MARKET MANIPULATION IN A HIGH-FREQUENCY CONTEXT

Colliding particle physics tools and financial market data

Marjolein E. Verhulst

Propositions

1. The large number of functionalities and rules in financial markets has made market surveillance extremely challenging.
   (this thesis)

2. Technological developments in financial markets advance faster than regulators can keep up.
   (this thesis)

3. All scientific research needs to involve multiple disciplines.

4. The narrow focus of top academic journals hinders scientific innovation by discouraging interdisciplinary research.

5. The investment by member states in CERN’s new proposed collider (The Future Circular Collider) accelerates research in other disciplines.

6. Holding social media influencers accountable reduces the spread of misinformation.

7. The individualization of society reduces empathy among people.

Propositions belonging to the thesis, entitled

Market manipulation in a high-frequency context: colliding particle physics tools and financial market data

Marjolein Eveline Verhulst

Wageningen, 27 September 2024

MARKET MANIPULATION IN A HIGH-FREQUENCY CONTEXT

Colliding particle physics tools and financial market data

Marjolein E. Verhulst

THESIS COMMITTEE

Promotors
Prof. Dr J.M.E. Pennings
Personal Chair at the Marketing and Consumer Behaviour Group
Wageningen University & Research

Dr C. Gardebroek
Associate Professor at the Agricultural Economics and Rural Policy Group
Wageningen University & Research

Other members
Prof. Dr B. Tekinerdogan, Wageningen University & Research
Prof. Dr G. Bagnarosa, Rennes School of Business, France
Prof. Mr H.M. Vletter-van Dort, Erasmus School of Law, Rotterdam
Dr S. Duivenvoorden, the Netherlands Authority for Consumers and Markets, Den Haag

This research was conducted under the auspices of the Graduate School Wageningen School of Social Sciences.

MARKET MANIPULATION IN A HIGH-FREQUENCY CONTEXT

Colliding particle physics tools and financial market data

Marjolein E. Verhulst

Thesis

submitted in fulfilment of the requirements for the degree of doctor
at Wageningen University
by the authority of the Rector Magnificus,
Prof. Dr C. Kroeze,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Friday 27 September 2024
at 10:30 a.m. in the Omnia Auditorium.

Marjolein E. Verhulst
Market manipulation in a high-frequency context: colliding particle physics tools and financial market data,
266 pages.

PhD thesis, Wageningen University, Wageningen, the Netherlands (2024)
With references, with summary in English and Dutch

DOI https://doi.org/10.18174/660427

Table of contents

Chapter 1

General introduction

"... in the longer run and for wide-reaching issues, more creative solutions tend to come from imaginative interdisciplinary collaboration."

– Robert J. Shiller, Nobel laureate in Economic Sciences (Nemko, 2016)

1.1 BACKGROUND

New technological developments in trading have followed rapidly ever since the digitalization of financial markets. Whereas trading previously occurred by open outcry in a physical trading pit, most financial markets shifted to electronic trading platforms in the 2000s. This transition, among other factors, has created the trading landscape as we know it today.

In electronic trading, buyers and sellers of a specific asset or instrument are registered in the electronic limit order book (LOB). The LOB is a central marketplace consisting of all prices and quantities (i.e., volumes) traders are willing to buy and sell for. The bid side of the LOB contains all buy orders, while the ask side contains all sell orders. Although many types of orders exist, the two main types are limit orders and market orders. Limit orders are used by traders willing to wait for their desired execution price. The prices of buy (sell) limit orders are lower (higher) than the current market price. Hence, these orders are resting in the LOB, waiting for the market price to reach the price of the limit order, i.e., the desired execution price. In other words, limit orders add liquidity to the market. Market orders are used by traders who wish to buy or sell immediately and are willing to take the market price at that moment. When a trader submits a buy (sell) market order, the market order executes immediately against resting sell (buy) limit orders. If the market order is large enough, it can run through the LOB and consume multiple price levels. Hence, market orders take liquidity from the market.

With the digitalization of the LOB, it became possible to automate trading actions, thus opening financial markets to new traders, including those using algorithms. The rise of algorithmic trading has made speed more important, as high-frequency trading (HFT) algorithms operate in nanoseconds (MacKenzie, 2021). Trading actions can be executed faster than ever before, and many actions can be automated, for example splitting one large order into multiple orders or selling when the price drops below a certain point. This also means that new types of market manipulation are now possible, occurring faster than visible to the naked eye. A manipulation type called 'spoofing', for example, can take place within a few seconds.

The digitalization of the LOB and HFT, among other developments, resulted in the availability of large amounts of data (known as 'big data') on financial markets. Markets with an infinite number of LOB levels that are open almost 24 hours per day, where trading activity sometimes occurs multiple times per millisecond and high-frequency traders operate in nanoseconds, produce massive amounts of data. All market activities are captured in market messages. Each time something occurs in the market – i.e., a trade, new order, deleted order or modified order – the exchange receives such a message. These messages consist of tags and values describing the market activity in detail and are stored according to a protocol. All messages together can be used to reconstruct the LOB, since the LOB comprises all market activity by traders. For illustration purposes, the crude oil outright futures market of the Chicago Mercantile Exchange (CME) Group generated 9,395,104,705 messages between July 2019 and June 2020, totaling to approximately 1748 GB of uncompressed message data.

1.2 PROBLEM STATEMENT

Identifying and studying specific events in financial big data is challenging, particularly for regulators seeking to detect market manipulation. Market manipulation has evolved, partially due to the increased usage of algorithms; it has become more sophisticated and can now occur in the blink of an eye, within and between markets and exchanges, and with a single trading account or multiple accounts. Even leaving aside the difficulties in detecting manipulation posed by big-data, defining what is – and what is not – illegitimate behavior is challenging enough on its own. So, in this massive pile of data that keeps growing by the second, where should regulators start searching for indications of market manipulation? And what characterizes market manipulation in these markets in the first place?

One type of market manipulation that has received significant attention in recent years is spoofing, which is the focus of this dissertation. Generally, spoofing involves placing a relatively large spoof order on one side of the LOB, aimed to induce traders to trade on the opposite side of the LOB. Once the objective of the spoofer is reached – for example, buying or selling at a better price – the spoof order is cancelled. This is what distinguishes a spoof order from a genuine order; the spoof order was never intended to be executed whereas a genuine order is. Limit orders are perceived as information in the market about traders' intentions to buy or sell. This information is valuable, as market participants can use it to make or adjust their trading decisions. Contrary to genuine limit orders, spoof orders are not intended to be executed. Thus, they (are used to) expose market participants to false and misleading information (Cartea et al., 2020; Dalko et al., 2020; Mendonça & De Genaro, 2020) meant to make these participants act in ways they otherwise would not. This behavior is illegal and problematic for the market.

Spoofing – and market manipulation in general – can severely harm the functioning of financial markets such as futures markets, the focus of this dissertation. Futures markets are a zero-sum game, meaning that any profit made by a manipulator constitutes a loss for the counterparty. Spoofing can create (short-term) artificial prices and volatility, and, when conducted frequently, can lead to long-term issues in the market. In the long run, spoofing can impact price accuracy (Fox et al., 2021) and reduce trading activity, since market participants find trading less profitable (Fox et al., 2021) and have less confidence in the integrity and fairness of markets (Coppler, 2015; Fox et al., 2021; Sanders, 2016). Among other effects, this results in less efficient markets, as prices no longer reflect the true value of the instrument (Canellos et al., 2016; MacKenzie, 2022). It harms the price-discovery functionality of futures markets – and consequently their underlying assets – as future prices can no longer be accurately predicted (Coppler, 2015). Liquidity may decrease as market participants become less confident in the market. Moreover, the additional volatility introduced by market manipulation can theoretically affect margin requirements, leading to more margin calls (Park & Abruzzo, 2016). This is problematic for market participants, such as hedgers, who need more funds to cover margin calls or who might not be able to meet these increased margins. Hence, spoofing affects the integrity, efficiency and functioning of futures markets and, in the worst-case scenario, may harm one of their main purposes: risk management. As such, spoofing can have serious consequences for the real economy.

Both the economy and society can suffer significantly if futures contracts lose their effectiveness as risk-management instruments. Futures contracts are available for a variety of industries, including agriculture, (crypto)currencies, energy, equity, interest rates and metals. Agents such as farmers, manufacturers, processors, investors and corporations from various sectors typically use these futures contracts to hedge their price risks. If futures markets were to stop functioning properly, these agents could no longer use futures contracts to manage their exposure to price risks, thus endangering their viability. Prices might then rise to account for this risk. For example, banks charging higher interest rates on new mortgages and loans as they cannot hedge the interest rate effectively. Bread, chocolate, coffee, sugar and other products containing wheat, corn, sugar, cocoa and soybeans would become more expensive as farmers, processors and manufacturers can no longer hedge the price risk of these commodities, thus increasing their cost of capital. Prices of animal products, such as meat and dairy, would also increase as animal feed often contains corn, wheat and soybeans. Companies operating overseas would also raise their prices to account for currency-related price risks. Thus, the impact of market manipulation would trickle down from industries to the wallets of citizens and affect the spending capacity of households. In addition, if futures markets required higher margin deposits due to additional price volatility, hedgers would be exposed to higher capital costs, which in turn would have a negative impact on the innovation capacity of firms. Higher volatility would also impact clearing houses, increasing their risk exposure. This is particularly problematic, as they are vital institutions in the financial system. Hence, detecting market manipulations like spoofing is essential for the continued stability and effectiveness of futures markets, which are crucial for the functioning of the economy and are key platforms for risk management. While spoofing can have serious consequences on multiple levels in the economy, its actual frequency and economic impact are poorly understood.

Several challenges emerge in attempting to identify market manipulation within financial big data. These challenges are outlined in more detail in section 1.5 but can generally be categorized into two main groups: financial big data challenges and market manipulation research challenges. Financial big data challenges revolve around the storage, processing, visualization and analysis of big data. These tasks can be quite challenging due to the volume, complexity and irregularity of the data. Moreover, besides the sheer volume of the data, methodological gaps exist as academics and regulators are limited to traditional analysis tools, which often cannot handle the complexities and nuances of modern trading. Many statistical tests, for example, become ineffective due to the sheer size of the data, which makes even very small differences highly statistically significant. As a result, it is more difficult to distinguish between a genuine statistically significant result and a significant result caused by the enormous number of observations. Market manipulation research challenges comprise, among others, the identification of market manipulation, its economic impact, and the relevant legal framework defining it. To identify anomalies indicating market manipulation, one needs to have a comprehensive understanding of the market to first establish what is considered 'normal' market activity, as well as a thorough understanding of what the characteristics are of market manipulation. However, defining 'normal' market behavior is complex in itself. The legal framework for market manipulation is broad and – depending on the jurisdiction – offers little guidance on market manipulation characteristics, finding spoofing evidence and sanctioning. This makes it difficult to differentiate between legitimate and illegitimate trading behavior using only market data, often creating a need for supporting evidence, such as chat logs. Moreover, trading technologies advance rapidly, outpacing the development of detection tools, which creates a perpetual game of cat and mouse. Due to the challenges in identifying market manipulation in financial big data, the frequency and impact of these activities remain largely unknown. Hence, all the challenges described here and in section 1.5 can be considered scientific gaps and practical gaps in the financial industry. Together, these challenges of analyzing high-frequency data and detecting market manipulation introduce a distinct issue: the need to effectively identify and analyze market manipulation in a high-frequency context.

1.3 OBJECTIVE AND RESEARCH QUESTIONS

This dissertation examines the identification and analysis of market manipulation in a high-frequency data context. The central research question this dissertation aims to answer is: *How can market manipulation in a high-frequency context be identified and analyzed?*

The main research question is divided into four sub-research questions:

1. How can we improve understanding of high-frequency markets and developments therein? (Chapter 2)
2. How can we improve understanding of market manipulation in a high-frequency context? (Chapter 3)
3. What are the characteristics of the market manipulation practice of 'spoofing' in a high-frequency context? (Chapter 4)
4. What is the frequency and impact on liquidity of the market manipulation practice of 'spoofing' in high-frequency markets? (Chapter 5)

While the empirical focus of this work is on U.S. futures markets, the methodologies can be applied to any market with a LOB. This includes stock, options, crypto and spot markets, but also consumer marketplaces that use a LOB.

1.4 INTERDISCIPLINARY APPROACH: PARTICLE PHYSICS TOOLS

In addressing complex (scientific) challenges, it can be beneficial to step outside one's comfort zone and take an interdisciplinary approach. In this dissertation, we reap the benefits from the decades of big data experience in particle physics. Interestingly, particle physics offers a unique perspective on the challenge described in section 1.2. Particle physics studies the basic elements that make up matter (CERN, 2023a). The European Organization for Nuclear Research (CERN) is the most prominent institution in particle physics and home to the world's largest particle accelerator, the Large Hadron Collider (LHC). CERN has decades of experience in storing, processing and handling big, high-frequency data. To illustrate, CERN's disk storage capacity passed the threshold of one exabyte – 1,000,000,000 gigabytes – in 2023 (Smith, 2023). All experiments at the LHC use the ROOT software framework – developed by CERN in cooperation with other parties (Brun & Rademakers, 1997). ROOT offers all the tools needed to store, process, analyze and statistically test big data (Antcheva et al., 2009; Brun & Rademakers, 1997). Moreover, particle physics has extensive experience in anomaly detection. CERN looks for unusual particles, or anomalies, in LHC data that do not fit the Standard Model of particle physics. These anomalies may lead to new, ground-breaking insights for physicists. Similarly, anomalies in the form of unusual trading behavior can be identified in LOB data. Hence, academics, regulators and other financial market stakeholder aiming to detect market manipulation can benefit significantly from the big data and anomaly detection experience of particle physicists.

This dissertation explores this unique collision of two worlds – particle physics and finance – and is part of project High Energy Physics Tools for Limit Order Book Analysis (HighLO). Project HighLO is a collaboration between CERN, Wageningen University & Research, Commodity Risk Management Expertise Centre (CORMEC) and Maastricht University. The project applies particle physics methods and tools to financial market data to identify malicious trading behavior, such as market manipulation, to better protect financial markets. By doing so, the project envisions to accelerate te development of market-manipulation detection tools and contribute to a better understanding of market manipulation. In turn, this ensures fairer and better functioning markets.

1.5 RESEARCH FRAMEWORK

Figure 1.1 shows the research framework of this dissertation. The academic challenges this dissertation covers are split between challenges related to financial big data (A) and market manipulation research (B).

A. Financial Big Data Challenges

Storing and processing (A.1) financial market data is challenging since it is voluminous and complex. Prior to storing the data, one needs to make the trade-off between storing parts of the data or all data. Storing parts of the data increases the risk of throwing away too much information – and potentially biasing results – while storing all of the data increases the complexity of finding events to study. Storage capacity can be a hurdle, as even small subsets of financial market data might be too big to save locally. Storage possibilities such as external hard drives or cloud storage can solve this but need to offer efficient access and retrieval mechanisms to support any data processing or analysis. Moreover, due to the complexity of the data, people with specific data (science) skills must be involved to disentangle the data into something comprehensible and analyzable. For example, the data in this dissertation is stored according to the Financial Information eXchange protocol, containing tags and values necessary to reconstruct the LOB. Certain tag-value combinations act as 'if-statements' and need to be carefully considered when processing and reconstructing the data. Datasets also differ as to their contents and sometimes need to be carefully aligned. For example, the reconstruction of the consolidated LOB requires the careful linking of the outright and implied LOB, demanding a thorough understanding of the nature of the data. Overall, the complete data-management process demands substantial computing capacity and cannot be executed locally.

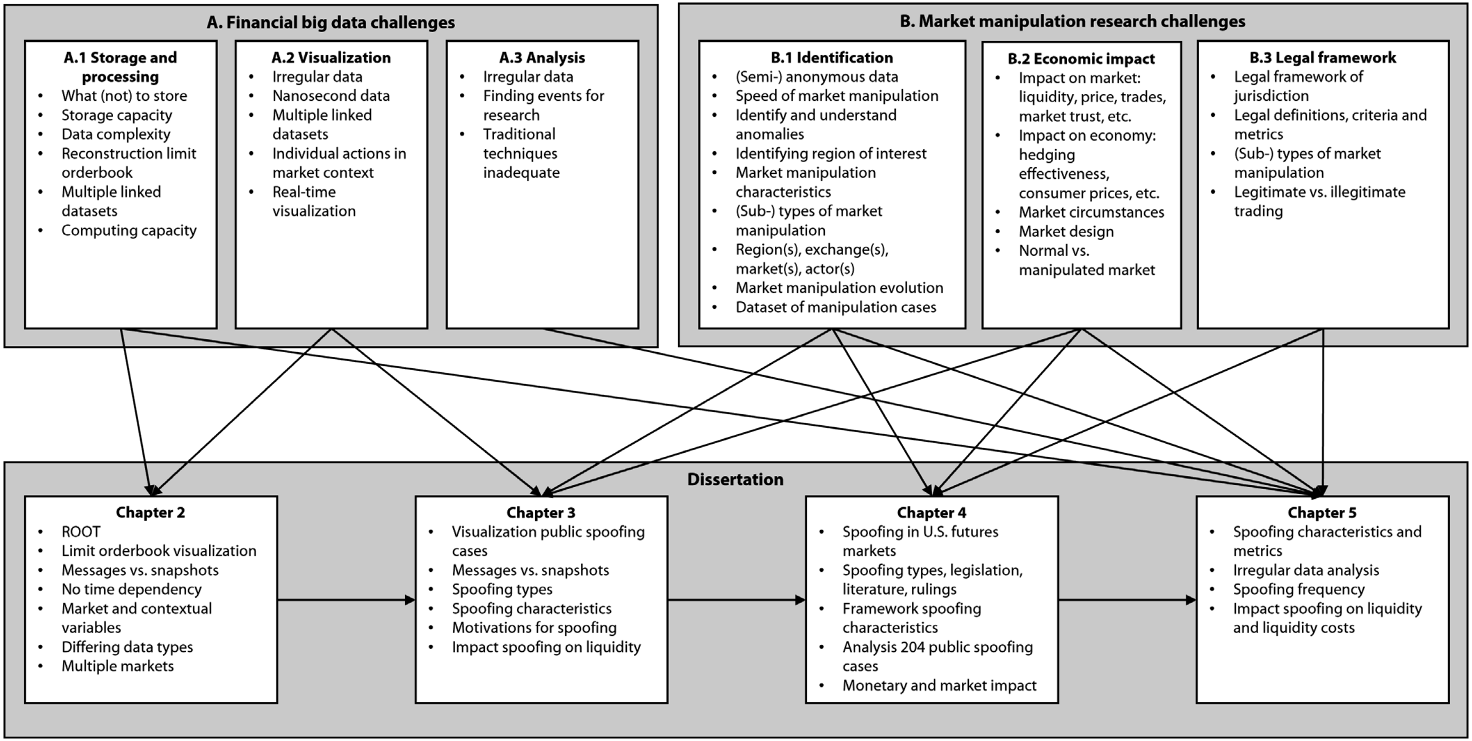


Figure 1.1 | Research framework of this dissertation.

Visualizing (A.2) high-frequency financial market data implies challenges, as the data is often timestamped at the nanosecond and highly irregular. This makes it difficult to visualize either a large time window – as there is too much data and details would become unnoticeable – or a small time window – as the surrounding (market) context of the time window would become obscured. To understand market behavior in its context, multiple datasets need to be carefully aligned in visualizations. For example, visualizing trades with the LOB requires alignment between a transaction dataset and a LOB dataset. Visualizing data in real time imposes additional challenges, as it requires an interface that conveys critical information effectively, so users can interact, understand and respond timely.

Finally, **analyzing** (A.3) financial-market big data has proven to be challenging. Traditional timeseries analysis tools rely on regular data, whereas market activity is highly irregular. Regular data has a single data point at fixed intervals, for example each second, hour, day or week. Market activity, on the other hand, generates irregular data as hundreds of actions sometimes occur within the same second, while at other times there may be no activity for several seconds. Moreover, finding events of interest is difficult, and small differences rapidly become statistically significant due to the magnitude of the data.

Overarching the financial big data challenges, the data management requires a seamless integration and interoperability between the storage, processing, visualization and analysis components. These systems need to work together harmoniously and require a cohesive, well-coordinated approach, so that each component not only addresses its specific challenges, but also complements and integrates well with the other components.

B. Market Manipulation Research Challenges

Besides challenges related to the data, several challenges arise in the context of market manipulation research. First, the **identification of market manipulation** (B.1). The type of order(book) data can range from fully anonymized datasets to those including specific IDs to market activity, each containing challenges in identifying market manipulation. For example, anonymized data complicates identification due to the lack of traceable actor information. Datasets with order IDs offer more granularity, but a single entity may operate under several IDs, obscuring the trail back to the original actor. Identifying events or anomalies that might indicate market manipulation is difficult, as market manipulation is often executed rapidly and can occur in the blink of an eye. In turn, finding those few seconds of market manipulation in days, weeks and years of data is cumbersome. More specifically, this could mean searching for a single market message – the one constituting market manipulation – among millions (depending on the timeline). Moreover, the differentiation between what constitutes an anomaly versus normal market behavior is complex, yet crucial in identifying market manipulation. The distinction between malicious anomalies – which may indicate manipulation – and benign ones is nuanced and requires a thorough understanding of 'normal' market characteristics, trading behavior and the legal framework, which are challenging concepts in themselves. Another layer of complexity concerns the undefined characteristics of market manipulation, combined with the numerous manipulation types, subtypes, evolutions of market manipulation over time, and their scope extending across markets, exchanges and regions. Identifying these widespread manipulation types demands access to multiple datasets, which is often hampered by the complexity in aligning datasets from different exchanges. Moreover, no centralized public dataset is available detailing all known manipulation cases, including timestamps of specific events, characteristics and impact on the market. Such a dataset could guide academics and financial stakeholders alike in identifying market manipulation.

Due to the above challenges in identifying market manipulation, it is also difficult to assess the **economic impact** (B.2) of market manipulation. The economic impact can be defined in terms of, for example, price impact or liquidity impact. Evaluating the economic impact requires the need to isolate the impact of market manipulation from other market variables, which is a challenging task. Moreover, gaining knowledge of how market circumstances or market design influence, discourage or encourage market manipulation requires a deep understanding of trading mechanisms, trading behavior and regulatory frameworks. As mentioned before, it also requires differentiation between a market in a manipulated state and one in a normal state. Assessing the long-term economic impact of market manipulation – not only directly on trading behavior, but also on factors such as market trust – is difficult since it requires different studies and datasets and, as mentioned before, isolation from other influencing variables. Market manipulation is often not confined to one market; its economic impact reaches beyond individual markets and can affect many facets of the global financial system. This shows how complex and intertwined the issues are when trying to assess the economic impact of market manipulation. This dissertation focuses on the economic impact in terms of liquidity. Liquidity is crucial for the functioning of markets, as it ensures that traders can buy/sell easily and quickly without significantly changing the market price. We do not consider other economic impacts of spoofing, such as its impact on price, market trust or hedging effectiveness.

The **legal framework** (B.3) adds to the complexities of market manipulation research, as market manipulation and its (sub-)types are often not captured in specific definitions with criteria and metrics. This makes it difficult to differentiate between legitimate and illegitimate trading based on these frameworks. Not only do the frameworks differ between countries, but also between financial products. For example, in the United States derivative and security markets have two different legal frameworks and some products, such as security futures products, are even regulated by both.

1.6 DISSERTATION OVERVIEW

Besides this introductory Chapter 1 and the concluding Chapter 6, this dissertation has four chapters that address the financial big data and market manipulation research challenges outlined in Figure 1.1. This section focuses on the chapters panel of Figure 1.1 and links the chapters to the respective financial big data (A) and market manipulation research (B) challenges.

Chapter 2 answers the first sub-research question: *"How can we improve understanding of high-frequency markets and developments therein?"* and addresses the storage, processing (A.1) and visualization (A.2) challenges of Figure 1.1. ROOT – the data-analysis tool for high-energy physics – is introduced to store and process financial market data, and the chapter offers a novel LOB-visualization tool. Different types of futures-market data are converted to ROOT files, which store and bundle data efficiently (addressing challenge A.1). This data conversion decreases the file size by approximately twenty times, resulting in reduced storage capacity needs and increased data accessibility. For example, the previously mentioned crude oil dataset in its original format is approximately 1748 GB in size but only 87.4 GB when converted to ROOT files. ROOT is then used to process the data and reconstruct the LOB. Next, a novel LOB-visualization tool is introduced that capitalizes on ROOTs histogram methodology, which is mainly used in particle physics analysis (addressing challenge A.2). The tool links multiple datasets and puts individual actions in the full LOB market context. It uses all information in the irregularly spaced message data and is not dependent on fixed time-intervals. Thus, for example, if one hundred market actions occur within one millisecond, it will not take averages but it will show all individual actions consecutively. The results illustrate how the visualization tool is easily adjustable to studying various topics in LOB research. For example, by visualizing other market characteristics alongside the LOB or multiple LOBs simultaneously to visually inspect correlation or spillover effects. The tool and its results are relevant for all financial market stakeholders, as it can be applied by all stakeholders in their day-to-day business. For example, traders can use the tool to monitor the market, and regulators can use it for market surveillance purposes. The tool enhances our understanding of high-frequency markets and the behavior in those markets, as it uses all market messages and puts individual trading actions in perspective of the market context.

To determine if the visualization tool for high-frequency markets is also effective in revealing market manipulation, **Chapter 3** applies the tool developed in Chapter 2 to one of the largest known spoofing cases to date. In doing so, Chapter 3 answers the second sub-research question: *"How can we improve understanding of market manipulation in a high-fre-*

quency context?". JPMorgan[1] settled in 2020 with the U.S. Commodity Futures Trading Commission (CFTC) over spoofing allegations for a record-breaking $920 million. The CFTC order outlining the charges against JPMorgan described nine specific examples of spoofing in U.S. treasury and metals markets between 2008 and 2016. Chapter 3 uses ROOT to store, process, and visualize market data from these nine spoofing examples. This includes market manipulation indicators and the impact of spoofing on market liquidity. The results improve our understanding of market manipulation in a high-frequency context, as they 1) demonstrate how various markets – and market indicators – respond to spoofing; 2) offer possible spoofing characteristics for identification purposes; 3) show how well-hidden spoofing can be; 4) provide insights into the complexities of the techniques required to recognize spoofing; 5) put a value on the miniscule price changes that makes spoofing economically viable; and 6) offer an alternative motivation for spoofing other than moving the price, in that spoofing can also be used to attract additional liquidity to the market rather than moving the price. The chapter demonstrates how high-frequency market data can be effectively visualized in the context of market manipulation, thereby enhancing the identification and understanding of, as well as the motivation for market manipulation. Hence, the results are particularly relevant for regulatory agencies and the compliance departments of exchanges. Chapter 3 addresses the visualization (A.2), identification (B.1) and economic impact (B.2) challenges outlined in Figure 1.1. The visualization (A.2) challenges are addressed in a similar manner as in Chapter 2, as Chapter 3 uses the same visualization tool. The chapter addresses identification (B.1) challenges by showcasing the rapid pace at which spoofing can unfold; the various actions spoofing involves; different subtypes of spoofing; and visualizations of markets that are affected by spoofing, including spoofing characteristics. The chapter offers guidance on how to characterize spoofing by way of variables – such as LOB-level specific volume and cancellations, liquidity costs and trade volume – and how to effectively visualize these variables. The economic impact (B.2) is specifically assessed by measuring the change in liquidity costs before, during and after spoofing.

Continuing the usage of real-life spoofing cases, **Chapter 4** dives deeper into spoofing by answering the third sub-research question: *"What are the characteristics of the market manipulation practice of 'spoofing' in a high-frequency context?".* All aspects of spoofing are delineated from both an economic and legal perspective, as it provides a comprehensive overview of spoofing types, legislation, academic literature and rulings. Together with expert knowledge from the *International Expert Group on Market Surveillance* (IMS Group)[2], these

1 Specifically, JPMorgan Chase & Co., JPMorgan Chase Bank, N.A., and J.P. Morgan Securities LLC.

2 The *International Expert Group on Market Surveillance* (IMS Group) is a collaboration between eighteen regulatory agencies across the world. It bridges the gap between high-energy physics, regulators and exchanges through co-creation and applying research findings in practice. At the time of writing, IMS Group consists of the Commodity Futures Trading Commission (CFTC), Chicago Mercantile Exchange Group (CME Group), Intercontinental Exchange (ICE) Futures Europe, European Securities and Markets Authority (ESMA), European Union Agency for the Cooperation of Energy Regulators (ACER),

aspects are combined into a conceptual framework consisting of spoofing dimensions and attributes. The framework helps to define the concept of spoofing, characterize spoofing, analyze spoofing cases and study legal responses to spoofing behavior. Using this conceptual framework, Chapter 4 studies and analyzes 204 spoofing cases, highlighting a variety of key spoofing behavior characteristics and legal responses. The developed conceptual framework and the results of Chapter 4 improve our understanding of spoofing and guides academics and practitioners into characterizing spoofing. Chapter 4 addresses all market manipulation research challenges (B) outlined in Figure 1.1. First, the results enhance our understanding in identifying (B.1) anomalies that can indicate spoofing. It highlights, among others, 1) various characteristics of spoofing; 2) subtypes of spoofing; 3) cross-exchange and -market spoofing; 4) spoofing executed by individuals and in collaboration; and 5) manual and algorithmic spoofing. Second, the results offer an overview of the economic impact (B.2) of spoofing on the market and its participants, including in monetary terms. Third, the legal framework (B.3) of these spoofing cases is delineated, and spoofing criteria and metrics are provided that were used by regulatory agencies to determine whether or not an order was legitimate (B.3). The comprehensive overview and conceptual framework of spoofing are aimed to accelerate research and insights on spoofing – and market manipulation in general – and relevant for all types of (financial) markets and their stakeholders.

Chapter 5 answers the final sub-research question: *"What is the frequency and impact on liquidity of the market manipulation practice of 'spoofing' in high-frequency markets?"*. It does so by identifying spoof orders in U.S. agricultural futures markets and assessing their impact on market liquidity. First, the spoofing characteristics from Chapter 4 are converted into metrics using spoofing cases from the CFTC. In other words, the legal framework (B.3) the CFTC uses is converted into criteria and metrics to define spoofing and differentiate illegitimate from legitimate trading behavior. These criteria and metrics are used to identify spoof orders (B.1) in six agricultural futures markets at CME Group. ROOT is used to store, process and filter the data reconstructing the LOB and linking multiple datasets in the process (A.1). More specifically, time windows are extracted around each identified spoof order, to measure changes in liquidity before, during and after spoofing. Next, the computing capacity (A.1) from the CERN Batch Service is used to conduct panel data regressions (A.3) and test the impact of spoof orders on three liquidity dimensions – tightness, depth and resiliency. The economic impact (B.2) is assessed in terms of liquidity costs. The fewest spoof orders were identified in the live cattle futures market (4,080 spoof orders) and the most in the soybean futures market (104,200 spoof orders). Results show that spoofing significantly affects

European Energy Exchange (EEX), Eurex Deutschland, Frankfurt Stock Exchange (FSE), Euronext Amsterdam, Netherlands Authority for the Financial Markets (AFM), Netherlands Authority for Consumers and Markets (ACM), Swiss Financial Market Supervisory Authority (FINMA), SIX Group, Commissione Nazionale per le Società e la Borsa (CONSOB), Borsa Italiana, Nord Pool and Autorité des Marchés Financiers (AMF).

liquidity dimensions, but responses depend on the market and are not consistent across markets. Liquidity dimensions tend to respond similarly within a single market: liquidity in the corn, soybean, soybean meal and soybean oil futures markets generally worsens after spoofing, and liquidity in the wheat and live cattle futures markets generally improves after spoofing. Moreover, results seem to suggest an inverse relationship between the frequency and the economic impact – in terms of liquidity costs – of spoofing. These results are relevant for 1) regulators wishing to enhance their surveillance systems using the spoofing criteria and metrics; 2) the decision-making process of exchanges regarding their market design to discourage spoofing; 3) market participants in their decision-making process when to or not to trade. Excluding the visualization (A.2) challenges, Chapter 5 addresses all financial big data (A) and market manipulation (B) challenges outlined in Figure 1.1. It combines all acquired knowledge from the previous chapters: particle physics tools are used to store, process and analyze data and spoof orders are identified using a framework embedded in law and economics to assess the impact of spoofing on market liquidity and its costs.

The dissertation is concluded with a general discussion in **Chapter 6**. The results from Chapters 2 to 5 are discussed in the context of the main research question. General theoretical and methodological reflections are provided, as well as implications for academics and practitioners, recommendations for future research and concluding remarks.

Chapter 2

When two worlds collide: using particle physics tools to visualize the limit order book

Marjolein E. Verhulst, Philippe Debie, Stephan Hageboeck,
Joost M.E. Pennings, Cornelis Gardebroek, Axel Naumann,
Paul van Leeuwen, Andres A. Trujillo-Barrera and Lorenzo Moneta

Published as:
Verhulst, M. E., Debie, P., Hageboeck, S., Pennings, J. M. E., Gardebroek, C., Naumann, A., van Leeuwen, P., Trujillo-Barrera, A. A., & Moneta, L. (2021). When two worlds collide: Using particle physics tools to visualize the limit order book. *Journal of Futures Markets*, 41(11), 1715-1734.

ABSTRACT

We introduce a methodology to visualize the limit order book (LOB) using a particle physics lens. Open-source data-analysis tool ROOT, developed by CERN, is used to reconstruct and visualize futures markets. Message-based data is used, rather than snapshots, as it offers numerous visualization advantages. The visualization method can include multiple variables and markets simultaneously and is not necessarily time dependent. Stakeholders can use it to visualize high-velocity data to gain a better understanding of markets or effectively monitor markets. In addition the method is easily adjustable to user specifications to examine various LOB research topics, thereby complementing existing methods.

Keywords: limit order book, visualization, particle physics, ROOT, liquidity

2.1 INTRODUCTION

The transition of financial and commodity exchanges from physical trading venues hosting open outcry auctions to predominantly electronic trading platforms (Hirsch et al., 2019), precipitated two major changes to financial markets. First, the limit order books (LOB) of electronic trading platforms became partially visible to market participants (see Appendix 2.A for an example of a LOB). The electronic LOB is a computerized system of all available demand and supply for securities and financial instruments (futures and options) at a specific time. All traders can submit orders to this platform with specifications, e.g., whether they wish to buy and/or sell, at which price and which quantity (Arzandeh & Frank, 2019). There are two main types of orders that can be submitted to the LOB: market orders and limit orders.[3] Market orders are immediately executed against the best bid or ask price. They consume liquidity and are placed by traders who immediately accept the market price (Hachmeister, 2007). Limit orders rest in the LOB and are executed at a pre-defined limit or at a better price. They provide liquidity and are placed by traders who can wait for their order to be executed (Hachmeister, 2007). While price discovery mainly occurs through limit orders, and these orders are much more numerous, market orders have a larger individual impact on the price (Brogaard et al., 2019). Second, the shift towards electronic trading platforms has transformed trading, in that algorithms can now trade, which has led to the emergence of high-frequency traders (HFT). Indeed, HFT and other forms of algorithmic trading now account for the majority of market turnover (Hirsch et al., 2019) and have introduced new challenges such as "Flash Crashes" (Aldridge & Krawciw, 2017; Bayraktar & Munk, 2018; CFTC-SEC, 2010; Golub et al., 2012; Kirilenko et al., 2017; Menkveld & Yueshen, 2019).

These changes create new challenges and opportunities for academics, regulators and industry participants alike, since they are faced with new high-frequency data that is more detailed and richer than ever before.[4] One approach to adjusting to the changes is to visualize the data. Visualization of the LOB helps stakeholders to provide the context of the market in which traders perform their actions and, hence, improve the understanding of the market they are analyzing and/or operating in and allows them, among others, to detect and identify anomalies. However, since LOB data is voluminous, complex in terms of structure and arriving at high frequencies, there is a need for a new way of thinking about storing, processing, visualizing and analyzing such data. This study attempts to address this issue by advocating the application of a visualization methodology commonly used in the particle physics literature to finance (see Appendix 2.B). Specifically, we try to answer the research question: to what extent can particle physics methodologies be used to visualize LOB data?

3 Market orders and limit orders are the main types, but variations exist. For example, limit orders can be hidden (hidden orders) and market orders can include special conditions such as whether to execute the volume completely or not at all (fill-or-kill) (Ranaldo, 2004).

4 For example, it takes the Securities and Exchange Commission five months to process and analyze two hours of LOB data (Gai et al., 2014).

The objective of this paper is twofold: first, to offer a novel visualization of the LOB, that is customizable by user specifications, using particle physics visualization tools. Second, to illustrate how the proposed visualization tool is easily adjustable to study various topics in LOB research, e.g., liquidity. Specifically, we use the open source data-analysis tool for high-energy physics ROOT, developed by CERN, among others, to reconstruct and visualize LOBs (Brun & Rademakers, 1997; CERN, 2018b).

This paper contributes to the existing literature by providing a visualization methodology that complements the existing visualizations with the following novel features: 1) LOB data can be visualized in different ways; either with time or messages on the x-axis, i.e., using "snapshots" or market activity; 2) the ability to visualize an extensive number of variables simultaneously; 3) the visualization of more complex concepts with separate but connected variables, such as liquidity; and 4) the visualization of multiple markets simultaneously. The methodology provides the ability to render LOB data in accordance with user specifications, making it visually easier to comb through LOB data and perform LOB data introspection. In addition, this paper introduces researchers to an open-source toolkit to store, process, generate statistics for, visualize and model LOB data. The focus of this paper is on the processing and visualization features of this toolkit.

2.2 LOB DATA AND RECONSTRUCTION

2.2.1 LOB Data

The LOB helps to explain the behavior of traders and informs theories of market microstructure and behavioral finance (Bhattacharya et al., 2018; Biais et al., 2010; Brolley, 2020; Buti et al., 2017; N. Chen et al., 2018; Chordia et al., 2019; Comerton-Forde et al., 2018; Dugast, 2018). The shift to electronic trading platforms has fundamentally changed market dynamics. LOB data contains more information and allows, among others, for more comprehensive measurements. For example, previous liquidity measurements took into account only one or two dimensions of liquidity whereas LOB data allows us to measure multiple liquidity dimensions simultaneously (Rösch & Kaserer, 2013). In addition, LOB data is stored according to various protocols, depending on the exchange. The Chicago Mercantile Exchange (CME Group), for example, uses the Financial Information eXchange (FIX) protocol, which provides the messages necessary to construct the LOB (FIXtrading, 2020). This means that academics have to reconstruct the LOB themselves, which may be challenging for various reasons.[5] Although regulators can get a better picture of the market using LOB data, its use also poses

5 Academics must understand the FIX protocol and the respective tags and values of messages to be able to reconstruct the LOB. Certain tag-value combinations, for example, act as "if-statements", which require careful consideration when reconstructing the LOB.

major challenges (Paddrik et al., 2016): the LOB can be difficult to reconstruct, visualize and analyze as it can be quite complex, due to, for example, various order types and different market microstructures across markets (Paddrik et al., 2016). In addition, high-frequency trading has a high velocity (nano- or milliseconds) and generates great amounts of data.

2.2.2 LOB Reconstruction: Messages vs. Snapshots

LOB data can be used to address many research questions, as it is usually high-frequency data and, thus, rich in information. Brogaard et al. (2014, 2019) highlight the additional information that is present in the LOB, and provide rationales as to why certain LOB data should be examined more carefully. For example, LOB data provides new information on trading strategies (of both HFTs and non-HFTs); (short-term) volatility; trading behavior around news announcements and imbalances in the LOB (Brogaard et al., 2014, 2019); front-running; manipulative strategies (Brogaard et al., 2019); and price discovery/efficiency in the LOB, including its levels and for different order types (Arzandeh & Frank, 2019; Brogaard et al., 2014, 2019; C. Cao et al., 2009). This paper builds on this notion by examining and comparing information present in LOB messages and in snapshots. Message data contains information the exchange receives about market activity.[6] It does not *contain* the LOB but can be used to *recreate* the LOB (FIXtrading, 2020). To the best of our knowledge, most LOB research does not use message data but already reconstructed LOB data. Studies that do use message data generally fail to provide information about the LOB reconstruction process, with a few exceptions (Arzandeh & Frank, 2019; Erenburg & Lasser, 2009). Moreover, little is known about handling, storing and processing the message data and reconstructing a LOB. The reconstruction process of the LOB is further outlined in section 2.3 and Appendix 2.C. This section highlights the differences between using messages and snapshots to re-create a LOB.

Orders arrive irregularly in the LOB, meaning that the messages received by the exchange are irregularly spaced over time. Snapshots of the LOB are needed to arrive at regular time intervals for analysis in a time-series framework. For example, if a snapshot size of one second is used, the last message within the one-second snapshot is used, i.e., only the net effect of all messages within one second is observed. The literature offers no uniform method to achieve the optimal snapshot size. Arzandeh and Frank (2019) calculated the optimal snapshot using the average duration of transaction price changes, following Engle and Russell (1998). Others, however, did not calculate the optimal snapshot size, choosing their intervals rather arbitrarily, anywhere between three seconds (Ito & Yamada, 2018), five seconds (Brogaard & Garriott, 2019), one minute (Gai et al., 2014; Hautsch & Horvath, 2019; Sinkovits et al.,

6 For example, part of a message can look as follows: 52=20150302150404453 107=ZCK5 269=0 270=392.5 271=8. This message was sent on March 2, 2015 at 15:04:04.453 (52=20150302150404453), it concerns the May 2015 corn futures contract (107=ZCK5) and mentions an update on the bid side of the LOB (269=0), whereby the volume at the price level of 392.5 dollar cents is updated to 8 futures contracts (270=392.5 and 271=8).

2014; Yao & Ye, 2018), three minutes (Hautsch & Horvath, 2019), five minutes (Kandel et al., 2012) and thirty minutes (Baruch et al., 2017).

The use of snapshots can be problematic for several reasons. First, if the snapshot size is too big, much of the data and information is lost in compression (i.e., into one snapshot), in that the snapshot only shows the aggregation of actions whereas their relative timings are lost. Conversely, if the snapshot size is too small, observations might repeat themselves – e.g., when there is little activity – introducing noise and problems such as heteroskedasticity in the dataset (Arzandeh & Frank, 2019). Second, high-speed trading produces high-frequency market data. In order to understand trading actions, it is beneficial to observe the same data granularity as the trading algorithms. Such nanosecond LOB data is highly granular input data. If this data is collapsed/aggregated into snapshots, much information is washed out, and analyses will suffer.

In this paper, every single message is used – i.e., no snapshots were taken,[7] except for comparison purposes. The analysis can handle message resolutions as detailed as those at the exchange itself, without the need for any aggregation. This means that there has been no further aggregation beyond the time precision and aggregation inherent in the CME data. Hence, all available information is used. Contrary to analyses that rely on the interpretation of snapshots, the use of highly granular data with irregular timing still allows for interpolations.

2.3 RESEARCH DESIGN: CME LOB DATA

Data consists of CME Group's proprietary market-depth dataset for all of 2015 for the U.S. Treasury Bond (T-Bond), corn, E-mini Dow Jones, crude oil, Henry Hub Natural Gas, soybeans, Chicago SRW wheat and rough rice futures markets. The files are in the CME Market Depth 3.0 (MDP) format, which provides the market messages required to recreate the LOB with millisecond precision. A market message is a set of tags and values which stores the data and metadata necessary to reconstruct the LOB. It is a sequential list of information about the LOB level without relation to other levels; this means that the preceding messages are needed to reconstruct the LOB (CME Group, 2020b).

Each file contains all contracts of the same futures contract ordered by message number and time of arrival. The LOB information is documented according to the FIX protocol (CME Group, 2020a). This protocol uses incremental updates of a data sequence to share exchange

7 Note that LOB reconstructions based on message data also use snapshots, in that the LOB is updated whenever a new message is received. These snapshots are based on irregular time intervals, however, contrary to the time-based snapshots, which use regular time intervals. Throughout the paper, we use the term "snapshot" to refer to these time-based snapshots, as opposed to message-based snapshots.

data with traders and regulators, either live or in batches. It does not store the LOB itself, but only the messages that can be used to recreate the LOB (FIXtrading, 2020).

