNotFoundException | UnsupportedEncodingException e) {
    e.printStackTrace();
}

/* Add header lines of field names to the files, and one data line
with a 5 character symbol name */
writerQBBO.println("symbol,timestamp,sequence,bid,ask");
writerQBBO.println("SUXXX," + (int)(currentTime * 1000) + "," + 0 + "," + 1.1 + "," + 1.2);

writerTrades.println("refdate,symbol,timestamp,sequence,quantity,price");
writerTrades.println(displayDate +",SUXXX," + (int)(currentTime * 1000) + "," + 0 + ",100," + 1.15);

/* Add some orders to the book... */
addOrder(5 * maxPrice / 9, true, true, 1);
addOrder(4 * maxPrice / 9, false, true, 1);
addOrder(5 * maxPrice / 9, true, true, 1);
addOrder(4 * maxPrice / 9, false, true, 1);
addOrder(5 * maxPrice / 9, true, true, 1);
addOrder(4 * maxPrice / 9, false, true, 1);

addOrder(6 * maxPrice / 9, true, true, 1);
addOrder(3 * maxPrice / 9, false, true, 1);
addOrder(6 * maxPrice / 9, true, true, 1);
addOrder(3 * maxPrice / 9, false, true, 1);
addOrder(6 * maxPrice / 9, true, true, 1);
addOrder(3 * maxPrice / 9, false, true, 1);
addOrder(6 * maxPrice / 9, true, true, 1);
addOrder(3 * maxPrice / 9, false, true, 1);

addOrder(7 * maxPrice / 9, true, true, 1);
addOrder(2 * maxPrice / 9, false, true, 1);
addOrder(7 * maxPrice / 9, true, true, 1);
addOrder(2 * maxPrice / 9, false, true, 1);
addOrder(7 * maxPrice / 9, true, true, 1);
addOrder(2 * maxPrice / 9, false, true, 1);

/* Run the simulation */
while (currentTime < lastTime){
    doNextEvent();
}

/* Debug Statistics Output */
System.out.println("Trades: " + middayTrades + " QBBO: " + middayQBBO + " AvgSpread: " + qspread/14400.0 + " Range: " + (middayBestAsk-middayBestBid) + " InvTot: " + middayInvadeTotal + " InvSuc: " + middayInvadeSuccess);
totalQBBO += middayQBBO;
/* End Debug Statistics Output */

/* Debug Statistics Quotes Output */
System.out.println("Invades: " + (qInvade / (qInvade + qBest + qOutside)) + " Best: " + (qBest / (qInvade + qBest + qOutside)) + " Outside: " + (qOutside / (qInvade + qBest + qOutside))");
/* End Debug Statistics Quotes Output */

/* Clean up the file output */
writerQBBO.close();
writerTrades.close();

/* Increment the day in case we run again */
calendar.add(Calendar.DATE,1);
}

/**
 * Performs the next event (limit order entry or cancel, or market order)
 * This method will update the current time to when that event occurs
 */
public void doNextEvent(){
    double soonestEventTime = lastTime; //start at the last possible event time
    int soonestEventType = 0; // EVENT_LIMIT, EVENT_CANCEL,
EVENT_INVASION, or EVENT_MARKET
int soonestEventIndex = 0; // index into the event array

/* Search for the soonest next event */
for(int i = minPrice; i <= maxPrice; i++){
    if (nextLimit[i] <= soonestEventTime){
        soonestEventTime = nextLimit[i];
        soonestEventType = EVENT_LIMIT;
        soonestEventIndex = i;
    }
    if (nextCancel[i] <= soonestEventTime){
        soonestEventTime = nextCancel[i];
        soonestEventType = EVENT_CANCEL;
        soonestEventIndex = i;
    }
    if (nextInvasion[i] <= soonestEventTime){
        soonestEventTime = nextInvasion[i];
        soonestEventType = EVENT_INVASION;
        soonestEventIndex = i;
    }
}
if (nextMarket <= soonestEventTime){
    soonestEventTime = nextMarket;
    soonestEventType = EVENT_MARKET;
    soonestEventIndex = 1; // not used
}