MDP data provides information about LOB levels, for example when a new price level is inserted in the LOB, the price or volume is changed at a particular level, or a price level is deleted from the LOB. Among other data, messages contain information about prices, bids, asks, trading/order quantities and the time of placing each limit and market order is placed in the platform, up to ten orders deep. Appendix 2.C contains a description of the message types in each file and how the LOB is reconstructed.

Table 2.1 provides some descriptive statistics regarding the number of messages per day, messages per second, the total volume in the LOB per message, the total LOB value per message and the total number of messages for the contract duration of the December T-Bond, corn, E-mini Dow Jones and crude oil futures contracts in 2015.

Table 2.1 | Descriptive statistics of the December T-Bond, Corn, E-mini Dow Jones and Crude Oil futures markets in 2015

	December T-Bond	**December Corn**	**December E-mini Dow Jones**	**December Crude Oil**
Time window	May 1 – Dec 21	Jan 1 – Dec 14	Jan 1 – Dec 17	Jan 1 – Nov 26
Total # messages	171,330,999	112,378,245	200,514,606	520,742,261
Messages per day				
Minimum	3	30	10	431,574
Mean	1,044,701	462,462	815,100	2,283,957
Median	407,078	425,095	43,622	2,073,330
Maximum	3,816,504	1,708,758	4,480,008	11,572,077
Messages per second				
Minimum	0	0	0	0
Mean	13	7	10	28
Median	0	0	0	2
Maximum	4,295	14,958	5,475	4,105
Total LOB volume per message				
Minimum	20	44	0	0
Mean	4,529	3,356	477	247
Median	4,729	2,859	519	136
Maximum	13,787	26,508	1,392	2,427
Total underlying LOB value ($) per message				
Minimum	2,963,906	822,188	0	0
Mean	705,829,909	65,422,050	40,895,393	12,071,469
Median	737,803,000	56,137,338	43,314,230	7,436,740
Maximum	2,157,002,250	497,274,150	123,262,490	110,284,130

Note: Numbers are rounded. Each contract has a different time window because of different start and expiration dates. The total LOB value per message (in dollars) is the total sum of price levels multiplied by their respective volume, converted to dollars (e.g., T-Bonds trade in points and corn in dollar cents). Trading days where most CME futures markets were closed were taken out of the sample: New Year's Day (January 1), Good Friday (April 3), the day before Independence Day (July 3) and Christmas (December 25). Zeros in the table may be due to the recent launch of the contract or the contract nearing its expiry date.

2.4 PARTICLE PHYSICS VISUALIZATION OF HIGH-FREQUENCY DATA: CERN'S ROOT

2.4.1 ROOT: Data storage

The software framework ROOT (Brun & Rademakers, 1997; CERN, 2018b) is used to reconstruct the LOB and create the visualizations. ROOT is developed by CERN, in conjunction with other parties, and is used to analyze large amounts of data, especially in particle physics. It is mainly written in C++ but is integrated with other languages such as Python, R and Mathematica (CERN, 2018b). All experiments at the LHC use ROOT to store and analyze their data. It is built to store large amounts of data – to date, the LHC project has stored more than one exabyte of data, i.e., 1,000,000,000 gigabytes – and process that data efficiently in a distributed setup (Tejedor & Kothuri, 2018). ROOT's strength lies in its bundling with Cling, a unique interactive C++ interpreter based on Clang and LLVM libraries (LLVM, 2021). This allows ROOT access to the best of both worlds: rapid development of code and more compiled optimizations for fast code execution. ROOT can be used to save, access and mine data, produce graphs, and it can run interactively or be used to build stand-alone applications (CERN, 2018a). Its key features include advanced data structures; reading and writing objects; graphics and visualization toolkit; and analysis modules.[8]

A ROOT file (TFile) is a data container to store and bundle different types of related data, such as raw data, metadata or graphics. It allows for fast reading, compressed data storage and data access over networks. The data from CME Group is stored in ROOT files to facilitate analyses. While converting the data to ROOT, no reconstruction of the LOB is performed, since storing the full LOB – including every message – would require more storage space than reconstructing it on demand (the latter is done while generating the plot).[9] The conversion to the ROOT file structure decreases the overall file size and increases the accessibility of the data. Subset data reconstructed in the ROOT format is about 15 times smaller than the corresponding raw data.[10]

8 A few examples of modules are the Toolkit for Multivariate Data Analysis with ROOT (TMVA) for multivariate machine learning, RooFit for data fitting and RooStats for statistical analysis. This allows users, among others, to model the expected distribution of events, use neural/deep networks, function discriminant analysis, support vector machines, multidimensional minimization, fitting, parametrization, and likelihood ratio tests for hypothesis testing (Antcheva et al., 2009; CERN, 2021a, 2021c, 2021b). Python and R bindings are automatically generated for the full framework, which reduces the learning curve and complexity of development. Furthermore, ROOT is backwards compatible which reduces challenges in using old code.

9 The LOB reconstruction for a full year is relatively fast and complements previous research on processing LOB data(Gai et al., 2014).

10 ROOT is built to work on a computing cluster, but the core program is relatively small and can run on almost any computer, with as little as 500MB of RAM. The algorithms developed for this research use a streaming architecture, meaning that the data is loaded in chunks, so that even terabytes of data can be processed on a computer with just 500MB of RAM. This also means that the algorithms are not limited by file size, since only a part of the data is loaded into memory at any point in time. The time complexity of the algorithm is linear towards the number of messages (O(n)). Code can reconstruct the LOB from message data at a rate of up to 180MB per second, or, depending on additional processing steps, between 200,000 to 500,000 messages per second.

A caveat of ROOT is that it is primarily written for the needs of particle physics and, hence, might not offer tools that are considered basic in other disciplines. Also, ROOT is backward compatible, which has advantages (e.g., "old" code still works with the current ROOT version) as well as disadvantages in the form of odd design choices and practices. For example, ROOT has its own classes (e.g., TList or THashList) where one would typically expect C++ Standard Template Library (STL) containers. This is because STL was still in its infancy when ROOT was developed and needed such classes. In addition, ROOT is not in any textbooks nor taught in computer/data science, which may act as a barrier to using it.

2.4.2 Visualization in ROOT

As particle physics analysis mostly involves statistical distributions, one of ROOT's main features for aggregation and visualization is the use of histograms. Histograms offer a natural technique of reducing data and visualizing distributions, summarizing millions of single measurements of a quantity under study by filling a few bins (i.e., histogram bars). Consequently, ROOT offers various histogram classes, such as TH1D, a one-dimensional histogram storing double-precision floating-point numbers, TH2D, its two-dimensional counterpart; as well as three- and N-dimensional histograms (TH3D, THn). The x-axis of one-dimensional histograms aggregates different value ranges of a quantity, while the y-axis shows event counts or densities for each range (Antcheva et al., 2009).

Histograms are not only used to count integers, but also to weight entries in the datasets. This is often necessary when analyzing simulations, since the probability of detecting a specific simulated event has to be taken into account to generate the correct frequency distributions. For example, traded volume can be visualized, where each trade fills a histogram bin and is weighted by the trade's volume. Hence, the total content of the bin is the sum of all trades' volumes for a given parameter range. ROOT histograms automatically compute the statistical (Poisson) uncertainty in each bin and take weights into account (Antcheva et al., 2009). The data in histograms can be further manipulated: histograms can, for example, be re-binned, smoothed, added, scaled, subtracted or projected to analyze their content. In finance, the use of binned axes – especially for the abscissa – is uncommon. In this paper, histograms are used to plot messages, but the aggregation tool is not used to visualize the reconstructed LOB: since each bin corresponds exactly to one message, there is no counting involved. In the next sections, we show the value of data visualization using ROOT in the high-frequency finance context of the LOB.

ROOT's graphical interface is able to plot advanced graphs and histograms. ROOT can display multiple panels in one window, each providing a coordinate system to plot objects. Since a window can host multiple panels, different quantities can be visualized simultaneously. In this paper, for example, a panel showing a two-dimensional weighted histogram is displayed together with one or more panels showing one-dimensional histograms, where all panels have identical x-axes but different y-axes. With this technique, for example, the evolution of many variables versus time can be displayed and interactively manipulated.[11]

2.5 PARTICLE PHYSICS VISUALIZATION OF THE LOB

In this study, a two-dimensional histogram is used to visualize the state of the LOB. We illustrate how the visualization can be adapted to preferences and research topics of interest. The dates and time windows visualized are selected for illustrative (visualization) purposes only; these visualizations are not meant to lead to any conclusions. Example code of the visualizations and online, zoomable versions of all the visualizations are made public in a GitHub repository (https://github.com/HighLO).

Figure 2.1 shows the LOB of the T-Bond futures market on November 12, 2015 between 09:00 AM and 10:00 AM CT. The x-axis of the histogram shows the message sequence number within the time window studied. The y-axis shows the range of prices between which the LOB moves. The histogram is filled by reconstructing the LOB for every message. For each message, the bins are filled with the volume at each LOB price level. This means that every bin is only filled once – contrary to the traditional, aggregating use of histograms. The midpoint, i.e., the middle of the first bid and first ask price, is visualized by a red line. The ask levels are above and the bid levels below the red line. In addition to the LOB histogram, the cumulative traded volume is shown at the beginning of the time window in a separate panel for every message. The visualization uses messages; however, time is also of importance as there are time-priority rules in place prioritizing orders that arrive first (Yao & Ye, 2018). Therefore, the progression of time accumulated since the beginning of the time window is shown in the lower panel of the figure for every message. A steeper (flatter) line signals a lower (higher) rate of messages and, thus, a lower (higher) level of market activity. In other words: a steeper (flatter) line signals more (less) time progression. Hence, our visualizations capitalize on both messages (market activity) as well as snapshots (time), in that both are visualized simultaneously. All panels are combined into one graph with the same x-axis, i.e., message number. The visualization can also be performed using snapshots instead of messages, i.e., with time on the x-axis. If, for example, the snapshot size is set at one second, the histogram only plots the LOB of the last message of each second.

11 Generating the visualizations in this paper takes around 40 seconds on a standard laptop, with most of the computation power going to extracting the FIX messages and reconstructing the order book. The code developed for the proposed visualization methodology runs over the data twice: 1) once to extract the statistics needed for the visualization, such as the minimum and maximum price that are used on the y-axis of the LOB and 2) once more to reconstruct the LOB. Note that the code can be altered to only run over the code once (dynamically increasing/decreasing the maximum and minimum price), in which it takes just 20 seconds to generate the plots. It can reconstruct the LOB from message data at a rate of up to 180MB per second, or, depending on additional processing steps, between 200,000 to 500,000 messages per second.

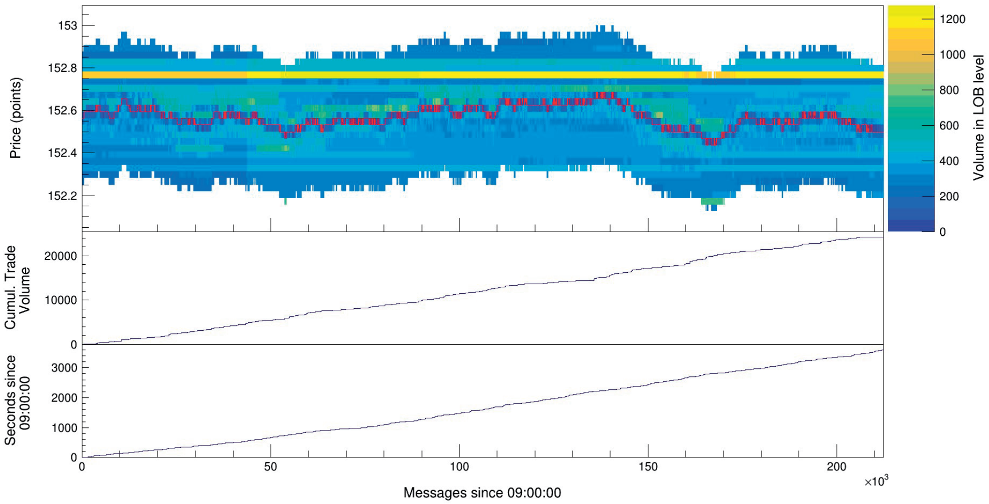


Figure 2.1 | December U.S. Treasury Bond (ZBZ5) LOB behavior (November 12, 2015 from 09:00 AM to 10:00 AM CT). The top panel shows the volumes at the 10 price levels on the bid and ask side of the LOB, respectively. Each unit on the x-axis is one message. The y-axis represents the price of the T-Bond in points. The color of each bin represents the volume in the LOB at that message and price. The scale ranges from blue to yellow, with the color becoming a brighter yellow as volume increases at that price level. The red line is the midpoint. The middle panel shows the cumulative trading volume for the selected time horizon. A steeper (flatter) line signals a higher (lower) rate of traded volume. The bottom panel shows how much time passes between messages reported by the exchange. A steeper (flatter) line signals a lower (higher) rate of messages, given that a steeper (flatter) line signals more (less) time progression. (data source: author's visualization of CME MDP 3.0 Market Data).

Data of the top 10 levels of the LOB was recorded in 2015. Hence, a total of 20 levels are plotted vertically for any message: the top 10 levels in the figures show the ask side and the bottom 10 levels show the bid side. Data outside these ranges still exists at the exchange, but traders did not see these levels and the exchange did not emit any messages for these price levels. Note that sometimes the last levels on both the bid and ask side are not completely filled. For example, the first and second levels may disappear, as either the volume at these levels was cancelled or a trade took place that consumed the volume of these levels. This would leave the LOB empty on the ninth and/or tenth level for a particular message.

Figure 2.1 shows that, on November 12 between 09:00 AM and 10:00 AM CT, the exchange received approximately 215,000 messages for the T-Bond futures market. The top panel shows that the LOB remained relatively stable, fluctuating between 152.1 and 153 points. Colors indicate the volume of the price level. For example, at the 152.75 points level, there was a consistently high volume throughout the time window, as indicated by the yellow color. This means that many sell orders were resting at this price level, waiting to be executed as soon as the price would reach this level. Cumulative trade volume rose at a steady pace (meaning that trades took place consistently), with a few spikes, e.g., at the 95,000, 135,000 and 140,000 message marks. Time progressed relatively stably, meaning that messages kept arriving at a regular rate, i.e., there were no periods with more or fewer messages arriving – which would have indicated more or less market activity.

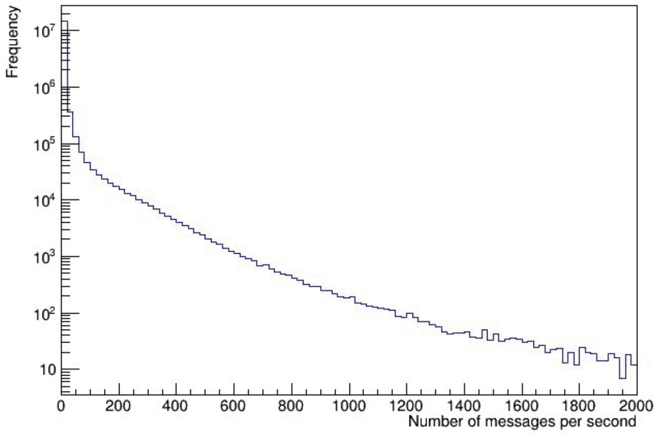


Figure 2.2 | Histogram of number of messages per second for the December Corn futures contract data in 2015 (ZCZ5). (data source: author's visualization of CME MDP 3.0 Market Data).

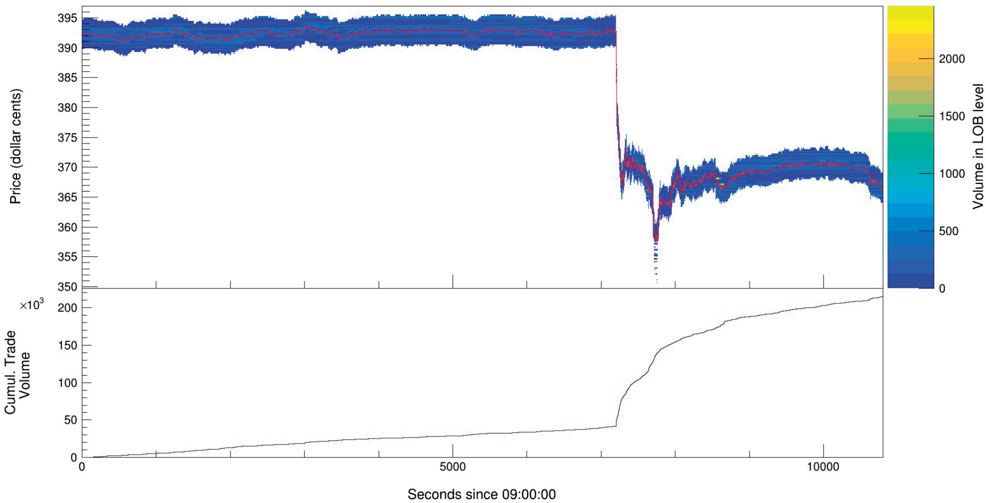


Figure 2.3 | December corn (ZCZ5) LOB behavior on an USDA announcement day using 5-second snapshots (August 12, 2015 from 09:00 AM to 12:00 AM CT). The top panel shows the volumes at the 10 price levels on the bid and ask side of the LOB, respectively. Each unit on the x-axis is one second. The y-axis represents the price of corn in dollar cents. The color of each bin represents the volume in the LOB at that second and price. The scale ranges from blue to yellow, with the color becoming a brighter yellow as volume increases at that price level. The red line is the midpoint. The bottom panel shows the cumulative trading volume per second. A steeper (flatter) line signals a higher (lower) rate of traded volume. (data source: author's visualization of CME MDP 3.0 Market Data).

In the sections below, the visualizations are modified to illustrate the benefits of using particle physics visualizations on LOB data. First, we illustrate the added value of using message numbers rather than time (i.e., snapshots) on the x-axis by visualizing the LOB of the December Corn contract for a U.S. Department of Agriculture (USDA) announcement day (Wednesday August 12, 2015). Second, we illustrate various ways to visualize trade volume and time by way of the December E-Mini Dow Jones futures contract. Third, two related markets – the December crude oil and December Henry Hub natural gas futures contracts – are plotted using the same timeframe. Fourth, the proposed methodology can help to visualize other variables as is demonstrated for liquidity in section 2.6. Finally, we show an additional visualization possibility by ROOT that compresses large amounts of data in a single figure in section 2.7.

2.5.1 Snapshot vs. Message

This section visualizes the LOB of the December Corn futures contract on a USDA announcement day, to illustrate the difference between using snapshots and messages in visualizations. First, Figure 2.2 shows a histogram of the December Corn futures contract to give an impression of how many messages arrive within one second. The x-axis shows the number of messages within one second; the y-axis is a logarithmic scale and shows how often a 'number of messages within one second' occurs in the dataset; each bar (or step/bin) represents 20 messages. Figure 2.2 shows that there are over ten million seconds that contain between 0-20 messages each. Next, there are more than 250,000 seconds that contain 20-40 messages each. The frequency of seconds decreases steadily as the number of messages per second increases, reaching a low of approximately 8 seconds that include 2000 messages each. Figure 2.2 thus illustrates that using one-second snapshots typically leads to a loss of information, as one second frequently contains multiple messages.

Next, the December Corn futures contract is visualized for a three-hour window on a USDA announcement day – August 12, 2015 from 09:00 AM to 12:00 AM CT – using both snapshots and messages. The announcement forecasted corn production to be 156 million bushels higher than projected in July. Supplies were forecasted to be 154 million bushels higher than projected in July, reaching a record of 15.5 billion bushels.

Figure 2.3 shows the LOB when using snapshots of five seconds and Figure 2.4 when using messages. The five-second snapshot was chosen arbitrarily for illustrative purposes. Figure 2.3 shows a steep drop at approximately 7000 seconds after 09:00 AM CT, exactly after the USDA report was announced, at 11:00 AM CT. The same drop can be seen in Figure 2.4 at approximately 210,000 messages after 09:00 AM CT. Simultaneously, trading increased significantly, as can be seen by the steep rise in cumulative trade volume in the second panel. Figure 2.3 shows two large LOB price drops: one immediately after the USDA announcement (at approximately 7000 seconds) and another one after the slight increase following the first drop (at approximately 7800 seconds). However, a different pattern emerges when visualizing all of the messages, as per Figure 2.4, rather than using snapshots. Instead of an immediate steep drop after the announcement, the LOB decreases (around the 210,000 message mark), after which it immediately recovers (around the 215,000 message mark) only to decrease further at a steady pace (around the 225,000 and the 305,000 message marks). Hence, Figure 2.4 shows three LOB price drops rather than the two drops identified in Figure 2.3. In addition, the increases and decreases after the steep LOB price drop are better visible in the message-based figure than in the snapshot-based figure. This is not due to zooming but to messages being compressed into seconds to create snapshots and, thus, to loss of information. The two figures illustrate that snapshots cannot reveal all of the information and market activity in the LOB: patterns can be better studied, in more detail, and visualizations can incorporate more information when using messages rather than when using snapshots (for example, orders that are added and subsequently almost immediately cancelled are not observable via snapshots but do become visible when using messages). Figure 2.4 also constitutes a novel method to visualize messages and time simultaneously, thus fully retaining any time-related information.

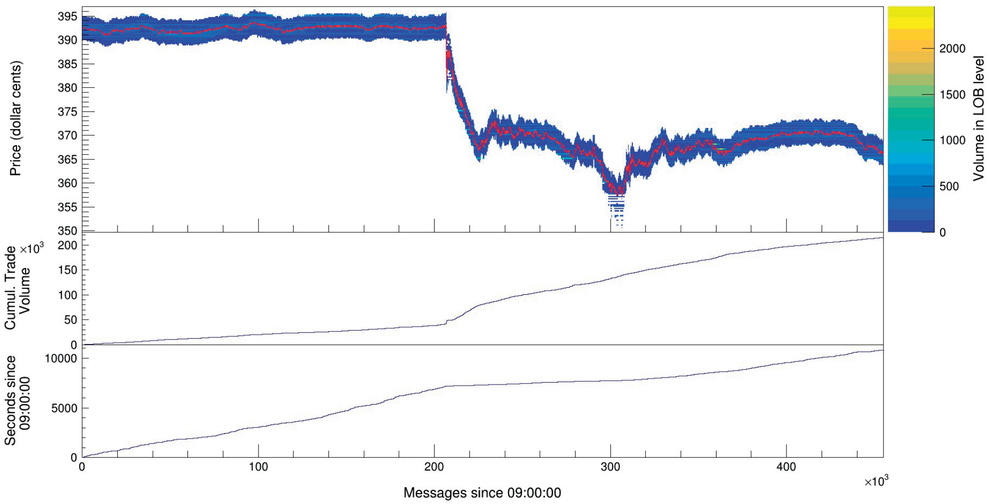


Figure 2.4 | December corn (ZCZ5) LOB behavior on an USDA announcement day using messages (August 12, 2015 from 09:00 AM to 12:00 AM CT). The top panel shows the volumes at the 10 price levels on the bid and ask side of the LOB, respectively. Each unit on the x-axis is one message. The y-axis represents the price of corn in dollar cents. The color of each bin represents the volume in the LOB at that message and price. The scale ranges from blue to yellow, with the color becoming a brighter yellow as volume increases at that price level. The red line is the midpoint. The middle panel shows the cumulative trading volume for the selected time horizon. A steeper (flatter) line signals a higher (lower) rate of traded volume. The bottom panel shows how much time passes between messages reported by the exchange. A steeper (flatter) line signals a lower (higher) rate of messages, given that a steeper (flatter) line signals more (less) time progression. (data source: author's visualization of CME MDP 3.0 Market Data).

2.5.2 Trade Volume and Time

The visualization can be modified to display many variables in various ways. Figure 2.5 illustrates this for trade volume and time by way of the December E-mini Dow Jones futures contract. In addition to cumulative trade volume, the middle panel shows trade volume per five-second snapshot in green bars. Note that this snapshot size was set by the authors. It is also possible to visualize the number of trades per message – though this would be less easily interpretable as there are more than 370,000 messages and, thus over 370,000 bars would have to be visualized. Around the 365,000 message mark, trade volume per snapshot rose to approximately 300 contracts. The LOB responded by increasing around the same message mark. This increase in trade volume is less discernable when looking at cumulative trade volume. Visualizing the same variable (i.e., trade volume) in various ways can thus result in different insights. In addition to seconds (since the start of the time window), time is visualized in 'messages per five-second snapshot'. This method of visualizing time gives

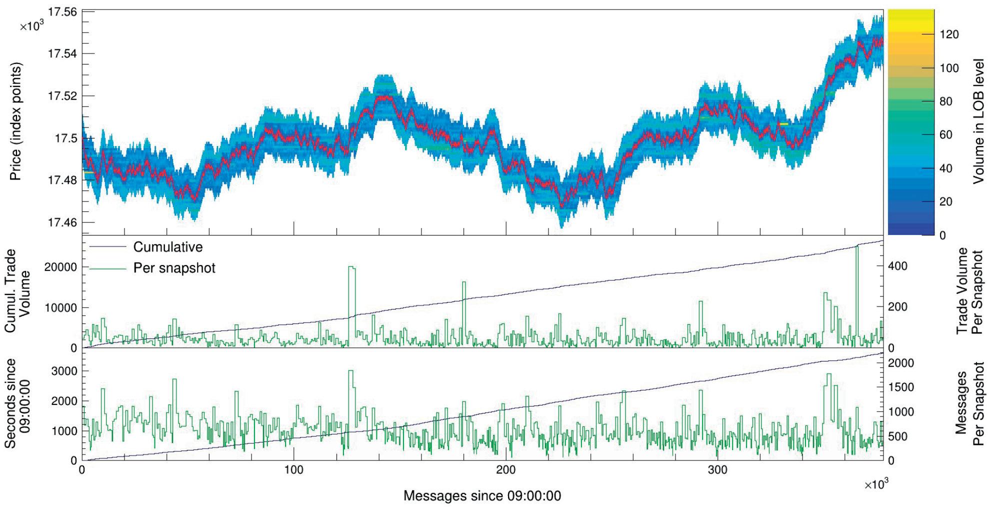


Figure 2.5 | December E-mini Dow Jones (YMZ5) LOB behavior (November 12, 2015 from 09:00 AM to 10:00 AM CT). The top panel shows the volumes at the 10 price levels on the bid and ask side of the LOB, respectively. Each unit on the x-axis is one message. The y-axis represents the price of the E-mini Dow Jones in index points. The color of each bin represents the volume in the LOB at that message and price. The scale ranges from blue to yellow, with the color becoming a brighter yellow as volume increases at that price level. The red line is the midpoint. The middle panel shows the cumulative trading volume for the selected time horizon and the trade volume per snapshot (5 seconds). The blue line indicates the cumulative trade volume on to the left y-axis; a steeper (flatter) line signals a higher (lower) rate of traded volume. The green bars show the total traded volume in a 5-second window on the right y-axis. The bottom panel shows how much time passes between messages reported by the exchange and how many messages occur within one snapshot of 5 seconds. A steeper (flatter) blue line signals a lower (higher) rate of messages, given that a steeper (flatter) blue line signals more (less) time progression. The blue line is related to the left y-axis. The green bars are related to the right y-axis and show the number of messages that occur within one snapshot of 5 seconds. (data source: author's visualization of CME MDP 3.0 Market Data).

a better impression when the exchange reports many messages. For example, around the 135,000 message mark, more messages occur within several snapshots, indicating more activity in the LOB. This is less visible in the visualization of the T-Bond market (Figure 2.1), which visualizes the number of seconds (since the start) of the plot. Depending on one's goals, variables can thus be visualized in various ways to study them from different perspectives and to gain more knowledge about LOB behavior.

Table 2.2 | Example of merging two LOBs for visualization purposes

$\mathbf{A_1}$	A_1	$\mathbf{A_2}$	$\mathbf{A_3}$	A_3
B_0	$\mathbf{B_1}$	B_1	B_1	$\mathbf{B_2}$

Note: Bold letters indicate the first time a message occurs.

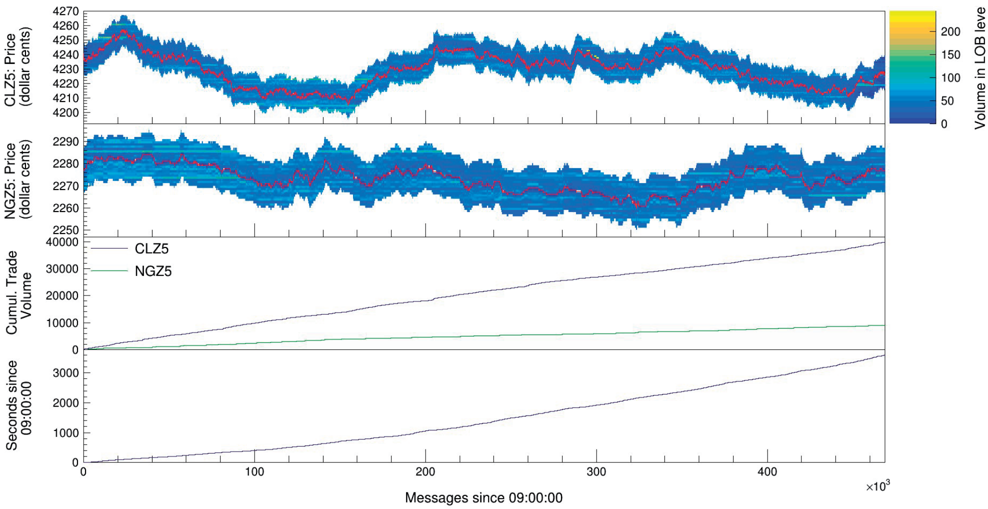


Figure 2.6 | December Crude Oil (CLZ5) and December Henry Hub Natural Gas (NGZ5) LOB behavior (November 12, 2015 from 09:00 AM to 10:00 AM CT). The top two panels show the volumes at the 10 price levels on the bid and ask side of the LOB, respectively, for the December crude oil futures contract (CLZ5) and the December natural gas futures contract (NGZ5), respectively. Each unit on the x-axis is one message. The y-axes represent the respective prices of crude oil and natural gas in dollar cents. The color of each bin represents the volume in the LOB at that message and price. The scale ranges from blue to yellow, with the color becoming a brighter yellow as volume increases at that price level. The red line is the midpoint. The third panel shows the cumulative trading volume for the selected time horizon for both markets. The blue line indicates the cumulative trade volume for crude oil and the green line for natural gas. A steeper (flatter) line signals a higher (lower) rate of traded volume. The bottom panel shows how much time passes between messages reported by the exchange. A steeper (flatter) line signals a lower (higher) rate of messages, given that a steeper (flatter) line signals more (less) time progression. (data source: author's visualization of CME MDP 3.0 Market Data).

2.5.3 Two Markets in One Visualization

The visualization can also be modified to show two (or more) LOBs simultaneously. However, careful attention must to be paid to how the LOBs are synchronized, as message numbers are not necessarily aligned between two markets. For example, a market with little activity may have 100 messages until 10:00 AM, whereas an active market may have 1000 messages in the same time window. To solve this, all messages are merged and ordered into a single time series. If more messages arrive in Market A, the LOB for Market B is repeated and not updated until a new message arrives for Market B. In other words, the last known LOB is retained until something changes (i.e., a new message arrives). Consider the following example of a merged list between Market A and Market B: [A_1, B_1, A_2, A_3, B_2]. The LOBs would be visualized as per Table 2.2.

Figure 2.6 illustrates the LOBs for the December crude oil (CLZ5) and December natural gas (NGZ5) futures contracts on November 12 between 09:00 AM and 10:00 AM CT. The top panel shows the LOB movements in both markets. More trading occurs in the crude oil futures market, with approximately 40,000 futures contracts being traded compared to less than 10,000 in the natural gas market. Furthermore, messages arrive at a regular pace, as shown by the steadily rising timeline in the bottom panel. Visualizing two LOBs within one graph can be beneficial when studying relationships between markets, the effect of an event on either market or the spillover effects from one market into another (e.g., regarding trades, volatility or liquidity).

2.6 PARTICLE PHYSICS ILLUSTRATION OF THE LOB: ILLUSTRATION FOR LIQUIDITY

The proposed visualization methodology can be adapted to many research topics. It helps answering research questions concerning, for example, market microstructure and optimal trading frequency (Du & Zhu, 2017). In this paper, it will be illustrated by way of liquidity, an important factor for the functioning of financial markets and one of the major topics in financial research (Capponi et al., 2019; Kerr et al., 2019; S. Li et al., 2019; Menkveld & Zoican, 2017; Peress & Schmidt, 2020; Trebbi & Xiao, 2019). Liquidity is illustrated since it is a concept with multiple dimensions, which makes it relatively complex to visualize in a single visualization. Liquidity consists of four dimensions: immediacy, tightness, depth and resiliency (Hasbrouck, 2017; Kyle, 1985). Most LOB visualizations in the literature show (some form of) liquidity, i.e., the volume in the market at different price levels ("depth") and the bid-ask spread ("tightness").

Table 2.3 | Comparison between various LOB visualizations

Authors	Aidov and Daigler (2015)		Paddrik et al. (2016)		Proposed visualization
Figure # in article	1	2	7	8	6 and 7
Message/time	Time	Time	Time	Time	Message
Time interval	5 minutes	Daily	100ms	ms	ms
Variables	Time, Depth, LOB level	Day, Bid and ask percentage quote, Bid and ask depth, Day and Night	Time, Price, Volume (per level per side), Bid and Ask	Time, Price stop loss and limit order book, Depth stop loss and limit order book, Bid and Ask side, Best bid, Best ask, (Cumulative) Volume, Most new order†, Most canceled†, Most modified†, Most aggressive†, Most passive†	Message number, Price, Volume (per level per side), Midpoint, Cumulative trade volume, Bid volume, Ask volume, Time
Volume	Bid and ask aggregated per level	Mean depth per day	Per level per side	Per level per side and cumulative volume	Per level per side, cumulative volume and total volume per side
Price	No	Relative	Absolute	Absolute and best bid and best ask	Absolute
Trades	No	No	No	No	Yes
Bid-ask spread	No	Relative	Can be deduced	Yes	Can be deduced or added
Midpoint	No	Relative	Can be deduced	Can be deduced	Yes

† Includes participant identification number and number of contracts.

Aidov and Daigler (2015) visualize the LOB in two ways. Figure 1 in their paper (column 2 in Table 2.3) visualizes the cumulative depth for each level across the book over time, i.e., the bid and ask volumes aggregated per level. It gives a better insight into how total volume behaves at each level and how it is distributed across time. However, the bid and ask side can have asymmetrical depths. By visualizing their cumulative depth, information might get lost, as it becomes impossible to disentangle the individual behaviors of the bid and ask side. In addition, no price levels are included. This makes it challenging to examine certain price-related volume patterns in the LOB, such as shifts in depth related to trades or a volatile market. Figure 2 in their paper (column 3 in Table 2.3) visualizes the shape of the LOB on one day for five levels on the bid and ask side. The visualization shows a quick and easy-to-understand snapshot of the LOB. However, averages of averages are taken (averaging depth across 5-minute intervals which are then averaged over the day), and it shows a "snapshot" of one day rather than the distribution over time. Hence, a lot of LOB information and behavior is lost. Paddrik et al. (2016) propose many visualizations for regulators to use, two of which are specific to the LOB. Figure 7 in their paper (column 4 in Table 2.3) visualizes the LOB with a heatmap of the depth at various price levels, i.e., different colors are used to indicate volume on the bid and ask side. It presents an instant picture of LOB behavior over time and the colors offer a quick overview of distributed volume. Figure 8 in their paper (column 5 in Table 2.3) represents a simultaneously animated visualization of the LOB and stop loss order book that includes many textual variables. Since trading takes place in nanoseconds, the animated information might change fast, making it less suitable for real-time animations and quick decision-making by market participants and regulators.

All visualizations would benefit from the inclusion of trades, as trades contain information, take liquidity from the LOB and can explain certain behavior. Informed traders, for example, prefer larger trades (Easley & O'Hara, 1987). This helps us understand LOB movements; for example, the LOB might respond to large trades, which would remain unobserved in any of the figures included in Table 2.3. Furthermore, all of these visualizations use time (i.e., snapshots), the disadvantages of which were discussed in section 2.2, rather than messages. While Figure 7 from Paddrik et al. (2016) shows the LOB in a heatmap format, in this paper, it is extended by the ability to render LOB data with user specifications, such that multiple variables can be added and visualized.

The methodology proposed in this paper complements the existing liquidity visualization tools. It visualizes all dimensions of liquidity and can easily be modified to display different and/or more variables. To demonstrate the relevance and power of the visualization, a liquid market – the Chicago SRW Wheat futures contract – is compared to a less liquid market, – the Rough Rice futures contract – on October 13, 2015, 9:00 AM – 12:00 AM CT. Figure 2.7 shows the Chicago SRW Wheat futures market, which is relatively liquid compared to the rough rice futures market, displayed in Figure 2.8. The visualization as illustrated in section 2.5 is modified to include more liquidity variables: total volume on the bid (yellow line) and ask (pink line) side. Other liquidity variables, such as the bid-ask spread or volume per level, can also be visualized.

With more than 180,000 messages compared to 4500 in the rough rice market, the wheat market shows more activity within the same time frame. The LOB moves smoothly and is compact in the wheat market, i.e., all orders in the LOB rest in adjacent price levels. In contrast, the rough rice market in Figure 2.8 shows frequent gaps – indicated by white spaces – where there is no volume for that specific price level. The wheat market displays a steady and smooth increase of cumulative trade volume and time progression, contrary to the rough rice market, where cumulative trade volume and time are subject to large spikes. This

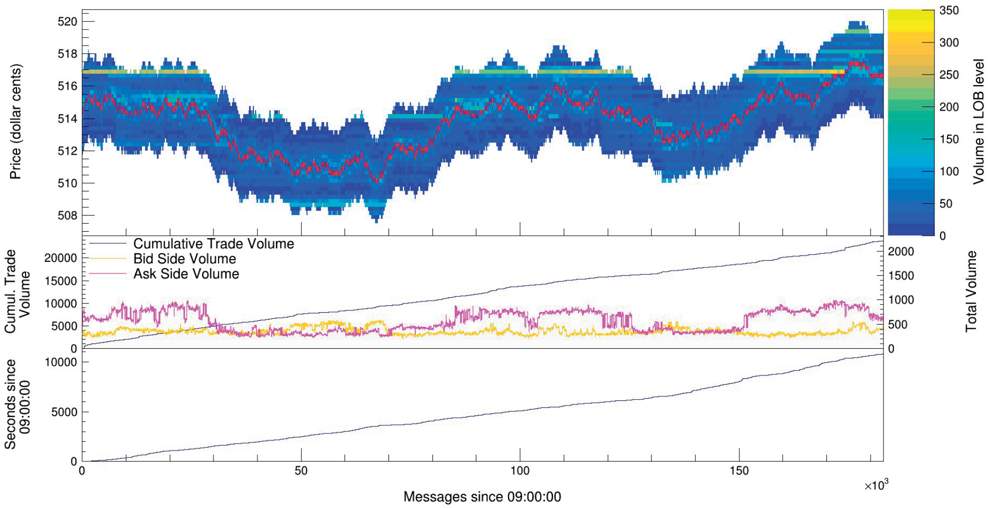


Figure 2.7 | December Chicago SRW Wheat (ZWZ5) LOB behavior (October 13, 2015 from 09:00 AM to 12:00 AM CT). The top panel shows the volume at the 10 price levels on the bid and ask side of the LOB, respectively. Each unit on the x-axis is one message. The y-axis represents the price of SRW wheat in dollar cents. The color of each bin represents the volume in the LOB at that message and price. The scale ranges from blue to yellow, with the color becoming a brighter yellow as volume increases at that price level. The red line is the midpoint. The middle panel shows the cumulative trading volume for the selected time horizon and the total volume on the separate sides of the LOB. The blue line indicates the cumulative trade volume on the left y-axis. A steeper (flatter) line signals a higher (lower) rate of traded volume. The pink line represents the total volume on the ask side and the yellow line the total volume on the bid side on the right y-axis. The bottom panel shows how much time passes between messages reported by the exchange. A steeper (flatter) line signals a lower (higher) rate of messages, given that a steeper (flatter) line signals more (less) time progression. (data source: author's visualization of CME MDP 3.0 Market Data).

means that messages arrive more irregularly in the rough rice market – and more regularly in the wheat market. As indicated by the flat line, the time between messages is shorter towards the end of the time window in the rough rice market. This indicates a higher frequency of messages and, thus, more activity in the LOB than before. This illustrates once again the advantage of using messages over snapshots: to create snapshots, messages and information are collapsed into regular time intervals, which might have led to different conclusions for the rough rice market. Therefore, the LOB behavior as revealed in Figure 2.8 might not have been observable by way of snapshots.

The proposed visualization methodology complements the other visualization techniques discussed in Table 2.3. All available information is embedded thanks to the use of messages rather than snapshots. It gives a quick overview of how volume is distributed, how active the market is, whether the bid-ask spread is tight or wide, what the midpoint price is and how trades affect each side of the market. Additional (liquidity) variables can be added to gain more insights.

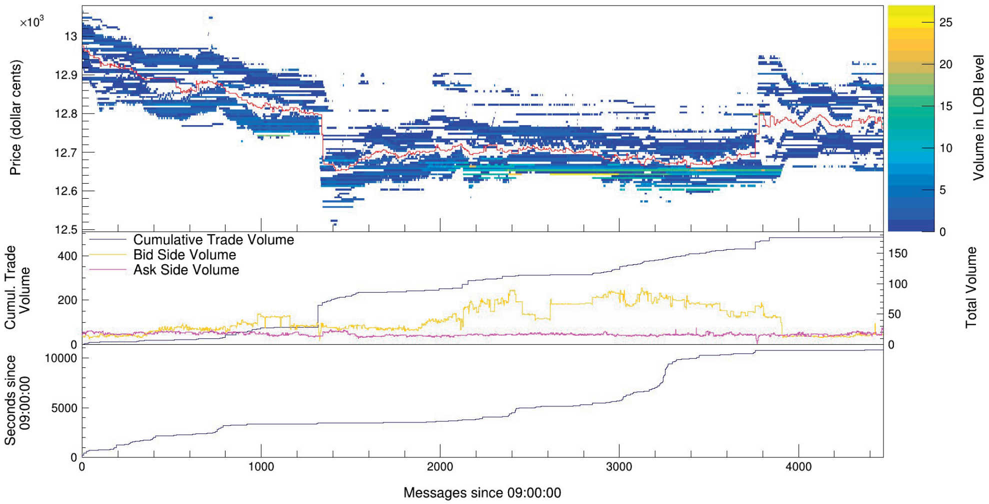


Figure 2.8 | December Rough Rice (ZRX5) LOB behavior (October 13, 2015 from 09:00 AM to 12:00 AM CT). The top panel shows the volume at the 10 price levels on the bid and ask side of the LOB, respectively. Each unit on the x-axis is one message. The y-axis represents the price of rough rice in dollar cents. The color of each bin represents the volume in the LOB at that message and price. The scale ranges from blue to yellow, with the color becoming a brighter yellow as volume increases at that price level. The red line is the midpoint. The middle panel shows the cumulative trading volume for the selected time horizon and the total volume on the separate sides of the LOB. The blue line indicates the cumulative trade volume on the left y-axis. A steeper (flatter) line signals a higher (lower) rate of traded volume. The pink line represents the total volume on the ask side and the yellow line the total volume on the bid side on the right y-axis. The bottom panel shows how much time passes between messages reported by the exchange. A steeper (flatter) line signals a lower (higher) rate of messages, given that a steeper (flatter) line signals more (less) time progression. (data source: author's visualization of CME MDP 3.0 Market Data).

2.7 ADDITIONAL VISUALIZATION POSSIBILITIES WITH ROOT: AN EXAMPLE

To illustrate ROOT's additional visualization possibilities, a different and new visualization is shown that combines large amounts of data in a single figure. Figure 2.9 shows the development of trade prices before and after a transaction, for the December E-mini Dow Jones futures contract of 2015.

Figure 2.9 is a 2-dimensional histogram, displaying trade price behavior before and after a transaction. It is illustrated using two parameters, the trigger and the relative impact (though more parameters can be included). The trigger is the event under study, in this example transactions (other examples of triggers include cancellations in the LOB, liquidity measurements, USDA announcements, news articles or the opening of the market, etc.). Transactions (triggers) are set at time zero on the x-axis. The price of a transaction is located

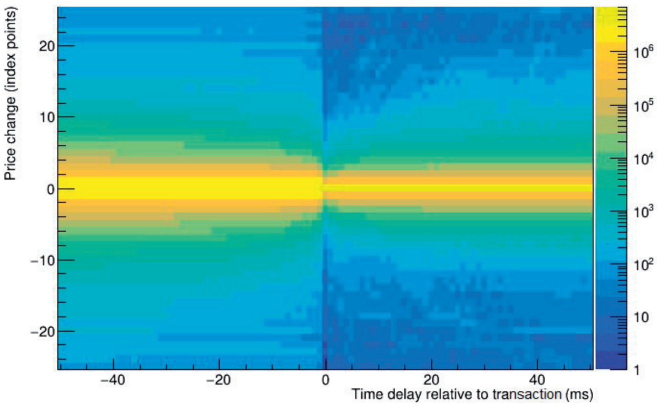


Figure 2.9 | Evolution of trade prices before and after a transaction for the December E-mini Dow Jones futures contract of 2015 (YMZ5). (data source: author's visualization of CME MDP 3.0 Market Data).

on the zero on the y-axis. All trades that take place in the specified time window are relative to the price level of the transaction (relative impact). For every transaction in the dataset, a time window is filtered of 50ms before and 50ms after the transaction. All trades that occur within this time window are counted for and put in their respective bins. Each bin represents the count frequency of timestamp-price combinations. ROOT counts the number of times that certain trade timestamp-price combinations occurs. The color of each bin represents the number count assigned to that bin. The logarithmic scale on the z-axis ranges from blue to yellow, with the color becoming a brighter yellow as the count increases.

To illustrate, 50ms after a transaction, approximately one million times a trade took place at 1 index point (E-mini Dow Jones) above the transaction. Figure 2.9 is a cumulative visualization, which allows for large datasets to be captured in a single plot: here it contains a total of 7,621,048 transactions. While many more inferences can be made from this figure, this is beyond the scope of this paper.

2.8 DISCUSSION AND CONCLUSION

The growing availability of LOB data poses new challenges to academics, industry, and regulators. LOB data is complex and can be difficult to reconstruct, visualize and analyze as it is irregularly spaced, high-speed and voluminous. This paper introduces a new method to visualize the LOB by looking at data through a particle-physics lens, which enables a better understanding of markets and the behavior in those markets, given that it takes all activities (messages) and preserves all of the information embedded in market participants' actions.

The proposed methodology to visualize the LOB complements the existing methods. While the existing methods are limited as to the number of variables they can display, the visualization proposed in this paper is richer and more easily adjustable with user specifications to many different types of research. It complements the previous literature in the following ways. First, all available messages are used and no snapshots are taken. This reveals the actual changes in the LOB, so that patterns in the LOB can be studied better. This is especially interesting when studying the behavior of HFTs, where the use of snapshots will make their behavior more opaque and harder to study, given that they trade within nanoseconds. Second, since the visualizations are based on messages, they are no longer limited by and dependent on time. Thus, given that the time variable is no longer the constraining factor, visualizations can be created that would otherwise have been impossible, and more relevant variables and their relationships can be shown. Third, the visualization shows the changing distribution of the LOB over time, including the trades that take place. This combination is important, as trades have an effect on the behavior of other traders and, thus, on the distribution of the LOB. Fourth, as illustrated with liquidity, the visualization is easily adjustable to include many variables. For example, it can display the informational content of the LOB, the behavior of HFTs in the market, changes in volatility, the moment(s) when price limits are hit or the cost of transactions over time. Moreover, variables can be visualized in various ways, as illustrated through trade volume and time. Fifth, two markets can be visualized at the same time to study their interconnectedness. For example, the LOB of wheat can be plotted along with that of corn, to study the effect of corn trades on the LOB of the wheat market. Moreover, it allows for cross-venue and cross-asset visualizations and all related markets can be visualized simultaneously: options, futures and spot markets for a specific product. Finally, the visualizations are created using ROOT, which has several advantages. It is multipurpose software that allows for more compact data storage and faster analyses. LOBs can be processed and analyzed relatively fast, which is beneficial for all financial stakeholders. Creating the visualizations is efficient and fast, and it allows users to interactively zoom, which will automatically update the plot. The technique is not limited to (agricultural) futures markets and can be used for any market with a LOB, such as equity, crypto and options markets.

The proposed methodology has several implications for financial market stakeholders. It offers an easy method to reconstruct the LOB and a useful tool for data exploration. As argued before, the visualization can be adjusted to study specific topics of interest. Academics can use the methodology to complement their research by adding visualizations with additional variables, such as the distribution of participants in the market, trade outliers or spoofing indicators. In addition, it is possible to visualize multiple markets in the same graph. For example, LOB volume differences at (similar) price levels can be used to study the varying LOB distributions in related markets. The synced LOBs facilitate future research in, among others, spillover effects, arbitrage trades and spread trading. This will help all parties involved to better understand markets. Regulators can use the methodology as a more advanced tool to monitor (related) markets and trace and study irregularities, such as market manipulation. It will help traders and controllers to monitor compliance. Furthermore, the methodology can help policymakers in regulating and designing the microstructure of markets. For example, the visualization can show the effect of larger and smaller price limits or adjustments on the tick size of the market. This is especially interesting when additional statistics are included in the visualization.

The proposed methodology carries several caveats. First, the visualization becomes hard to read when there are many messages. In this case, the program will skip the visualization of a predefined number of messages. For example, if there are 100 messages and the skip is set to ten messages, the 10^{th}, 20^{th}, 30^{th}, etc. message will be plotted. However, due to the density of the messages, this is not visible to the eye. Second, adding multiple variables may be counterproductive, in that they may cluttering the visualization and thus make it harder to read and understand. Finally, academics and industry participants are used to seeing time being plotted on the x-axis, rather than in a separate graph. Therefore, it may take some adjustment and learning to get used to the new way of visualizing the LOB, where messages replace time on the x-axis.

From a methodological point of view, new metrics can be developed that can be applied to LOB visualization, for example, to measure the distribution of the LOB underlying the visualization. Future studies can explore whether certain distributions of the LOB are different from the “average” or benchmark LOB distribution. Furthermore, the use of particle anomaly detection may be helpful in tracing anomalies in LOB markets. Studies of irregularities and malpractices, such as spoofing, can apply this visualization technique to identify these anomalies. Future research on application aspects might focus on using this visualization method with different topics of interest, such as liquidity, volatility, market participant distribution, HFT behavior, large trades, market resilience, transaction costs, LOB information content, endogenous and exogenous shocks and spillover effects between markets.

Chapter 3

Unraveling the JPMorgan spoofing case using particle physics visualization methods

Marjolein E. Verhulst, Philippe Debie, Joost M.E. Pennings,
Cornelis Gardebroek, Axel Naumann, Stephan Hageboeck,
Andres A. Trujillo-Barrera, Lorenzo Moneta and Paul van Leeuwen

Published as:

Debie, P., Gardebroek, C., Hageboeck, S., van Leeuwen, P., Moneta, L., Naumann, A., Pennings, J. M. E., Trujillo-Barrera, A. A., & Verhulst, M. E. (2023). Unravelling the JPMorgan spoofing case using particle physics visualization methods. *European Financial Management*, 29(1), 288-326.

ABSTRACT

On September 29, 2020, JPMorgan was ordered to pay a settlement of $920.2 million for spoofing the metals and Treasury futures markets from 2008 to 2016. We examine these cases using a visualization method developed in particle physics (CERN) and the messages that the exchange receives about market activity rather than time-based snapshots. This approach allows to examine multiple indicators related to market manipulation and complement existing research methods, thereby enhancing the identification and understanding of, as well as the motivation for, market manipulation. In the JPMorgan cases, we offer an alternative motivation for spoofing than moving the price.

Keywords: spoofing, limit order book, visualization, particle physics, high-frequency trading

3.1 INTRODUCTION

"A LITTLE RAZZLE DAZZLE TO JUKE THE ALGOS..." wrote a JPMorgan Treasury trader in a chat message in November 2012, after successfully tricking high-frequency traders and moving the market (Schoenberg & Robinson, 2020). Fast forward to the year 2020, and JPMorgan (JPM) had to pay a record-breaking settlement of $920.2 million for manipulating the precious metals and Treasury markets (CFTC, 2020e; Michaels, 2020). Specifically, JPM[12] admitted to spoofing the gold, silver, platinum, palladium, Treasury note and Treasury bond futures markets[13] between 2008 and 2016.

Spoofing has been illegal under the Dodd-Frank Act since 2010 and is defined as: *"bidding or offering with the intent to cancel the bid or offer before execution"* (United States, 2010). Spoofers manipulate the displayed order volume[14] (hereafter referred to as "order volume") in the limit order book (LOB) to persuade market participants to trade in the spoofer's desired direction (Dalko & Wang, 2018). The LOB shows the order volume at various price levels. However, it presents incomplete information to market participants. For example, market participants do not know what type of order is submitted, the actual volume of an iceberg order and whether a reduction in volume is due to a cancellation or an order execution (Dalko & Wang, 2018). Spoofers can take advantage of this market microstructure by introducing conditions that can influence the decisions of other traders (Mendonça & De Genaro, 2020).

One of the basic types of spoofing involves the spoofer wanting to buy at a lower price than the current price (Dalko & Wang, 2018): a relatively small genuine order (i.e., an order intended to be executed) is placed on the bid side and a relatively large spoof order (i.e., an order not intended to be executed) is placed on the opposite side – the ask side – of the LOB. Market participants then act on the newly created imbalance in the LOB and move the market in the direction of the genuine order's price, often by way of herd behavior (Dalko & Wang, 2018). Shortly after placing the spoof order, or once the genuine order has been executed, the large spoof order is cancelled, and the imbalance created is gone. The result is that the spoofer was able to buy at a lower price (CFTC, 2020e; Dalko & Wang, 2018). Other types of spoofing include, but are not limited to, layered spoofing, layered spoofing with collapsing and spoofing with vacuuming and flipping (Neurensic, 2016). Spoofing can be hard to identify as it may, for example, take place within a single market, between corre-

12 JPMorgan Chase & Company and its subsidiaries, JPMorgan Chase Bank and J.P. Morgan Securities LLC.
13 These futures contracts were/are traded on the Commodity Exchange, Inc. (COMEX), the New York Mercantile Exchange (NYMEX) and the Chicago Board of Trade (CBOT).
14 Contrary to hidden order volume, which can be the case with iceberg orders, an iceberg order is an order whereby only a fraction of the total order is displayed in the LOB and the rest is not visible to other market participants (Buti & Rindi, 2013).

lated markets (e.g., soybean futures and soybean oil futures), between different calendar contracts (e.g., the March and September contracts of E-mini S&P 500 futures), between derivatives (e.g., gold futures and gold options), between exchanges and by one party or by multiple parties. Moreover, spoofing concerns the trading intention to cancel before execution, and “intention” is difficult to capture in market data.

Spoofing is harmful to markets and their participants for numerous reasons. Spoofers intentionally distort the available information that traders use to make decisions. This makes non-spoofing market participants vulnerable as they are misguided by false buy and/or sell liquidity figures (Dalko & Wang, 2018). This negatively impacts the price-formation process and hence distorts the price (Dalko & Wang, 2020b; Mendonça & De Genaro, 2020). It also creates additional volatility in price, trading volume and order volume, which negatively impacts the stability of the market (Dalko & Wang, 2020b). Moreover, its effects can spill over into interconnected markets, making them inefficient too (Mendonça & De Genaro, 2020).

Over the course of eight years, JPM placed hundreds of thousands of spoof orders resulting in $172,034,790 in gains. Conversely, however, these orders harmed the market and its participants, causing $311,737,008 in market losses (CFTC, 2020e). Since this only represents identified spoofing by one firm, the real damage caused by spoofing across all markets is likely to be much greater, making this a serious problem for all stakeholders. The current supervisory systems are not adequate and effective enough to detect such illegal trading behavior, given that 1) JPM’s supervision system failed to detect manipulative practices, such as spoofing, until 2014 (CFTC, 2020e); and 2) it took the Commodity Futures Trading Commission (CFTC) three to eleven years after the spoofing occurred to file charges against JPM, and many of the spoofing instances were probably discovered thanks to secured documents and computer communication.

Using a visualization methodology developed in particle physics by the European Organization for Nuclear Research (CERN) (Antcheva et al., 2009; CERN, 2018b; Verhulst et al., 2021), we describe the LOB in a novel way, providing new insights into the JPM spoofing case. Specifically, we visualize all spoofing examples as documented in the CFTC report (CFTC, 2020e). It contributes to the literature as follows. First, we offer guidance on how to characterize spoofing by way of variables and how to effectively visualize these variables. Second, we offer an alternative motive for spoofing, namely attracting liquidity rather than changing the price. To the best of our knowledge, this has not been reported before. Third, we provide insight into how spoofing is conducted and how (in)visible it is to other market participants. Fourth, while previous LOB visualizations were solely time based (i.e., using snapshot intervals of, for example, five seconds), this study complements these visualizations with the original messages about traders’ market activity as sent to the exchange. Moreover, it illustrates how high-frequency LOB data can be effectively visualized. This novel way of visualizing high-frequency data can contribute to new insights in future research and inspire further analyses among stakeholders. Companies such as JPM, for example, can use the methodology to enhance and refine their surveillance programs and internal control systems, and regulators, such as the CFTC, can use it to enhance their understanding of manipulative trading practices.

3.2 LITERATURE REVIEW

Empirical literature on spoofing is scarce, particularly due to constraints in obtaining (LOB) data that matches the purpose of the research (Lee et al., 2013; Linton & Mahmoodzadeh, 2018; Putniņš, 2012). Several studies have tried to detect spoofing in markets by using order data (Lee et al., 2013; Zhai et al., 2017). This data differs from LOB data, in that order data comprises the submitted, cancelled, and modified orders of individual traders, whereas LOB data constitutes all these orders and shows the LOB visible to all market participants. For example, LOB data reveals the best bid and ask prices and total volumes belonging to specific price levels in the LOB (Mendonça & De Genaro, 2020). Although order data contains more information on individual orders (provided that it is not aggregated), studies attempting to detect spoofing with order data have omitted to reconstruct the LOB, let alone visualize it.

LOB data is nevertheless needed to understand the current state of the market, which influences trading and spoofing decisions. It helps to identify higher-level patterns or parameters related to spoofing, e.g., imbalances between the bid and ask volumes (Cartea et al., 2020). To the best of our knowledge, there are only a handful of researchers who studied spoofing using LOB data. Mendonça and De Genaro (2020) generated one-minute LOB snapshots from order data and used both datasets to detect spoofing on the Brazilian Stock Exchange. Leangarun et al. (2016) tried to detect, among others, spoofing in three NASDAQ stock markets by training neural networks and using one-minute LOB intervals.

However, these papers, as well as other LOB-related papers (e.g., Biais, Bisière & Spatt, 2010; and Menkveld & Yueshen, 2019), lack visualizations of the LOB. Visualizations help academics, industry participants and regulators to better understand the market; they allow them, among other things, to identify and understand anomalies such as spoofing (Verhulst et al., 2021). LOB visualization literature is thus scarce (Aidov & Daigler, 2015; Paddrik et al., 2016). In addition, visualizations that do exist are often time-based and thus have limitations: since orders arrive irregularly, order data and LOB data are irregularly spaced over time. To achieve regular time intervals, time-based visualizations and time-series analyses use snapshots of the LOB. As a result, information is lost since the information is being aggregated. In addition, the literature provides no uniform method to achieve optimal snapshot size (Verhulst et al., 2021). Snapshot sizes that have been used in LOB analyses so far are: five minutes (Chordia et al., 2019; Kahraman & Tookes, 2017), one minute (Hautsch & Horvath, 2019; Mendonça & De Genaro, 2020; Yao & Ye, 2018), ten seconds (Cont et al., 2014), five seconds (Brogaard & Garriott, 2019), three seconds (Ito & Yamada, 2018) and one second (Battalio et al., 2016; Brogaard et al., 2019; Colliard & Hoffmann, 2017; Dugast, 2018). This lack of uniformity can be explained by the ever-increasing velocity (size) of data. At the start of the 21st century, one day of message data was comparable in size to 30 years of daily data (Dacorogna et al., 2001). Ten years later, data velocity had increased tenfold (Fabozzi et al., 2011). With today's high-frequency orders, one-second intervals can contain thousands of orders and action/reaction cycles of algorithms, hence increasing the need for a high resolution. Past research has identified benefits of high-frequency trading (HFT) for market participants. Brogaard (2010) shows that HFT adds substantially to the price discovery process and Brogaard et al., (2014) find that HFT facilitates price efficiency by trading in the direction of permanent price changes and in the opposite direction of transitory pricing errors. Hasbrouck (2018) examines high-frequency quoting and finds, among others, a positive relation between competition and quote volatility. He indicates that his analysis is directed at a broad classification of quote volatility and does not rule out occurrences of quote stuffing or spoofing (Hasbrouck, 2018). Here, we exclusively focus on spoofing as an example of HFT, and as such, this paper may contribute to literature that examines the role and impact of HFT on financial markets. Moreover, it complements past literature on LOBs and existing visualizations by applying visualization methodologies from particle physics to message-based LOB data.

3.3 DATA AND METHODOLOGY

Data consists of the Chicago Mercantile Exchange (CME) Group's proprietary market-depth dataset for all spoofing examples reported by the CFTC (CFTC, 2020e). The files are in the CME Market Depth 3.0 format, which provides messages about market activity.[15] These messages can be used to recreate the LOB with millisecond precision. The open-source ROOT software framework, developed by CERN, among others, to analyze the massive data generated in the Large Hadron Collider, is used to reconstruct and visualize the LOB (Brun & Rademakers, 1997; CERN, 2018b; Verhulst et al., 2021). ROOT is used in particle physics to save, access and mine data, among other applications, as well as to generate visualizations (CERN,

15 The data contains messages on LOB level changes (i.e., a new price level is inserted or deleted, or the volume is changed at a level). If two traders add volume at the same price level in the same millisecond, for example, there will be a single message about the aggregated volume addition, rather than two separate messages (i.e., one for each trader).

2018a). Large amounts of data can be stored and processed efficiently in a distributed setup (Tejedor & Kothuri, 2018).

The CFTC (2020e) reported nine specific examples of spoofing and manipulation by JPM, including the associated markets, dates, timestamps (Central Time), volume orders and prices. We discuss the nine examples according to their spoofing strategies: 1) "traditional" spoofing, i.e., there is a displayed genuine order and a single spoof order; 2) spoofing with iceberg orders, i.e., the genuine order is an iceberg order with displayed and hidden volumes and a single spoof order; 3) layered spoofing, i.e., there is a displayed genuine order and multiple spoof orders at various price levels; and 4) layered spoofing with iceberg orders, i.e., the genuine order is an iceberg order with displayed and hidden volumes and there are multiple spoof orders at various price levels. The results section discusses only one example per spoofing category, and meaningful differences will be noted. Figures and tables for all spoofing examples not discussed in this paper are available in Appendix 3.A.

Time windows in which the spoofing examples took place are visualized using ROOT's graphing facilities (Brun & Rademakers, 1997; CERN, 2018b). For readability, only the top ten bid and ask levels are visualized from the consolidated limit order book.[16] First, we will show the LOB for a single spoofing example using two snapshot sizes employed in previous literature: a five-second snapshot (Brogaard & Garriott, 2019) and a one-second snapshot (Battalio et al., 2016; Brogaard et al., 2019; Colliard & Hoffmann, 2017; Dugast, 2018). Subsequently, we visualize the LOB message by message, rather than time-based visualizations which are customary in the existing literature. We use messages as they have (almost) the same data granularity as trading algorithms, and we demonstrate that these visualizations show what is actually happening in the market. Second, we highlight one spoofing example for each spoofing category by enriching the visualizations with variables that may further characterize spoofing behavior. We provide a unique visualization of the LOB in the relevant time window, showing: 1) the prices and volumes of all LOB levels; 2) midpoint prices; 3) the number of messages received by the exchange; 4) cumulative trade volume and individual trades including their respective prices; 5) volumes of the first bid and ask levels; 6) cancelled volume on the first bid and ask levels; and 7) bid and ask side liquidity.

Liquidity is measured by the Adverse Price Movement (APM) of the Exchange (or Xetra) Liquidity Measure (Gomber et al., 2015; Gomber & Schweickert, 2002; Sensoy, 2019). APM bid (APM ask) represents the execution costs in basis points (bps) of a trader who immediately wants to sell (buy) a dollar value and takes liquidity from the bid (ask) side by submitting

16 Several LOBs from the spoofing examples contain more than ten levels because of the implied LOB, but these are not visualized as they are generally further away from the top ten bid and ask levels. Visualizing all levels would make the visualizations unreadable in a paper version.

market orders. A lower APM indicates that the cost of trading is low and, therefore, liquidity is high (Gomber & Schweickert, 2002). For each message, the total LOB dollar value is calculated by multiplying the LOB prices with their respective volumes. The mean dollar value is calculated for the respective month in which the spoofing example took place and is used for the APM calculation. To test whether significant changes in liquidity occur before, during and after spoofing, the data is split into three parts for each spoofing example: "before" represents the time up until the spoof order was added; "during" the period from when the spoof order was added until it was cancelled; and "after" the time following the cancellation of the spoof order. Five different time windows are used: 1) the same time window as the duration of the spoof (i.e., identical to the "during" part); 2) ten seconds; 3) thirty seconds; 4) one minute; and 5) five minutes. Normality is assumed under the central limit theorem. Levene's test indicated variances are not equal, resulting in the use of Welch's *t*-tests to measure if liquidity was significantly different before, during and after the spoof for all five time windows. APMs for the *t*-tests are calculated per 10-millisecond snapshot.

3.4 RESULTS

First, a single JPM spoofing case is used to showcase the benefits of using message-based visualizations rather than time-based visualizations. Subsequently, one JPM spoofing case is discussed per spoofing category. The spoofing actions as identified by the CFTC (CFTC, 2020e) are described in detail, followed by LOB visualizations of these actions in subsections to facilitate the reading and interpretation of the figures. We examine four dimensions – trades, volume, cancellations, and liquidity – to show how they behave during a spoof.

3.4.1 Snapshots vs. Message-Based Visualizations

The spoofing of JPM in the September 2015 Ultra T-Bond is used to illustrate the benefits of message-based visualizations. In summary, this spoofing involved one iceberg genuine order with one contract displayed and 199 contracts hidden on the bid side, and a single ask spoof order of 100 contracts. More details of this particular spoof are discussed in section 3.4.3. The contract is visualized during the spoofing time window on June 30, 2015 between 08:45:40 and 08:46:10.[17] Rather than using the quoted five-decimal prices, prices in the visualizations are rounded to two decimals for readability. Figure 3.1 shows the behavior of the Ultra T-Bond LOB using five-second snapshots. The top panel shows the ten ask (bid) levels above (below) the midpoint price, as indicated by the red horizontal line. The colors show the volumes at each price level. The various spoofing actions as identified by the CFTC

17 Interactive data visualizations can be included for each figure to let readers interact and engage with our research. Code can also be made available to readers.

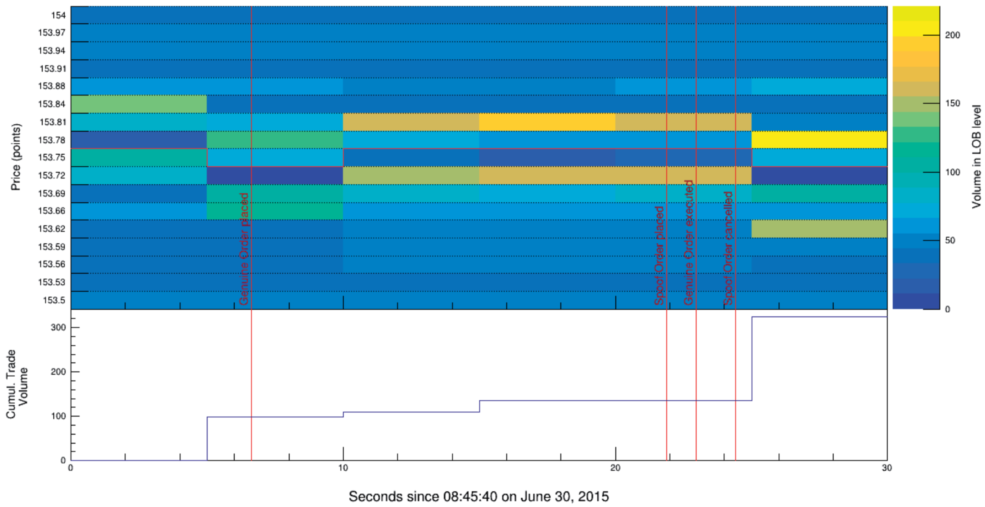


Figure 3.1 | Ultra T-Bond September 2015 LOB using 5-second snapshots. This figure visualizes the Ultra T-Bond September 2015 LOB using 5-second snapshots. The *top* panel shows the volumes at the individual bid and ask levels between prices of 153.5 and 154 points. Each unit on the x-axis is one second. The y-axis represents the price of the Ultra T-Bond in points. The color represents the volume at each price level of the LOB for each five-second snapshot. The scale ranges from blue to yellow, with the color becoming a brighter yellow as volume increases at that price level. The red line is the midpoint. The *bottom* panel shows the cumulative trade volume per second. A steeper (flatter) line signals a higher (lower) rate of traded volume. The red vertical lines signal when the JPM spoofing activities took place, from left to right: when the genuine iceberg order was placed, when the spoof order of 100 contracts was placed, when the first contract of the genuine order was executed and when the spoof order was cancelled.

(CFTC, 2020e) are marked by vertical red lines. The bottom panel visualizes cumulative trade volume. Figure 3.1 demonstrates that the spoofing remains invisible when using high-frequency data and visualizing it using five-second snapshots. The volume remains relatively constant at the individual bid and ask levels, and the midpoint price is also relatively constant. The placing and cancelling of the spoof order happened within the same snapshot interval, leaving the addition and subtraction of 100 contracts invisible. Cumulative trade volume increases in a staircase pattern at the end of every five-second snapshot. The only visible spoofing-related action in Figure 3.1 is the significant increase in trading volume 25 seconds into the time window, i.e., the 51 contracts from the genuine order that were executed. However, one would not know that this was spoofing from merely looking at this figure.

Figure 3.2 is identical to Figure 3.1 but uses one-second snapshots instead of five-second snapshots. Contrary to Figure 3.1, the spoofing activities are visible in Figure 3.2. The addition and cancellation of the spoof order are now visible as a yellow bar at the first ask level, whereas they were not in Figure 3.1. Furthermore, trading volume spikes when the genuine order is executed, and increases more gradually. However, this visualization cannot convey

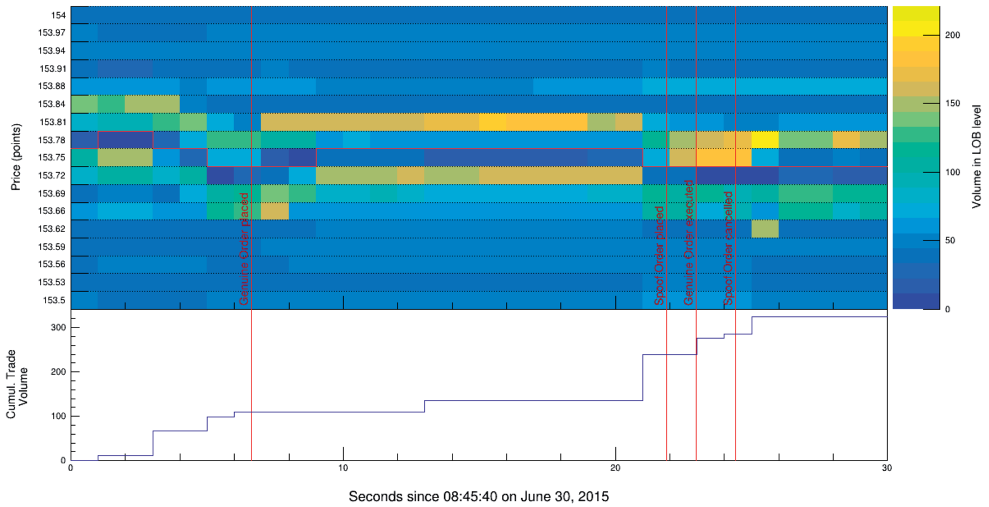


Figure 3.2 | Ultra T-Bond September 2015 LOB using 1-second snapshots. This figure visualizes the Ultra T-Bond LOB September 2015 using 1-second snapshots. The *top* panel shows the volumes at the individual bid and ask levels between prices of 153.5 and 154 points. Each unit on the x-axis is one second. The y-axis represents the price of the Ultra T-Bond in points. The color represents the volume at each price level of the LOB for each one-second snapshot. The scale ranges from blue to yellow, with the color becoming a brighter yellow as volume increases at that price level. The red line is the midpoint. The *bottom* panel shows the cumulative trade volume per second. A steeper (flatter) line signals a higher (lower) rate of traded volume. The red vertical lines signal when the JPM spoofing activities took place, from left to right: when the genuine iceberg order was placed, when the spoof order of 100 contracts was placed, when the first contract of the genuine order was executed and when the spoof order was cancelled.

the exact timing of the spoofing activities. For example, the spoof order was cancelled at 08:46:04.418, but the visualization's one-second resolution shows it as having been cancelled "some time between 08:46:04 and 08:46:05". Due to this lower resolution, the spoofing order appears to have been active for a longer time period than it actually was, as shown by the yellow bar after the vertical red line that reads "Spoof Order cancelled".

Visualizing the LOB using one-millisecond snapshots would solve the problem of the delay between trading action and visualization, as the granularity of the visualization equals that of the timestamps in the raw data (i.e., one millisecond). However, the exchange frequently receives multiple messages, i.e., changes to the LOB, within the same millisecond. Hence, information may be lost, as changes within the same millisecond are aggregated and not individually visible. Therefore, Figure 3.3 visualizes the LOB using messages instead of time-based snapshots. An additional panel is added to the bottom of Figure 3.3 to indicate how time passes between messages (blue line) and when one second has passed (green line). A steeper (flatter) blue line signals a lower (higher) rate of messages, given that a steeper (flatter) line signals more (less) time progression. The LOB volume shows more information

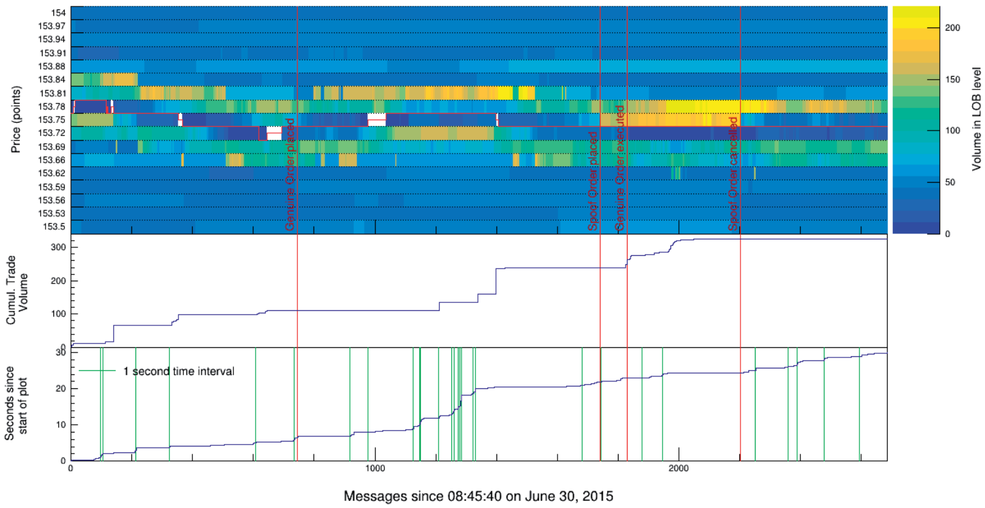


Figure 3.3 | Ultra T-Bond September 2015 LOB using messages. This figure visualizes the Ultra T-Bond September 2015 LOB using messages received by the exchange. The *top* panel shows the volumes at the individual bid and ask levels between prices of 153.5 and 154 points. Each unit on the x-axis is one message. The y-axis represents the price of the Ultra T-Bond in points. The color represents the volume at each price level of the LOB for each message. The scale ranges from blue to yellow, with the color becoming a brighter yellow as volume increases at that price level. The red line is the midpoint. The *middle* panel shows the cumulative trade volume per message. A steeper (flatter) line signals a higher (lower) rate of traded volume. The *bottom* panel shows how much time passes between messages reported by the exchange. A steeper (flatter) blue line signals a lower (higher) rate of messages, given that a steeper (flatter) line signals more (less) time progression. The green vertical lines indicate when one second has passed. The red vertical lines signal when the JPM spoofing activities took place, from left to right: when the genuine iceberg order was placed, when the spoof order of 100 contracts was placed, when the first contract of the genuine order was executed and when the spoof order was cancelled.

on (small) volume changes than the figures before. The addition and subtraction of small volumes might be an indication of algorithms "probing" for other algorithms and hidden liquidity (Bongiovanni et al., 2006; Chakrabarty & Shaw, 2008). These 'probes' become visible when using (almost) the same data granularity as trading algorithms, i.e., using messages rather than millisecond snapshots. Visualizations based on messages show what is actually happening in the market. In addition, they make the effect of executing an iceberg order more visible. The JPM spoofing in the Ultra T-Bond market consisted of an genuine iceberg order, and this becomes visible in the cumulative trade volume panel, once the first contract of the genuine order is executed. Many trades take place within the same millisecond, which would be aggregated (into one trade) in a snapshot-based visualization. However, Figure 3.3 shows that trade volume accumulated slower in this event, as the iceberg order was executed one contract at a time. This information was not visible in the previous visualizations and can help to understand spoofing behavior.

3.4.2 Traditional Spoofing

Two futures contracts are part of the "traditional spoofing" category: the March 2010 and December 2011 contracts from the 10-Year T-Note market. This section only discusses and presents results for the December 2011 contract, as both contracts show similar results. Table 3.1 shows the spoofing actions of the December 2011 contract, which took a total of 3.749 seconds. The spoof consisted of the placement of one genuine order on the first level of the bid side and a single large spoof order on the first ask level.

Table 3.2 shows the state of the LOB one millisecond prior to the first spoofing action, providing insight into what would have happened if JPM had placed the genuine order as a market order rather than a limit order. The spoofing involved buying 50 contracts at 129.578125 points, with a total underlying value of $6,478,906.25 (one point equaling $1000). Had the same number of contracts been bought through a market order, JPM would have bought at 129.594 points, representing a total underlying value of $6,479,700. Excluding trading

Table 3.1 | Spoofing actions on September 27, 2011 in the 10-Year T-Note December 2011 futures market

Time	Order type	LOB side	Action	Price (points)	Volume
14:03:54.205	Genuine order	Bid	Add	129.578125	50
14:03:57.636	Spoof order	Ask	Add	129.59375	3000
14:03:57.671	*Complete genuine order executed*				
14:03:57.954	Spoof order	Ask	Cancel	129.59375	3000

Note: This table presents the various spoofing actions JPM took on September 27, 2011 in the 10-Year T-Note December 2011 futures market. Per spoof action, the table reports the timestamp (*Time*), whether it concerned a genuine or spoof order (*Order type*), the LOB side the spoof action occurred on (*LOB side*), whether the order from the spoof action was added or cancelled (*Action*), the price level affected by the spoof action (*Price (points)*) and the volume related to the spoof action (*Volume*).

Table 3.2 | LOB state one millisecond before placement of the genuine order from the 10-Year T-Note December 2011 spoof

Bid volume	Bid price (points)	Level	Ask price (points)	Ask volume
431	129.578	1	129.594	640
1889	129.562	2	129.609	1415
1742	129.547	3	129.625	1593
1720	129.531	4	129.641	1201
1648	129.516	5	129.641	1201
1893	129.5	6	129.641	1201
1041	129.484	7	129.641	1201
979	129.469	8	129.703	953
592	129.453	9	129.719	699
1081	129.438	10	129.734	658

Note: This table reports the state of the LOB one millisecond before the genuine order from the 10-Year T-Note December 2011 spoof was added. It shows the prices and volumes of each level on the bid and ask side.

costs, JPM thus succeeded in buying the contracts $793.75 cheaper through spoofing than without spoofing.

Notably, in both traditional spoofing cases, the genuine order was placed on the first level of the bid side. Hence, these spoofing actions might not have been used to move the price – as otherwise the genuine order would have been placed on a deeper level of the LOB[18] – but to attract more liquidity to the market in order to sell at the price of the first bid level. At the time the genuine order was placed, the first bid level already comprised 431 contracts. Hence, due to the price-time-priority rule, 431 contracts had to be sold at 129.578 points first, before the 50 contracts of the genuine order could be sold. This hypothesis – the motivation of this spoof being to attract liquidity – is further examined in section 3.4.2.4 and section 3.4.6.

3.4.2.1 Traditional Spoofing: Visualization of the LOB and Trades around Spoofing

Figure 3.4 shows the behavior of the LOB and trades around the spoofing of the December 2011 contract between 13:03:45 and 13:04:05. The top panel shows the last traded price (blue line) and the occurrence of individual trades (gray lines). The second and third panel visualize the LOB and cumulative trade volume, respectively. The bottom panel shows the number of messages reported by the exchange in the relevant time window and, hence, the amount of time that passes between messages.

The second panel in Figure 3.4 shows that, when the genuine order was added, individual LOB levels contained volumes of between 500 to 2500 contracts.[19] When the spoof order of 3000 contracts was placed, volume on the first ask level increased significantly, as indicated by the bright yellow color. This increase in volume remained in the LOB during the execution of the genuine order and ended when the spoof order was cancelled. The addition of the spoof order, the execution of the genuine order and the cancellation of the spoof order all occurred within the same second, as indicated by the space between the green vertical lines.

The top panel in Figure 3.4 shows that when the genuine bid order was placed at 129.578 points, the last traded price was also 129.578 points. This illustrates once more that the goal of this spoof may not have been to move the price, but to attract more liquidity, so as to increase the chance of fully executing the genuine bid order of 50 contracts.[20]

18 In this event, spoofing would be used to move the price in the desired direction and push it through the first level(s) of the LOB to get a better price than the current best bid/ask.

19 Volume in The March 2010 LOB was considerably higher, as most levels contained volumes of between 1500 to 3500 contracts.

20 In contrast to the March 2010 contract, where the last traded price (118.281 points) was higher when the genuine order was added (118.266 points).

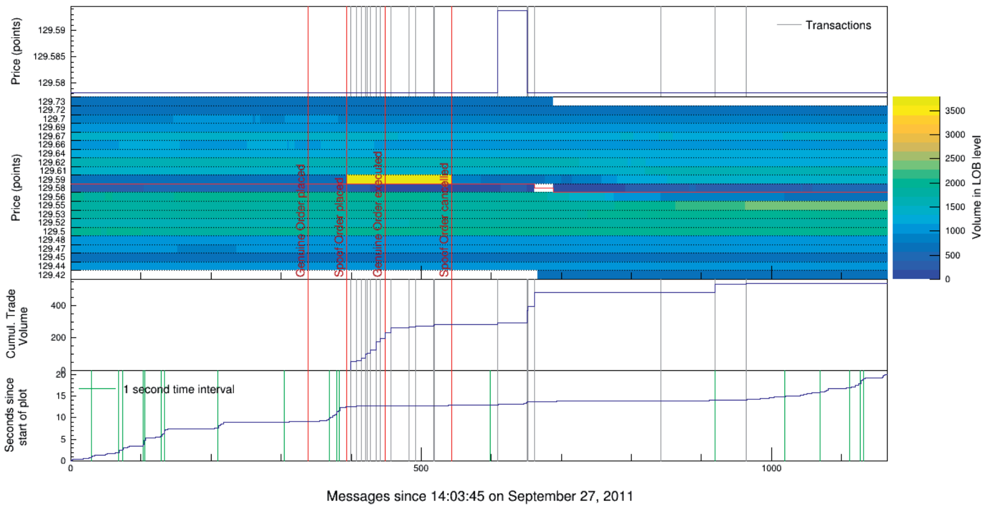


Figure 3.4 | 10-Year T-Note December 2011 LOB and trade behavior around the spoof of September 27, 2011. This figure visualizes the LOB and trade behavior around the spoof of September 27, 2011 in the 10-Year T-Note December 2011 futures market. The first panel shows the price of the last trade that took place (blue line) and when a trade took place (gray line). The second panel shows the volumes at the individual bid and ask levels between prices of 129.42 and 129.73 points. Each unit on the x-axis is one message. The y-axis represents the price of the 10-Year T-Note in points. The color represents the volume at each price level of the LOB for each message. The scale ranges from blue to yellow, with the color becoming a brighter yellow as volume increases at that price level. The red line is the midpoint. The third panel shows the cumulative trade volume per second. The fourth panel shows how much time passes between messages reported by the exchange. A steeper (flatter) blue line signals a lower (higher) rate of messages, given that a steeper (flatter) line signals more (less) time progression. The red vertical lines signal when the JPM spoofing activities took place, from left to right: when the genuine order was placed, when the spoof order of 3000 contracts was placed, when the genuine order was executed and when the spoof order was cancelled.

This will be further explored in section 3.4.2.4 and section 3.4.6. The last traded price remained constant at 129.578 points during all spoofing actions. The cumulative trade volume panel in Figure 3.3 shows that no trades took place in the time window until the genuine order and spoof order were placed.[21] After the spoof order was placed, a staircase pattern emerged. Our data shows that this was caused by the genuine bid order not being executed at once but being split into smaller executed trades. After the genuine order was fully executed, cumulative trade volume continued to increase – albeit at a lower volume – and remained constant (i.e., no trades occurred) right before and after the cancellation of the spoof order.

3.4.2.2 Traditional Spoofing: Visualization of Volume around Spoofing

Figure 3.5 visualizes the volume changes on the first bid and ask levels around the time of the spoof. When the genuine order was added, volume on the first bid and ask levels was

21 This does not mean no trades occurred in the market during that day, but that no trades occurred in the visualized time window until the spoof order was placed.

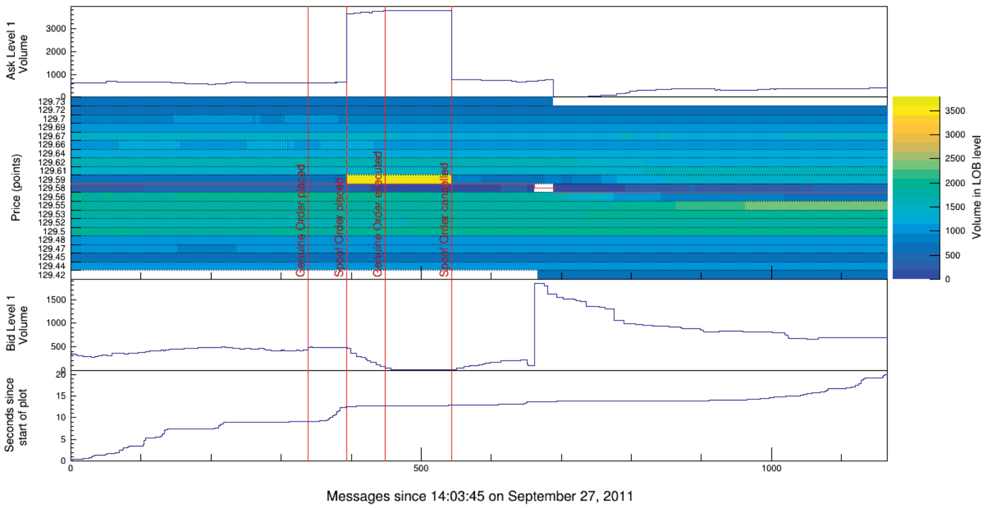


Figure 3.5 | 10-Year T-Note December 2011 first-level volume behavior around the spoof of September 27, 2011. This figure visualizes first-level bid and ask volume behavior around the spoof of September 27, 2011 in the 10-Year T-Note December 2011 futures market. The *first* panel shows the volume of the best ask level. The *second* panel shows the volumes at the individual bid and ask levels between prices of 129.42 and 129.73 points. Each unit on the x-axis is one message. The y-axis represents the price of the 10-Year T-Note in points. The color represents the volume at each price level of the LOB for each message. The scale ranges from blue to yellow, with the color becoming a brighter yellow as volume increases at that price level. The red line is the midpoint. The *third* panel shows the volume of the best bid level. The *fourth* panel shows how much time passes between messages reported by the exchange. A steeper (flatter) blue line signals a lower (higher) rate of messages, given that a steeper (flatter) line signals more (less) time progression. The red vertical lines signal when the JPM spoofing activities took place, from left to right: when the genuine order was placed, when the spoof order of 3000 contracts was placed, when the genuine order was executed and when the spoof order was cancelled.

stable at approximately 480 and 640 contracts, respectively. Volume increased significantly by 3000 contracts on the first ask level when the spoof order was added. Between the spoof order being added and the genuine order being executed, the volume on the first bid level decreased gradually. This is attributed to trades being executed and taking volume from the bid level, as shown in the third panel in Figure 3.4. After the genuine order was executed, volume on the first bid level decreased to two contracts. When the spoof order was cancelled, volume on the first ask level decreased significantly by 3000 contracts to 771 contracts, and volume on the first bid level gradually increased.