currentTime = soonestEventTime; // set time to event being processed

if(soonestEventType == EVENT_LIMIT){
    if(r.nextBoolean()){
        // new Ask limit order entry
        index = ridingBestAsk + soonestEventIndex;
        // determine LOB index from best ask offset
        if(index <= maxPrice){
            // only
            add if its inside the book (protect array out of bounds)
            addOrder(index, true, true, soonestEventType); // add the order
            if(nextCancel[index] >= lastTime){
                // if no cancellation pending for this index
                updateNextCancel(index); // then
            }
        }
    }
}
or else{
    // new Bid limit order entry
    index = ridingBestBid - soonestEventIndex;
    // determine LOB index from best bid offset
    if(index >= minPrice){
        // only
        add if its inside the book (protect array out of bounds)
        addOrder(index, false, true, soonestEventType); // add the order
        if(nextCancel[index] >= lastTime){
            // if no cancellation pending for this index
            }
updateNextCancel(index); // then
add a pending cancellation at this index
}
}
updateNextLimit(soonestEventIndex); // always
update the next limit order entry time for this offset
}
else if (soonestEventType == EVENT_INVASION){
    int invasionDepth = soonestEventIndex; // soonestEventIndex is always 0 here for now
    if (r.nextBoolean()) {
        // new Ask spread invading limit order entry
        index = ridingBestAsk - (invasionDepth + 1); // determine LOB index from best ask offset
        if (index > minPrice) {
            // only
            add if its inside the book (protect array out of bounds)
            addOrder(index, true, true, soonestEventType); // add the order
            if (nextCancel[index] >= lastTime) {
                // add a pending cancellation at this index
                updateNextCancel(index);
            }
        }
    }
    else{
        // new Bid spread invading limit order entry
        index = ridingBestBid + (invasionDepth + 1); // determine LOB index from best ask offset
        if (index <= maxPrice) {
            // only
            add if its inside the book (protect array out of bounds)
            addOrder(index, false, true, soonestEventType); // add the order
            if (nextCancel[index] >= lastTime) {
                // add a pending cancellation at this index
                updateNextCancel(index);
            }
        }
    }

updateNextInvasion(soonestEventIndex); // always
update the next spread invading limit order entry time for this offset
}

/**
 * Performs the next event (limit order entry or cancel, or market order)
 * This method will update the current time to when that event occurs
 */
public void doNextEvent(){
    double soonestEventTime = lastTime; //start at the last possible event time
    int soonestEventType = 0; // EVENT_LIMIT, EVENT_CANCEL, EVENT_INVASION, or EVENT_MARKET
    int soonestEventIndex = 0; // index into the event array (
int index;

/* Search for the soonest next event */
for (int i = minPrice; i <= maxPrice; i++){
    if (nextLimit[i] <= soonestEventTime){
        soonestEventTime = nextLimit[i];
        soonestEventType = EVENT_LIMIT;
        soonestEventIndex = i;
    }
    if (nextCancel[i] <= soonestEventTime){
        soonestEventTime = nextCancel[i];
        soonestEventType = EVENT_CANCEL;
        soonestEventIndex = i;
    }
    if (nextInvasion[i] <= soonestEventTime){
        soonestEventTime = nextInvasion[i];
        soonestEventType = EVENT_INVASION;
        soonestEventIndex = i;
    }
}
if (nextMarket <= soonestEventTime){
    soonestEventTime = nextMarket;
    soonestEventType = EVENT_MARKET;
    soonestEventIndex = 1; // not used
}