3.4.2.3 Traditional Spoofing: Visualization of Cancellations around Spoofing

Figure 3.6 visualizes the cancellations on the first ask (top panel) and bid levels (third panel) around the time of the spoof. Cancellations on the first ask level remained close to zero until the cancellation of the spoof order, only to increase significantly after the spoof order of 3000 contracts was removed from the LOB. Cancellations on the first bid level remained

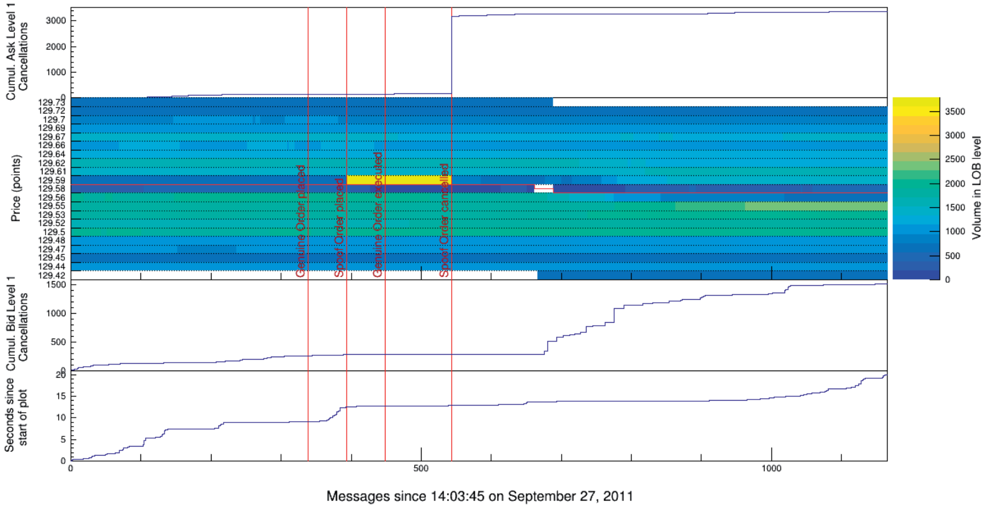


Figure 3.6 | 10-Year T-Note December 2011 first-level cancellation behavior around the spoof of September 27, 2011. This figure visualizes cumulative first-level bid and ask cancellation volume around the spoof of September 27, 2011 in the 10-Year T-Note December 2011 futures market. The *first* panel shows the cumulative volume of cancellations of the best ask level. The *second* panel shows the volumes at the individual bid and ask levels between prices of 129.42 and 129.73 points. Each unit on the x-axis is one message. The y-axis represents the price of the 10-Year T-Note in points. The color represents the volume at each price level of the LOB for each message. The scale ranges from blue to yellow, with the color becoming a brighter yellow as volume increases at that price level. The red line is the midpoint. The *third* panel shows the cumulative volume of cancellations of the best bid level. The *fourth* panel shows how much time passes between messages reported by the exchange. A steeper (flatter) blue line signals a lower (higher) rate of messages, given that a steeper (flatter) line signals more (less) time progression. The red vertical lines signal when the JPM spoofing activities took place, from left to right: when the genuine order was placed, when the spoof order of 3000 contracts was placed, when the genuine order was executed and when the spoof order was cancelled.

constant at a cumulative cancellation volume of around 300 contracts during all spoofing actions.

3.4.2.4 Traditional Spoofing: Visualization of Liquidity around Spoofing

The first and third panel in Figure 3.7 show the behavior of liquidity costs on the ask and bid side, respectively, around the December 2011 spoof. On the ask side, liquidity costs were relatively stable at around 4.7 bps up until the spoof order was placed. When the spoof order was added, liquidity costs drastically decreased to approximately 2.2 bps. After the cancellation of the spoof order, liquidity costs returned to approximately the same level as before the spoof order was placed. Bid-side liquidity costs remained relatively stable between 3.9 and 4.2 bps during all spoofing actions by JPM.[22]

22 The March 2010 contract shows more fluctuations in liquidity on the bid and ask sides than the December 2011 contract, as other volume not related to the JPM spoofing example was repeatedly shifted between the tenth bid and tenth ask level.

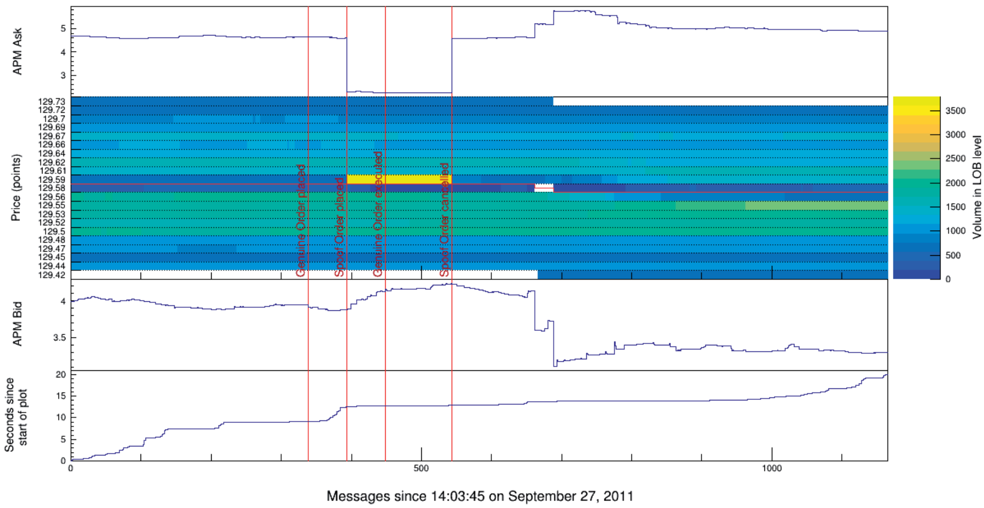


Figure 3.7 | 10-Year T-Note December 2011 bid and ask APM behavior around the spoof of September 27, 2011. This figure visualizes bid and ask liquidity costs (APM) behavior around the spoof of September 27, 2011 in the 10-Year T-Note December 2011 futures market. The first panel shows the APM of the ask side. APM measures the liquidity costs (in basis points) of a trader who wants to buy or sell a specific dollar value by submitting market orders. The second panel shows the volumes at the individual bid and ask levels between prices of 129.42 and 129.73 points. Each unit on the x-axis is one message. The y-axis represents the price of the 10-Year T-Note in points. The color represents the volume in each price level of the LOB for each message. The scale ranges from blue to yellow, with the color becoming a brighter yellow as volume increases at that price level. The red line is the midpoint. The third panel shows the APM for the bid side. The fourth panel shows how much time passes between messages reported by the exchange. A steeper (flatter) blue line signals a lower (higher) rate of messages, given that a steeper (flatter) line signals more (less) time progression. The red vertical lines signal when the JPM spoofing activities took place, from left to right: when the genuine order was placed, when the spoof order of 3000 contracts was placed, when the genuine order was executed and when the spoof order was cancelled.

Welch's *t*-tests were used to test whether liquidity differed significantly between the periods before, during and after the spoofing. Results are reported in Table 3.3. In any time window, liquidity costs before the spoofing were higher than during the spoofing. In other words, liquidity was lower before than during the spoofing and improved during the spoof. After the spoof ended, liquidity costs significantly increased and, hence, liquidity was significantly lower after than during the spoof. Up to 30 seconds after the spoof ended, liquidity costs were higher than before the spoof. In other words, liquidity was significantly worse after the spoof than before.[23]

3.4.3 Traditional Spoofing with Iceberg Orders

Two futures contracts are part of the "traditional spoofing with iceberg orders" category: the Silver March 2014 and Ultra T-Bond September 2015 contracts. This section only discusses results for the Ultra T-Bond September 2015 contract. Table 3.4 outlines the spoof-

23 Results differ for the March 2010 contract, as can be seen in the Appendix 3.A.

Table 3.3 | Mean ask liquidity costs (bps) around the 10-Year T-Note December 2011 spoof for different time windows

Time window	Before vs. during	During vs. after	Before vs. after
Spoof duration	4.587 > 2.250***	2.250 < 4.620***	4.587 < 4.620***
10 seconds	4.605 > 2.250***	2.250 < 4.838***	4.605 < 4.838***
30 seconds	4.623 > 2.250***	2.250 < 4.632***	4.623 < 4.632*
1 minute	4.700 > 2.250***	2.250 < 4.680***	4.700 > 4.680***
5 minutes	4.644 > 2.250***	2.250 < 4.945***	4.644 < 4.945***

Note: The table reports the mean liquidity costs (bps, measured by APM) around the spoof in the 10-Year T-Note December 2011 market for different periods and various time windows. *Before* represents the time up until the spoof order was added; *during* the period from when the spoof order was added until it was cancelled; and *after* the time following the cancellation of the spoof order. Five different time windows are used, the *Spoof duration* time window being 0.310 seconds. A lower APM indicates that liquidity costs are low and, hence, liquidity is high. Welch's *t*-tests were used to test for mean differences between the periods. Significance at the 0.1%, 1% and 5% (two-tailed) levels is indicated by ***, ** and *, respectively.

Table 3.4 | Spoofing actions on June 30, 2015 in the Ultra T-Bond September 2015 futures market

Time	Order type	LOB side	Action	Price (points)	Volume
08:45:46.627	Genuine order	Bid	Add	153.71875	1 displayed 199 hidden
08:46:01.891	Spoof order	Ask	Add	153.75	100
08:46:02.979	*First contract of genuine order executed*				
08:46:04.288	*Last contract of genuine order executed (51 of 200 contracts executed)*				
08:46:04.418	Spoof order	Ask	Cancel	153.75	100
08:46:08.074	Genuine order	Bid	Cancel	153.71875	149

Note: This table presents the various spoofing actions JPM took on June 30, 2015 in the Ultra T-Bond September 2015 futures market. Per spoof action, the table reports the timestamp (*Time*), whether it concerned a genuine or spoof order (*Order type*), the LOB side the spoof action occurred on (*LOB side*), whether the order from the spoof action was added or cancelled (*Action*), the price level affected by the spoof action (*Price (points)*) and the volume related to the spoof action (*Volume*).

ing actions JPM undertook in the Ultra T-Bond September 2015 market on June 30, 2015. All spoofing actions lasted for 21.447 seconds and consisted of a single genuine and spoof order. The genuine order was an iceberg order on the first bid level and consisted of one visible contract and 199 hidden contracts. The spoof order was placed on the first ask level and consisted of 100 contracts.

Table 3.5 shows the state of the LOB one millisecond before JPM's first spoofing action in the Ultra T-Bond market. The spoofing involved buying 51 contracts at 153.71875 points, representing a total underlying value of $7,839,656.25 (one point equaling $1000). If the JPM trader had executed their genuine order with market orders, they would have bought 51 contracts at 153.75 points, representing a total underlying value of $7,841,250. Hence, due to spoofing, JPM bought the contracts $1593.75 cheaper, excluding trading costs. Assuming

Table 3.5 | LOB state one millisecond before placement of the genuine order from the Ultra T-Bond September 2015 spoof

Bid volume	Bid price (points)	Level	Ask price (points)	Ask volume
30	153.71875	1	153.75000	69
121	153.68750	2	153.78125	127
104	153.65625	3	153.81250	52
37	153.62500	4	153.84375	40
45	153.59375	5	153.87500	65
42	153.56250	6	153.90625	43
41	153.53125	7	153.93750	45
54	153.50000	8	153.96875	51
47	153.46875	9	154.00000	38
47	153.43750	10	154.03125	46

Note: This table reports the state of the LOB one millisecond before the genuine order from the Ultra T-Bond September 2015 spoof was added. It shows the prices and volumes of each level on the bid and ask side.

the JPM trader wanted the full genuine order executed, i.e., buy 200 contracts rather than 51 contracts, the gains would have been larger. In that situation, a market order of 200 contracts would have 'run up' the LOB: they would have bought 69 contracts at 153.75 points; 127 contracts at 153.78125 points and four contracts at 153.8125 points. The total underlying value using market orders would have been \$30,754,218.75, which is \$10,468.75 more than the total underlying value of buying 200 contracts in the spoofing scenario (\$30,743,750). JPM might have placed an iceberg order or initiated the spoofing actions not to move the price, but to attract more liquidity to avoid running up the LOB and incur liquidity costs. This will be further explored in section 3.4.3.4 and section 3.4.6.

3.4.3.1 Traditional Spoofing with Iceberg Orders: Visualization of the LOB and Trades around Spoofing

Figure 3.8 visualizes the behavior of the LOB and information about trades around the spoofing in the Ultra T-Bond September 2015 contract, on June 30, 2015 between 08:45:40 and 08:46:10. When the genuine order was added, most of the volume in the LOB was concentrated on the second and third bid levels and the first two ask levels. Once the spoof order of 100 contracts was placed, volume increased significantly on the first ask level, as indicated by a bright yellow color. After the genuine order was executed, volume on the first bid level decreased, indicated by ever darker shades of blue. In contrast, more volume was added on the second ask level. Volume on the first ask level was significantly lower once the spoof order was cancelled.

The top panel in Figure 3.8 shows that the price of the genuine order and the last traded price were identical (153.71875 points) at the time of placing the genuine order. Hence, the spoof order may have been used to attract more liquidity to the price of the genuine order,

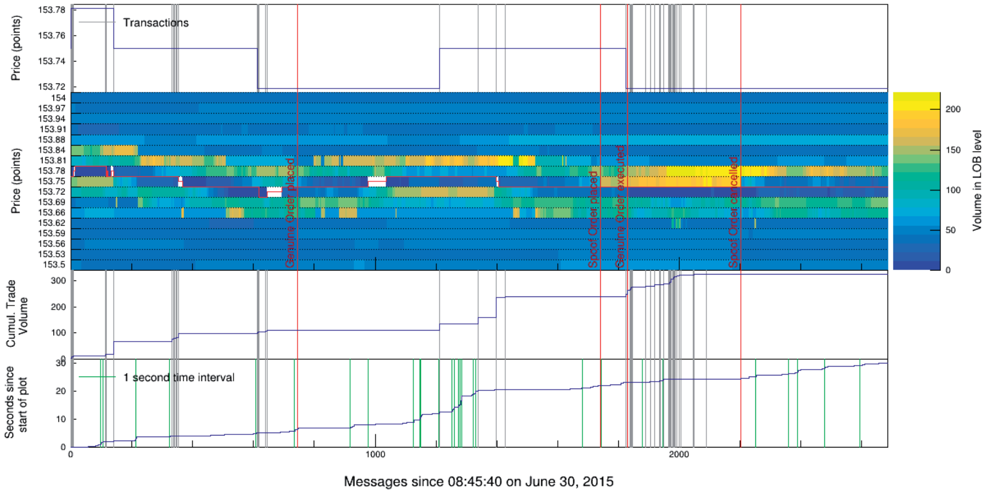


Figure 3.8 | Ultra T-Bond September 2015 LOB and trade behavior around the spoof of June 30, 2015. This figure visualizes the LOB and trade behavior around the spoof of June 30, 2015 in the Ultra T-Bond September 2015 futures market. The *first* panel shows the price of the last trade that took place (blue line) and when a trade took place (gray line). The *second* panel shows the volumes at the individual bid and ask levels between prices of 153.5 and 154 points. Each unit on the x-axis is one message. The y-axis represents the price of the Ultra T-Bond in points. The color represents the volume at each price level of the LOB for each message. The scale ranges from blue to yellow, with the color becoming a brighter yellow as volume increases at that price level. The red line is the midpoint. The *third* panel shows the cumulative trade volume per second. The *fourth* panel shows how much time passes between messages reported by the exchange. A steeper (flatter) blue line signals a lower (higher) rate of messages, given that a steeper (flatter) line signals more (less) time progression. The red vertical lines signal when the JPM spoofing activities took place, from left to right: when the genuine iceberg order was placed, when the spoof order of 100 contracts was placed, when the first contract of the genuine order was executed and when the spoof order was cancelled.

which will be further explored in section 3.4.3.4 and section 3.4.6. When the spoof order was placed, the last traded price was 153.75 points, and shortly after the placement – 1.087 seconds later – it decreased to the price level of the genuine order, to stay there for the remainder of the visualized time window. The cumulative trade volume panel in Figure 3.8 provides more information about the trading patterns of iceberg orders: while previous trades showed staircase patterns, the spoofing-related trades are more gradual because of the associated iceberg order. This order only executes one trade at a time, whereby each trade is recorded in a separate message.[24] Hence, in this case, visualizing trades based on messages provides more insight in the type of order and trade. Furthermore, it can provide additional insights in the type of trader. For example, an algorithm could also have produced the same type of trade pattern, as algorithms trade in nanoseconds and can therefore rapidly execute market orders in a short time window.

24 The iceberg order of the Silver March 2014 spoof used five visible contracts and, hence, five contracts at a time can be executed. This caused cumulative volume to increase in a staircase pattern rather than gradually, as it did in the Ultra T-Bond September 2015 contract.

3.4.3.2 Traditional Spoofing with Iceberg Orders: Visualization of Volume around Spoofing

Figure 3.9 visualizes the changes in volume on the first bid and ask levels around the spoofing in the Ultra T-Bond September 2015 contract. When the genuine order was added, volume on the first bid and ask levels changed regularly, which can be attributed to a new first price level being added or removed from the LOB. When the spoof order was added, volume on the first ask level increased by 100 contracts and kept increasing gradually until the spoof order was removed. Volume on the first bid level remained relatively stable when the spoof order was added and dropped when the genuine order was executed. At this point, it remained between one to ten contracts until the spoof order was cancelled and shortly after.

3.4.3.3 Traditional Spoofing with Iceberg Orders: Visualization of Cancellations around Spoofing

Figure 3.10 shows the cumulative cancellations on the first bid and ask levels around the spoof. In general, the volume cancelled on the first bid and ask levels is small. Up until when

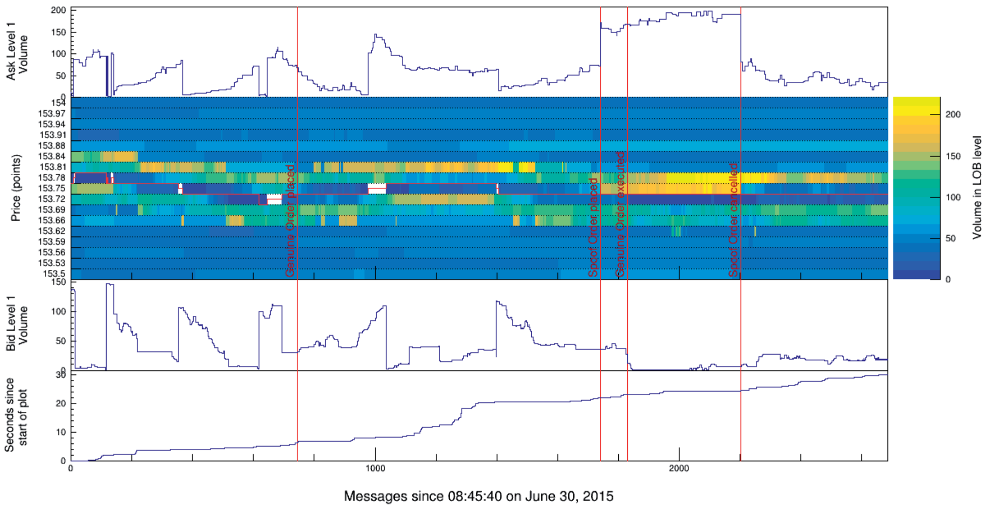


Figure 3.9 | Ultra T-Bond September 2015 first-level volume behavior around the spoof of June 30, 2015. This figure visualizes first-level bid and ask volume behavior around the spoof of June 30, 2015 in the Ultra T-Bond September 2015 futures market. The *first* panel shows the volume of the best ask level. The *second* panel shows the volumes at the individual bid and ask levels between prices of 153.5 and 154 points. Each unit on the x-axis is one message. The y-axis represents the price of the Ultra T-Bond in points. The color represents the volume at each price level of the LOB for each message. The scale ranges from blue to yellow, with the color becoming a brighter yellow as volume increases at that price level. The red line is the midpoint. The *third* panel shows the volume of the best bid level. The *fourth* panel shows how much time passes between messages reported by the exchange. A steeper (flatter) blue line signals a lower (higher) rate of messages, given that a steeper (flatter) line signals more (less) time progression. The red vertical lines signal when the JPM spoofing activities took place, from left to right: when the genuine iceberg order was placed, when the spoof order of 100 contracts was placed, when the first contract of the genuine order was executed and when the spoof order was cancelled.

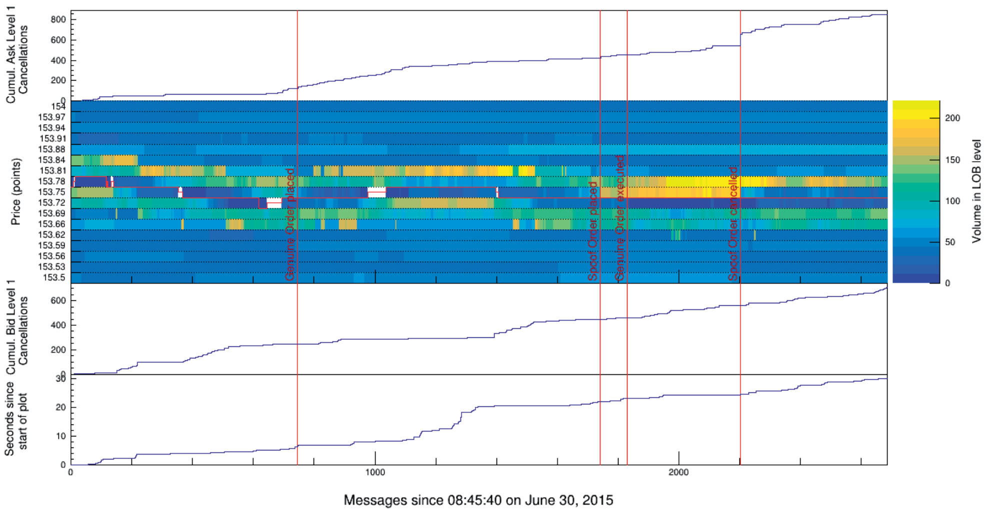


Figure 3.10 | Ultra T-Bond September 2015 first-level cancellation behavior around the spoof of June 30, 2015. This figure visualizes cumulative first-level bid and ask cancellation volume around the spoof of June 30, 2015 in the Ultra T-Bond September 2015 futures market. The *first* panel shows the cumulative volume of cancellations of the best ask level. The *second* panel shows the volumes at the individual bid and ask levels between prices of 153.5 and 154 points. Each unit on the x-axis is one message. The y-axis represents the price of the Ultra T-Bond in points. The color represents the volume at each price level of the LOB for each message. The scale ranges from blue to yellow, with the color becoming a brighter yellow as volume increases at that price level. The red line is the midpoint. The *third* panel shows the cumulative volume of cancellations of the best bid level. The *fourth* panel shows how much time passes between messages reported by the exchange. A steeper (flatter) blue line signals a lower (higher) rate of messages, given that a steeper (flatter) line signals more (less) time progression. The red vertical lines signal when the JPM spoofing activities took place, from left to right: when the genuine iceberg order was placed, when the spoof order of 100 contracts was placed, when the first contract of the genuine order was executed and when the spoof order was cancelled.

the spoof order was cancelled, cancellations on the first ask level were increasing gradually. When the spoof order was cancelled, it increased significantly by 100 contracts. Cancellations on the first bid level continued to gradually increase in the visualized time window. Figure 3.10 complements Figure 3.9, in that Figure 3.10 explains whether the shifts in Figure 3.9 should be attributed to cancellations or to other causes.

3.4.3.4 Traditional Spoofing with Iceberg Orders: Visualization of Liquidity around Spoofing

Figure 3.11 visualizes the bid and ask liquidity costs around the spoof in the Ultra T-Bond September 2015 contract. Before the spoof order was placed, liquidity costs on the ask side fluctuated between 9.5 and 13 bps. Immediately when the spoof order was placed, ask liquidity costs dropped from 10.38 bps to 7.97 bps and further decreased to approximately

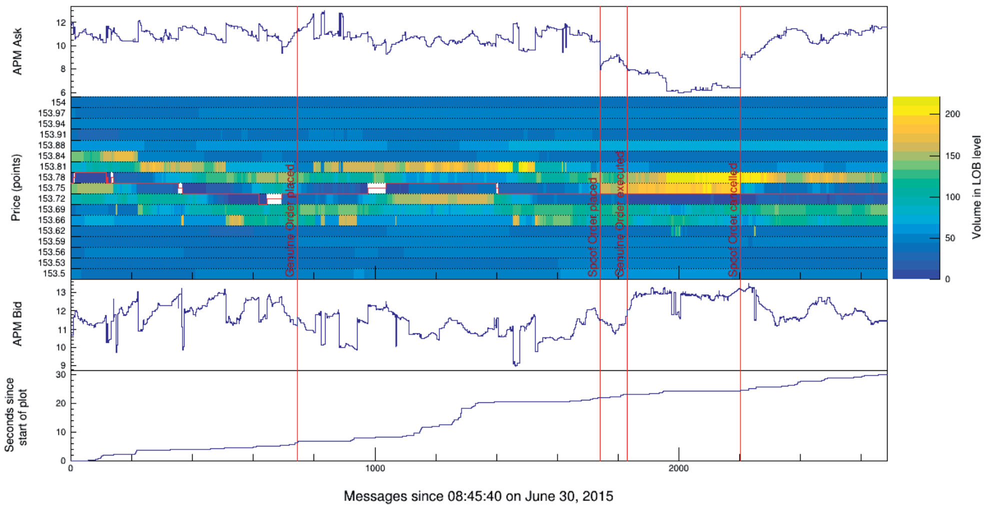


Figure 3.11 | Ultra T-Bond September 2015 bid and ask APM behavior around the spoof of June 30, 2015. This figure visualizes bid and ask liquidity costs (APM) behavior around the spoof of June 30, 2015 in the Ultra T-Bond September 2015 futures market. The *first* panel shows the APM of the ask side. APM measures the liquidity costs (in basis points) of a trader who wants to buy or sell a specific dollar value by submitting market orders. The *second* panel shows the volumes at the individual bid and ask levels between prices of 153.5 and 154 points. Each unit on the x-axis is one message. The y-axis represents the price of the Ultra T-Bond in points. The color represents the volume in each price level of the LOB for each message. The scale ranges from blue to yellow, with the color becoming a brighter yellow as volume increases at that price level. The red line is the midpoint. The *third* panel shows the APM for the bid side. The *fourth* panel shows how much time passes between messages reported by the exchange. A steeper (flatter) blue line signals a lower (higher) rate of messages, given that a steeper (flatter) line signals more (less) time progression. The red vertical lines signal when the JPM spoofing activities took place, from left to right: when the genuine iceberg order was placed, when the spoof order of 100 contracts was placed, when the first contract of the genuine order was executed and when the spoof order was cancelled.

6 bps right before the spoof order was cancelled. After the spoof order was cancelled, ask liquidity costs fluctuated between 9 and 12 bps in the visualized time window. Compared to the ask side, bid side liquidity costs were relatively more volatile, fluctuating between 9 and 13 bps.

Table 3.6 shows the test results for whether liquidity costs were significantly different before, during and after the spoofing. Irrespective of the time window, liquidity costs were higher before and after the spoof than during the spoof. In other words, liquidity was better during the spoof than before and after. When comparing liquidity costs before and after the spoof, the results differ per time window. Liquidity was better 2.52 seconds after the spoof than before the spoof. For each subsequent time window, the results are mixed.

Table 3.6 | Mean ask liquidity costs (bps) around Ultra T-Bond September 2015 spoof for different time windows

Time window	Before vs. during	During vs. after	Before vs. after
Spoof duration	10.695 > 7.981***	7.981 < 10.267***	10.695 > 10.267***
10 seconds	10.545 > 7.981***	7.981 < 10.851***	10.545 < 10.851***
30 seconds	10.873 > 7.981***	7.981 < 10.907***	10.873 = 10.907
1 minute	10.935 > 7.981***	7.981 < 10.799***	10.935 > 10.799***
5 minutes	13.284 > 7.981***	7.981 < 10.369***	13.284 < 10.369***

Note: The table reports the mean liquidity costs (bps, measured by APM) around the spoof in the Ultra T-Bond September 2015 market for different periods and various time windows. *Before* represents the time up until the spoof order was added; *during* the period from when the spoof order was added until it was cancelled; and a*fter* the time following the cancellation of the spoof order. Five different time windows are used, the *Spoof duration* time window being 2.52 seconds. A lower APM indicates that liquidity costs are low and, hence, liquidity is high. Welch's *t*-tests were used to test for mean differences between the periods. Significance at the 0.1%, 1% and 5% (two-tailed) levels is indicated by ***, ** and *, respectively.

Table 3.7 | Spoofing actions on July 20, 2009 in the T-Bond September 2009 futures market

Time	Order type	LOB side	Action	Price (points)	Volume
07:47:13.597	Genuine order	Ask	Add	116.171875	100
07:47:17.098	Spoof layer 1	Bid	Add	116.078	300
07:47:17.847	Spoof layer 2	Bid	Add	116.094	300
07:47:18.583	Spoof layer 3	Bid	Add	116.109	300
07:47:19.379	Spoof layer 4	Bid	Add	116.125	300
07:47:20.212	Spoof layer 5	Bid	Add	116.141	300
07:47:21.020	Spoof layer 6	Bid	Add	116.156	300
07:47:21.036	*Complete genuine order executed*				
07:47:22.039	Spoof layer 6	Bid	Cancel	116.156	300
07:47:22.064	Spoof layer 5	Bid	Cancel	116.141	300
07:47:22.064	Spoof layer 4	Bid	Cancel	116.125	300
07:47:22.064	Spoof layer 3	Bid	Cancel	116.109	300
07:47:22.067	Spoof layer 2	Bid	Cancel	116.094	300
07:47:22.303	Spoof layer 1	Bid	Cancel	116.078	300

Note: This table presents the various spoofing actions JPM took on July 20, 2009 in the T-Bond September 2009 futures market. Per spoof action, the table reports the timestamp (*Time*), whether it concerned a genuine or spoof order (*Order type*), the LOB side the spoof action occurred on (*LOB side*), whether the order from the spoof action was added or cancelled (*Action*), the price level affected by the spoof action (*Price (points)*) and the volume related to the spoof action (*Volume*).

3.4.4 Layered Spoofing

Four futures contracts are part of the "layered spoofing" category: the Silver March 2012, Silver May 2014, Gold April 2014 and T-Bond September 2009 contracts. This section only discusses results for the T-Bond September 2009 contract. Table 3.7 shows the spoofing actions by JPM in the T-Bond September 2009 contract (CFTC, 2020e), which lasted for a total of 8.706 seconds. The spoof consisted of one genuine order with a volume of 100 contracts at the second ask level[25] and six spoof orders with a volume of 300 contracts.

Table 3.8 | LOB state one millisecond before placement of the genuine order from the T-Bond September 2009 spoof

Bid volume	Bid price (points)	Level	Ask price (points)	Ask Volume
59	116.141	1	116.156	55
85	116.125	2	116.172	62
90	116.109	3	116.188	180
163	116.094	4	116.203	102
79	163.078	5	116.219	105
116	116.062	6	116.234	108
75	116.047	7	116.25	61
184	116.031	8	116.266	204
42	116.016	9	116.281	233
35	116	10	116.297	41

Note: This table reports the state of the LOB one millisecond before the genuine order from the T-Bond September 2009 spoof was added. It shows the prices and volumes of each level on the bid and ask side.

Table 3.8 shows the state of the T-Bond September 2009 contract on July 20, 2009 one millisecond before the genuine order was placed. JPM sold 100 contracts at 116.171875 points, amounting to a total underlying value of $11,617,187.5. Had JPM submitted their genuine order as a market order rather than a limit order, it would have consumed the first and part of the second bid level. In that scenario, JPM would have sold 59 contracts at 116.141 points and 41 contracts at 116.125 points, representing a total underlying value of $11,613,444. Hence, JPM sold their contracts for $3743.5 more through spoofing, excluding trading costs.

3.4.4.1 Layered Spoofing: Visualization of the LOB and Trades around Spoofing

Figure 3.12 shows the visualization of the LOB and trades around the JPM spoofing in the T-Bond September 2009 market on July 20, 2009 from 07:47:10 to 07:47:30. The second panel shows that when the genuine order was added, individual levels contained approximately between 50 to 250 contracts. Most volume was concentrated on the third, eighth and ninth ask levels and on the fourth and eighth bid levels. Spoof orders of 300 contracts were placed on six different levels and, as indicated by the bright yellow color, were relatively large compared to the volumes on these levels. The spoof orders were placed from the lower to the higher levels in the LOB, i.e., from level six to level one. Conversely, spoof orders were cancelled from the higher to the lower levels in the LOB, i.e., from level one to level six. Hence, the spoof orders closest to the top of the LOB were active for the shortest amount of time. The execution of the genuine order and the cancellation of all spoof orders occurred within the same second, as indicated by the green vertical lines in the lower panel.

25 The genuine orders for the Silver March 2012, Silver May 2014 and Gold April 2014 contracts were all placed on the first rather than the second ask level.

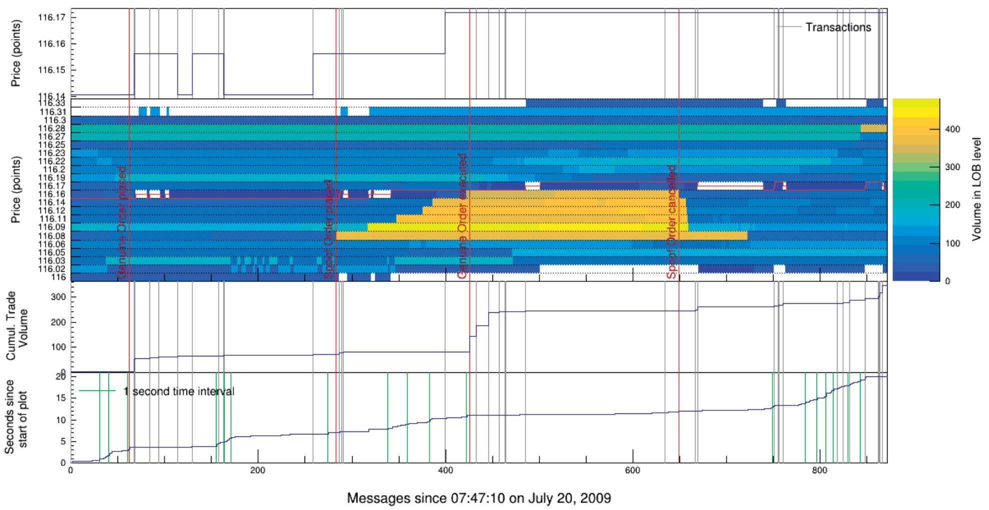


Figure 3.12 | T-Bond September 2009 LOB and trade behavior around the spoof of July 20, 2009. This figure visualizes the LOB and trade behavior around the spoof of July 20, 2009 in the T-Bond September 2009 futures market. The *first* panel shows the price of the last trade that took place (blue line) and when a trade took place (gray line). The *second* panel shows the volumes at the individual bid and ask levels between prices of 116 and 116.33 points. Each unit on the x-axis is one message. The y-axis represents the price of the T-Bond in points. The color represents the volume at each price level of the LOB for each message. The scale ranges from blue to yellow, with the color becoming a brighter yellow as volume increases at that price level. The red line is the midpoint. The *third* panel shows the cumulative trade volume per second. The *fourth* panel shows how much time passes between messages reported by the exchange. A steeper (flatter) blue line signals a lower (higher) rate of messages, given that a steeper (flatter) line signals more (less) time progression. The red vertical lines signal when the JPM spoofing activities took place, from left to right: when the genuine order was placed, when the first spoof order of 300 contracts was placed, when the genuine order was executed and when the first spoof order was cancelled.

The top panel in Figure 3.12 shows that, when the genuine order was placed at 116.171875 points (rounded 116.172 points), the last traded price was 116.141 points. Hence, the goal of this spoof might have been to move the price up towards the ask price of the genuine order.[26] Before the first spoof order was placed, the last traded price moved between the highest bid (116.141 points) and lowest ask (116.156 points).[27] This illustrates which side triggers the trade: a trader wanting to buy and taking the lowest ask, or a trader wanting to sell and taking the highest bid. Shortly after the fifth spoof order was placed, the trade price increased to 116.172 points and the genuine order was executed. The cumulative trade panel in Figure 3.12 shows that, before the genuine order was executed, one large trade

26 The price of the genuine order was equal to the last traded price in the case of the Silver May 2014 spoof.

27 The last traded price of the Silver May 2014 contract did not move during the visualized time window (from 08:18:35 to 08:18:50).

occurred (shortly after the genuine order was added) while the other trades were relatively small. At the time of the execution of the genuine order and after, larger trades were executed, as indicated by the staircase pattern.

3.4.4.2 Layered Spoofing: Visualization of Volume around Spoofing

Figure 3.13 visualizes the changes in volume on the second levels around the spoof of the T-Bond September 2009 contract. When the genuine order was added, the second ask level consisted of 62 contracts and the second bid level of 85 contracts. Both volumes remained relatively constant within these price levels until the first spoof order was added. Large fluctuations in the second ask level were mainly attributable to a changing bid-ask spread and, hence, changing second ask price level. Once the spoof order was placed on the second bid level, around the 380 message mark, the volume increased significantly by 300 contracts. Although the price level of the second bid level changed around the 480 and 500 message mark, the volume on the second bid level continued to be high as 300 contracts were

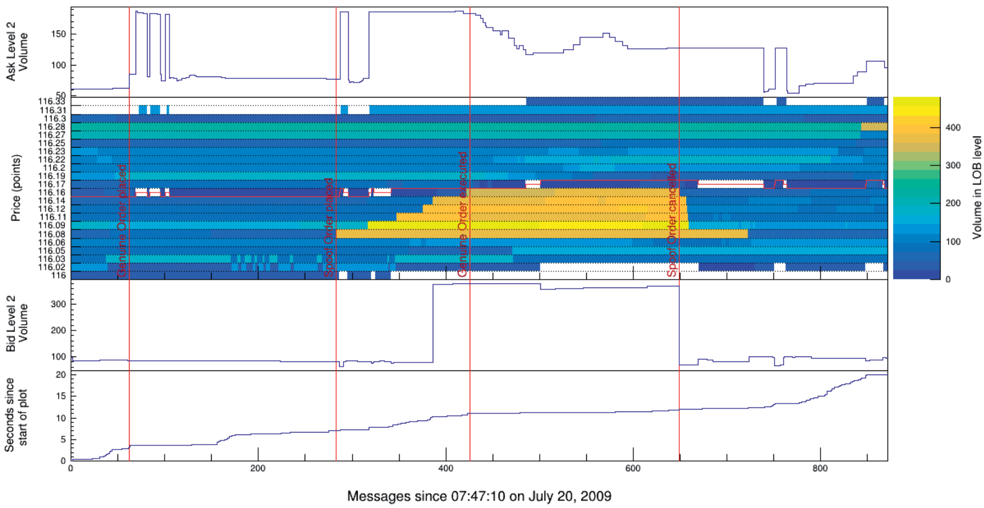


Figure 3.13 | T-Bond September 2009 second-level volume behavior around the spoof of July 20, 2009. This figure visualizes second-level bid and ask volume behavior around the spoof of July 20, 2009 in the T-Bond September 2009 futures market. The *first* panel shows the volume of the second ask level. The *second* panel shows the volumes at the individual bid and ask levels between prices of 116 and 116.33 points. Each unit on the x-axis is one message. The y-axis represents the price of the T-Bond in points. The color represents the volume at each price level of the LOB for each message. The scale ranges from blue to yellow, with the color becoming a brighter yellow as volume increases at that price level. The red line is the midpoint. The *third* panel shows the volume of the second bid level. The *fourth* panel shows how much time passes between messages reported by the exchange. A steeper (flatter) blue line signals a lower (higher) rate of messages, given that a steeper (flatter) line signals more (less) time progression. The red vertical lines signal when the JPM spoofing activities took place, from left to right: when the genuine order was placed, when the first spoof order of 300 contracts was placed, when the genuine order was executed and when the first spoof order was cancelled.

added to multiple layers by the JPM trader. Once the spoof order was cancelled, the volume decreased significantly by 300 contracts.[28]

3.4.4.3 Layered Spoofing: Visualization of Cancellations around Spoofing

Figure 3.14 shows the cancellations on the second bid and ask levels around the spoof in the T-Bond September 2009 market. Cancellations on the second ask level gradually increased in the visualized time window, the largest cancellations being approximately ten contracts in one message. Cumulative cancellations on the second bid level remained under 40≈contracts up until the cancellation of the first spoof order. When the first spoof order was cancelled, it significantly increased by 300 contracts, after which it continued to gradually increase at a slower pace.

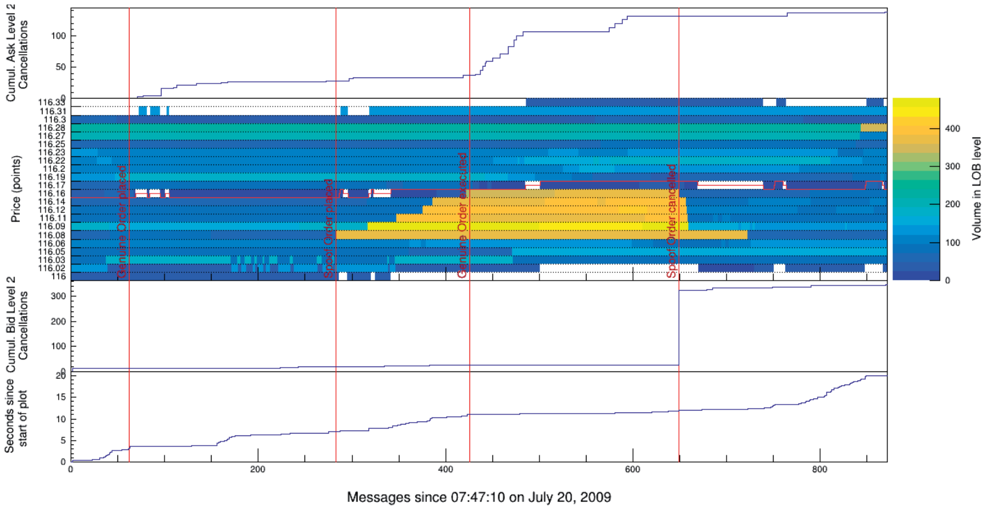


Figure 3.14 | T-Bond September 2009 second-level cancellation behavior around the spoof of July 20, 2009. This figure visualizes cumulative second-level bid and ask cancellation volume around the spoof of July 20, 2009 in the T-Bond September 2009 futures market. The *first* panel shows the cumulative volume of cancellations of the second ask level. The *second* panel shows the volumes at the individual bid and ask levels between prices of 116 and 116.33 points. Each unit on the x-axis is one message. The y-axis represents the price of the T-Bond in points. The color represents the volume at each price level of the LOB for each message. The scale ranges from blue to yellow, with the color becoming a brighter yellow as volume increases at that price level. The red line is the midpoint. The *third* panel shows the cumulative volume of cancellations of the second bid level. The *fourth* panel shows how much time passes between messages reported by the exchange. A steeper (flatter) blue line signals a lower (higher) rate of messages, given that a steeper (flatter) line signals more (less) time progression. The red vertical lines signal when the JPM spoofing activities took place, from left to right: when the genuine order was placed, when the first spoof order of 300 contracts was placed, when the genuine order was executed and when the first spoof order was cancelled.