currentTime = soonestEventTime; // set time to event being processed

if (soonestEventType == EVENT_LIMIT){
    if (r.nextBoolean()){ // new
        Ask limit order entry
            index = ridingBestAsk + soonestEventIndex; //
        determine LOB index from best ask offset
            if (index <= maxPrice){ // only
                add if its inside the book (protect array out of bounds)
                    addOrder(index, true, true, soonestEventType); // add
                the order
                        if (nextCancel[index] >= lastTime){ // if no
                    cancellation pending for this index
                        updateNextCancel(index); // then
                    add a pending cancellation at this index
                        }
                    }
    }
    else{ // new
        Bid limit order entry
            index = ridingBestBid - soonestEventIndex; //
        determine LOB index from best bid offset
            if (index >= minPrice){ // only
                add if its inside the book (protect array out of bounds)
                    addOrder(index, false, true, soonestEventType); // add
                the order
                        if (nextCancel[index] >= lastTime){ // if no
                    cancellation pending for this index
                        }
            }
    }
}
updateNextCancel(index); // then
add a pending cancellation at this index
}
}
updateNextLimit(soonestEventIndex); // always
update the next limit order entry time for this offset
}
else if(soonestEventType == EVENT_INVASION){
    int invasionDepth = soonestEventIndex; // soonestEventIndex is always 0 here for now
    if(r.nextBoolean()){
        // new
        Ask spread invading limit order entry
        index = ridingBestAsk - (invasionDepth + 1); //
        determine LOB index from best ask offset
        if(index >= minPrice) {
            // only
            add if its inside the book (protect array out of bounds)
            addOrder(index, true, true, soonestEventType); // add
            the order
            if(nextCancel[index] >= lastTime){ // if no
                add a pending cancellation at this index
            }
        }
    }
    else{
        // new
        Bid spread invading limit order entry
        index = ridingBestBid + (invasionDepth + 1); //
        determine LOB index from best ask offset
        if(index <= maxPrice){
            // only
            add if its inside the book (protect array out of bounds)
            addOrder(index, false, true, soonestEventType); // add
            the order
            if(nextCancel[index] >= lastTime){ // if no
                add a pending cancellation at this index
            }
        }
    }
    updateNextInvasion(soonestEventIndex); // always
update the next spread invading limit order entry time for this offset
}

/**
 * Updates the time of the next cancellation at a given LOB index
 * @param bin The LOB index of the pending cancellation
 */
public void updateNextCancel(int bin){
    double currentQuantity;
    int distance;
    if ((bin >= 0) && (bin < tickBins)) {
        // range check
```java
currentQuantity = Math.abs(currentLOB[bin]); // current number of orders on the book (whether buy or sell)
if (bin >= bestAsk){
    distance = bin - bestAsk; // find distance from best ask (to determine theta)
    if((currentQuantity == 0) || (distance < 0)){
        nextCancel[bin] = lastTime + 1; // no expiration if no limit orders at this price
    }
} else{
    nextCancel[bin] = currentTime + (-Math.log(r.nextDouble()) / (theta[distance] * currentQuantity));
}
else if (bin <= bestBid) {
    distance = bestBid - bin; // find distance from best ask (to determine theta)
    if((currentQuantity == 0) || (distance < 0)){
        nextCancel[bin] = lastTime + 1; // no expiration if no limit orders at this price
    }
} else{
    nextCancel[bin] = currentTime + (-Math.log(r.nextDouble()) / (theta[distance] * currentQuantity));
}
else{
    nextCancel[bin] = lastTime + 1; // no expiration if in the spread
}
}

/** Adds (or removes) an order to the book
 * @param price the tickbin of the order
 * @param sellOrder if limitEntry is true: true if sell, false if buy
 * @param limitEntry true if limit or market, false if cancel
 * @param orderType used to determine if a trade or not
 */
public void addOrder(int price, boolean sellOrder, boolean limitEntry, int orderType) {
    int bestPrice; // temporary storage to find best bid and best ask
    boolean updateRiding = false; // currently unused (riding variables always updated -- no sticky prices)

    if(limitEntry){
        if(sellOrder){
            if(price >
                bestBid){ // place new sell limit order
                if(r.nextDouble() <=
                    specialLOBSize){ // add to the partial limit order book only sometimes
                    specialLOB[price]++;
                }
            }
        }
    }
}
```
currentLOB[price]++;  
// always add to the global limit order book

if((orderType == EVENT_INVASION) && (invadeCal && invadeCalUse)) {  
  // if this order is supposed to be inside the spread
  invadeSuccess++;  
  // and auto-calibration is currently active, track success
  /* Debug Statistics Calculation 4 */
  if((currentTime > 39600.0) && (currentTime < 54000.0)) {
    middayInvadeSuccess++;  
    // also track success during mid-day for static purposes
  }
  /* End Debug Statistics Calculation 4 */
  /* Debug Statistics Quotes Calculation */
  if (price < bestAsk) { qInvade++; }
  else if (price == bestAsk) { qBest++; }
  else { qOutside++; }
  /* End Debug Statistics Quotes Calculation */
}
else if (orderType != EVENT_INVASION) {  
  // market order sell -- Trade (ignore spread invasion limit order entries)
  if(bestBid > minPrice) {  // range protection
    if(r.nextDouble() <= (double)specialLOB[bestBid]/(double)currentLOB[bestBid]) {
      specialLOB[bestBid]++;  
      // remove from the partial limit order book only sometimes (if there are orders)
      currentLOB[bestBid]++;  
      // always remove from global limit order book
    }
    else {  // we really don't want to get here... if we do,
      then tickBins should be increased (but this protects the simulation from crashing)
      currentLOB[minPrice]++;  
    }
  }
  else {  
    invadeCal = true;  // if an spread invading order fails,
    then we can start using auto-calibration
  }
}
else {  
  /* buy Order */
  if(price < bestAsk) {  // place new buy limit order
    if(r.nextDouble() <= specialLOBSIZE) {  // add to the partial limit
      specialLOB[price]--;  
    }
    currentLOB[price]--;  
    // always add to the global limit order book
  }
}
if(orderType == EVENT_INVASION) && (invadeCal && invadeCalUse)){ // if this order is supposed to be inside the spread
    invadeSuccess++;
    // and auto-calibration is currently active, track success
    /* Debug Statistics Calculation 5 */
    if((currentTime > 39600.0) && (currentTime < 54000.0)){
        middayInvadeSuccess++;
    // also track success during mid-day for static purposes
    }
    /* End Debug Statistics Calculation 5 */
}
/* Debug Statistics Quotes Calculation */
if(price > bestBid) { qInvade++; }
else if (price == bestBid) { qBest++; }
else { qOutside++; }
/* End Debug Statistics Quotes Calculation */
else if (orderType != EVENT_INVASION){ // market
    order buy -- Trade (ignore spread invasion limit order entries)
    if(bestAsk < maxPrice){ // range protection
        if(r.nextDouble() <=
            (double)specialLOB[bestAsk]/(double)currentLOB[bestAsk]){ // remove
            specialLOB[bestAsk]--;
        }
    } else{ // we really don't want to get here... if we do,
        then tickBins should be increased (but this protects the simulation from
        crashing)
            currentLOB[bestAsk]--;
    }
} else{
    invadeCal = true; // if an spread invading order fails,
    then we can start using auto-calibration
    }
} else{ /* cancellation */
    if(currentLOB[price] > 0){ // Sell order cancellation
        if(r.nextDouble() <=
            (double)specialLOB[price]/(double)currentLOB[price]){ // remove from the partial
            specialLOB[price]--;
        }
    } else if(currentLOB[price] < 0){ // Buy order cancellation
        if(r.nextDouble() <=
            (double)specialLOB[price]/(double)currentLOB[price]){
specialLOB[price]++; // remove from the partial
limit order book only sometimes (if there are orders)
} currentLOB[price]++; // always remove from the
global limit order book
} // If no orders to cancel
at this price, do not do anything