28 Due to a frequently changing bid-ask spread in the visualized time window, the first bid and ask volumes fluctuated more in the Gold April 2014 contract than in the other spoofing examples.

3.4.4.4 *Layered Spoofing: Visualization of Liquidity around Spoofing*

The ask and bid liquidity costs around the T-Bond September 2009 contract are visualized in Figure 3.15. The liquidity costs on the bid side fluctuated between 6.5 and 10 bps before the first spoof order was added and continued to decrease with every additional spoof order added, reaching their lowest point at 1.9 bps before stabilizing at approximately 3 bps. After all spoof orders were cancelled, the bid liquidity costs fluctuated between 4 and 9 bps. The ask liquidity costs fluctuated between 5.2 and 7.4 bps, and reached their lowest point in the visualized time window during the spoof.[29]

Results from the Welch's *t*-tests for the T-Bond September 2009 spoof are reported in Table 3.9. The bid liquidity costs were significantly higher before and after the spoof than

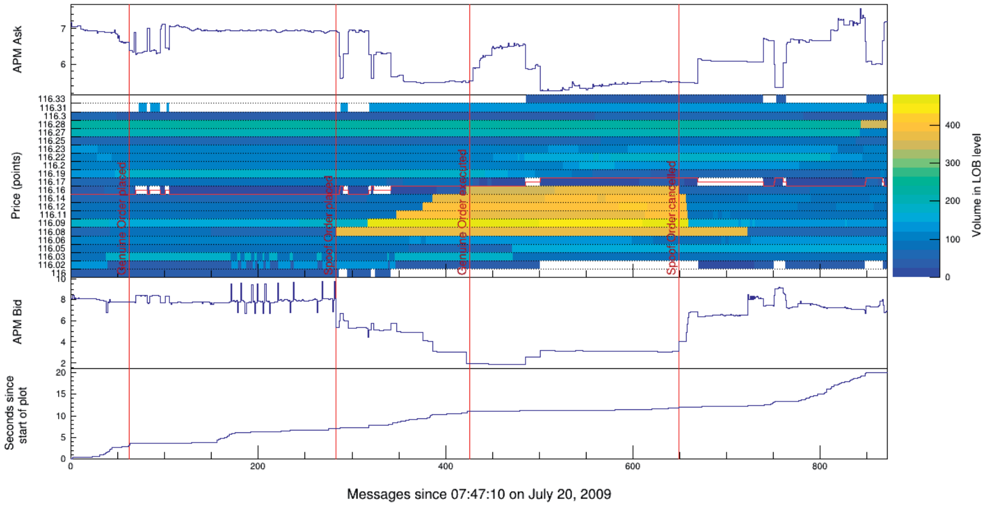


Figure 3.15 | T-Bond September 2009 bid and ask APM behavior around the spoof of July 20, 2009. This figure visualizes bid and ask liquidity costs (APM) behavior around the spoof of July 20, 2009 in the T-Bond September 2009 futures market. The *first* panel shows the APM of the ask side. APM measures the liquidity costs (in basis points) of a trader who wants to buy or sell a specific dollar value by submitting market orders. The *second* panel shows the volumes at the individual bid and ask levels between prices of 116 and 116.33 points. Each unit on the x-axis is one message. The y-axis represents the price of the T-Bond in points. The color represents the volume in each price level of the LOB for each message. The scale ranges from blue to yellow, with the color becoming a brighter yellow as volume increases at that price level. The red line is the midpoint. The *third* panel shows the APM for the bid side. The *fourth* panel shows how much time passes between messages reported by the exchange. A steeper (flatter) blue line signals a lower (higher) rate of messages, given that a steeper (flatter) line signals more (less) time progression. The red vertical lines signal when the JPM spoofing activities took place, from left to right: when the genuine order was placed, when the first spoof order of 300 contracts was placed, when the genuine order was executed and when the first spoof order was cancelled.

29 The other spoofing examples in this category all showed a similar downward patterns in liquidity costs.

during the spoof, regardless of the time window. Hence, liquidity improved during the spoof. Up until 30 seconds after the spoof, the liquidity costs were significantly lower than during the spoof.

3.4.5 Layered Spoofing with Iceberg Orders

One futures contract is part of the "layered spoofing with iceberg orders" category: the Platinum July 2016 contract. Table 3.10 outlines the spoofing actions JPM took in the Platinum market on June 22, 2016 (CFTC, 2020e). All spoofing actions lasted for a total of 6.76 seconds

Table 3.9 | Mean bid liquidity costs (bps) around the T-Bond September 2009 spoof for different time windows

Time window	Before vs. during	During vs. after	Before vs. after
Spoof duration	7.788 > 4.284***	4.284 < 7.424***	7.788 > 7.424***
10 seconds	7.882 > 4.284***	4.284 < 7.340***	7.882 > 7.340***
30 seconds	6.723 > 4.284***	4.284 < 5.980***	6.723 > 5.980***
1 minute	5.954 > 4.284***	4.284 < 6.603***	5.954 < 6.603***
5 minutes	8.265 > 4.284***	4.284 < 7.791***	8.265 > 7.791***

Note: The table reports the mean liquidity costs (bps, measured by APM) around the spoof in the T-Bond September 2009 market for different periods and various time windows. *Before* represents the time up until the spoof order was added; *during* the period from when the spoof order was added until it was cancelled; and a*fter* the time following the cancellation of the spoof order. Five different time windows are used, the *Spoof duration* time window being 5.2 seconds. A lower APM indicates that liquidity costs are low and, hence, liquidity is high. Welch's *t*-tests were used to test for mean differences between the periods. Significance at the 0.1%, 1% and 5% (two-tailed) levels is indicated by ***, ** and *, respectively.

Table 3.10 | Spoofing actions on June 22, 2016 in the Platinum July 2016 futures market

Time	Order type	LOB side	Action	Price	Volume
02:14:33.935	Genuine order	Ask	Add	$981.8	1 displayed 19 hidden
02:14:35.926	Spoof layer 1	Bid	Add	$981.2	5
02:14:36.072	Spoof layer 2	Bid	Add	$981.4	5
02:14:36.214	Spoof layer 3	Bid	Add	$981.6	5
02:14:36.374	Spoof layer 4	Bid	Add	$981.6	5
02:14:36.519	Spoof layer 5	Bid	Add	$981.6	5
02:14:36.520	*Four contracts of genuine order executed*				
02:14:36.678	Spoof layer 6	Bid	Add	$981.6	5
02:14:36.824	Spoof layer 7	Bid	Add	$981.6	5
02:14:37.006	Spoof layer 8	Bid	Add	$981.6	5
02:14:37.407	Spoof layer 3-8	Bid	Cancel	$981.6	30
02:14:38.063	Spoof layer 2	Bid	Cancel	$981.4	5
02:14:40.695	Spoof layer 1	Bid	Cancel	$981.2	5

Note: This table presents the various spoofing actions JPM took on June 22, 2016 in the Platinum July 2016 futures market. Per spoof action, the table reports the timestamp (*Time*), whether it concerned a genuine or spoof order (*Order type*), the LOB side the spoof action occurred on (*LOB side*), whether the order from the spoof action was added or cancelled (*Action*), the price level affected by the spoof action (*Price*) and the volume related to the spoof action (*Volume*).

Table 3.11 | LOB state one millisecond before placement of the genuine order from the Platinum July 2016 spoof

Bid volume	Bid price	Level	Ask price	Ask volume
2	$981.7	1	$982.2	4
4	$981.6	2	$982.3	2
3	$981.4	3	$982.5	2
6	$981.3	4	$982.6	1
6	$981.2	5	$982.7	2
8	$981.0	6	$982.8	2
4	$980.9	7	$983.0	7
4	$980.8	8	$983.1	3
2	$980.7	9	$983.2	7
3	$980.6	10	$983.3	1

Note: This table reports the state of the LOB one millisecond before the genuine order from the Platinum July 2016 spoof was added. It shows the prices and volumes of each level on the bid and ask side.

and consisted of 1) a genuine iceberg order on the first ask level, with one contract displayed and nineteen hidden; and 2) eight spoof orders with a volume of five contracts each.

Table 3.11 shows the state of the LOB one millisecond prior to adding the genuine iceberg order. JPM sold four contracts for $981.80, with a total underlying value of $196,360. Had they sold these four contracts with a market order, they would have sold two contracts for $981.7 and two contracts for $981.6, with a total underlying value of $196,330. Hence, excluding trading costs, JPM received $30 more by using a limit order and spoofing the market. Assuming that JPM wanted the full genuine iceberg order executed, i.e., wanted to sell twenty rather than four contracts, the gains would have been larger. In that case, the spoofing would have resulted in JPM selling at an underlying value of $981,800. Using a market order of volume twenty, the order would have run down the LOB and consume the first four bid levels. In that case, JPM would have sold at a total underlying value of $981,400, which would have been $400 less than with spoofing, excluding transaction costs.

3.4.5.1 Layered Spoofing with Iceberg Orders: Visualization of the LOB and Trades around Spoofing

The behavior of the LOB and trades around the spoofing of the Platinum July 2016 contract is visualized in Figure 3.16 between 02:14:25 and 02:14:45. When the genuine order was added on the first ask level, the volume on the individual LOB levels was low at between zero to ten contracts, as visualized in the second panel. The bid-ask spread was wider before the genuine order was added than after: $0.4 and $0.1, respectively. This may illustrate that the spoof orders were used by JPM to attract more liquidity to the market, thereby tightening the bid-ask spread. This will be further explored in section 3.4.5.4 and section 3.4.6. The first spoof order was placed at the sixth bid level, the second spoof order at the fourth bid level

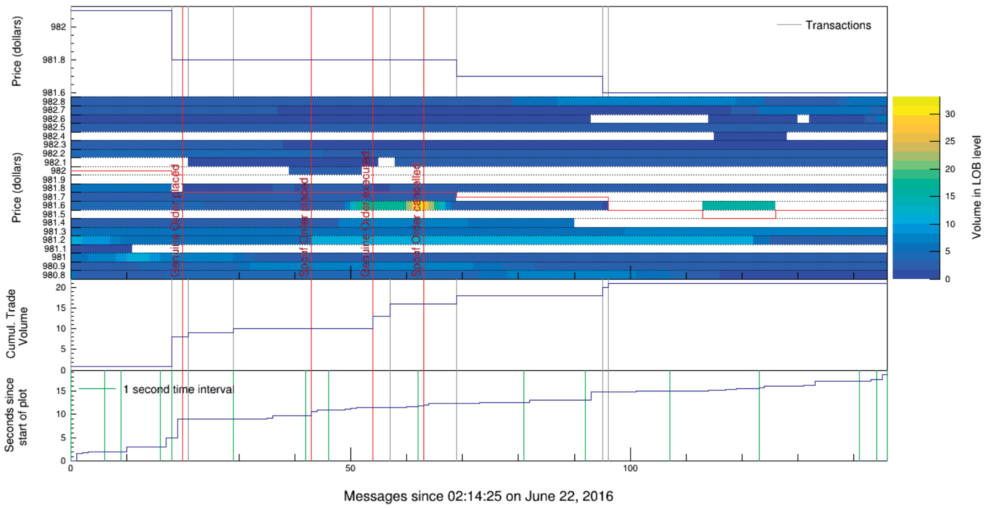


Figure 3.16 | Platinum July 2016 LOB and trade behavior around the spoof of June 22, 2016. This figure visualizes the LOB and trade behavior around the spoof of June 22, 2016 in the Platinum July 2016 futures market. The *first* panel shows the price of the last trade that took place (blue line) and when a trade took place (gray line). The *second* panel shows the volumes at the individual bid and ask levels between prices of $980.8 and $982.8. Each unit on the x-axis is one message. The y-axis represents the price of Platinum in dollars. The color represents the volume at each price level of the LOB for each message. The scale ranges from blue to yellow, with the color becoming a brighter yellow as volume increases at that price level. The red line is the midpoint. The *third* panel shows the cumulative trade volume per second. The fourth panel shows how much time passes between messages reported by the exchange. A steeper (flatter) blue line signals a lower (higher) rate of messages, given that a steeper (flatter) line signals more (less) time progression. The red vertical lines signal when the JPM spoofing activities took place, from left to right: when the genuine iceberg order was placed, when the first spoof order of 5 contracts was placed, when the first contract of the genuine order was executed and when the first spoof order was cancelled.

and the third to eighth spoof orders at the second bid level. This is visualized in Figure 3.16 by a color change on the respective level from blue to a lighter blue, green or yellow. Spoof orders were still being added one second after four contracts from the genuine order were executed, and the cancellations of the spoof orders started another second later.

The top panel in Figure 3.16 shows that, when the genuine order at price $981.8 was added, a transaction occurred in the same millisecond at a trade price of $981.8. Before this transaction, the last traded price was $982.1. For the duration of JPM's spoofing actions, the transaction price remained at $981.8. Cumulative trade volume increased steadily after the genuine order was placed.

3.4.5.2 Layered Spoofing with Iceberg Orders: Visualization of Volume around Spoofing

Figure 3.17 visualizes the volume changes in the second bid and ask levels around the time of the spoof. When the genuine order and the first spoof order were added, the volume on

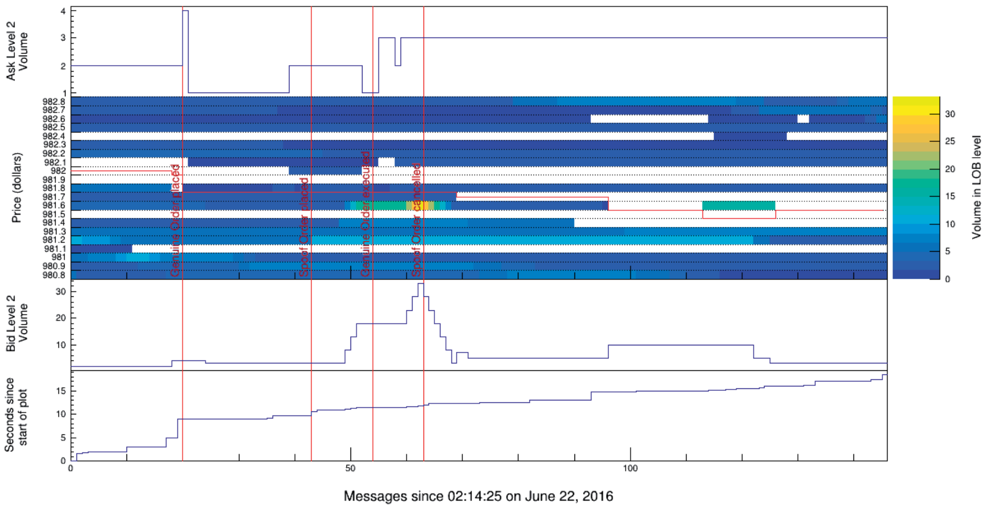


Figure 3.17 | Platinum July 2016 second-level volume behavior around the spoof of June 22, 2016. This figure visualizes second-level bid and ask volume behavior around the spoof of June 22, 2016 in the Platinum July 2016 futures market. The *first* panel shows the volume of the second ask level. The *second* panel shows the volumes at the individual bid and ask levels between prices of \$980.8 and \$982.8. Each unit on the x-axis is one message. The y-axis represents the price of Platinum in dollars. The color represents the volume at each price level of the LOB for each message. The scale ranges from blue to yellow, with the color becoming a brighter yellow as volume increases at that price level. The red line is the midpoint. The *third* panel shows the volume of the second bid level. The *fourth* panel shows how much time passes between messages reported by the exchange. A steeper (flatter) blue line signals a lower (higher) rate of messages, given that a steeper (flatter) line signals more (less) time progression. The red vertical lines signal when the JPM spoofing activities took place, from left to right: when the genuine iceberg order was placed, when the first spoof order of 5 contracts was placed, when the first contract of the genuine order was executed and when the first spoof order was cancelled.

the second bid and ask level was low at two contracts on each side. Once spoof orders were added on the second bid level, an upward staircase pattern emerged. After the first spoof order is cancelled, the same staircase pattern emerged but downwards. The height of the steps shows that the added and subtracted volumes were identical, i.e., five contracts per step.

3.4.5.3 *Layered Spoofing with Iceberg Orders: Visualization of Cancellations around Spoofing*

Cancellations on the second bid and ask levels around the spoof in the Platinum July 2016 contract are visualized in Figure 3.18. During the visualized time window, zero contracts were cancelled on both the bid and ask side when the genuine order was placed. Between the first spoof order being placed and being cancelled, cumulative cancellations amounted to one contract on the bid side and three contracts on the ask side. Once the first spoof order was cancelled, another upward staircase pattern emerged on the bid side with iden-

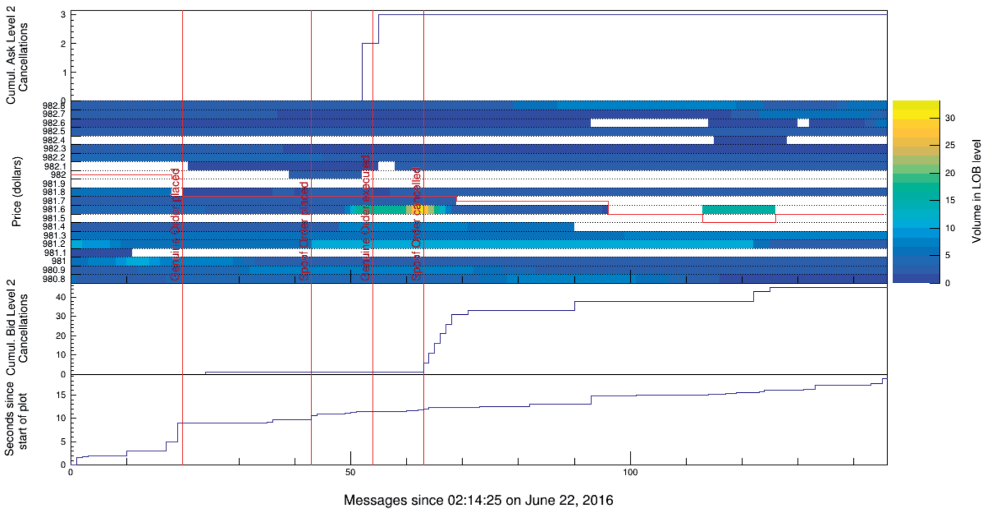


Figure 3.18 | Platinum July 2016 second-level cancellation behavior around the spoof of June 22, 2016. This figure visualizes cumulative second-level bid and ask cancellation volume around the spoof of June 22, 2016 in the Platinum July 2016 futures market. The *first* panel shows the cumulative volume of cancellations of the second ask level. The *second* panel shows the volumes at the individual bid and ask levels between prices of $980.8 and $982.8. Each unit on the x-axis is one message. The y-axis represents the price of the Platinum in dollars. The color represents the volume at each price level of the LOB for each message. The scale ranges from blue to yellow, with the color becoming a brighter yellow as volume increases at that price level. The red line is the midpoint. The *third* panel shows the cumulative volume of cancellations of the second bid level. The *fourth* panel shows how much time passes between messages reported by the exchange. A steeper (flatter) blue line signals a lower (higher) rate of messages, given that a steeper (flatter) line signals more (less) time progression. The red vertical lines signal when the JPM spoofing activities took place, from left to right: when the genuine iceberg order was placed, when the first spoof order of 5 contracts was placed, when the first contract of the genuine order was executed and when the first spoof order was cancelled.

tical heights of the steps, indicating that the cancellations had identical volumes. After all spoof orders from JPM were cancelled on the second bid level, cancellations continued in the visualized time window, albeit less frequently. The second level on the ask side showed no cancellations in the time window one second after the genuine order was executed.

3.4.5.4 Layered Spoofing with Iceberg Orders: Visualization of Liquidity around Spoofing

Figure 3.19 shows the ask and bid APM around the spoofing in the Platinum July 2016 market. Apart from one relatively large decrease, the liquidity costs on the bid side were relatively stable between 5 and 7 bps. Once the first spoof order was placed, liquidity costs decreased stepwise with each additional spoof order. Liquidity costs decreased from approximately 5.25 to 1.5 bps. Similarly, when the first spoof order was cancelled, liquidity costs increased stepwise with each spoof order cancelled. Ask side liquidity costs fluctuated between 8 and 10.5 bps during all JPM spoofing actions.

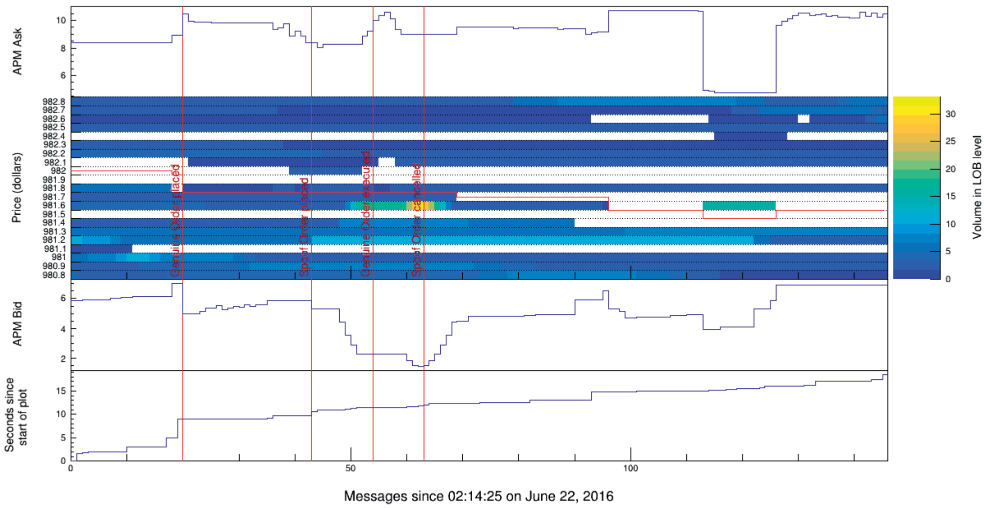


Figure 3.19 | Platinum July 2016 bid and ask APM behavior around the spoof of June 22, 2016. This figure visualizes bid and ask liquidity costs (APM) behavior around the spoof of June 22, 2016 in the Platinum July 2016 futures market. The *first* panel shows the APM of the ask side. APM measures the liquidity costs (in basis points) of a trader who wants to buy or sell a specific dollar value by submitting market orders. The *second* panel shows the volumes at the individual bid and ask levels between prices of $980.8 and $982.8. Each unit on the x-axis is one message. The y-axis represents the price of Platinum in dollars. The color represents the volume in each price level of the LOB for each message. The scale ranges from blue to yellow, with the color becoming a brighter yellow as volume increases at that price level. The red line is the midpoint. The *third* panel shows the APM for the bid side. The *fourth* panel shows how much time passes between messages reported by the exchange. A steeper (flatter) blue line signals a lower (higher) rate of messages, given that a steeper (flatter) line signals more (less) time progression. The red vertical lines signal when the JPM spoofing activities took place, from left to right: when the genuine iceberg order was placed, when the first spoof order of 5 contracts was placed, when the first contract of the genuine order was executed and when the first spoof order was cancelled.

Table 3.12 shows the results of the Welch's *t*-tests used to test whether liquidity costs were significantly different before, during and after the spoof. For all different time windows, liquidity costs were higher before the spoof than during the spoof, meaning that liquidity increased during the spoof. Similarly, liquidity costs were lower during the spoof than after the spoof for all time windows. In other words, liquidity was better during the spoof than after the spoof. Moreover, when comparing the liquidity costs before and after the spoof, liquidity was better before than after the spoof, as the liquidity costs after the spoof were higher than before.

3.4.6 Liquidity as a Motivation for Spoofing

In previous sections, we proposed an alternative explanation for the use of spoofing, namely attracting liquidity rather than moving the price. Table 3.13 summarizes for each spoofing example identified by the CFTC, whether our results correspond to the motivation of attracting more liquidity. The second column of Table 3.13, "Genuine order: placed on first level",

Table 3.12 | Mean bid liquidity costs (bps) around the Platinum July 2016 spoof for different time windows

Time window	Before vs. during	During vs. after	Before vs. after
Spoof duration	5.969 > 4.419***	4.419 < 6.763***	5.969 < 6.763***
10 seconds	6.045 > 4.419***	4.419 < 6.829***	6.045 < 6.829***
30 seconds	6.402 > 4.419***	4.419 < 7.024***	6.402 < 7.024***
1 minute	6.948 > 4.419***	4.419 < 8.896***	6.948 < 8.896***
5 minutes	9.225 > 4.419***	4.419 < 10.543***	9.225 < 10.543***

Note: The table reports the mean liquidity costs (bps, measured by APM) around the spoof in the Platinum July 2016 market for different periods and various time windows. *Before* represents the time up until the spoof order was added; *during* the period from when the spoof order was added until it was cancelled; and *after* the time following the cancellation of the spoof order. Five different time windows are used, the *Spoof duration* time window being 4.76 seconds. A lower APM indicates that liquidity costs are low and, hence, liquidity is high. Welch's *t*-tests were used to test for mean differences between the periods. Significance at the 0.1%, 1% and 5% (two-tailed) levels is indicated by ***, ** and *, respectively.

Table 3.13 | Liquidity as motivation for spoofing for each JPM spoofing example

Spoofing example	Genuine order: placed on first level	Genuine order: price identical to last traded price	Increase of liquidity after the spoof
Traditional spoofing			
10-Year T-Note December 2011	Yes	Yes	No
10-Year T-Note March 2010	Yes	No	No
Traditional spoofing with iceberg orders			
Ultra T-Bond September 2015	Yes	Yes	Yes
Silver March 2014	Yes	No	No
Layered spoofing			
Silver March 2012	Yes	No	No
Silver May 2014	Yes	Yes	Yes
Gold April 2014	Yes	No	Yes
T-Bond September 2009	No	No	Yes
Layered spoofing with iceberg orders			
Platinum July 2016	Yes	No	No

Note: This table reports for each spoofing example three indicators for the motivation to use spoofing to attract liquidity. "Yes" ("No") indicates that results conform (do not conform) to attracting liquidity. *Genuine order: placed on first level* indicates if the genuine order is placed on the first level. *Genuine order: price identical to last traded price* indicates if the price of the genuine order was identical to the last traded price. *Increase of liquidity after the spoof* shows if liquidity immediately after the spoof (*Spoof duration*) was better than before the spoof.

corresponds to the situation in which JPM seeks to attract more liquidity by placing the genuine order on the first bid or ask level – as, otherwise, they would have placed the genuine order deeper in the LOB and would have used the spoof to push the price through the first level(s) and hence get a better price than before. The third column of Table 3.13, "Genuine order: price identical to last traded price", conforms to the situation when the price of the genuine order is identical to the last traded price. The fourth column of Table 3.13, "Increase of liquidity after the spoof", shows whether liquidity is better immediately after the spoof than before the spoof. We use the "Spoof duration" time window to determine this for each spoofing example.

Table 3.13 shows that there are cases in which attracting liquidity seems to be the motivation for spoofing. The Ultra T-Bond September 2015 and Silver May 2014 spoofing examples have all indicators point towards attracting liquidity as the motivation behind the spoof. In these cases, JPM was successful at attracting more liquidity: even after the spoof orders were cancelled, liquidity was higher after than before the spoof. Hence, JPM may have spoofed the market to keep prices stable and bait more traders into trading against their preferred price. In the other spoofing examples, one or two indicators confirm the motivation of attracting liquidity, i.e., there is no spoofing example with all spoofing indicators being "No".

3.5 CONCLUSION

This research delved deeply into the JPM spoofing case and visualized their spoofing strategies from different angles. Using messages as its primary component, rather than time-based snapshots, a novel visualization methodology was used from particle physics to identify the JPM spoofing cases. This methodology allows researchers to study high-frequency data at a particular point in time (in our case, the time window of the spoofing), while also placing this data in the perspective of the market environment, i.e., the entire LOB and related variables such as trades, bid and ask volumes, cancelled volume and liquidity. In other words, the message-based approach allows for the simultaneous visualization of activities in the LOB as well as surrounding activities, (re)actions and market output (e.g., price changes, liquidity). The time axis can be dynamically compressed or inflated to show the full details of the spoof, while leaving ample space for the state of the LOB before and after the spoofing activities. This visualization method 1) shows how well-hidden spoofing can be; 2) provides insights in the complexity of the techniques required to recognize spoofing; and 3) puts a value on the minuscule price changes that make spoofing economically viable. We analyze the JPM spoofing examples as identified by the CFTC in detail with numerous characteristics and, in some cases, propose an alternative explanation of why JPM spoofed the market. Rather than move the market to their benefit by inducing short-term price trends, their intention may sometimes have been to attract liquidity, so as to buy or sell numerous futures contracts without having to bear the financial consequences of an illiquid market (i.e., incur costs for trading in a less-than-perfectly liquid market). These visualizations offer a glimpse of the patterns, techniques, time scales, and motivations of the spoofer, thus yielding invaluable information for fraud detection. Messages are visualized in a unique way and help to retrieve more retrospective information about patterns in the LOB at the time when a trader spoofed the market. Reconstructing and visualizing the LOB is key to detecting spoofing, since raw data presents an incomplete overview that does not show orders or changes in the market in relation to its context. Environmental and contextual variables are needed to understand order and market behavior as a whole. However, the data and visualizations alone are not sufficient to identify (new types of) spoofing.

Gained spoofing insights and the visualizations have implications for all stakeholders. Both academics and industry participants gain a better understanding of various types of spoofing and how the market behaves during spoofing. New insights into the motives of market manipulation will help academics to model market behavior in, for example, agent-based modelling. The provided visualization demonstrates how high-frequency LOB data can be effectively visualized and why message-based visualizations contain more information than time-based visualizations. Both academics and industry participants can use these visualizations and adjust them to any variable of interest. Regulators and exchanges gain a different perspective on spoofing as they can now observe all market activity, rather than have to resort to aggregated market activity. Moreover, the visualizations can enhance and refine surveillance programs.

The visualization approach in this research may encourage and inspire future researchers to use more diverse LOB visualization methodologies. Future research might focus on which types of spoofing can be visualized and which go undetected. Moreover, large portions of trading in equity markets are nowadays driven by algorithms. Future research could examine how visualizations may help to control potential spoofing activities by algorithmic trading. Also, the proposed visualization allows for an alternative explanation of spoofing as a means to attract liquidity. We did not know the true intentions of JPM and can only speculate on their intentions. To further examine the motivation of spoofing, further research can focus on in-depth interviews; behavioral and experimental studies to identify the set of motivations for spoofing; and the relationship between spoofing and liquidity costs. Furthermore, this research may motivate future research into the development of theoretical frameworks that can help us to better understand anomalies and market manipulation in financial markets. Finally, the use of iceberg orders in spoofing may trigger a debate about the visibility of orders to regulators and market participants. Future research may have to address whether the use of iceberg orders is fair, whether these orders facilitate manipulative practices, such as spoofing, and whether it still makes sense to allow them in a modern trading environment with algorithmic traders. The message-based visualizations proposed in this research may contribute to this debate.

Chapter 4

Spoofing in U.S. futures markets: an interdisciplinary approach

Marjolein E. Verhulst and Joost M.E. Pennings

ABSTRACT

Spoofing is a manipulative practice that can severely harm the functioning of markets. Economically, the characteristics and impact of spoofing on markets are scarcely studied. Legally, the spoofing statute is broad and complex. This paper delineates all aspects of spoofing from both an economic and legal perspective by providing a comprehensive overview of spoofing types, legislation, literature, rulings and a conceptual framework of spoofing dimensions and attributes. This framework is used to analyze 204 U.S. spoofing cases in futures markets and highlights the nuances of spoofing. Although the empirical focus of this paper is on U.S. futures markets, the conceptual framework is generic to other markets, trading mechanisms, jurisdictions and it can be adapted to include other market-manipulation types.

Keywords: spoofing, illegitimate behavior, law, economics, market regulation, market manipulation

4.1 INTRODUCTION

"Spoofing is illegal—pure and simple,"

– CFTC Chairman Heath P. Tarbert (CFTC, 2020a)

Market manipulation in futures markets – and financial markets in general – has evolved since trading shifted from physical trading pits to electronic trading platforms. Whereas traders could always associate a person with a certain market activity in the open outcry, trading on electronic platforms is anonymous (MacKenzie, 2022). Electronic platforms allow for new market participants, such as algorithms, to enter the market and trading functionalities to be automated. In continuous trading markets, traders are brought together in a centralized marketplace: the limit order book (LOB). The LOB displays all prices traders are willing to buy and sell for, including their respective volumes. However, the LOB represents an incomplete display of market activity as it only displays visible submitted orders to the market (Dalko et al., 2020). For example, executed orders, cancelled orders, modified orders, order types, the number of participants behind aggregated order volume and the hidden volume from iceberg orders are not directly observable in the LOB. The changes caused by electronic platforms, as well as their design, among others, have impacted market manipulation in that 1) the incomplete display of the LOB can be used to the manipulators' advantage (Dalko et al., 2020); 2) market participants do not know who is behind an action, potentially making it harder to differentiate legitimate from illegitimate trading (MacKenzie, 2022); and 3) manipulative actions can now be automated and executed within (fractions of) seconds (Lin, 2017). Conversely, however, the audit trail makes it easier to identify manipulation on electronic platforms than in open outcry, thus deterring participants from manipulation attempts (IMS Group, personal communication, February 13, 2023).

One type of market manipulation that has received significant attention over the past years is spoofing. Although already illegal before 2010, spoofing in derivative markets was specifically prohibited by the Dodd-Frank Act of 2010, where it is defined as: "*bidding or offering with the intent to cancel the bid or offer before execution*" (United States, 2010, p.1739).[30] Spoofers use these bid or ask orders with the intention to improve the terms of their intended transaction (Fox et al., 2021), for example a better price or faster execution. There are different types of spoofing, but in general, spoofers take advantage of the incomplete LOB display and behavioral biases such as herding (Dalko et al., 2020). Spoofing can induce more

30 European law uses a narrower definition of spoofing by including that large orders are submitted to execute a trade on the opposite side of the large order. It is defined as follows in European law in the Commission Delegated Regulation (EU) 2016/522 under Annex II Section 1(5)(e): *"Submitting multiple or large orders to trade often away from the touch on one side of the order book in order to execute a trade on the other side of the order book. Once the trade has taken place, the orders with no intention to be executed shall be removed — usually known as layering and spoofing. […]"* (European Commission, 2015).

price volatility (Dalko et al., 2020) or attract more market participants to the market (Debie et al., 2022). Although a common misconception, spoofing is not only carried out by algorithms and high-frequency traders in an automated fashion but, as this paper demonstrates, is often also performed manually. While the focus of this paper is on futures markets, spoofing can occur in any market with a limit order book. Given the frequency of enforcement actions against spoofing and the magnitude of their reported gains, spoofing is believed to occur frequently enough to cause concern (Fox et al., 2021).

Spoofing – and market manipulation in general – can severely harm the functioning of markets. As futures markets are a zero-sum game, any profit made by the spoofer means a loss for the counterparty. Fast traders such as high-frequency traders are believed to bear these costs, since spoofing often takes little time (Fox et al., 2021). Spoofing can create artificial prices, making futures markets less efficient since prices no longer reflect the true value of the instrument (Canellos et al., 2016; MacKenzie, 2022). In turn, the price-discovery functionality of futures markets – and consequently their underlying asset – suffers as future prices cannot accurately be predicted (Coppler, 2015). However, Fox et al. (2021) argue that the direct effect of spoofing on price accuracy is short-lived and not an important direct consequence of spoofing. They argue that it can have an indirect effect on long-run price accuracy and that all market participants face the indirect consequences of frequent spoofing in the market. Market participants respond to spoof orders just as they would to the arrival of good or bad news. Hence, it is hard to distinguish spoofing from informed trading as these signals get muddied. As a result, liquidity suppliers may experience adverse selection losses and, in defense, offer wider spreads. These wider spreads lead to increased liquidity costs, which can potentially result in less market participants finding it profitable to supply liquidity and reduced trading activity (Fox et al., 2021). Moreover, when markets are not perceived as fair, it affects the confidence of traders in the integrity (Coppler, 2015; Fox et al., 2021; Sanders, 2016) as well as the liquidity and efficiency of these markets (Coppler, 2015; Fox et al., 2021; Mark, 2019; Sahu, 2022) and it may reduce participation in these markets (Fox et al., 2021). Spoofing can also induce additional price volatility (Lee et al., 2013), which can theoretically affect margin requirements and lead to more margin calls (Park & Abruzzo, 2016). This is problematic for market participants, such as hedgers, who might not be able to meet these increased margins. Hence, spoofing affects the integrity, efficiency and functioning of markets and, in the worst-case scenario, may harm one of the main purposes of futures markets: risk management. Hence, spoofing can have serious consequences to the real economy.

Regulators and exchanges attempt to protect the markets and their integrity by sanctioning and fining spoofing offenders. While spoofing in derivative markets is specifically mentioned and prohibited by the Dodd-Frank Act, it has not been defined for securities markets, where spoofing is characterized as a manipulative practice (Canellos et al., 2016).[31,32] This difference in wording means that different evidence is required for the same practice; whereas derivative regulators have to prove the *intent to cancel* an order before execution, securities regulators have to prove *manipulation* (Canellos et al., 2016). However, in more recent years, spoofing in securities markets has also been prosecuted using other theories. For example, the U.S. Securities and Exchange Commission has prosecuted spoofing under 'fraudulent interstate transactions'[33] (e.g., SEC, 2021), and the U.S. Department of Justice has prosecuted spoofing under 'wire fraud statutes'[34] (e.g., DOJ, 2022). The division between derivative and security-market law can be confusing for market participants, especially when dealing with security futures products (narrow-based security indices), which are regulated both as securities and futures contracts (FINRA, 2022; Sanders, 2016). Moreover, as will be discussed later, it has been argued that the spoofing statute is broad and complex (Coppler, 2015).

The goal of this paper is to gain a better understanding of the economic and legal aspects of spoofing in markets with an empirical focus on U.S. futures markets. The economic literature on spoofing generally makes no distinction between various types of spoofing, its intended impact on the market and the goal of spoofing. Also, spoofing is defined broadly in legislation. We delineate all aspects of spoofing by providing a comprehensive overview of the various types of spoofing, spoofing legislation, academic literature on spoofing, spoofing strategy elements, existing spoofing cases and a conceptual framework for all stakeholders involved to study spoofing. Specifically, the conceptual framework encompasses dimensions and attributes that help define the concept of spoofing: the legal responses to spoofing as well as characteristics of spoofing behavior. It was designed by using spoofing legislation, academic literature and expert knowledge from the *International Expert Group on Market Surveillance*[35] (IMS Group). Although this paper focuses on U.S. futures markets,

31 Contrary to U.S. law, the European Market Abuse Regulation (MAR) and Delegated Regulation (EU) 2016/522 apply both to European derivative and security markets (IMS Group, personal communication, March 3, 2023).

32 Specifically, Section 10(b) of the Securities Exchange Act of 1934, which states, *"To use or employ, in connection with the purchase or sale of any security registered on a national securities exchange or any security not so registered, or any securities-based swap agreement any manipulative or deceptive device or contrivance in contravention of such rules and regulations as the Commission may prescribe as necessary or appropriate in the public interest or for the protection of investors."* (U.S. Congress, 1934).

33 Specifically, Section 17(a)(1): *"to employ any device, scheme, or artifice to defraud"*; and Section 17(a)(3): *"to engage in any transaction, practice, or course of business which operates or would operate as a fraud or deceit upon the purchaser"* in the Securities Act of 1933 (U.S. Congress, 1933).

34 Wire fraud in the 18 U.S. Code §1343 is defined as *"Whoever, having devised or intending to devise any scheme or artifice to defraud, or for obtaining money or property by means of false or fraudulent pretenses, representations, or promises, transmits or causes to be transmitted by means of wire, radio, or television communication in interstate or foreign commerce, any writings, signs, signals, pictures, or sounds for the purpose of executing such scheme or artifice, shall be fined under this title or imprisoned not more than 20 years, or both. If the violation occurs in relation to, or involving any benefit authorized, transported, transmitted, transferred, disbursed, or paid in connection with, a presidentially declared major disaster or emergency (as those terms are defined in section 102 of the Robert T. Stafford Disaster Relief and Emergency Assistance Act (42 U.S.C. 5122)), or affects a financial institution, such person shall be fined not more than $1,000,000 or imprisoned not more than 30 years, or both."* (U.S. Code, 2023b).

35 The *International Expert Group on Market Surveillance* consists of the CFTC, CME Group, ICE Futures Europe, European Securities and Markets Authority (ESMA), European Union Agency for the Cooperation of Energy Regulators (ACER), European

the methodology used in designing the conceptual framework is generic to other markets – e.g., spot and stock markets – and other jurisdictions such as European markets. Moreover, the conceptual framework is not limited to spoofing; the methodology can be extended and the dimensions and attributes can be modified to include other market-manipulation types. Note, however, that the conceptual framework is not a legal framework, nor does this paper provide or contribute to a new legal framework that can be used in various jurisdictions. This paper uses the conceptual framework to analyze 204 U.S. orders and disciplinary notices. We are the first to investigate the nuances of and differences between various spoofing strategies from an economic and legal perspective, in an attempt to make spoofing better understood. Important to note is that we do not conduct a legal analysis. This paper lies on the intersection of law and economics, capitalizing on the economic and legal aspects of spoofing.

The disciplines of law and economics interact as the law drives economic behavior, which in turn drives the law. The laws on spoofing are broadly defined and complex (Coppler, 2015) and do not provide precise definitions, criteria or metrics to identify spoofing (Pennings et al., 2020). This makes it difficult to differentiate between legitimate and illegitimate behavior on a legal basis (Sar, 2017). Regulators do not know what to identify as spoofing or which (combinations of) attributes and levels to set (Pennings et al., 2020). Moreover, spoofing evolves over time and increases in complexity, making detection a challenging task (Mark, 2019). This paper bridges law, economics and practice by being the first to provide a conceptual framework consisting of a set of spoofing dimensions. Regulators can use this conceptual framework to facilitate a systematic discussion and turn the spoofing dimensions into spoofing metrics, which can be used, for example, in manipulation-detection algorithms. In turn, these spoofing metrics can assist the interaction between economics and law as they can be used to refine (the interpretation of) the legal framework (Pennings et al., 2020). This affects the economic significance and the application of the law, for example by leading to rulings, the effect of the law on the market (behavior) and the economic impact, such as market damages. The first step is a conceptual framework, which this paper provides.

The literature and research on spoofing and its impact on futures markets is scarce (Mendonça & De Genaro, 2020). Few papers empirically study known spoofing orders and disciplinary notices, hereinafter referred to as "spoofing cases", and no coherent overview of public spoofing cases exists. Putniņš (2020) calls for a more comprehensive dataset of manipulation cases to overcome the limitations of existing empirical studies. This research answers this call and provides the first detailed overview of spoofing cases in U.S. futures

Energy Exchange (EEX), Eurex Deutschland, Frankfurt Stock Exchange (FSE), Euronext Amsterdam, Netherlands Authority for the Financial Markets (AFM), Netherlands Authority for Consumers and Markets (ACM), Swiss Financial Market Supervisory Authority (FINMA), SIX Group, Commissione Nazionale per le Società e la Borsa (CONSOB) and Borsa Italiana.

markets since the Dodd-Frank Act took effect. As such, it is the first paper that provides a complete inventory of public spoofing cases and discusses their dimensions and attributes by introducing a novel conceptual framework. A total of 204 spoofing cases by the Commodity Futures Trading Commission (CFTC), Chicago Mercantile Exchange (CME Group) and the Intercontinental Exchange Futures U.S. (ICE) are summarized and discussed. All detailed spoofing examples outlined in these spoofing cases are provided in Appendix 4.B to make them more accessible and easier to study.

The remainder of the paper is organized as follows. Section 4.2 discusses related literature on spoofing in futures markets and background information. Section 4.3 outlines the conceptual framework, followed by the research design in section 4.4. Results are presented in section 4.5 and section 4.6 concludes with a discussion.

4.2 RELATED LITERATURE AND BACKGROUND

4.2.1 Literature on Spoofing in Futures Markets

Several theoretical works seek to understand what encourages and discourages spoofing, and whether they can differentiate spoofing from legitimate high-frequency trading (HFT) activity. Martínez-Miranda et al. (2016) simulated spoofing and pinging behavior within a reinforcement learning framework, to discover what encourages traders to engage in these manipulative activities. They find that, in bullish markets, spoofing is an optimal investment strategy for traders trying to maximize investment growth, but it can be discouraged by regulators through counteractive mechanisms such as fines. Using agent-based simulations and game-theoretic analyses, X. Wang et al. (2018) studied various market environments to establish whether a cloaking mechanism could mitigate spoofing. Symmetrical cloaking conceals orders for a specific number of price levels. They find that the cloaking mechanism can significantly mitigate spoofing in certain market environments and that the mechanism is not easily circumvented. Yang et al. (2012) used a multi-agent approach to simulate the E-mini S&P 500 futures market and an inverse reinforcement learning algorithm to distinguish legitimate HFT activity from HFT spoofing strategies. Their algorithm identifies HFT spoofing strategies with an accuracy of at least 92%. Li and Yang (2022) used an agent-based model to study the impact of spoofing on other market participants. Their simulation consisted of different types of market participants, including fundamentalists, chartists, zero-intelligence and spoofing agents. They find that spoofing increases price volatility, prolonging the price-discovery process. Moreover, fundamentalists lose money during the spoofing but can profit during the price-recovery process, while chartists lose money both when the spoofing agent profits and when the price-recovery process starts.

One qualitative study on spoofing was conducted by MacKenzie (2022). As part of a broader study, he interviewed 337 financial-market participants, four lawyers directly involved in spoofing cases, surveillance specialists, regulators and exchange staff on the topic of spoofing. He finds, among others, that 1) spoofing already existed in trading pits and was "considered to be good brokerage"; 2) high-frequency traders have different views on the illegitimacy of spoofing: some find it "self-evidently illegitimate" while others consider it "normal" trading behavior; 3) spoofing is not unitary, nor unambiguously defined and 4) there is no drawing of the boundary regarding spoofing as law and morality interact and definitions are not black and white.

Empirical studies of known spoofing cases are limited. Debie et al. (2022) conducted an in-depth examination of the spoofing by JPMorgan Chase & Co., JPMorgan Chase Bank, N.A., and J.P. Morgan Securities LLC (hereinafter together referred to as "JPMorgan"). JPMorgan spoofed in futures markets between 2008 and 2016 and settled with the CFTC for a record-breaking $920.2 million in 2020. Using a method developed in particle physics, spoofing examples from the CFTC were visualized, including various market-manipulation indicators to better understand the market behavior surrounding spoofing. They find, among other things, that JPMorgan spoofed not only to move the price, but also to attract more liquidity. Unrelated to U.S. futures markets, Leonard, Cao, Haas, and Mocek (2020) discussed and visualized two UK spoofing cases: 1) the spoofing by Michael Coscia in the Brent Futures market at the ICE Futures Europe exchange; and 2) the spoofing by Da Vinci Invest Ltd. involving the Aquarius Platinum Ltd. stock at the London Stock Exchange.

4.2.2 Rules Prohibiting Spoofing

The CFTC uses the spoofing definition from the Dodd-Frank Act to prosecute spoofing.[36] The Dodd-Frank Act introduced the definition of spoofing in Section 747, and is an amending act to the U.S. Commodity Exchange Act (CEA). It is currently in force in Section 6c(a)(5) of the CEA (U.S. Code, 2023a). The remainder of this paper will refer to the spoofing-related sections in the Dodd-Frank Act and CEA as "the Dodd-Frank Act". CME Group and ICE incorporated the definition in their own rulebook under respectively rules 575 and 4.02, as outlined in Table 4.1. Unlike the Dodd-Frank Act, CME Group and ICE not only prohibit the cancellation of an order before execution, they also ban the modification of orders to avoid execution (Sar, 2017).

Despite the fact that the Dodd-Frank Act specifically defines spoofing, many academics and industry stakeholders argue that the definition is too broad (e.g., Coppler, 2015; Mac-

36 Since it is unclear which rules the SEC uses to prosecute spoofing in security futures markets (SEC, personal communication, September 15, 2022), this paper focuses solely on spoofing orders at the CFTC, CME Group and ICE.

Table 4.1 | Rules prohibiting spoofing in U.S. futures markets

Source	Rule and description prohibiting spoofing
Dodd-Frank Act (United States, 2010)	**SEC. 747. ANTIDISRUPTIVE PRACTICES AUTHORITY.** […] "(5) DISRUPTIVE PRACTICES.—It shall be unlawful for any person to engage in any trading, practice, or conduct on or subject to the rules of a registered entity that— [...] ''(C) is, is of the character of, or is commonly known to the trade as, 'spoofing' (bidding or offering with the intent to cancel the bid or offer before execution).
CME Group Rulebook (CBOT, 2022; CME, 2022; NYMEX, 2022)	**575. DISRUPTIVE PRACTICES PROHIBITED** All orders must be entered for the purpose of executing bona fide transactions. Additionally, all nonactionable messages must be entered in good faith for legitimate purposes. No person shall enter or cause to be entered an order with the intent, at the time of order entry, to cancel the order before execution or to modify the order to avoid execution; [...]
ICE Rulebook (ICE Futures U.S., 2022)	**Rule 4.02 Trade Practice Violations** [...] (l) Engage in any other manipulative or disruptive trading practices prohibited by the Act or by the Commission pursuant to Commission regulation, including, but not limited to: (1) Entering an order or market message, or cause an order or market message to be entered, with: (A) The intent to cancel the order before execution, or modify the order to avoid execution;

Note: This table outlines the rules and descriptions prohibiting spoofing in the Dodd-Frank Act and the rulebooks of CME Group and ICE. Only relevant sections of the rules are included.

Kenzie, 2022; and Sar, 2017). In particular, Coppler (2015) argues that the anti-spoofing statute is impermissibly vague, and thereby calling it a statute that *"fails to provide a person of ordinary intelligence fair notice of what is prohibited, or is so standardless that it authorizes or encourages seriously discriminatory enforcement"* (United States v. Williams, 2008, p.2). Statutes can be deemed impermissibly vague, for example, when they can be interpreted to encompass other acts, when they fail to provide a standard for enforcement or when terms are not defined in any applicable regulations or interpretive releases (Coppler, 2015). Under the present anti-spoofing statute of the Dodd-Frank Act, legitimate trading may constitute spoofing, making it difficult to distinguish legitimate from illegitimate trading (Sar, 2017) and potentially exposing market participants to prosecution using laws that do not provide a fair warning (Coppler, 2015). For example, HFT market makers can be seen as spoofers as they place several orders at multiple exchanges and cancel all of these orders as soon as the first order is executed (Leuchtkafer, 2015). Even compliance officers within the same company may find it difficult to distinguish legitimate from illegitimate trading. A case in point is the CFTC order against the Bank of Nova Scotia (CFTC, 2020d), where spoofing was brought to the attention of two compliance officers. One compliance officer said it was *"pretty obvious"* and problematic, while another did not find it problematic and stated *"[W]hat is being seen* *may look like potential layering or spoofing, but based on the fact [that] we are talking [about] 1 lots, we believe he is just adjusting his exposure to the marketplace."* (CFTC, 2020d). Determining legality can be challenging, especially when individual actions are legal but become problematic when they are repeated in a specific pattern or under certain conditions. For instance, a trader may place an order on one side of the market and, due to a sudden change in information, cancel that order to place it on the opposite side. While this action in itself is perfectly legal, repeating and mirroring this pattern of actions can make it illegal (Fox et al., 2021).

Specifically, Coppler (2015) argues that the Dodd-Frank Act is problematic for several reasons. First, the statute states that it is unlawful to engage in any activity that is "of the character of, or is commonly known to the trade as, 'spoofing'". However, it does not explicitly state what constitutes "of the character of" spoofing; hence, it might include other, legitimate, trading and leaves enforcers with broad discretion. Second, without a clear demarcation of what constitutes spoofing and what does not, the definition of spoofing remains complex and, as such, not defined exactly, leaving regulators freedom to interpret the term. The definition also fails to specify *when* the intent to cancel is required, leaving room for ambiguous situations when a trader changes their mind about their intent to execute an order. Third, spoofing is mainly proven through circumstantial evidence regarding trading patterns, which may result in discriminatory enforcement. HFTers, for example, cancel 95-98% of their orders as part of their strategy, which may be mistaken for spoofing. Hence, trading patterns alone cannot distinguish spoofing from normal trading activities (Coppler, 2015). However, more recently, U.S. courts have ruled that the anti-spoofing statute in the Dodd-Frank Act is not unconstitutionally vague. See, for example, *United States v. Coscia*, 866 F.3d 782 (7th Cir. 2017). In this case, the court ruled that 1) *"a statute is not vague simply because it requires law enforcement to exercise some degree of judgment. Bell*, 697 F.3d at 462*"* (United States v. Coscia, 2017, p.11); and 2) the spoofing definition in the Dodd-Frank Act *"imposes clear restrictions on whom a prosecutor can charge with spoofing; prosecutors can charge only a person whom they believe a jury will find possessed the requisite specific intent to cancel orders at the time they were placed."* (United States v. Coscia, 2017, p.11).

4.2.3 Spoofing Strategies

As discussed in section 4.2.2, many practices can fall under the definition of spoofing. The literature is inconsistent in documenting spoofing as several elements are mixed up, e.g., the effect of spoofing on the market and the goal of the spoofer. Hence, it is important to make several distinctions when discussing spoofing strategies. We propose the three spoofing elements in Figure 4.1 to disentangle spoofing strategies, clear any confusion and facilitate discussions on spoofing.

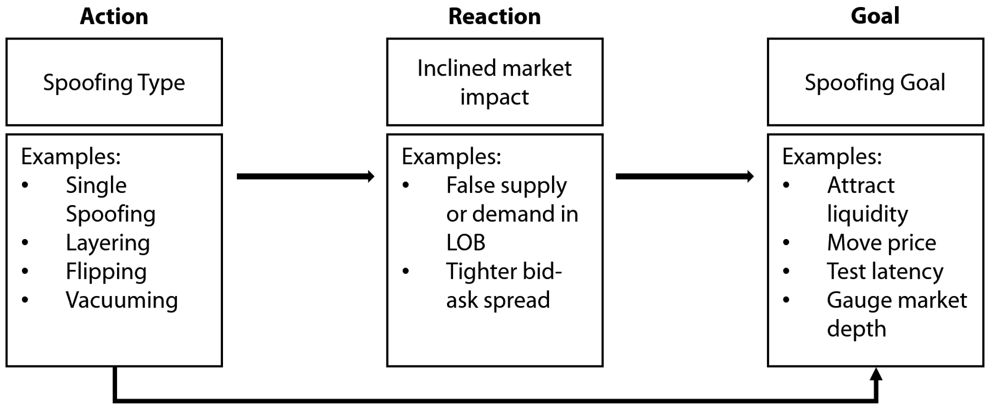


Figure 4.1 | Spoofing strategy elements. This figure shows the various elements of a spoofing strategy. *Action* is the spoofing type used; *reaction* is the spoofer's desired impact on the market; and *goal* is the objective the spoofer wishes to achieve.

We distinguish between action, reaction and goal. The *action* is the type of spoofing used, for example layering, flipping or vacuuming – all of which are discussed in detail in Appendix 4.A. *Reaction* is the spoofer's desired impact on the market, for example achieving a tighter bid-ask spread or creating a false sense of supply or demand in the LOB. The *goal* is the objective the spoofer wishes to achieve, for example a price movement or increased liquidity. Some spoofing strategies include all elements (i.e., the top two arrows): action, reaction and goal. For example when a spoofer uses spread-squeeze spoofing to tighten the bid-ask spread in the market, in order to buy or sell at a better price. Other spoofing strategies may only use the action and goal elements (i.e., the bottom arrow) and do not require an impact on the market. This is the case, for example, when a spoofer uses flash spoofing to test latency or gauge the market depth at certain LOB levels. The latter strategies are also often referred to as pinging. Note that spoofing strategies are often repeated by switching sides in the LOB. For example, a trader spoofs on the ask side to buy contracts at a lower price and then reverses their strategy by spoofing the bid side to sell their recently bought contracts at a higher price, and vice versa.

The distinction between 'action', 'reaction' and 'goal' is important and necessary. Lumping different spoofing strategies together can have serious consequences to statistical significances and conclusions drawn in literature. While some conclusions may be true for a specific spoofing type, market reaction or spoofing goal, this does not have to be true for all spoofing strategies, and this is overlooked when all spoofing strategies are combined. Distinguishing the various elements can help tailor economic tools for identifying wrongdoing and better understand what impact certain spoofing strategies have on the market. This, in turn, can help identifying the most harmful types of spoofing strategies and allow regulators to be better equipped to target certain spoofing strategies over others.

The focus of this paper is on spoofing strategies that use all elements in Figure 4.1, i.e., the action, reaction and goal element. Spoofing strategies that do not seek to impact the market, i.e., that only include the action and goal elements in Figure 4.1 (bottom arrow), are not considered. An overview of various spoofing types – i.e., the "action" element in Figure 4.1– including examples is provided in Appendix 4.A.

4.3 CONCEPTUAL FRAMEWORK

To analyze the spoofing cases, a conceptual framework was designed by means of an iterative process as outlined in Figure 4.2. The attributes that describe and help define the concept of spoofing were extracted from the literature and the spoofing cases by the CFTC, CME Group and ICE, until no new attributes could be identified and the list was saturated. This list was converted into a conceptual framework, which was tested by the industry. Specifically, the *International Expert Group on Market Surveillance* provided feedback to finalize the conceptual framework, presented in Table 4.2. All spoofing cases were analyzed on the dimensions and attributes of this conceptual framework. Researchers and regulators can use this conceptual framework to have a systematic discussion about spoofing from an economic and market perspective as well as from a legal and enforcement perspective.

The conceptual framework in Table 4.2 encompasses nine dimensions which resulted from the iterative process: case information, affected market, spoofing: general, genuine order, spoof order, genuine vs. spoof order, market impact, monetary action and data. Some of the dimensions relate directly to the nature of the spoofing (e.g., 'genuine vs. spoof order') while other dimensions relate to the legal document (e.g., 'case information' and 'monetary action'). Hence, the conceptual framework can be used to summarize legal responses to spoofing

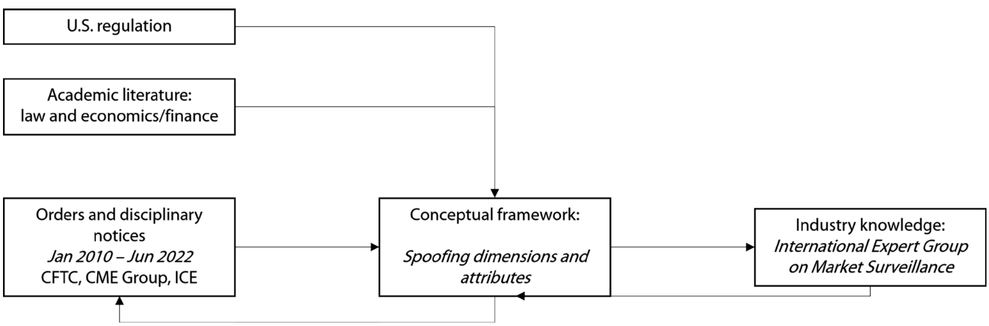


Figure 4.2 | Research process to design the conceptual framework. This figure shows the research process that was used to design the conceptual framework. The conceptual framework consists of spoofing dimensions and attributes extracted from U.S. regulation; academic literature; orders and disciplinary notices; and industry knowledge.

behavior as well as summarize and characterize spoofing behavior itself. Researchers and industry stakeholders can select dimensions and attributes depending on their objectives, as using all 80 attributes might be cumbersome. Note that the action, reaction and goal elements from Figure 4.1 are included in the conceptual framework, whereby 'spoofing type' (attribute 27), for example, refers to the action, the 'market impact' dimension (attributes 65 to 76) refers to the reaction and the 'spoofing goal' (attribute 31) refers to the goal.

Ultimately, the complete conceptual model should be considered to present a full picture of spoofing. Dimensions and attributes should not be studied in isolation but as connected to one another and, if possible, to a trader's individual trading patterns. For example, a large order cancellation, by itself, might not be a good indicator of spoofing, but combined with other factors – such as trades on the other side of the cancellation – extreme order patterns can provide more insights (Sar, 2017).

Table 4.2 | Conceptual framework to analyze spoofing cases

Attribute		Description	Hypothetical examples
Case Information			
1.	Enforcer	Name of exchange or regulatory agency	Regulator A
2.	Applicable legal framework	Which legal framework is used	Dodd-Frank Act
3.	Case number	Identification number of document	01-2345
4.	Publication year document	Year the document was published	2016
5.	Defendants	Name(s) of person and/or company	Jane Doe
6.	Company type	Type of company of the defendant	Processor
7.	Relevant period/days	The period in which the defendant(s) spoofed	Mar 1, 2018 – Sep 25, 2020
Affected Market			
8.	Exchange	At which exchange is the affected financial instrument traded	Exchange B
9.	Financial instrument*	Type of financial instrument	Futures
10.	Market category	General market	Equity
11.	Underlying asset/commodity	Underlying asset/commodity	Corn
12.	Delivery contract (month and year)	Month and year when the affected contract matures	March 2013
13.	Market/trading mechanism	The market mechanism used for trading	Central limit order book
14.	Matching algorithm	The matching process to fill orders	First-in-first-out
15.	Contract unit	Quantity of the underlying commodity/asset for a single contract	112,000 pounds
16.	Price quotation	Currency in which the contract is traded	U.S. cents per pound
17.	Tick size/minimum price change	The minimum possible price fluctuation	$10.00

Table 4.2 | Continued

Attribute	Description	Hypothetical examples
18. Available delivery months	All available delivery months of the contract	March, June and September
19. Settlement method	How the contract is settled at expiration	Cash
20. Trading hours	When the market is open for trading	Sunday – Friday: 5:00 p.m. – 4:00 p.m. CT
21. Position limit	Maximum number of contracts a trading entity can hold	60,000 contracts
22. Price limit	The type of price limit in place and its boundaries	Circuit breaker
23. Other market specifics	Other market characteristics specific to the affected market that may affect spoofing behavior	Speed bump
Spoofing: General		
24. Number of times spoofed	The total number of times the defendant(s) spoofed during the relevant period	1500
25.1 Algorithmic or manual	Was the spoofing conducted manually or algorithmically? If algorithmically, what type of algorithm?	Algorithmically; Deep Learning
25.2 Automated features	If manual spoofing, did the defendant(s) use any automated features/systems	Yes, spreading genuine orders
26.1 Order splitter	Was an order-splitter functionality used	Yes
26.2 Order splitter for genuine or spoof orders	Was the order splitter used to place genuine or spoof orders	Genuine Orders
27. Spoofing type	What kind of spoofing was conducted	Layered
28. Number of spoof layers	If multiple layers were used to spoof, how many layers were used	4-6
29. Reverse spoofing strategy	Was the spoofing strategy repeated multiple times by switching sides in the LOB	Yes
30. Net position after spoofing within one day	What was the spoofer's net position after a trading day	Long position of 100 contracts
31. Spoofing goal	The objective the spoofer wishes to achieve	Trade larger quantities
32. Individual or coordinated	Did the defendant spoof individually or in coordination with others	Coordinated
33.1 Spoofing in the same or correlated market	Were the genuine and spoof orders placed in the same market, or were the genuine orders placed in one market and the spoof orders in a correlated market	Correlated
33.2 Type of correlated market	If correlated market, which market	Maturity Month
34. During day or night	Did the defendants spoof during day or night trading sessions	Night

Table 4.2 | Continued

Attribute	Description	Hypothetical examples
35. During settlement period	Did the spoofing occur during the settlement period or closing price procedure	Yes
36. Multiple accounts or account IDs	Did the defendant use multiple accounts to disguise their spoofing	Yes
37. Spoofing evidence	How was "the intent to cancel an order before execution" demonstrated	Chat messages
Genuine Order		
38. Total number of genuine orders	The total number of genuine orders placed during the relevant period	2000
39. Market or limit order	Were the genuine orders market or limit orders	Limit
40. Type of market/limit order	The order type that is used for the genuine order	Fill or kill
41. Size (volume)	How much volume did a single genuine order contain on average	1 contract
42. At limit order book level(s)	At which level(s) were the genuine orders placed	2
43. Iceberg order	Were the genuine orders iceberg orders	Sometimes
44. Placement before or after spoof order	Were the genuine orders placed before or after the spoof orders	Before
Spoof Order		
45. Total number of spoof orders	The total number of spoof orders placed during the relevant period	3200
46. Type of order	The order type that is used for the spoof order	All or none
47. Flashing spoof orders	Were spoof orders flashed	Sometimes
48. Size (volume)	How much volume did a single spoof order contain on average	50-100
49. Ratio volume spoof orders vs. affected price level(s) (%)	How much volume (%) did the spoof orders account for at the affected price levels	Spoof orders accounted for 60% of the volume on the affected price levels
50. Value of spoof orders ($)	What was the average value ($) of the spoof orders	$5,000,000
51. At limit order book level(s)	At which price level(s) were the spoof orders placed	2-4
52. Mean volume increase at spoofing levels (%)	How much did volume increase at the level(s) the spoof orders were added	200%
53. Iceberg order	Was the iceberg functionality used for spoof orders or for other large orders	Only for large non-spoof orders
54. (Partial) cancellation of total spoof orders (%)	What percentage of spoof orders resulted in a (partial) cancellation	99%
55. Cancellation time (seconds) after placement	How fast were spoof orders cancelled (seconds) after	10 seconds

Table 4.2 | Continued

Attribute		Description	Hypothetical examples
56.	Cancellation time (seconds) after genuine order	How fast were spoof orders cancelled after execution of the genuine order	2 seconds
57.	Cancellation order	In what order were the layered spoof orders cancelled	Descending bid price levels
58.	Average number of modifications per spoof order**	How many times was a spoof order modified on average	20
59.	Hit rate large orders vs. spoof orders (%)	How many spoof orders versus large orders were executed (%)	Large Orders: 80% Spoof Orders: 1%
Genuine vs. Spoof Order			
60.	Iceberg order	Was the iceberg functionality used for genuine, spoof or large non-spoof orders	Only for genuine and large non-spoof orders
61.	Hit/fill rate (%)	How many genuine and spoof orders were executed (%)	40% genuine vs. 1% spoof
62.	Cancellation rate and time (seconds)	How many genuine and spoof orders were cancelled and their cancellation time	10 genuine orders, average 3 min; and 300 spoof orders, average 10 seconds
63.	Ratio number of spoof orders vs. genuine orders	How many spoof orders were placed compared to genuine orders	8:1
64.	Ratio volume spoof orders vs. genuine orders	How large were the spoof orders compared to the genuine orders	80:1
Market Impact			
65.	Market losses ($)	How much did the market lose ($) because of spoofing	$1,000,000
66.	Qualitative market damages	How was the market damaged in terms of qualitative indicators, such as trust and integrity	Qualitative analysis
67.	Profit defendant(s)	How much did the defendant(s) gain ($) because of spoofing	$7,000,000
68.	Limit order book balance or imbalance due to spoofing	Did the spoofing create a limit order book balance or imbalance	Imbalance
69.	Bid-ask volume ratio before, during and after the spoofing	How did the bid-ask volume ratio of the top 10 levels change before, during and after the spoofing	Before: 100:110 During: 100:200 After: 90:100
70.	Market effect in behavior	How did traders change their behavior due to the spoofing	More cancellations
71.	Liquidity	How did liquidity change before, during and after spoofing	Quantitative analysis
72.	Market quality	How did market quality change before, during and after spoofing	Quantitative analysis
73.	(Price) volatility	How did volatility change before, during and after spoofing	Quantitative analysis
74.	Market activity	How did market activity change before, during and after spoofing	Quantitative analysis
75.	Information content	How did the information content of the limit order book change before, during and after spoofing	Quantitative analysis
76.	Price trend	How was the price trend before, during and after the spoofing	Upwards, downwards, downwards

Table 4.2 | Continued

Attribute	Description	Hypothetical examples
Monetary Action		
77. Monetary penalty	How high was the monetary penalty imposed on the defendant(s)	$200,000
78. Profit disgorgement	How much profit did the defendant(s) have to disgorge	$50,000
Data		
79. Detailed example	Are detailed examples available including dates, timestamps and contract specifics	Yes, for November 12, 2013
80. Provided market data	Is (raw) market data provided	Yes, limit order book tables

Note: This table details attributes that help define the concept of spoofing, extracted from academic literature and spoofing cases by the CFTC, CME Group and ICE. The attributes are categorized into nine dimensions in a conceptual framework: 1) *Case Information* describes general information about the spoofing case; 2) *Affected Market* consists of attributes describing the market targeted by the spoofing; 3) *Spoofing: General* details general attributes about the spoofing strategy; 4) *Genuine Order* contains attributes characterizing the genuine order placed by the spoofer; 5) *Spoof Order* consists of attributes describing the spoof orders; 6) *Genuine vs. Spoof Order* compares the attributes of the genuine and spoof orders; 7) *Market Impact* details attributes that describe the response of the market around the execution of the spoofing strategy; 8) *Monetary Action* describes the penalty and profit disgorgements defendants had to pay; and 9) *Data* details whether the spoofing case included detailed examples and/or market data.
* Since this paper focuses on spoofing in futures markets, this attribute is called "financial instrument". However, spoofing can also occur in, for example, spot products that are not considered to be financial instruments. Hence, this attribute should be renamed when using the conceptual framework with other assets.

4.4 RESEARCH DESIGN

The conceptual framework from section 4.3 was used to analyze U.S. spoofing cases, to delineate and gain a better understanding of all aspects of spoofing. Publicly available spoofing cases were collected from the websites of the CFTC, CME Group and ICE.[37]

Spoofing-related orders are published as an attachment to press releases on the CFTC's website. The keywords "spoofing", "layering", "squeeze", "vacuum" and "manipulation" were used separately to search the CFTC's press releases. The results of each search were checked manually to see whether they related to spoofing. This search resulted in 62 cases by the CFTC. Note that all spoofing orders at the CFTC were used, regardless of their legal consequences. Moreover, dates in this paper refer to when the order was published by the CFTC, not to when legal action was taken, e.g., a settlement.

Two robustness checks were performed for the CFTC spoofing cases. An additional automated search of all press releases was conducted on the CFTC's website, since their website

37 Specifically the following URLs: https://www.cftc.gov/PressRoom/PressReleases, https://www.cmegroup.com/tools-information/advisorySearch.html and https://www.ice.com/futures-us/notices.

stores press releases in a uniform format with sequential identifiers in the URL. This makes it possible to download all press releases quickly and rigorously. A Python script was used to download all 2852 press releases since January 2010 – the year the Dodd-Frank Act came into effect – which were checked on the following keywords: "spoof"; "layer"; "flip"; "squeez"; "manipulat"; and "vacuum". These keywords were matched using the lowercase base form of the words in both the title and text of the press release, e.g., "spoof" would match "Spoofing" and "spoofer". This search resulted in 264 unique press releases, which were reevaluated and cross-referenced with the manual search. Three additional documents were identified from an already discovered spoofing case: the non-prosecution agreements of Jeremy Lao, Daniel Liao and Shlomo Salant from the Citigroup Global Markets spoofing case in 2017. Hence, the original manual search captured 95% of the robustness check, and a final total of 62 spoofing cases were identified at the CFTC. In addition, Lexis Uni and WestlawNext were used to examine spoofing cases by the CFTC and this did not result in additional cases being identified. Based on these numbers, we concluded a successful search of all documents and expected a similar hit rate for the websites of CME Group and ICE.

Spoofing-related disciplinary notices were searched for on the CME Group website using the keyword "575", as spoofing falls under CME Group's rule 575. This rule has been effective from September 15, 2014 (CME Group, 2014). Spoofing practices before this date were prohibited *"under other Exchange rules, including, but not limited to, Rules 432.T. ("to engage in dishonorable or uncommercial conduct"), 432.B.2. ("to engage in conduct or proceedings inconsistent with just and equitable principles of trade"), and 432.Q. ("to commit an act which is detrimental to the interest or welfare of the Exchange or to engage in any conduct which tends to impair the dignity or good name of the Exchange")"* (CME Group, personal communication, April 29, 2022). Since it is beyond the scope of this paper to investigate all disciplinary actions on spoofing-related contents under these rules, only spoofing cases by CME Group after September 15, 2014 are included. A total of 115 spoofing cases were collected from CME Group.

Disciplinary notices from ICE are available from 2014 on; no notices are publicly available from before 2014. No keywords were used to search for spoofing-related notices as the number of notices was limited. Subsequently, all notices were checked manually for a relationship with spoofing and, if any, included and used for analysis. This resulted in 27 spoofing cases by ICE.

We adhered to the following procedure to examine spoofing cases. First, we identified the relevant section of a ruling for this research. These sections are the summary, introduction and facts for CFTC cases; the exchange rules, findings and penalty for CME Group cases; and the exchange rules, summary, products and penalty for ICE cases. Other sections were not included in this paper as they do not concern the traders' actions and actual events that occurred. Second, spoofing cases were not included if the description was ambiguous or if it was unclear whether they related to spoofing. Third, the contents of the remaining spoofing cases were analyzed and all attributes present in the conceptual framework were extracted. A detailed overview per spoofing case of all these attributes can be found in online Appendix 4.C. For analysis purposes, spoofing cases were in some cases merged to avoid double counting, for example when multiple spoofing cases documented the same spoofing or had the same case number.

Table 4.3 summarizes the analyzed spoofing documents. A total of 204 spoofing orders and disciplinary notices were collected: 62 (30.4%) orders by the CFTC, 115 (56.4%) disciplinary actions by CME Group and 27 (13.2%) disciplinary actions by ICE. Figure 4.3 shows the year in which these 204 spoofing cases were published. Note that this does not mean the spoofing took place in that year, nor that any sanctions were necessarily imposed that year. A significant increase in spoofing cases followed after 2015. On the one hand, this can be attributed to the data-collection process as CME Group added a specific rule prohibiting spoofing from mid-2014 on. On the other hand, it can be attributed to the fact that the first high-frequency

Table 4.3 | Descriptive statistics of the spoofing cases by the CFTC, CME Group and ICE, from January 2010 to June 2022

	CFTC	CME Group	ICE
Date first spoofing case	December 12, 2012	September 22, 2014	March 11, 2016
Date last spoofing case	April 14, 2022	April 22, 2022	June 29, 2022
Number of spoofing docs	62	115	27

Note: Date first spoofing case shows the first identified spoofing case by the respective regulator/exchange after the Dodd-Frank Act came into effect. *Date last spoofing case* is the last spoofing case that was added for analysis before July 1, 2022. *Number of spoofing docs* is the total number of orders/disciplinary notices found. Note that spoofing cases by CME Group only include cases brought under Rule 575, which was effective after September 15, 2014. Before this date, spoofing was brought under multiple rules, which were excluded from this analysis.

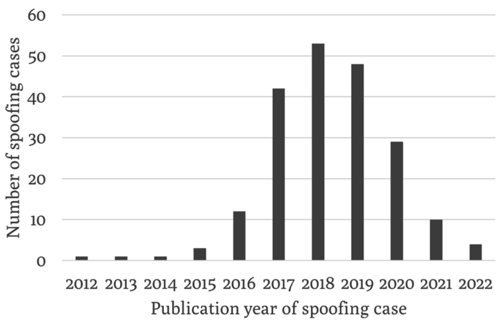


Figure 4.3 | Publication year of spoofing cases by the CFTC, CME Group and ICE from January 2010 to June 2022. This figure shows the number of spoofing cases in futures markets published each year by the CFTC, CME Group and ICE since the introduction of the Dodd-Frank Act.

trader – Michael Coscia – was convicted of spoofing by a jury in 2015, which clarified the definition of spoofing for the industry (Polansek, 2015). The decreasing number of spoofing cases after 2018 may be caused by the fact that it may take several years to detect and identify spoofing and then take legal action. For example, JPMorgan spoofed between 2008 and 2016, but the CFTC order was published four to twelve years later, in 2020.

4.5 RESULTS: ANALYZED SPOOFING CASES USING THE CONCEPTUAL FRAMEWORK

Spoofing cases will be discussed based on the conceptual framework dimensions outlined in section 4.3: Case Information, Affected Market, Spoofing: General, Genuine Order, Spoof Order, Genuine vs. Spoof Order, Market Impact, Monetary Action and Data. Not all attributes of these dimensions will be discussed as not all attributes were mentioned in the cases. Moreover, not all spoofing cases contain the same information, and, to avoid double counting, not all results subsections total to 204 spoofing cases. Note that no in-text spoofing-related citations are made, other than naming the defendants. All spoofing cases and collected data can be found in the online Appendix 4.C. In addition, in-depth discussions will mainly concern CFTC spoofing cases as these provided the most details.

4.5.1 Case Information

Defendants. The 204 identified spoofing cases involve a total of 134 individuals and 42 companies, including banks, brokers, clearing firms, hedge funds, investment firms, trading firms, wholesalers, distributors and affiliates of refineries. Some of the brokers involved placed spoof orders on behalf of their customers without their authorization (e.g., ICE Futures U.S., 2021a).

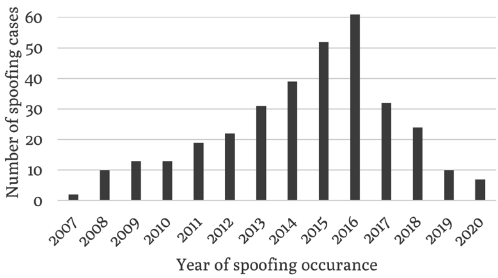
Relevant Period. The relevant period describes the time window when the defendants of the spoofing cases executed their spoofing strategies. Spoofing strategies in the spoofing cases were executed from 2007 to 2020. As Figure 4.4 shows, the year 2007 saw the lowest amount of identified spoofing, with only two spoofing cases, while 2016 saw the highest amount, with 61 spoofing cases. The decrease in identified spoofing after 2016 may be caused by the fact that investigations take time, meaning there is a natural lag, or previous sanctions may have sent a proper message to rogue traders (IMS Group, personal communication, March 6, 2023).

4.5.2 Affected Market

Market Category. Six distinctive market categories were affected by spoofing between 2007 and 2020: the metals, agricultural, energy, equity-index, interest-rate and FX markets.

Table 4.4 shows that the metals market appeared the most in the spoofing cases, followed by the agriculture, energy, equity-index, interest-rate and FX markets.

Underlying commodity. Table 4.5 divides the market categories into specific underlying commodity markets. Panel A shows that, in agricultural markets, most spoofing occurred in the soybean, wheat and live-cattle futures markets. Most of the spoofing cases in the energy markets (Panel B) occurred in the crude-oil, natural-gas and RBOB gasoline futures markets. Panel C shows that, in the metals market, gold, silver and copper showed the most cases of spoofing. Only one spoofing case mentioned the FX market, specifically the Nikkei/Yen futures market (Panel D). The E-mini S&P 500 appeared the most in the equity market (Panel E), followed by the E-mini Nasdaq-100 and the E-mini Dow ($5) futures markets. The interest-rate market (Panel F) showed the most spoofing cases in the Eurodollar futures market, followed by the 5- and 10-Year T-Note, and the 2- and 30-Year T-Note. Note that the summation of the number of spoofing cases per commodity market in Table 4.5 may differ



* This figure uses all the spoofing cases by the CFTC, CME Group and ICE between January 2010 and June 2022.

Figure 4.4 | Years during which spoofing occurred in the spoofing cases* by the CFTC, CME Group and ICE. This figure shows the time window in years when the defendants executed their spoofing strategies, as specified in the spoofing cases by the CFTC, CME Group and ICE. The first incidence of spoofing in this time frame was in 2007.

Table 4.4 | Market category affected by spoofing from spoofing cases by the CFTC, CME Group and ICE, from January 2010 to June 2022

	Number of spoofing cases
Metals	125
Agriculture	80
Energy	48
Equity index	40
Interest rates	15
FX	2

Note: This table shows the market categories and the number of times these markets were identified as being affected by spoofing in the spoofing cases by the CFTC, CME Group and ICE.

Table 4.5 | Futures markets affected by spoofing from spoofing cases by the CFTC, CME Group and ICE, from January 2010 to June 2022

Product	# Spoofing Cases	Product	# Spoofing Cases	Product	# Spoofing Cases
A. Agriculture		***B. Energy***		***E. Equity Index***	
Canola	3	Brent Crude	1	CBOE Volatility Index	2
Class III Milk	1	Crude Oil	21	CAC40 Index	1
Cocoa	5	Heating Oil	3	E-mini Dow ($5)	4
Coffee C	4	Henry Hub Natural Gas	2	E-mini Nasdaq-100	9
Corn	3	Henry LD1 Fixed Price	1	E-mini Russell 2000 Index	1
Cotton No. 2	6	Natural Gas	10	E-mini S&P 500	17
Feeder Cattle	2	NY Harbor ULSD	3	E-mini S&P MidCap 400	1
Lean Hog	4	RBOB Gasoline	7	MSCI Emerging Markets Index	1
Live Cattle	7	**Total**	**48**	Russell 2000 Index	1
Lumber	1	***C. Metals***		Russell 2000 Index Mini	1
Oats	1	Copper	17	**Total**	**38**
Rough Rice	1	Gold	42	***F. Interest Rates***	
Soybean	15	Palladium	13	2-Year T-Note	1
Soybean Meal	4	Platinum	14	5-Year T-Note	2
Soybean Oil	5	Silver	36	10-Year T-Note	2
Sugar No. 11	5	**Total**	**122**	30-Year T-Note	1
Wheat	11	***D. FX***		Eurodollar	3
Total	**78**	Nikkei/Yen	1	**Total**	**9**

Note: The table shows the futures markets, ordered by market category, and the number of times these markets were identified as being affected by spoofing, in the spoofing cases by the CFTC, CME Group and ICE.

from Table 4.4 as some cases only mention the market category, without stating the specific underlying market.

4.5.3 Spoofing: General

Algorithmic vs. Manual Spoofing. Some researchers associate spoofing with high-frequency trading (e.g., Li & Yang, 2022; and Mark, 2019) as it is usually executed within seconds. However, of the 34 spoofing cases that mentioned how the spoof orders were placed, 31 were placed manually, two used algorithms and one case used both manual and algorithmic strategies.

Manual traders did sometimes use automated features to make it easier to execute their spoofing strategy. Andre Flotron used an automated trading tool he controlled that placed additional genuine orders conditional on prior genuine orders being filled. Kevin Crepeau used an automated spread feature to enter small genuine orders. Krishna Mohan and Kamaldeep Gandhi used a similar feature, the order splitter, to place multiple randomly sized genuine and spoof orders. These features made it easier to disguise spoofing from market participants (CFTC, 2018h).

Spoofers also exploited features from trading front ends. For example, Krishna Mohan used the proprietary trading system of the firm he worked at – SuperGUI – which provided tools that decreased the amount of time and the number of mouse clicks needed to execute strategies. SuperGUI made it possible, among others, for traders to 1) personalize default settings, such as the default visible quantity for iceberg orders; 2) pre-set the left and right mouse buttons to a specific order size; 3) input orders by clicking price levels on a trading ladder screen; 4) use an order splitter and 5) automate repetitive tasks using pre-programmed hotkeys (CFTC, 2018d). Moreover, Igor Oystacher exploited the avoid-orders-that-cross functionality in a commercially available trading platform to execute a flipping strategy. This functionality prevents orders from the same trader to match with each other and is used to prevent another manipulation form called "wash trading". For illustration purposes, assume that Trader A has a buy order resting at $10. If Trader A were to enter a market order to sell for $10, this order would normally execute against their own buy order at $10. However, the avoid-orders-that-cross function prevents this from happening and automatically cancels the resting buy order for $10 when Trader A adds a market order to sell for $10. Hence, using this functionality, traders can execute flipping strategies by placing a large spoof order on one side of the market and adding a genuine order on the opposite side of the market for the same price as the spoof order. Using this functionality, Igor Oystacher was able to automatically and immediately cancel his resting spoof orders by entering an aggressive genuine order at a single push of the button (CFTC, 2015a).

Algorithmic spoofing strategies were used by Andrei Sakharov and Navinder Sarao. Andrei Sakharov's algorithm placed a single large spoof order on the opposite side of a smaller genuine order. Once the genuine order was executed, the algorithm would cancel the spoof order. Navinder Sarao used an algorithm developed by Jitesh Thakkar and Edge Financial Technologies Inc. for his layering spoofing strategy. The algorithm would simultaneously place four to six spoof orders on separate levels, starting at least three to four levels from the best price level. As the market moved, the algorithm would simultaneously move the spoof orders to keep them three to four levels away from the best price. This ensured that the spoof orders remained visible to market participants while also minimizing the risk of execution. One cycle of placing, modifying and canceling orders through the layering algorithm lasted less than six minutes. The algorithm used several functions such as the ability to 1) cancel an order when the market gets close and change the definition of "close"; 2) enter multiple orders at different price levels in one click; 3) remove an order if there is not a certain number of orders beneath it; 4) automatically modify resting orders up and down by one lot each time volume is added to the same price level; and 5) remove an order immediately if it is being partially executed. The latter two functionalities were used to ensure that the orders by Navinder Sarao remained at the back of the queue – i.e., behind the resting orders of the other market participants – thus decreasing the chances of his orders getting executed whilst making them appear genuine.

Table 4.6 | Identified spoofing types from spoofing cases by the CFTC, CME Group and ICE, from January 2010 to June 2022

	Number of spoofing cases
Single spoofing	63
Layered spoofing	94
Flipping	4
Spread squeezing	3
Vacuuming	1

Note: This table shows the number of times spoofing types were mentioned or described in the spoofing cases by the CFTC, CME Group and ICE.

Spoofing Type. Table 4.6 shows that the majority of spoofing cases that mention a spoofing type referred to layered and single spoofing. Besides using spoofing to create a false appearance of market depth, Anuj Singhal also used his layered spoof orders to convey a false sense of increased volatility, by intentionally and repeatedly modifying his orders up and down the LOB. In some layered spoofing cases, the placement and cancellation order of the layers were discussed. Spoofers placed the layers from lower to higher bid levels (or higher to lower ask levels) and would cancel them in reverse, i.e., the highest bid (lowest ask) would be cancelled first since these layers had a high risk of execution (e.g., CFTC, 2019a and 2020d; and Debie et al., 2022).

Although the focus of this paper is on spoofing that uses all elements – action, reaction, goal – from Figure 4.1, two spoofing cases mention "flash spoofing", which does not necessarily encompass the "reaction" element. Not much is known about flash spoofing since it is only defined in the CFTC order against Navinder Sarao: *"... Defendants "flashed" large lot orders in a variety of lot sizes in the Order Book that were quickly canceled with no intention of these orders resulting in trades (Flash Spoofing)."* (CFTC, 2015b, p.3). Contrary to other spoofing types, genuine orders do not seem to be used in flash spoofing. Rather, spoof orders are rapidly added and cancelled – "flashed" – making speed key. Flash spoofing was used by Navinder Sarao as a separate strategy and in combination with his layering algorithm to amplify the impact of the layering algorithm. He used three distinctive volumes for his flash spoofing orders: 118-lots and 289-lots (the "118/289-lot spoofing"), and 2000-lots (the "2000-lot spoofing"). Typically, two or three of the 118/289-lot spoof orders were placed at the same price level on the ask side, two or three ticks from the best ask price. This created extreme momentary ask-side imbalances. During the relevant period[38], Navinder Sarao placed 1728 of these spoof orders with an approximate notional value of $26.5 billion. About 80% of the 118/289-lot

38 The relevant period in this case was from April 2010 to January 2012, July 2012 to June 2014 and September 2014 to April 2015.

spoof orders were placed while the layering algorithm was active. He flashed the 2000-lot orders as a separate strategy and typically placed these at the best bid or ask level. It is notable that, although "flash spoofing" involves rapid placement and cancellation of spoof orders, Navinder Sarao executed this spoofing strategy manually. Additionally, one trader from Mizuho Bank Ltd. flashed spoof orders of 500 or 700 contracts to test the reaction of the market, as the trader wanted to hedge at a future time.