/* Auto-calibration of spread invasion */
if((orderType == EVENT_INVASION) && (invadeCal && invadeCalUse)){ //
if auto-calibration is active and enabled
if (invadeTotal/invadeSuccess < 20.0) invadeTotal++; //
increase the attempts count (but cap it to prevent really slow simulations)
/* Debug Statistics Calculation 6 */
if((currentTime > 39600.0)&&(currentTime < 54000.0)){
    middayInvadeTotal++;
}
/* End Debug Statistics Calculation 6 */
}

/* update best bid/ask */
for(bestPrice = minPrice; bestPrice < maxPrice; bestPrice++){
    if(currentLOB[bestPrice] > 0) break;
}
bestAsk = bestPrice;
for(bestPrice = bestAsk; bestPrice > minPrice; bestPrice--){
    if(currentLOB[bestPrice] < 0) break;
}
bestBid = bestPrice;

ridingBestBid = bestBid; // this was used for sticky prices -- right
now it just passes the value through
ridingBestAsk = bestAsk; // this was used for sticky prices -- right
now it just passes the value through
/* Debug Statistics Calculation 1 */
if((currentTime > 39600.0)&&(currentTime < 54000.0)&&(bestBid <
middayBestBid))
    middayBestBid = bestBid;
if((currentTime > 39600.0)&&(currentTime < 54000.0)&&(bestAsk >
middayBestAsk))
    middayBestAsk = bestAsk;
/* End Debug Statistics Calculation 1 */

/* Determine best bid and ask for the partial limit order book which
may differ from the global book */
for(bestPrice = bestAsk; bestPrice < maxPrice; bestPrice++){
    if(specialLOB[bestPrice] > 0) break;
}
int specialBestAsk = bestPrice;
for(bestPrice = bestBid; bestPrice > minPrice; bestPrice--){
    if(specialLOB[bestPrice] < 0) break;
}
int specialBestBid = bestPrice;
/* write to qbbo csv (if best bid or best ask changed)
 * symbol,timestamp,sequence,bid,ask
 * BCOVVV,46969210,8246533,20.40000000,20.85000000
 */
if((specialBestBid != lastSpecialBestBid) || (specialBestAsk != lastSpecialBestAsk))
    lastSpecialBestBid = specialBestBid; // keep track of partial book best bid between added orders
    lastSpecialBestAsk = specialBestAsk; // keep track of partial book best ask between added orders
writerQBBO.println(symbol + "", (int)(currentTime * 1000) + "," + sequenceNumber + "," + (priceOffset+specialBestBid)/100.0 + "," + (priceOffset+specialBestAsk)/100.0);
/* Debug Statistics Calculation 2 */
if((currentTime > 39600.0)&&(currentTime < 54000.0)){
    middayQBBO++;
    qspread += (double) (specialBestAsk - specialBestBid) *
    (currentTime - lastTimeQspread);
    lastTimeQspread = currentTime;
}
/* End Debug Statistics Calculation 2 */
/* write to trades csv (if a market order was done)
 * refdate,symbol,timestamp,sequence,quantity,price
 * 2012-11-05 00:00:00,BCOVVV,35356389,97978,1245,11.22000000
 */
if(orderType == EVENT_MARKET){ // if a trade occurred
    writerTrades.println(displayDate + "," + symbol + "," + (int)(currentTime * 1000) + "," + sequenceNumber + ",100," + (priceOffset+price)/100.0);
/* Debug Statistics Calculation 3 */
    if((currentTime > 39600.0)&&(currentTime < 54000.0)){
        middayTrades++;
    }
/* End Debug Statistics Calculation 3 */
}
sequenceNumber++; // increment the sequence for csv output
}
public StochasticDynSim() {
    /* No initialization here */
}
References


173


Locke, J. (1699). Some considerations of the consequences of lowering the interest rate and raising the value of money. pp. 63–64. Reprinted in *Several papers relating to money, interest and trade, etc.* New York, NY: Augustus M. Kelly. (Original work published 1692).


179


Walras, L. (1874) Éléments d'économie politique pure, ou théorie de la richesse sociale [Elements of pure economics, or the theory of social wealth]. Transl. W. Jaffé. Publication location unknown.