Individual vs. Coordinated Spoofing. Thirteen out of 166 spoofing cases mention traders who coordinated their spoofing actions: 1) Stephen Gola and Jonathan Brims, who worked for Citigroup Global Markets Inc.; 2) Heet Khara and Nasim Salim; 3) James Vorley, Cedric Chanu and David Liew, who worked at different locations for Deutsche Bank AG; 4) Krishna Lakhani and Shiv Agarwal, who worked for Arab Global Commodities DMCC; 5) traders at Merrill Lynch Commodities Inc.; 7) James Harkness and Steven Loulousis; and 8) Wang Jian and Wang Yiwu. Traders of the same firm could see the resting orders of their colleagues, whether these colleagues were in the same office or in a different location. Typically, one trader would place genuine orders while another would place spoof orders if the genuine order was not executed fast enough. Although unrelated to spoofing, David Liew was the only trader who also coordinated his market manipulation with a trader at a different firm, so as to trigger stop-loss orders from other market participants.

Spoofing in the Same vs. in a Correlated Market. Spoofing does not have to be limited to a single market. For example, spoofers can place genuine orders in one market (Market A) while placing spoof orders in a correlated market (Market B). Consequently, the spoofer aims to entice market participants in Market A to move in the direction of the genuine order by creating a false impression of market depth in Market B. Sixty-two spoofing cases mention whether the spoofing occurred in the same or in a correlated market. The majority of these spoofing cases – 52 cases – took place in the same market, while ten cases took place in multiple, correlated markets.

Traders who spoofed in correlated markets not only did so in correlated futures markets at the same exchange (see CME Group, 2017a), but also at different exchanges and in different countries. For example, Michael Franko took advantage of the short-term correlation between the copper prices at COMEX and LME. He placed spoof orders in the COMEX market to affect genuine orders in the LME market and vice versa. Others took advantage of the correlation between different maturity contracts or calendar spreads (e.g., CFTC 2020b; and CME Group, 2019). Moreover, correlated markets were not only limited to correlated futures markets, but also involved correlated cash and options markets. Stephen Gola and Jonathan Brims from Citigroup Global Markets Inc. coordinated their spoofing in correlated treasury futures and cash markets. Besides spoofing in the same futures market, David Skudder also employed a spoofing strategy in the soybean futures and options markets. He placed a genuine order in the soybean options market and spoof orders on the opposite side in the soybean futures market. In this strategy, there was no requirement for the spoof orders to create an imbalance between the futures and options markets. He created the imbalance only in the futures market, to put pressure in the direction of his genuine order in the options market.

Spoofing During U.S. Day vs. Night Trading Session. Spoofing outside of regular day-trading hours can have advantages and make the spoofing strategy easier to execute. Contrary to active daytime sessions, trading volume and volatility are significantly lower during night-trading sessions (CFTC, 2018d, 2020c). Hence, spoof orders have a larger impact on the bid-ask balance and are more likely to have the desired impact. They can be active for longer without execution and smaller spoof orders can be used, which translates into smaller financial risk (CFTC, 2018d, 2020c). Seven spoofing cases mention the spoofing taking place at non-regular trading hours. James Harkness, Jiongsheng Zhao, Krishna Mohan, Steven Loulousis and traders from Deutsche Bank Securities Inc. spoofed during overnight hours. Traders from Arab Global Commodities DMCC spoofed after business hours, and Hin Pang Kou also spoofed during Asian and European trading hours. In some cases, trading in night-trading sessions provided proof that these traders had the intention to cancel their (spoof) orders before execution. For example, Krishna Mohan executed almost 95% of his spoofing strategy during overnight sessions. The difference in large order placements during daytime and overnight sessions also demonstrated the wrongful intent of Jiongsheng Zhao, who placed 98% of his large orders (≥50 contracts, containing spoof and non-spoof orders) during overnight sessions and only 2% during daytime sessions. Smaller orders (<50 contracts) were placed 71% of the time during overnight sessions and 29% during daytime sessions. In their order, the CFTC argued that, had Jiongsheng Zhao wanted to execute his spoof orders, he would have placed them during daytime sessions when they were more likely to execute, but he purposefully avoided placing large orders during daytime sessions.

Spoofing Using Multiple Accounts. Denis Pospelov, Ge Shuai, Vincent Ngo and traders from Arab Global Commodities DMCC and Belvedere Trading LLC used different accounts or operator IDs (Tag50) in an attempt to hide their spoofing. Eric Moncada used different accounts in the name of BES Capital LLC and Serdika LLC, not for spoofing purposes but for wash trading.

Spoofing Evidence. The intent to cancel an order before execution was demonstrated based on the traders' actions that were visible in the data and supporting evidence. Intent was demonstrated in at least seven different ways using hard data. First, spoofing patterns were predictable and executed numerous times, independent of market conditions, which demonstrated that it was a predetermined strategy not dependent on market changes (CFTC, 2015a, 2018d, 2018e, 2020b). Second, the number of executed genuine and executed spoof orders – the fill rate – demonstrates a different execution intent between these orders. As will be discussed in more detail below, on average, the fill rate was above 35% for genuine orders, compared to under 1% for spoof orders (CFTC, 2015a, 2018d, 2022). Genuine orders from Jiongsheng Zhao, for example, were 180 times more likely to get executed than his spoof orders, demonstrating he successfully avoided the execution of spoof orders. David Skudder's spoof orders were also compared to similarly sized genuine orders. While his spoof orders had a fill rate of less than 1%, the fill rate for his similarly sized genuine orders was 10.6%. From a legal point of view, the fill rate alone is not strong enough an argument: it must be placed in the context of the traders' strategies. The fill rate needs to be combined with trading strategies that are meant to keep the fill rate low for spoof orders (IMS Group, personal communication, March 5, 2023), such as entering spoof orders behind existing orders and/or on price levels further away from the best levels, fast modifications and cancellations of the spoof orders and significant volume increases on the affected price levels due to the spoof orders. Third, Krishna Mohan's intention only to execute genuine orders was demonstrated because he had placed more than 450 genuine orders at the exact same price level and side as the spoof order within one second after the spoof order was cancelled. Fourth, as discussed above, executing the spoofing strategy during night sessions for over 95% of the time was indicative of wrongful intent on the part of Krishna Mohan and Jiongsheng Zhao. Fifth, traders used various tactics to conceal their manipulative activity from the market. For example, iceberg orders were often used for genuine orders to conceal traders' true interest and, in combination with spoof orders, cause a significant imbalance (CFTC, 2018d). Moreover, as mentioned above, traders used multiple accounts or tools to place many randomly sized orders, making it seem as if multiple traders were adding orders. Sixth, spoof orders and genuine orders had a different cancellation rate and cancellation time. As will be discussed below, spoof orders resulted in (partial) cancellation for more than 95% of the time, while genuine orders were rarely cancelled. Spoof orders were cancelled fast to protect them from execution, and these consistently fast cancellation times reflect the intent to cancel from the start (CFTC, 2018e, 2020b). Seventh, some spoofers used a limited number of large order sizes outside their spoofing strategies. For example, Jiongsheng Zhao almost always used large orders for his spoofing strategy and not for legitimate trading, and Roman Banoczay Jr. rarely used large orders in the four years before he began spoofing.

Supporting evidence such as internal and external chats, phone calls, e-mails and web meetings were used in several spoofing cases to demonstrate spoofing and the intent to cancel orders before execution (CFTC, 2015b, 2017b, 2018b, 2018f, 2018g, 2018c, 2019b, 2020e). For example, a Merrill Lynch Commodities Inc. trader wrote in a chat message on November

16, 2010 [sic]: *"guys the algos are really geared up in here. [I]f you spoof this it really moves .... "* (CFTC, 2019b, p.3). A UBS trader wrote on April 30, 2010 [sic]: *"u gotta be quick with spoofs cause everyone else knows the trick too ... except for smaller shops ... and algos of course."* (CFTC, 2018g, p.4). Krishna Mohan had several documents on his Google Drive describing spoofing strategies. For example, one document entitled "Smart Stuffing book.(order cancel replace)" contains the following instructions [sic]:

> *"allow trader to control which side they want to show bluff. trader decide each level how many contracts should be put in. and contracts should be break into multiple smaller contracts in random size which overall average order size match with the entire books average size. we can also add 30,40,50 contracts into the book to mimic market makers. never throw easily detected size like 100/200/250 or 60s like 9ner.*
>
> *Only put bluffing orders in when book exceed 400 contract. and flipped for more than 500 ms.*
>
> *every 10 second add 1 lot in the existing order and minus 1 lot to make sure every order of ours will be at the back of the queue. make a butt called (refresh bluffing)*
>
> *All those orders are not mean to be traded. if for any reason some one flush big size and bluff order get filled, order will be sent out immediately to the price it get filled try to scratch out." (CFTC, 2018c, p.15)*

Moreover, the design and development of an algorithm can also prove the intent to cancel an order before execution. For example, the "Back-of-Book" functionality, developed by Jitesh Thakkar for Navinder Sarao, continuously modified orders to remain at the back of the price level queue and automatically cancelled orders once they were partially executed. According to the CFTC order against Jitesh Thakkar, he understood that this functionality would be used to place spoof orders and cancel them before they were executed (CFTC, 2018c, p.9-10).

4.5.4 Genuine Order

Number of Genuine Orders in Relevant Period. Four out of 166 cases included the number of genuine orders placed in the relevant period (CFTC, 2018e, 2018d, 2020b, 2022). The average number of genuine orders per month was calculated for each case by dividing the relevant periods in months by the number of genuine orders placed. This yielded an average of 20 to 1200 genuine orders placed per month. Krishna Mohan placed the most genuine orders, with 2400 orders in two months in the E-mini Dow ($5) and E-mini NASDAQ 100 futures markets, while David Skudder placed the fewest genuine orders, with 1077 orders in 4 years and 7 months in the soybean futures and options markets.

Genuine Order Size. Fifteen out of 168 spoofing cases specify the size of the genuine orders, eleven of which involved a genuine order of one to ten contracts. The remaining spoofing cases mention genuine order sizes of 1-40 contracts, 1-50 contracts and 20-100 contracts. The latter were placed using the iceberg functionality, meant to hide a large proportion of the order from market participants.

Limit vs. Market Order. When specified, most genuine orders in spoofing schemes were placed using a limit order. Specifically, 48 of the 167 spoofing cases mention genuine orders being placed as limit orders. In one spoofing case, aggressive market orders were used for genuine orders as part of the flipping strategy (CFTC, 2019d). Other cases did not specifically state what type of genuine orders were used.

Placement at LOB Level. Six out of seven spoofing cases mention that the spoofers placed the genuine order at or near the best level, i.e., levels one to three. The remaining spoofing case used spread squeezing and placed genuine orders within the bid-ask spread. Hence, they did not place genuine orders at the existing best level, but at a better price.

Iceberg Order. Approximately 15.5% of the 168 spoofing cases mention the use of iceberg orders for genuine orders. Iceberg orders were used to hide the spoofers' true buy/sell interests and to cause a larger imbalance in combination with the spoof orders (CFTC, 2018d). Typically, the visible part of the iceberg order would be one contract, or less than 25% of the actual order's size (e.g., CFTC, 2018a; and CME Group, 2018). Although iceberg orders were frequently used and are useful to hide the spoofer's true interest, they hinder full execution under the "First In, First Out" (FIFO) matching algorithm. Once the visible portion of an iceberg order is executed, the newly visible portion moves to the end of the queue and only gets executed once the existing volume gets executed or removed. Despite spoof orders not having this limitation – since iceberg orders were rarely used for spoof orders – genuine orders were executed much more than spoof orders, as will be discussed below (CFTC, 2018d).

Placement Genuine Order Before/After Spoof Order. As Table 4.7 shows, the majority of genuine orders was placed before the spoof order was placed. In six spoofing cases, however, the traders placed their genuine orders after submitting spoof orders. In five spoofing cases, traders placed genuine orders both before and after the spoof orders.

Table 4.7 | Distribution of genuine orders placed before or after the spoof order from spoofing cases by the CFTC, CME Group and ICE, from January 2010 to June 2022

	Number of spoofing cases (out of 166 cases)
Genuine order before spoof order	36
Genuine order after spoof order	6
Both	5

Note: This table shows the number of spoofing cases by the CFTC, CME Group and ICE that described the genuine order as being placed before, after or before and after the spoof order.

Table 4.8 | Number of times spoofed and total spoof orders in spoofing cases by the CFTC, CME Group and ICE, including relevant period between January 2010 and June 2022

Defendant(s) spoofing case	Relevant period	Number of times spoofed in relevant period	Total number of spoof orders in relevant period	Average number of spoof orders per month
David Skudder, Global AG LLC and Nesvick Trading Group LLC	Sep 2014 - Mar 2019	Futures scheme: 276 Options scheme: 279	Futures scheme: 293 Options scheme: 299	Futures scheme: 5 Options scheme: 5
Igor B. Oystacher and 3 Red Trading LLC	Dec 2011 - Jan 2014	1316	5207	200
Jiongsheng Zhao and Propex Derivatives Pty Ltd.	Jul 2012 - Mar 2017	2300	3100	54
Krishna Mohan	Nov 2013 - Dec 2013	1500	36,300	18,150
Roman Banoczay Jr., Roman Banoczay Sr. and BAZUR Spol. S.R.O.	Jan 2018 - Feb 2018	2000+	19,100	9,550
Citigroup Global Markets Inc., Stephen Gola and Jonathan Brims	Jul 2011 - Dec 2012	NA	2500+	139
Eric Moncada, BES Capital LLC and Serdika LLC	Oct 6, 12, 14, 19, 26, 27, 29 and 30, 2009	NA	710	Average per day: 89

Note: This table shows the spoofing cases by the CFTC, CME Group or ICE that indicated the number of times defendants spoofed over a period of time. The first column lists the name(s) of the defendant(s) in the spoofing case. The second column describes the period in which the defendant(s) spoofed. The third and fourth column show the number of times the defendant(s) spoofed and the total number of spoof orders placed in the relevant period. The last column states the average number of spoof orders per month by dividing the total number of spoof orders by the relevant period in months. The non-prosecution agreements with Jeremy Lao, Daniel Liao and Shlomo Salant included the total number of times spoofed in the relevant period but were not included in this table as the documents did not contain further information. They spoofed 80, 30 and 35 times respectively, from July 16, 2011 to December 31, 2012.

4.5.5 Spoof Order

Number of Times Spoofed and *Number of Spoof Orders in Relevant Period.* Spoofing cases that mention the number of times defendant(s) spoofed in the relevant period and the number of spoof orders they used are summarized in Table 4.8. The cases in Table 4.8 show that multiple spoof orders (layers) were often used when executing spoofing strategies. For example, Krishna Mohan spoofed 1500 times in the relevant period, using a total of 36,300 spoof orders and an average of 24 spoof orders per spoofing occurrence. Roman Banoczay Jr. started spoofing in 2018 after suffering significant losses, that is, more than $1 million in three days and almost all profits of the previous year. In the second half of January 2018, he spoofed between 6 to 61 times per day (an average of 27 times per day). From February 2 to 11, he boosted his spoofing activity, spoofing consistently between 76 to 332 times per day (averaging 200 times per day). On February 12, 2018, he spoofed the market 700 times in a single day.

Size (Volume). Several spoofing cases discuss the absolute spoof-order volume, while other spoofing cases refer to the spoof-order volume relative to the genuine-order volume. Absolute spoof-order volumes ranged from one contract to two thousand contracts. Twelve out of 166 spoofing cases mention the size of the spoof order being between one to 50 contracts. Marc Michelotti used multiple spoof orders of size one, seeking to narrow the bid-ask spread as part of a spread-squeeze spoofing strategy. Traders of Heraeus Metals New York LLC were restricted by a maximum order size of 20 contracts, which is why their spoof orders were of volume 20. Sixteen out of 166 spoofing cases mention spoof-order sizes of between 60 and 700 contracts. Citigroup Global Markets Inc., Stephen Gola and Jonathan Brims used spoof orders with a volume of more than one thousand contracts, and Navinder Sarao frequently used spoof orders of volume 2000. Most spoofing cases mentioning relative volumes, i.e., seven cases, describe the spoof orders as having a five times higher volume than the (visible) genuine orders. Krishna Mohan's spoof orders were nine times larger than his genuine orders and comprised at least 40 more contracts than his genuine orders. The largest relative volume discrepancy mentioned concerned the spoofing case of Deutsche Bank Securities Inc., where the spoof-order's volume was twenty times that of the genuine order.

Placement at LOB Level. Ten out of 166 spoofing cases mention that spoof orders were placed within the first three levels of the LOB. Spoof orders placed between levels four to six were generally part of the layered spoofing strategy (involving three spoofing cases). While these spoof orders were placed behind existing orders – reducing the risk of execution – other spoof orders created new LOB levels. These spoof orders were placed within the bid-ask spread and were used in a spread-squeeze spoofing strategy.

Volume Increase at Spoofed Levels. Spoofing significantly increases the resting volume at the affected levels. The spoof orders by Michael Franko multiplied volume at the affected price level by fifty. Igor Oystacher's spoofing increased individual price-level volume by 1877%, 99%, 1710%, 1696% and 880-996% in respectively the copper, crude oil, natural gas, VIX and E-mini S&P 500 futures markets.

Iceberg Order. Using the iceberg functionality for spoof orders is not obvious. After all, spoof orders are often used to significantly increase buying/selling pressure, meaning that the spoof volume should be visible to the market. Nevertheless, two out of 166 spoofing cases mention the use of iceberg spoof orders. Eric Moncada placed 710 large-lot orders, four of which used the iceberg function. The fact that he refrained from using the iceberg functionality for the other 706 large orders illustrated that he had no intention to execute those four orders. Also, on 17 September, 2014, Kamaldeep Gandhi placed an iceberg spoof order in the December 2014 E-mini S&P 500 futures contract. This spoof order consisted of 250 contracts, with nineteen contracts visible to the market and 231 remaining hidden through the iceberg functionality.

(Partial) Cancellation Rate. The lowest cancellation rate mentioned in a spoofing case concerned the CFTC order against Navinder Sarao: approximately 95% of his flash spoof orders were cancelled. The highest cancellation rate – of 100% – was mentioned in the spoofing case of Jiongsheng Zhao and Propex Derivatives Pty Ltd. The differences between the cancellation rates of spoof and genuine orders will be discussed in more detail in section 4.5.6.

Cancellation Time. In general, spoof orders are cancelled relatively quickly to avoid the risk of getting executed. Ten out of 167 spoofing cases mention the cancellation time of spoof orders. Five of these cases mention the spoof orders being cancelled within two seconds after placement (e.g., Jiongsheng Zhao and Krishna Mohan), and two cases mention a cancellation time of between four to seven seconds. David Skudder used the longest cancellation time in his futures (options) spoofing scheme, cancelling spoof orders within 30 seconds, with a median cancellation time of 10.38 (11.71) seconds. Cancellation times in the aforementioned spoofing cases were independent of whether the genuine order was executed. In other spoofing cases, cancellation times depended on the execution of the genuine order. For example, In-Ho Hwang cancelled his spoof orders within five seconds of the genuine order being executed, and Krishna Mohan cancelled 60% (84%) of his spoof orders at the first level within one second (two seconds) after his last genuine order was executed. Cancellation times also depended on how close the orders were to the top of the LOB. Spoof orders closer to the best bid/ask were cancelled faster (CFTC, 2018d) because they faced a higher risk of execution. Hence, in the event of layered spoofing, spoof orders are cancelled in order from the highest to the lowest levels of the LOB. This can be seen, for example, in the visualization of JPMorgan's spoofing by Debie et al. (2022).

Number of Modifications Per Spoof Order. Spoofers can modify their spoof orders for various reasons. Geneva Trading USA LLC, Roman Banoczay Jr. and David Skudder modified their spoof orders away from the top levels to avoid execution before cancellation. Anuj Singhal repeatedly modified his layered spoof orders up and down the LOB to create a false sense of increased volatility. Navinder Sarao's layering algorithm modified spoof orders thousands of times, at an average of 161 modifications per order – compared to an average of one modification per order for other traders. His algorithm would modify an order each time the market price changed, so as to ensure that his spoof orders stayed at least three ticks away from the best price. His total order modifications represented 60% of the total ask-side order modifications on twelve specific spoofing days. On May 6, 2010 – the day of the 2010 Flash Crash – his layering algorithm performed 19,000 modifications on six orders with underlying values of between $170-200 million before they were cancelled. Note that the modification of spoof orders can also serve to implicitly cancel them: spoof orders can be modified to a price level far away from the best price levels, which does not officially cancel the spoof order but makes its execution nearly impossible (IMS Group, personal communication, March 3, 2023).

4.5.6 Genuine vs. Spoof Order

Differences between the characteristics of genuine and spoof orders are often used as evidence for different intent behind these orders. While, in theory, genuine and spoof orders can be compared on every dimension of the conceptual framework, this subsection compares them on select dimensions.

Hit/Fill Rate. Five out of 166 spoofing cases describe the proportion of genuine orders versus spoof orders executed during spoofing strategies. When David Skudder placed spoof and genuine orders in the same futures market, his genuine orders had a hit rate of 21.1%, while his spoof orders had a hit rate of 0.22%. When he placed genuine orders in the options market and spoof orders in the futures market, the hit ratio of his genuine orders was 7.39%, compared to 0.79% for his spoof orders. Hit rates for Igor Oystacher were different for each market, but on average, the hit rate for his genuine orders was 54.5% versus 0.83% for his spoof orders. Krishna Mohan's genuine orders had a hit rate of 39%, compared to less than 1% for his spoof orders. Roman Banoczay Jr. had a hit rate of 69% for his genuine orders and less than 2% for his spoof orders. Jiongsheng Zhao's genuine orders were more than 180 times more likely to be executed than his spoof orders. All these cases demonstrate that genuine orders were much more likely to get executed, constituting evidence that the intent to execute differed between genuine and spoof orders.

Cancellation Rate. Multiple spoofing cases mention that the cancellation of a genuine order was rare (e.g., Jiongsheng Zhao and Roman Banoczay Jr.). The CFTC order against Igor Oystacher provided a comparison of cancellation rates between spoof and genuine (flip) orders in different markets, as shown in Table 4.9. In all markets, spoof orders were cancelled significantly more often than genuine flip orders, with the lowest spoof-order cancellation rate being 98.13% – in the crude oil futures market – and the highest genuine-order cancellation rate being 62.47% – in the VIX futures market.

Cancellation Time. In the spoofing cases, the cancellation times of spoof orders are compared with those of genuine orders and non-event orders, i.e., (large) orders that are cancelled but are not related to any spoofing strategy. Table 4.10 summarizes these comparisons for the various spoofing cases. In general, spoof orders are active in the LOB for a significantly shorter time than genuine, large-lot and non-event orders. This can – and in some cases did – reveal a different intent behind the spoof orders.

Table 4.9 | Cancellation rates of spoof vs. genuine orders in the Igor Oystacher spoofing case by the CFTC in 2015

	Mean cancellation rate	
Futures market	**Spoof orders**	**Genuine (flip) orders**
Copper	99.11%	55.70%
Crude Oil	98.13%	30.17%
Natural Gas	99.49%	49.24%
VIX	99.06%	62.47%
E-mini S&P 500	99.43 – 99.83%	27.81 – 37.90%

Note: This table shows the mean cancellation rates of spoof orders versus genuine orders in the Igor Oystacher spoofing case by the CFTC (CFTC, 2015a).

Table 4.10 | Median cancellation time for spoof, genuine, large-lot and non-event orders in spoofing cases by the CFTC, from January 2010 to June 2022

	Median cancellation time (sec)			
Spoofing case	**Spoof orders**	**Genuine orders**	**Large-lot orders**	**Non-event orders**
Jiongsheng Zhao and Propex Derivatives Pty Ltd.	0.737	106	NA	NA
Roman Banoczay Jr. & Sr. and BAZUR Spol. S.R.O.	6.23	22.7	48.6	NA
Krishna Mohan: E-mini Dow ($5)	1.7	7.5	NA	8.3
Krishna Mohan: E-mini Nasdaq-100	1.9	7.7	NA	4.6
David Skudder: Futures Scheme	10.38	NA	NA	180
David Skudder: Options Scheme	11.71	NA	NA	180

Note: This table shows the mean cancellation times in seconds for spoof, genuine, large-lot and non-spoofing event ("non-event") orders in spoofing cases by the CFTC.

4.5.7 Market Impact

Market Losses. Four of the 61 CFTC spoofing cases contain information on the financial damage to markets in dollars. Jiongsheng Zhao and Propex Derivatives Pty Ltd. caused market losses of $464,300 during the time they spoofed, representing an average of $8146 per month for the entire spoofing period. Tower Research Capital LLC spoofed for 22 months, causing market damages of $32,593,849, averaging $1,481,539 per month. The Bank of Nova Scotia (CFTC, 2020d) damaged the market by $6.6 million, equaling an average $64.078 per month. Finally, JPMorgan spoofed for nine years, causing market losses of $311,737,008, averaging to $2,886,454 per month. Note that the CFTC's exact method of calculating market losses is not publicly known. Hence, we do not know over which time period the market losses were calculated, nor whether these losses merely concern the damage incurred by the spoofing victim(s) or also include the additional losses suffered by other market participants due to the changes in market conditions.

Profit Defendant(s). Five of the 61 CFTC spoofing cases describe the profits the defendant(s) made using spoofing strategies. Panther Energy Trading LLC and Michael Coscia made a profit of $1.4 million during the three months they spoofed. Heet Khara and Nasim Salim also spoofed for three months, making $200,000 in profits in one of the two accounts they used. Navinder Sarao spoofed for 800 days, from April 2010 to January 2012, and made a total of $40 million, averaging to $1,818,182 per month. JPMorgan profited $172,034,790 (an average of $1,592,915 per month) in the nine years they used their spoofing strategy. During the eight most intense days of spoofing, Roman Banoczay Jr. made $332,000, an average of $41,500 per day. As with the market losses, the CFTC's profit calculation method has not been publicly disclosed, so we do not know whether these are realized profits or, for example, unspent costs from being able to buy cheaper due to spoofing.

LOB Balance or Imbalance. Thirty out of 168 spoofing cases mention that spoofing caused imbalance in the LOB. The spoofing case of James Vorley and Cedric Chanu, for example, describes the market as being slightly imbalanced before the spoof, with 58% more bids than asks. During the spoof, however, the imbalance increased significantly to 207% more bids than asks. Navinder Sarao used his layering algorithm to create an *"extreme momentary sell-side order imbalance"* (CFTC, 2015b, p.18). For example, on May 5, 2010, he placed five spoof orders totaling 2500 contracts with a value of $146.3 million on the ask side. On the day of the 2010 flash crash – May 6, 2010 – his ask spoof orders totaled to 3600 contracts, which almost equaled the entire bid-side volume. Although David Skudder created LOB imbalances in his futures scheme, notably this was not a requirement for his options scheme. Recall that, in his options scheme, David Skudder placed a genuine order in the options market and a spoof order in the futures market. While there was no requirement for the spoof order to create an imbalance between the options and futures markets, the spoof order did create an imbalance in the futures market to put pressure on the options market. It is striking that one spoofing case – by Roman Banoczay Jr. of February 12, 2018 involving the Crude Oil March 2018 contract – describes the LOB as having become both balanced and imbalanced due to spoofing. Excluding the spoof orders by Roman Banoczay Jr., the LOB had 69.6% more ask volume than bid volume, with 290 contracts on the ask side and 171 contracts on the bid side. The spoof orders by Roman Banoczay Jr. eliminated this imbalance, bringing the LOB back in balance with a 290 ask volume and 291 bid volume. Despite its importance in understanding spoofing, this small detail has not been documented anywhere else. It illustrates that spoofing is about injecting new (false) information in the market, not necessarily about causing an imbalance between buy and sell volumes. In other words, in a market that is consistently imbalanced, the sudden balancing of that market is considered "new and unexpected information", on which market participants will act.

Market Effect in Behavior and *Spoofing Goal.* Spoofing creates a false impression of greater buyer/seller interest – or market depth – and book pressure (e.g., CFTC, 2017a, 2018g and 2019f), resulting in misinformation (CFTC, 2019a), pressure on the market (ICE Futures U.S., 2021b) artificial prices (CFTC, 2020e), volatility and an (im)balanced LOB (CFTC, 2015b, 2020b). The goal of creating this effect on the market is to have other market participants trade against genuine orders at better prices, in larger quantities, sooner and at times they would probably not have traded otherwise (CFTC, 2018a, 2019c).

4.5.8 Monetary Action

Monetary actions include the penalties, disgorgements and restitutions imposed by the CME Group, ICE and the CFTC. Since CFTC spoofing cases do not necessarily have legal or monetary consequences, an additional search was conducted to link monetary actions by the CFTC to these spoofing cases. This section only includes the spoofing cases that men-

Table 4.11 | Descriptive statistics of monetary penalties in spoofing cases by CME Group, ICE and the CFTC from January 2010 to June 2022

Variable		**Value**
Total number of monetary penalties		180
Minimum amount		$2,500
Maximum amount		$436,431,811
Mean amount		$3,772,604.16
Distribution of monetary penalties		
Category	**Range**	**Number of spoofing cases**
1	≤ $100,000	125
2	$100,000.01 – $1,000,000	32
3	$1,000,000.01 ≤	23

Note: This table shows the total number of spoofing cases by CME Group, ICE and the CFTC that list a monetary penalty including descriptive statistics of that penalty.

tioned any monetary action. Note that some monetary actions are not solely imposed for spoofing and include penalties and disgorgements for other violations.

Monetary Penalty. Table 4.11 shows the descriptive statistics of the monetary penalties. A total of 180 penalties are documented in the spoofing cases, the lowest being $2,500 and the highest being $436,431,811. The monetary penalties have been divided into three categories to gain more insight into their distribution. The majority of penalties falls into the first category; 125 penalties were less or equal to $100,000. CME Group issued 104 from these 125 penalties (83.2%) and a total of five penalties were issued to companies. CFTC issued larger penalties, as the majority of category two (81.25%) and category three (100%) comprises penalties by the CFTC. Seventeen from the 32 penalties in category two were issued to individuals, with Igor Chernomzav receiving the highest penalty of $750,000 by the CFTC. Penalties in category three differ greatly in size: thirteen penalties were below $3,000,000; nine penalties were between $11,000,000 and $42,000,000; and the highest penalty of $436,431,811 was issued to JPMorgan by the CFTC. The highest penalty issued to an individual was the penalty of $25,743,174.52 from the CFTC to Navinder Sarao. Note that this overview excludes CME Group's monetary penalties under Rule 432 – such as the $600,000 fine against Panther Energy Trading LLC (CME Group, 2013a) and the $200,000 fine against Michael Coscia (CME Group, 2013b) – making the monetary penalties in Table 4.11 an underestimation. Moreover, note that the penalty amount does not only depend on the specifics of the case, but also on the legal framework applied (IMS Group, personal communication, March 5, 2023).

Profit Disgorgement. Since futures markets are a zero-sum game, profit disgorgements provide an indication of how much harm was inflicted on the market by means of spoofing. Descriptive statistics of profit disgorgements are provided in Table 4.12. Note that the profit disgorgements in Table 4.12 are an underestimation, since CME Group's spoofing actions under non-575 rules – such as the profit disgorgement of $1,312,947.02 against Michael Coscia and Panther Energy Trading LLC (CME Group, 2013b and 2013a) – are not included. The minimum amount of profit disgorged was $198.28 and the maximum was $172,034,790 (Panel A). However, the relevant period has not been taken into account, making it hard to compare these profit disgorgements. Hence, profit disgorgements have been divided by the months in which the spoofing occurred – i.e., the relevant period – to get a notion of the average profit defendants made per month (Panel C). On average, spoofing yielded $80,547.31 per month, with a minimum of $49.57 and a maximum of $1,592,914.72.

A better indication of how much harm was inflicted on the market by means of spoofing is the restitution amount. Five spoofing cases by the CFTC mentioned a restitution amount: 1) JPMorgan was ordered to pay $311,737,008 restitution; 2) The Bank of Nova Scotia was ordered to pay $6,622,190 restitution in 2020; 3) Merrill Lynch Commodities Inc. had to pay $2,364,585 in restitution; 4) Tower Research Capital LLC was ordered to pay $32,593,849 restitution; and 5) Propex Derivatives Pty Ltd. Had to pay $464,300 restitution.

4.5.9 Data

Detailed Examples. Many spoofing cases handled by the CFTC provide detailed examples of when the spoofing took place, including the timestamps up to the millisecond; the market and maturity month; and the spoofing actions. All 60 examples are summarized in Appendix 4.B.

Provided Market Data. The CFTC orders by Krishna Mohan and Jiongsheng Zhao provide detailed tables of the provided examples. These tables contain the first two to three LOB levels – including prices and volumes – as well as the specific actions taken by the spoofers in the markets.

Table 4.12 | Descriptive statistics of profit disgorgements in spoofing cases by CME Group, ICE and the CFTC from January 2010 to June 2022

A. Total profit disgorgement		
Variable		**Value**
Total number of profit disgorgements		44
Minimum amount		$198.28
Maximum amount		$172,034,790
Mean amount		$5,012,278.77
B. Distribution of monetary total profit disgorgement		
Category	**Range**	**N**
1	≤ $100,000	37
2	$100,000.01 – $15,000,000	6
3	$15,000,000.01 ≤	1
C. Average profit disgorgement per month		
Variable		**Value**
Total number of profit disgorgements		41
Minimum amount per month		$49.57
Maximum amount per month		$1,592,914.72
Mean amount per month		$80,547.31

Note: This table shows descriptive statistics of the monetary disgorgements listed in spoofing cases by CME Group, ICE and the CFTC. Panel A shows the total number of spoofing cases describing a profit disgorgement, including the minimum, maximum and mean dollar amounts. Panel B categorizes the profit disgorgements from the spoofing cases to get a notion of their distribution. Panel C divides the profit disgorgement in each spoofing case by its respective relevant period (in months) to show the average profit disgorgement per month.

4.6 DISCUSSION AND CONCLUSION

This paper studies spoofing through an interdisciplinary approach using an economic and legal perspective. We provide a comprehensive overview of spoofing, including academic literature, regulation, elements of spoofing strategies, types of spoofing and an overview of public spoofing cases. The paper proposes a conceptual framework based on the academic literature, actual spoofing cases and the expert knowledge at IMS Group. The framework consists of spoofing dimensions and attributes that help define the concept of spoofing and analyze spoofing cases. It can be used to study characteristics of spoofing and legal responses to spoofing behavior. Using the conceptual framework, 204 spoofing cases – identified by the CFTC, CME Group and ICE between January 2010 and June 2022 – were studied and summarized. In doing so, the paper highlights a variety of key spoofing behavior characteristics and legal responses to spoofing. It shows, among others, how diverse and complex spoofing is: 1) spoofing is conducted across different markets, instruments and exchanges; 2) traders collaborate to spoof; 3) rather than creating imbalances, spoofing may create (falsely) balanced markets (CFTC, 2020b); and 4) spoofing is executed both manually and algorithmically. The empirical results are biased towards the CFTC spoofing cases, which provide more details than the CME Group and ICE cases. Moreover, the conceptual framework developed is based on the public information about spoofing cases. As such, the spoofing characteristics discussed in these cases are not exhaustive and may be purposefully vague to discourage market participants from attempting to reverse engineer surveillance techniques (IMS Group, personal communication, March 4, 2023). In this paper, we examined identified instances of spoofing and reported on the results of these spoofing cases. This does not mean, however, that those results can be generalized; the observation, for example, that the identified spoofing cases involved more manual than algorithmic spoofs does not mean that spoofing in general is committed manually more often than algorithmically; it merely means that more manual spoofers were caught spoofing. Further research is needed before making such generalizations.

This paper aspires to accelerate research and insights on spoofing – and market manipulation in general – for all types of (financial) markets and their stakeholders. Currently, the spoofing-strategy elements ‘reaction’ and ‘goal’ (Figure 4.1) are often mixed up, creating confusion when discussing spoofing. Law and regulation tend to confound these two elements, which translates into economists also mixing them up, for example when trying to identify the sort of damage done by spoofing. The spoofing-strategy elements (Figure 4.1) help to decrease this confusion, fostering clarity, unambiguousness and consistency, thus ensuring that everyone is on the same page when discussing spoofing. Although the focus of this paper is on spoofing in U.S. futures markets, the conceptual framework developed (Table 4.2) for analyzing spoofing cases is generalizable and not limited to spoofing, the U.S.

markets or futures markets. The methodology to develop the conceptual framework can be extended, and dimensions and attributes can be modified for other manipulation types, such as insider trading, banging the close and money laundering. It is not limited to U.S. markets and can be applied to markets outside the U.S. too. The conceptual framework is also relevant for other markets, such as stock, crypto and spot markets, and can be applied to trading mechanisms other than continuous trading, such as the call market. Currently, practitioners and academics do not explicitly take into account important attributes, such as liquidity and market microstructure, when studying spoofing. Moreover, the comprehensive list of attributes from the conceptual framework can be incorporated into monitoring systems and (cross-market) detection algorithms to improve them. In turn, individual attributes can be studied as to their importance in identifying spoofing. A select number of attributes may be needed to flag events as suspicious, which the regulators can use to further study these events in detail. The attributes can also be used to identify new types of spoofing that have, for example, similar effects on the market but use a different strategy. Moreover, the conceptual framework allows for easy comparison between cases and can act as a 'scorecard' and overall 'score of suspicion' in future cases (IMS Group, personal communication, March 3, 2023). Legal and regulatory stakeholders can use the conceptual framework to gain a better understanding of and more grip on spoofing. Regulations such as the Dodd-Frank Act may be improved and refined, so as to make them less broad and make it easier to differentiate legitimate from illegitimate trading behavior. Doing so would benefit all market stakeholders.

Although market microstructure attributes are important considerations when studying spoofing, they were hardly ever mentioned in the spoofing cases by the CFTC, CME Group and ICE. Insights from this paper foster the discussion about the role of market microstructure in encouraging and discouraging spoofing. First, iceberg orders can be used for genuine orders to hide the spoofers' true buy/sell interest. One might argue whether this functionality should be programmed into a trading platform or whether it facilitates manipulative practices such as spoofing. However, if trading platforms did not offer the iceberg functionality, traders could program it into their algorithms themselves. Second, the 'avoid-orders-that-cross' functionality used by Igor Oystacher for his spoofing was pre-programmed into the front end that he used, demonstrating that such functionalities can be abused and exploited – even if they are meant to prevent manipulation. The impact of independent trading front ends on market manipulation requires more research. Third, spoofing can be less risky, and hence more attractive, during off-peak hours or night sessions when trading volume and volatility are lower. It may be interesting to discuss and examine the relationship between opening hours and market manipulation. Lastly, while price limits are sometimes used to counter market manipulation (Deb et al., 2010; Kim & Park, 2010), they also induce manipulative and destructive practices by large investors (T. Chen et al., 2019).

The conceptual framework provided in this paper offers a starting point for more extensive market-manipulation research. Statistical analysis can be conducted on the dimensions and attributes, to answer questions such as 1) are certain types of spoofing more prevalent in specific markets?; 2) does the order type of the genuine order (e.g., market or limit order) correlate with spoof order volume, spoof cancellation rate/time or profits?; 3) does one type of spoofing (e.g., single spoofing) tend to see particular sizes of spoof orders?; 4) does the cancellation time of the spoof order differ between spoofing types?; and 5) do regulators impose larger penalties for certain types of spoofing? The spoofing strategy elements ('action', 'reaction' and 'goal') as outlined in Figure 4.1 encourages further research into these elements. For example, the following research questions can be studied: 1) are certain spoofing types more correlated with certain spoofing goals?; 2) are certain intended market reactions more harmful to the market than others?; and 3) are certain intended market reactions more associated with particular spoofing goals? Moreover, spoofing often involves creating an order imbalance in the market. Research is called for to examine the circumstances and conditions under which order imbalances are informative and generate a response in the market (IMS Group, personal communication, March 6, 2023). Future research might focus on the impact of spoofing – and other manipulative tactics – on the market and the effect of changes in the market microstructure on manipulative practices. The market-design characteristics mentioned above may be a starting point. While we focused on spoofing, other manipulative practices could be studied using a similar approach, conceptual framework and outline as presented in this paper. Spoofing can occur in any market with an order book, even with other trading mechanisms. Future research could focus on other types of markets, other trading mechanisms and other countries. These research directions help policy makers, lawyers, regulators and economists.

Chapter 5

The impact of spoofing on U.S. agricultural futures markets

Marjolein E. Verhulst, Cornelis Gardebroek,
Joost M.E. Pennings and Philippe Debie

ABSTRACT

Agricultural futures markets are crucial in the global economy but face challenges, including the disruptive market-manipulation practice called spoofing. This paper explores the frequency and impact of spoofing on agricultural futures markets. Based on spoofing cases published by the CFTC, we define spoofing characteristics and identify spoof orders in the corn, wheat, soybean, soybean meal, soybean oil and live cattle futures markets. Panel data regressions are used to assess the impact of spoofing on liquidity dimensions, using nine liquidity measures. Generally, liquidity declines after spoofing in the corn, soybean, soybean meal and soybean oil markets. Conversely, liquidity improves in the wheat and live cattle futures markets. The findings suggest an inverse relationship between spoofing frequency and the impact on liquidity costs. This highlights the complex dynamics between market manipulation and liquidity and adds to the dialogue on the impact of market manipulation on markets.

Keywords: spoofing, liquidity, market manipulation, detection, limit order book

5.1 INTRODUCTION

Financial markets are crucial in the global economy but encounter numerous challenges that threaten their functioning, including the market-manipulation practice known as spoofing. While many practices can fall under the definition of spoofing, the most well-known spoofing strategies use one or multiple spoof order(s) to create a false sense of supply or demand in the market, to benefit genuine orders on the opposite side of the spoof order(s). These strategies use spoof orders (action) to create a visible change in the market (reaction), in order to buy or sell at a better price or in a higher quantity (goal) than would have been possible without spoofing (Verhulst & Pennings, 2023). Regardless of how spoofing strategies are defined in regulations or what evidence is needed to demonstrate spoofing, certain trading patterns in market data can be indicative of spoofing. Although it has been over 13 years since spoofing was prohibited under the Dodd-Frank Act (United States, 2010), and many regulatory enforcements related to spoofing have been carried out (see Verhulst and Pennings, 2023), little is known about the frequency and the impact of spoofing on the liquidity of futures markets, which is the focus of this study.

In recent years, there has been an increase in research focusing on spoofing in financial markets. Topics include, but are not limited to: 1) the market microstructure and spoofing (Lee et al., 2013); 2) what encourages/discourages spoofing (Cartea et al., 2020; Martínez-Miranda et al., 2016; X. Wang et al., 2018, 2021; Williams & Skrzypacz, 2020); 3) the susceptibility of OTC FX markets to spoofing (Stenfors & Susai, 2021); 4) cross-market spoofing (Stenfors, Dilshani, et al., 2023; Stenfors, Doraghi, et al., 2023; Williams & Skrzypacz, 2020); 5) the dynamics between non-manipulative and manipulative traders (Wang et al., 2020, 2021); 6) the dynamics between regulators and manipulative traders (Wang & Wellman, 2020); 7) the impact of spoofing on prices, price discovery and volatility (Brogaard et al., 2022; Chen & Hsieh, 2023; Wang, 2015; Williams & Skrzypacz, 2020); and 8) modeling and detecting spoofing using econometric, machine learning or statistical physics approaches (Y. Cao et al., 2014, 2015; Do & Putnins, 2023; Leangarun et al., 2016; H. Li et al., 2023; Martínez Miranda et al., 2019; Mendonça & De Genaro, 2020; Tao et al., 2022; Zhai et al., 2017, 2018).

Measuring the frequency of spoofing in futures markets is challenging, as its detection is complex. Literature on spoofing detection exists but, to the best of our knowledge, does not address the actual frequency of spoofing in these markets. Moreover, a limited number of studies addresses the impact of spoofing on liquidity. Wang (2015) studied the characteristics, profitability, determinants and the impact of spoofing on the Taiwan Capitalization Weighted Stock Index futures market. Among others, they studied the impact of spoofing on two liquidity measures: the percentage effective spread and trading volume. They found that spoofing induces more trading volume and increases the spread, regardless of whether the spoof order is placed on the bid or ask side, and of the type of investor who is spoofing. The increased spread is also confirmed by Williams and Skrzypacz (2020). Brogaard et al. (2022) studied the effect of spoofing on the market quality in Canadian equity markets, measuring liquidity with time-weighted quoted spreads, volume-weighted effective spreads and volume-weighted realized spreads. They find that successful spoofing increases the effective and realized spreads but decreases the quoted spreads. Spoofing increases spreads as market makers face more adverse selection, causing them to widen the spread to cover these costs (Brogaard et al., 2022; Wang, 2015; Williams & Skrzypacz, 2020). Finally, Debie et al. (2022) examined the Commodity Futures Trading Commission (CFTC) spoofing case of JPMorgan Chase & Company and its subsidiaries, JPMorgan Chase Bank and J.P. Morgan Securities LLC (hereinafter referred to as "JPMorgan"). Using a visualization method from particle physics, they showed the effects on liquidity – as measured by the Adverse Price Movement (APM) – before, during and after JPMorgan spoofed futures markets. Their research demonstrated that attracting liquidity can be a motivation to spoof the market and that, in some cases, liquidity was indeed better after the spoof than before. Although not specifically investigating market manipulation or spoofing, other closely related literature explores the impact of (unforeseen) events on liquidity in agricultural markets (e.g., Couleau et al., 2020; and He et al., 2021).

This paper analyzes the effect of spoofing on liquidity, with a focus on agricultural futures markets. The following two research questions are addressed: 1) how frequently does spoofing occur in specific agricultural futures markets?; and 2) how is liquidity in these futures markets affected before, during and after spoofing? We focus on the single spoofing strategy and estimate the impact of this strategy on liquidity in the corn, wheat, soybean, soybean meal, soybean oil, and live cattle futures markets. To answer the first research question, we identify spoof orders using the spoofing cases published by the CFTC. Based on these CFTC cases, we define spoofing characteristics and use a selection of these characteristics to filter potential spoof orders from the data. We do not know with certainty that these are actual spoofing cases, as trading strategies can only be labelled 'spoofing' by regulatory agencies. To answer the second question, we use nine liquidity measures in panel data regressions to study the impact of spoofing on liquidity. Timeseries windows are extracted around each identified spoof order, acting as the units in the panel. Dummy variables distinguish observations before, during or after the spoof, allowing us to analyze whether and how liquidity changes before, during and after spoof orders. Results are summarized per liquidity dimension: tightness, depth and resiliency.

Most of the literature suggests liquidity decreases after spoofing. Decomposing liquidity into dimensions, previous literature has shown that the tightness dimension, often measured by the bid-ask spread, generally worsens due to spoofing, while the percentage-effective spread, volume-weighted effective spread and realized spread all widen due to spoofing (Brogaard et al., 2022; Wang, 2015; Williams & Skrzypacz, 2020). One exception is the time-weighted quoted spread, which tightens (Brogaard et al., 2022). Hence, we expect the tightness dimension to decrease after spoofing in agricultural futures markets. The impact of spoofing on the depth dimension has, to the best of our knowledge, not been studied yet. We expect the depth dimension to be similar or worse after spoofing. That is, the false sense of supply/demand that spoof orders create induces other market participants to add orders on the same side as the spoof order, often in the form of herd behavior (Dalko et al., 2020; Montgomery, 2016). We expect the induced market participants to also remove their orders when the spoof order is removed – partially due to this herd behavior and partially due to them recognizing that the order was not genuine – thus returning depth to its pre-spoof levels. Successful spoofing attempts include the execution of genuine orders, which take depth from the opposite side of the spoof order. Hence, after spoofing, we expect depth to be similar on the *same* side as the spoof order but worse on the *opposite* side of the spoof order. Debie et al. (2022) provide anecdotal evidence for the effect of spoofing on the resiliency dimension. Thirty seconds – the same time window as used in this paper – after JPMorgan spoofed the market, liquidity decreased in five out of nine instances, remained constant in one instance, and increased in three instances. Although their results differ per futures markets and constitute anecdotal evidence only, we expect resiliency in agricultural futures markets to show a similar response as most JPMorgan instances, i.e., to decrease after spoofing.

This paper relates to the research by Brogaard et al. (2022) and Do and Putnins (2023), but there are several conceptual and methodological differences: We use distributions of the state of the limit order book (LOB), using anomalies in these distributions as a characteristic of spoofing. This also means that, rather than merely creating an imbalance in the market, spoof orders can cause markets to become imbalanced, balanced or more imbalanced. Our approach focuses on the single spoofing strategy and is independent of pairing genuine and spoof orders. We extend our analysis beyond the tightness dimension to also include the depth and resiliency dimensions, as well as more comprehensive measures for liquidity taking into account the full LOB.

This paper contributes to the literature by giving an indication of the frequency of spoofing in agricultural futures markets. It also provides a deeper understanding of the impact of spoofing on various liquidity dimensions, where previous research has mainly focused on the tightness dimension – i.e., the bid-ask spread – and has scarcely considered other liquidity dimensions. Surprisingly – and contrary to previous research showing that spoofing affects liquidity negatively – we find that, in relatively small agricultural futures markets, spoofing improves liquidity. Hence, we add to existing research by showing the nuanced effects of spoofing on liquidity and how it may benefit some market participants. Practitioners can use the insights to become more informed on the effects of spoofing on the market, helping them to decide, for example, in which markets (not) to trade. Regulators can use our methodology to identify spoof orders to enhance their surveillance systems. The empirical focus of this paper is on the U.S. futures markets, but the methodology is applicable to any market in which spoofing may occur. Hence, this paper facilitates conducting additional research into the impact of manipulation on markets.

5.2 SPOOFING CHARACTERISTICS

5.2.1 CFTC Spoofing Examples

The spoofing cases as identified by the CFTC are used to define the characteristics of spoofing. Whenever the CFTC takes legal action – including against spoofing – they publish a press release with an order (i.e., a legal document) containing the details of the violation. Verhulst and Pennings (2023) analyzed 204 of these U.S. spoofing orders issued by the CFTC, the Chicago Mercantile Exchange (CME) and the Intercontinental Exchange, combining their analysis with academic literature, the legal framework and industry expertise into a conceptual framework that characterizes spoofing into dimensions and attributes. These CFTC orders sometimes contain specific examples of the spoofing strategies, including timestamps, prices, volumes and other specifics. Verhulst and Pennings (2023) provide an overview of 58 CFTC examples in their appendix. We capitalize on their framework and the examples provided to specify the characteristics, so as to identify spoof orders and assign threshold values. Specifically, 28 out of the 58 CFTC examples concerned the single spoofing strategy. These 28 examples were used to extract spoof order characteristics, and a select number were turned into metrics. For example, the characteristic 'spoof duration' – referred to as 'cancellation time' in the framework by Verhulst and Pennings (2023) – is assigned the threshold value of one second based on the CFTC examples. This means that orders that are in the LOB for less than one second are flagged as potential spoof orders if they also meet the other criteria. Note that only six out of the 28 spoofing examples concern agricultural futures markets, but given the absence of any alternatives, we have included all examples, regardless of market. An overview of the characteristics can be found in online Appendix 5.A for all 28 examples. Note that we were able to elaborate certain spoofing examples of the CFTC in more detail, as we were familiar with the market data of these examples from previous research.

In this paper we focus on a spoofing strategy that is used to create a false sense of supply or demand in the LOB: single spoofing (Verhulst & Pennings, 2023). Note that we are unable to identify which genuine order(s) belong to spoof orders within a spoofing strategy, as the dataset we use – market by order (MBO) data – contains order IDs but not market participant IDs. This means that we can track single orders during their life cycle, but we do not know if multiple orders come from the same market participant. Since only the spoof orders are illegal – as there is no intention for them to be executed – and it is not possible (yet) to link genuine orders to spoof orders in anonymized data, we focus on quantifying spoof orders only, rather than counting both the spoof and genuine orders. In doing so, we are also implicitly taking cross-market spoofing into account. In cross-market spoofing, a spoofer places a genuine order in Market A and a spoof order in a correlated Market B. Since we exclusively focus on identifying spoof orders – regardless of any corresponding genuine orders in the same or different markets – we do not depend on pairing genuine orders to spoof orders. This allows us to capture spoof orders that are part of same-market, cross-market and even cross-exchange spoofs.

Before going into the details of quantifying the single spoofing strategy, the next section first discusses the LOB state metric used.

5.2.2 State of the Limit Order Book: the Imbalance Ratio

The focus of this paper is on spoofing strategies that create a false sense of supply or demand in the LOB. Hence, it is necessary to quantify the state of the LOB in terms of volume on the bid and ask sides. Doing so allows us to measure how the state of the LOB changes due to (spoof) orders. We express the state of the LOB in a single number by applying the imbalance ratio as described by Stenfors et al. (2023), which is a variation of the volume imbalance by Cartea et al. (2020). We refer to these ratios as the LOB state:

$$LOB\ state = \frac{\sum_{i=1}^{n} b_i - \sum_{i=1}^{n} a_i}{\sum_{i=1}^{n} b_i + \sum_{i=1}^{n} a_i} \tag{5.1}$$

where n is the number of price levels from the LOB included in the calculation, and b_i and a_i are the bid and ask volumes at the ith price level, respectively. The LOB state returns a normalized value between -1 and 1, where -1 indicates the market only has volume on the ask and no volume on the bid side and is, thus, a severe ask-pressured market; 1 indicates the market only has volume on the bid and no volume on the ask side, or a severe bid-pressured market; and 0 is a market with no dominant side that puts pressure on the market.

5.2.3 Single Spoofing Characteristics

Table 5.1 shows the descriptive statistics of several spoof order characteristics of the 28 CFTC single spoofing strategy examples. Spoof order sizes differ greatly, with the smallest spoof order being 33 contracts and the largest being 3000 contracts. This can be attributed to spoofing examples covering various markets of different volumes. For example, the spoof order of size 33 was added in the gold futures market, while the spoof order of size 3000 was added in the 10-Year T-Note futures market. Naturally, order size also depends on liquidity at the time the spoof event takes place. Markets with large volumes in the LOB require relatively larger spoof order volumes to be noticed by other market participants. Contrary, smaller spoof order volumes are sufficient in markets with thin levels in the LOB. Spoof orders in the CFTC examples were cancelled as fast as a minimum of 0.318 seconds after being added and were, on average, active in the market for 7.515 seconds. Table 5.1 also shows that the spoof orders were mainly placed at the first level of the LOB and generally no deeper than the seventh level. This may be due to the fact that, for some futures markets, the information content of the LOB is higher for levels closer to the best bid and ask prices (e.g., Cao et al., 2009; and Arzandeh & Frank, 2019). Hence, the spoof order may be considered more informative when placed closer to the top of the book. Note that a spoof level of zero means that the spoof order was placed within the spread, i.e., at a lower (higher) price level than the current best ask (bid) level. When spoof orders are placed on existing levels, they are placed at the back of the queue. Hence, the risk of execution is mitigated. The minimum volume resting immediately before a spoof order was six contracts and the maximum was 1219 contracts. This, again, also depends on how liquid the market is, in which the respective spoof order was placed. On average, spoof orders increased volume on the respective LOB *side* by 64.58% and on the affected price *level* in the LOB by 357.73%. The spoof orders in the examples from the CFTC were used to execute genuine orders. Table 5.1 shows that, after reaching their objectives – i.e., execution of the (last) genuine order – spoof orders were cancelled relatively fast with a mean of 3.756 seconds.

Table 5.1 | Descriptive statistics of single spoof orders in 28 examples from the CFTC between January 2010 and June 2022.

	n	Mean	Median	Std	Min	Max
Order size (contracts)	43	515.86	201	717.06	33	3000
Duration (seconds)	42	7.515	1	19.552	0.318	120
Level placement	33	1.97	1	1.65	0	6
Pending volume level (contracts)	8	310.25	151.5	416.56	6	1219
Volume increase LOB side	11	64.58%	60%	44.30%	4.5%	161.29%
Volume increase top 5 levels	8	357.73%	124.99%	545.23%	53.59%	1666.67%
Cancellation time after last genuine order executed (seconds)	23	3.756	1.6	6.686	0.13	29.854

Note: This table shows several descriptive statistics for a set of spoof order characteristics. These characteristics were extracted from spoofing examples outlined in CFTC documents between January 2010 and June 2022, as compiled by Verhulst and Pennings (2023). The list is not exhaustive, and the number of observations (*n*) can differ from the 28 examples, due to the fact that some examples do not specify all details and some examples include multiple single spoofing strategies executed consecutively. *Order size* is the size of the spoof order measured in contracts, *Duration* is the time (in seconds) the spoof order was active in the limit order book (LOB) before it was cancelled, *Level placement* is the price level to which the spoof order was added, *Pending volume level* is the volume already resting at the price level affected by the spoof order, *Volume increase LOB side* is the increase in volume in percentages of the entire LOB side affected by the spoof order, *Volume increase top 5 levels* is the increase in volume of the first 5 levels on the LOB side affected by the spoof order, *Cancellation time after last genuine order executed* shows how fast the spoof order was cancelled after the last genuine order was executed.

Since a spoof order in the single spoofing strategy seeks to create a false sense of supply/demand in the LOB, it must be noticeable by other market participants and relatively large. Moreover, spoof orders are not meant to be executed, so they are cancelled relatively fast to avoid this risk. To distinguish a spoof order from a regular order, we have selected from Verhulst and Pennings (2023) key characteristics that reflect these features, while including a new characteristic that captures the magnitude of the false supply/demand created by the spoof order. Using the examples in the CFTC orders, we quantify these characteristics into variables with threshold values – indicating when something is a spoof order or not. Note that these variables should not be considered in isolation, as only their combination can be indicative of spoofing.

5.2.3.1 Spoof Duration

Since spoof orders are not meant to be executed, we include the spoof order's duration, i.e., the time from adding to cancelling the spoof order. We use the one-second median from the CFTC examples, as depicted in Table 5.1. Contrary to their intention, spoof orders are sometimes nevertheless partly executed because, for example, they are placed too close to the top of the LOB. Although this is a possibility, we exclude partial execution, thus only studying 'successful' spoofing attempts – as partially executed spoof orders can be interpreted as a 'failed' spoofing attempt. Moreover, note that spoof orders can be implicitly cancelled by modifying their size or modifying them towards deeper levels in the LOB before eventually explicitly cancelling them (Verhulst & Pennings, 2023). We focus on spoof orders that are explicitly cancelled, rather than modified and then cancelled.

5.2.3.2 Placement

Spoof orders must be placed close to the top of the LOB for visibility. Based on the CFTC examples, we use LOB levels one to five: more levels than the mean (level 1) and fewer levels than the maximum level reported (level 6), to include sufficient levels for spoof orders to be placed at. We assume that single spoof orders are not used to create new best bid/ask levels, a practice often associated with, for example, the spread-squeezing spoofing strategy (Verhulst & Pennings, 2023). Hence, we exclude orders that create new levels within the spread.

5.2.3.3 LOB State

We use equation (5.1) in section 5.2.2 to measure the state of the LOB each time an order is added to the LOB. To determine the effect of a spoof order on the LOB state, the first fifty levels of the bid and ask sides of the LOB are considered, totaling to one hundred LOB levels.[39]

39 When measuring the LOB state, it is necessary to include sufficient levels, so as to diminish the effect of changes in the LOB state being attributable to levels disappearing or reappearing. When considering ten LOB levels, for example, any large differences in the LOB state might in fact be due to the addition (deletion) of a new (old) level, i.e., the tenth level disappearing (or reappearing) in the LOB. In this event, it would not have been the order that impacted the LOB state but

Next, the difference in the LOB state between two consecutive actions is calculated when an order is added. This allows us to perceive the impact of a single order on the LOB state. A distribution is made of all these LOB-state differences within one trading day, yielding one distribution of LOB-state differences per day. We focus on orders that fall in the tails of these distributions, specifically the one percentile at both ends. These orders cause more of a change in the LOB state than the other 98% of added limit orders in the market. The one percentile was derived by applying the approach to market data from known CFTC spoofing cases available to the authors from previous research. This is discussed in more detail in Appendix 5.B.

Spoof order size is important, as it should be relatively large compared to the rest of the volume in the LOB to elicit a response from market participants. However, what is considered a 'large volume' is highly dependent on the market. Using the LOB-state differences, we implicitly take spoof order size into account, in that the focus lies on orders that cause the greatest changes in the LOB state. This eliminates the need to include order size as a separate variable in quantifying the single spoofing strategy.

In summary, we define a spoof order as an order that adheres to all the following criteria: it is an order that 1) is added to the LOB and cancelled within one second; 2) is not (partially) executed; 3) is not modified and then cancelled; 4) is placed on one of the top five LOB levels; 5) does not create a new best bid or ask level; and 6) causes a change in the LOB state larger than that caused by the other 98% of added orders in the market for that day. For illustration purposes, since criterion 6 uses various parameters, Appendix 5.C shows the number of spoof orders identified in the corn futures dataset when criteria 1 to 5 are held constant and the parameters of criterion 6 vary.

5.3 DATA AND METHODOLOGY

5.3.1 Data

Data is obtained from the CME Group MBO dataset from July 2019 to June 2020. The original dataset contains all CME Group futures and options markets, but this paper focuses on the following agricultural futures markets: corn, wheat (Chicago soft red winter wheat), soybean, soybean meal, soybean oil and live cattle. MBO data contains individual actions by traders timestamped at the nanosecond, such as trades and the submission, modification and cancellation of orders. These actions are captured in market messages - consisting

the shift in price levels. To maximally eliminate this effect, sufficient LOB levels must be included. In this study, we are considering 50 levels on either side of the LOB.

of tags and values - which are formatted according to the Financial Information eXchange (FIX) protocol. It is possible to reconstruct the full depth of the LOB using all market activity messages. The ROOT software framework, mainly used in the particle physics domain, is used for efficient storage and processing of the MBO data (Brun & Rademakers, 1997; CERN, 2018b; Verhulst et al., 2021). Developed by the European Organization for Nuclear Research (CERN) and other parties, ROOT is particularly well suited for managing big data, a common characteristic in particle physics research. The FIX data files are transformed into ROOT files for efficient storage and access. To illustrate this: the soybean market represents the largest dataset in this study, totaling to 213 GB of messages encoded in FIX format. ROOT makes it possible to run the analysis on its compressed data objects, reducing all soybean market messages to 10.6 GB. The full LOB is reconstructed from all messages using ROOT (Verhulst et al., 2021). This paper uses both the individual order messages and the reconstructed LOB to quantify spoofing and measure its impact on liquidity. Table 5.2 shows descriptive statistics of the data. Appendix 5.D shows the LOB state and changes herein for all market activity in the datasets, i.e., the LOB state is calculated for each message, whereby the change reflects the difference in the LOB state between two consecutive messages.

To get a single timeseries per market, we use nearby futures contracts and roll over on the first day of the expiration month (de Boer et al., 2022). We focus on the day trading sessions and, similar to Foucault et al. (2007) and Shkilko and Sokolov (2020), remove the first and last five minutes of these sessions to exclude the opening and closing behavior of the day. The data is checked for spoof orders by extracting orders that include the characteristics described in section 5.2. For each spoof order identified, a timeseries window is extracted starting 30 seconds before the addition of the spoof order and ending 30 seconds after its cancellation. Taking the various time windows surrounding spoof orders as units, this produces a balanced data panel dataset.

Table 5.3 shows the descriptive statistics for the identified spoof orders per market. These characteristics can be linked to Table 5.2, which describes similar variables for the complete market. We observe some differences between the descriptive statistics for the full market (Table 5.2) and for the spoof orders (Table 5.3). For example, the average volume added through spoof orders is considerably larger than through average market activity. That is, the average order size of spoof orders (all market activity) are 179.09 (8.6) contracts in the corn market, 17.37 (2.1) contracts in the wheat market, 51.05 (3) contracts in the soybean, 26.35 (2.9) contracts in the soybean meal, 21.69 (3) contracts in the soybean oil and 10.57 (1.8) contracts in the live cattle markets. Moreover, using the number of spoof orders as a fraction of the total added orders in the market yields estimates of the percentages of spoof orders added per market: 0.10% in the corn, 0.06% in the wheat, 0.17% in the soybean, 0.07% in the soybean meal, 0.06% in the soybean oil and 0.02% in the live cattle futures markets.

Table 5.2 | Descriptive statistics of the complete CME Group dataset from July 2019 to June 2020.

		Corn	Wheat	Soybean	Soybean meal	Soybean oil	Live cattle
A. Market activity							
Number of messages		505,735,938	481,979,943	1,126,276,473	1,026,169,596	837,227,460	215,553,628
Number of added order messages		38,576,541	36,983,734	61,845,072	58,399,395	59,186,894	17,600,032
Number of modified order messages		21,241,029	15,264,147	59,182,016	80,399,931	37,988,459	7,053,859
Number of cancelled order messages		38,145,651	36,928,033	61,624,180	58,352,713	59,157,810	17,540,571
Number of trade messages		13,628,921	7,826,131	13,287,083	7,260,351	8,458,430	4,921,419
B. Average volume per message							
Added orders		8.6	2.1	3	2.9	3	1.8
Modified orders		20.3	4	3.9	2.9	3.2	2.3
Cancelled orders		8.2	2	3	2.8	2.9	1.5
Trades		7.7	3.1	3.6	3.1	3.2	2.2
C. Spoofing estimation characteristics							
Order age in seconds	Mean	2672.46	1238.66	1290.08	3340.21	883.404	614.424
	Median	108.23	102.512	113.278	147.967	110.888	101.425
	St.dev.	19988.9	11495.8	10528	10841.5	6986.47	7821.6
Price level placement	Mean	2.4	2.2	2.1	2	2.9	1.8
	Median	0.3	0.4	0.2	0.4	0.4	0.3
	St.dev.	8	6.2	7	5.8	9.2	5.7
LOB state	Mean	-0.16	0.03	-0.26	0.52	0.45	-0.32
	Median	-0.18	0.05	-0.28	0.59	0.50	-0.33
	St.dev.	0.27	0.26	0.24	0.31	0.31	0.33
LOB state change	Mean	0.002	0.002	0.003	0.004	0.003	0.002
	Median	0.003	0.003	0.005	0.006	0.004	0.004
	St.dev.	0.010	0.012	0.011	0.011	0.012	0.010

Note: This table shows descriptive statistics for all contracts in the corn, Chicago SRW wheat, soybean, soybean meal, soybean oil and live cattle futures markets of CME Group between July 2019 and June 2020. *Panel A* shows the total number of messages and splits these into messages for orders that are added, modified, cancelled and traded. *Panel B* shows the average volume per message for orders that are added, modified, cancelled and traded. *Panel C* shows the mean, median and standard deviation for characteristics related to the assessment of spoofing: the *order age* is the time an order is active in the limit order book (LOB), capped at a maximum of five days old (this includes 99.99% of the data), regardless of whether the order is cancelled or modified or resulted in a transaction; the *price level placement* is the price level to which orders are added; the *LOB state* is calculated per message and is the total bid volume minus the total ask volume, divided by the sum of the total bid and ask volumes; and the *LOB state change* is the difference in LOB state between two consecutive messages.

Table 5.3 | Descriptive statistics for presumed spoof orders in CME Group data from July 2019 to June 2020.

	Corn	Wheat	Soybean	Soybean meal	Soybean oil	Live cattle
Number of spoof orders	39,347	20,635	104,200	43,444	37,002	4080
Bid spoof orders	19,293	10,884	53,059	20,461	18,387	1690
Ask spoof orders	20,054	9751	51,141	22,983	18,615	2390
Mean (SD) spoof duration in seconds	0.127 (0.242)	0.252 (0.357)	0.084 (0.209)	0.117 (0.238)	0.131 (0.255)	0.199 (0.317)
Mean (SD) spoof order size in contracts	179.09 (136.77)	17.37 (15.07)	51.05 (33.78)	26.35 (23.43)	21.69 (18.79)	10.57 (11.99)
Mean (SD) LOB price level placement	1.19 (0.53)	1.63 (1)	1.22 (0.50)	1.33 (0.76)	1.29 (0.73)	1.6 (1.16)
Mean (SD) LOB state change	0.0004 (0.0055)	0.0003 (0.0041)	0.0021 (0.0097)	-0.0003 (0.0069)	-0.0006 (0.005)	0.0002 (0.0057)

Note: This table shows descriptive statistics for the presumed single-spoofing strategies in the corn, wheat, soybean, soybean meal, soybean oil and live cattle futures markets based on data from CME Group between July 2019 and June 2020. *Spoof duration* is the total time the spoof order was active in the limit order book (LOB), i.e., from its addition to its cancellation; *spoof order size* is the volume of the spoof order in number of futures contracts; *price level placement* is the price level to which spoof orders were added; and *LOB state change* is the change the spoof order causes in the LOB state, whereby the LOB state is calculated by subtracting the total ask volume from the total bid volume, and dividing this by the summation of the total bid and ask volumes.

These percentages are an underestimate, as the total number of added orders in the market (Table 5.2) is determined using *all* data, while the number of spoof orders (Table 5.3) is determined over a *subset* of the data with contracts that roll over. Also note that the LOB state change in Table 5.2 is calculated for *all* messages, while the LOB state change in Table 5.3 is only calculated for messages *added* to the LOB.

5.3.2 Model

Panel data regressions are used to measure the impact of spoofing on liquidity. The spoof orders are the cross-sectional dimension, and the observations (messages) within the corresponding time windows are the time-series dimension of panel data. We assume each time window has unique characteristics and, hence, we use fixed-effects models. *F*-tests for poolability confirm this assumption.

5.3.2.1 Dependent Variables: Liquidity

Liquidity is measured across three dimensions – tightness, depth and resiliency – and measures are used that can be calculated for each data point in time, i.e., measures that do not take averages over a time window. **Tightness** is measured using the bid-ask spread (*Spread*), which is calculated for each message by subtracting the best bid price from the best ask price. **Depth** is measured using six variables that either sum volume or orders of specific price levels per LOB side. The volume variables differentiate between the top of the LOB and deeper levels of the LOB, since previous research suggests that LOB levels differ as to how informative they are (e.g., Arzandeh & Frank, 2019b). This results in a measure for the first five levels for the ask ($AskDepth_{1\text{-}5}$) and bid ($BidDepth_{1\text{-}5}$) sides, and a measure for the sixth to tenth levels for the ask ($AskDepth_{6\text{-}10}$) and bid ($BidDepth_{6\text{-}10}$) sides. In addition, the number of unique orders on the first ten levels on the ask (*AskOrders*) and bid (*BidOrders*) sides are summed. Finally, the APM measurement (Gomber & Schweickert, 2002) includes the tightness, depth and immediacy dimensions and, when measured over time, covers **resiliency**. APM bid (ask) represents the execution costs (in bps) of a trader who wants to sell (buy) a certain dollar value, by submitting market orders and taking liquidity from the bid (ask) side. A lower (higher) APM indicates lower (higher) execution costs and, hence, higher (lower) liquidity (Gomber & Schweickert, 2002). We adjust the equations from Gomber et al. (2015) to make them applicable to our data and this study:[40]

$$APM_{A,t}(V) = 10.000\,\frac{\bar{P}_{A,t}(V) - MQ_t}{MQ_t} \tag{5.2}$$

$$APM_{B,t}(V) = 10.000\,\frac{MQ_t - \bar{P}_{B,t}(V)}{MQ_t} \tag{5.3}$$

where $APM_{A,t}(V)$ and $APM_{B,t}(V)$ are the respective adverse price movements in basis points for the ask and bid side at time t; V is the order size in dollars; $\bar{P}_{A,t}(V)$ and $\bar{P}_{B,t}(V)$ are the respective quantity-weighted average execution prices for the ask and bid side at time t; and MQ_t is the quote midpoint at time t. Since this measurement requires a constant dollar value, we calculate the dollar value for each market by multiplying the first fifty bid and ask prices with their respective volumes for each message and use the mean dollar value from all these messages in a year.[41] $APM_{A,t}$ and $APM_{B,t}$ are included in the model as dependent variables, totaling to nine different equations for these liquidity measures.

5.3.2.2 Explanatory Variables

Dummy variables are created to categorize observations into having occurred before, during or after the spoof. The 'during' (*DuringSpoof*) and 'after' (*AfterSpoof*) dummy variables are included in the equations, setting the 'before' period as the reference category. *T*-tests allow for testing between the 'before' and 'during' and 'before' and 'after' periods, whereas

40 In particular, Gomber et al. (2015) use the Exchange Liquidity Measure, or Xetra Liquidity Measure (XLM), to calculate liquidity costs for a buy order ($XLM_{B,t}$) and sell order ($XLM_{S,t}$) at time *t*. Since we are interested in the impact a buy (sell) order has on the LOB, we reverse the equations: a buy (sell) order has an adverse price movement on the ask (bid) side of the order book. Hence, the equation for $XLM_{B,t}$ is, in our case, the equation for $APM_{A,t}$, and the equation for $XLM_{S,t}$ is, in our case, the equation for $APM_{B,t}$. Moreover, the original equations used *P*, rather than . We use the latter, following Hachmeister (2007), to better reflect that the average execution price is quantity weighted.

41 This results in a mean dollar value of approximately $7,665,200 for the corn market, $1,129,550 for the wheat market, $5,149,050 for the soybean market, $764,492 for the soybean meal market, $67,431.6 for the soybean oil market and $144,795 for the live cattle market.

a Wald test is used to test for significant differences between the 'during' and 'after' coefficients.

To determine the optimal number of lags, R-squared values are examined for various liquidity measure lags. A notable difference in R-squared values was observed when comparing models with no lags to those with two lags. The transition from two lags to four lags showed minimal variation in coefficients and R-squared values. Hence, two lags are included in all models.

In addition to the dummy and lagged variables, a variable is included to control for the irregularly spaced data. Variable *dTimestamp* is the time difference between two consecutive messages. Our data – and high-frequency data in general – are irregularly spaced, but the panel data regressions treat these data as if they were regularly spaced. The *dTimestamp* variable implicitly tests whether the irregular time intervals between messages matter and may thus offer new insights into handling irregular high-frequency data.

Combining all variables, we derive the following equation for the fixed-effects panel data regression:

$$Liquidity_{it} = \beta_1 DuringSpoof_{it} + \beta_2 AfterSpoof_{it} + \beta_3 dTimestamp_{it} + Liquidity_{i,t-1} + Liquidity_{i,t-2} + \gamma_i + \varepsilon_{it} \tag{5.4}$$

where $Liquidity_{it}$ is one of the nine liquidity measures for each time-window unit i at time (message) t; $DuringSpoof_{it}$ and $AfterSpoof_{it}$ are the dummy variables indicating whether an observation occurred during or after the spoof order in unit i at time t, respectively; $dTimestamp_{it}$ is the difference between two messages in unit i at time t; $Liquidity_{i,t-1}$ and $Liquidity_{i,t-2}$ are the first and second lag of the dependent liquidity variable, respectively; γ_i is the fixed effect for each unit i; and ε_{it} is the error term for unit i at time t.

5.3.2.3 Bid versus Ask Spoof Orders

Research has demonstrated that bid and ask side liquidity can behave asymmetrically, especially during crisis periods (Cenesizoglu & Grass, 2018; Sensoy, 2019; Tripathi et al., 2020). Hence, in addition to splitting liquidity measures by bid and ask side, we also differentiate between buy and sell spoof orders. This results in three separate panel data regressions for each liquidity measure: 1) a restricted model with no differentiation between bid and ask spoof orders; 2) an unrestricted model with only bid spoof orders; and 3) an unrestricted model with only ask spoof orders. A likelihood ratio test is conducted to test whether there are significant differences between the coefficients of the unrestricted bid and ask spoof-order models. This paper reports results and regression tables for the unrestricted bid and ask spoof-order models, as all likelihood ratio tests indicated a significant difference between the coefficients of bid and ask spoof orders. Results of the restricted model are available from the authors upon request.

Conducting three panel data regressions per liquidity measure for nine liquidity measures and six futures markets results in a total of 162 regression models.

5.3.3 Computational Tools and Resources

ROOT is used to store and process the MBO data. CERN's SWAN (CERN, 2023b) is used to create and execute Python scripts to clean the processed MBO data and conduct less memory-intensive panel data regressions. The CERN Batch Service is used to execute the computational scripts to process the MBO data, filter the time windows around the identified spoof orders and conduct the more memory-intensive panel data regressions. This computing environment enables us to process and analyze big data more efficiently and with sufficient memory. For this research, we used approximately 3696 core hours to process the MBO data, run the memory-intensive panel data regressions and perform robustness checks.

5.4 RESULTS

This section discusses how liquidity dimensions are affected by spoofing and the impact of spoofing on liquidity costs. The results focus on the change in liquidity after spoofing, compared to the period preceding spoofing, as discussing all 162 regressions is beyond our scope. Appendix 5.E provides a full overview of what happens to liquidity before, during and after spoofing per futures market. Regression tables supporting the results are included in Appendix 5.F; additional output – such as from the Wald test – is available upon request.

5.4.1 Impact of Spoofing on Liquidity Dimensions

Table 5.4 shows, for each futures market, how liquidity dimensions change after spoofing, compared to before spoofing. Note that the table does not reflect whether individual liquidity measures increase or decrease but how liquidity itself increases or decreases. For example, an increase in the bid-ask spread, APM bid and APM ask indicates a decrease in liquidity, so the table will report a decrease. Results from Table 5.4 are discussed for each liquidity dimension and for each futures market. To improve readability, no distinction is made between bid and ask spoof orders, unless there is a clear difference in the results.

Impact of spoofing on liquidity dimensions. Table 5.4 shows no consistency in how liquidity dimensions respond to spoofing, as this highly depends on the futures market in which the spoofing occurs. We expected the **tightness** dimension to decrease based on previous lit-

Table 5.4 | Change in liquidity 30 seconds after spoofing in CME Group futures markets between July 2019 and June 2020.

	Bid spoof orders						**Ask spoof orders**					
	Corn	**Wheat**	**Soybean**	**Soybean meal**	**Soybean oil**	**Live cattle**	**Corn**	**Wheat**	**Soybean**	**Soybean meal**	**Soybean oil**	**Live cattle**
Tightness												
Bid-ask spread	–			–	–	+	–		–	–		+
Depth												
Bid depth $_{1–5}$	+	+			–	+		+	+	–		+
Bid depth $_{6–10}$				–	–	+	–	–		–	–	+
Ask depth $_{1–5}$		+		+	–	+	+	+	+	+		+
Ask depth $_{6–10}$	–		–		–	+		+	–	–	–	+
Bid orders	–	+	–	–	–	+	–	+	–	–	–	+
Ask orders	–	+	–	–	–	+	–	+	–	–	–	+
Resiliency												
APM bid	+	+		–	–	+	–		–	–	–	+
APM ask	–	–	–	–	–	+			–	–	–	+

Note: This table shows the direction in which liquidity measures and dimensions change, 30 seconds after spoofing, in CME Group futures markets between July 2019 and June 2020. The '+' sign indicates that liquidity improves, and the '–' sign indicates that liquidity deteriorates after spoofing. Note that the signs do not indicate the direction of the liquidity measure, but what this direction means for liquidity itself. For example, since a decrease in the bid-ask spread indicates improved liquidity, it gets a '+' sign in the table.

5

erature. Table 5.4 shows no consistent response as, without differentiating between bid and ask spoof orders, the tightness dimension increases in 17%, decreases in 50% and remains constant in 33% of cases. Similarly, the **depth** dimension increases in 38%, decreases in 43% and remains constant in 19% of cases if we do not differentiate between bid and ask spoof orders. We expected depth to decrease on the opposite side of the spoof order and remain constant on the same side of the spoof order. Depth on the opposite side of the bid (ask) spoof order[42] increases in 33% (33%) of the cases, decreases in 44% (50%) and remains constant in 22% (17%) of cases. Depth on the same side of the bid (ask) spoof order[43] increases in 33% (50%), decreases in 39% (39%) and remains constant in 28% (11%) of cases. Hence, results are not in line with our expectations, and the depth dimension does not respond consistently across markets to spoofing. The **resiliency** dimension shows more consistency, to a certain extent, behaving similar to our expectations: we expected resiliency to decrease after spoofing, and Table 5.4 shows that this is true in 58% of cases, while resiliency increases and remains constant for 25% and 17% of cases, respectively. This is similar to the anecdotal evidence gathered by Debie et al. (2022), who found that, depending on the futures market, APM responds differently but mostly decreases 30 seconds after spoofing.

In summary, we cannot conclude that liquidity dimensions respond consistently to spoofing *across* markets. This is, however, not the case when we examine liquidity dimensions *within* a market as the next section discusses.

Impact of spoofing on liquidity for each futures market. In general, the tightness and depth dimensions of the **corn futures market** are worse after spoofing, as liquidity decreases for these dimensions. The depth dimension shows an exception for two liquidity measures, which increase rather than decrease after spoofing: liquidity on the first five bid (ask) levels increases after bid (ask) spoof orders are used. This indicates that spoof orders in the corn futures market attract more volume to the side to which they are added – bid (ask) spoof orders attract more bid (ask) volume to the top five levels – and this volume persists for at least 30 seconds after the spoof order is removed. This result is in line with Debie et al., (2022) and can be attributed, for example, to herd behavior, where market participants copy others (Dalko & Wang, 2018). The impact of spoofing on the resiliency dimension depends on the type of spoof order (bid or ask) and the LOB side. Resiliency of the bid side increases after spoofing with bid orders, while resiliency on the ask side decreases. On the other hand, when ask spoof orders are used, resiliency on the bid side decreases, while remaining constant on the ask side.

42 The percentages are calculated based on ask depth (ask depth $_{1–5,}$ ask depth $_{6\text{-}10}$ and ask orders in Table 5.4) for bid spoof orders and bid depth (bid depth $_{1–5,}$ bid depth $_{6\text{-}10}$ and bid orders in Table 5.4) for ask spoof orders.

43 The percentages are calculated based on bid depth (bid depth $_{1–5,}$ bid depth $_{6\text{-}10}$ and bid orders in Table 5.4) for bid spoof orders and ask depth (ask depth $_{1–5,}$ ask depth $_{6\text{-}10}$ and ask orders in Table 5.4) for ask spoof orders.

In the **wheat market** most liquidity dimensions improve after spoofing. Tightness is unaffected, but depth improves after spoofing. Resiliency outcomes vary here too: resiliency remains unaffected by ask spoof orders, while bid side resiliency improves and ask side resiliency worsens when bid spoof orders are used. Overall, most liquidity measures and dimensions in the wheat market improve after spoofing.

Liquidity dimensions in the **soybean market** differ in direction. Tightness worsens when ask spoof orders are used but remains constant when bid spoof orders are used. In general, depth and resiliency worsen. Notable is the increase in volume at the top of the bid and ask side when ask spoof orders are used, as these are the only depth liquidity measures in the soybean market that improve after spoofing.

All liquidity dimensions in the **soybean meal market** worsen after spoofing. Specifically, all significant changes show that liquidity worsens after spoofing, with two exceptions: more ask volume is added to the top of the LOB for both bid and ask spoof orders, meaning that, after spoofing, more market participants are willing to sell.

Like the soybean and soybean meal market, the **soybean oil market** is negatively affected by spoofing. The tightness dimension is negatively affected by bid spoof orders, and similarly, depth and resiliency deteriorate, regardless of whether bid or ask spoof orders are used.

Surprisingly, all liquidity dimensions of the **live cattle market** are constant in their response and improve after spoofing, regardless of whether bid or ask spoof orders are used.

In summary, spoofing significantly affects the liquidity dimensions – as most liquidity measures show a significant change – but the direction of the change highly depends on the commodity market. Liquidity dimensions within a single market exhibit a fairly consistent pattern in how they are affected by spoofing, with most liquidity measures within a single market pointing in the same direction. Liquidity in the corn, soybean, soybean meal and soybean oil futures market generally worsens after spoofing. Surprisingly, liquidity in the wheat and live cattle market improves after spoofing.

5.4.2 Impact of Spoofing on Liquidity Costs

Using the APM bid and ask, we can calculate the impact of spoof orders on trades occurring shortly after the spoof. That is, APM calculates the changed liquidity costs (in dollars) of a transaction within 30 seconds after a spoof order is cancelled. This calculation is exemplified using the corn futures market, while the results for all futures markets are presented in Table 5.5. Note that this section focuses exclusively on the economic impact of spoofing on liquidity costs, omitting its impact on transaction prices. In the corn futures market, the average midpoint price[44] for trades between July 2019 and June 2020 was 377 U.S. cents

44 We do not use the average *trade* price due to the way trade messages work at CME Group. To illustrate this, imagine trader

Table 5.5 | Liquidity costs 30 seconds after the single spoofing strategy in CME Group futures markets between July 2019 and June 2020.

Liquidity costs		Corn	Wheat	Soybean	Soybean meal	Soybean oil	Live cattle
Trade one contract after one spoof event							
Bid spoof orders	Bid side	–$0.04 (–0.00019%)	–$0.02 (–0.00006%)	–	$0.03 (0.00011%)	$0.01 (0.00005%)	–$0.25 (–0.00058%)
	Ask side	$0.03 (0.00017%)	$0.03 (0.00012%)	$0.02 (0.00004%)	$0.02 (0.00006%)	$0.01 (0.00007%)	–$0.12 (–0.00027%)
	Total LOB	–$0.004 (–0.00002%)	$0.02 (0.00006%)	$0.02 (0.00004%)	$0.05 (0.00017%)	$0.02 (0.00012%)	–$0.37 (–0.00085%)
Ask spoof orders	Bid side	$0.07 (0.00039%)	–	$0.04 (0.00009%)	$0.01 (0.00004%)	$0.02 (0.00012%)	–$0.16 (–0.00037%)
	Ask side	–	–	$0.02 (0.00005%)	$0.03 (0.00009%)	$0.03 (0.00018%)	–$0.09 (–0.0002%)
	Total LOB	$0.07 (0.00039%)	–	$0.06 (0.00014%)	$0.04 (0.00013%)	$0.05 (0.0003%)	–$0.25 (–0.00057%)

Note: This table shows the change in liquidity costs 30 seconds after spoofing in CME Group futures markets between July 2019 and June 2020. It shows the change in liquidity costs in dollars (percentage of underlying value) when a trader buys/sells one contract after a single spoofing event. Insignificant changes are marked by '–'. The respective underlying values for a single contract are $18,850; $26,150; $44,800; $30,200; $18,000 and $43,200 for the corn, wheat, soybean, soybean meal, soybean oil and live cattle futures market. Dollar values in the table are rounded to the second decimal.

per bushel.[45] The corn contract unit is 5000 bushels and, hence, the underlying value of one contract totals to $18,850.[46] Table 5.E.1 in Appendix 5.E shows the coefficients for APM bid and ask when bid and ask spoof orders are used. After bid spoof orders are used in the corn futures market, liquidity costs on the bid side decrease by 0.019 bps (0.00019%) and liquidity costs on the ask side increase by 0.017 bps. Hence, if a trader buys (sells) one contract within 30 seconds after a bid spoof order is cancelled – taking liquidity from the ask (bid) side – liquidity costs are $0.032 higher ($0.036 lower) per contract than without spoofing. We identified 19,293 bid spoof orders and 20,054 ask spoof orders in the corn futures market between July 2019 and June 2020 (see Table 5.3). Hence, if a single contract is bought (sold) after each bid spoof order, total liquidity costs of all these spoof orders are $618.24 higher ($690.98 lower) than without spoofing. These are minimum estimates assuming only one contract is traded after spoofing. However, taking the average (rounded) added volume per order[47] from Table 5.2 yields an increase (decrease) in liquidity costs of $5,564.20 ($6,218.81) for buying (selling) nine contracts after all the bid spoof orders.

Applying the same calculation to the other markets, Table 5.5 shows the results for the corn, wheat, soybean, soybean meal, soybean oil and live cattle markets assuming one contract is traded after one spoof order. It shows that liquidity costs decrease for the bid side of the corn and wheat market when bid spoof orders are used and for both sides of the live cattle market when using bid and ask spoof orders. A decrease in liquidity costs indicates improved liquidity after the use of spoof orders. Liquidity costs increase for all other cases, with the exception of insignificant changes marked by '–'. While the dollar values of liquidity costs are relatively similar between markets, liquidity costs show more variation when expressed relative to their underlying values. For example, a $0.02 increase in liquidity costs per contract in the soybean meal market represents 0.00006% of its underlying value, while this same dollar value constitutes 0.00012% of the underlying value in the soybean oil market. Relative to its underlying value, corn shows the highest increase in liquidity costs (0.00039%) on the bid side when ask spoof orders are used. The live cattle futures market shows the highest decrease in liquidity costs (–0.00058%) on the bid side when bid spoof orders are used. It is striking that, despite the soybean market having the highest frequency

A adds an order of volume ten, which gets executed for 400 U.S. cents against three separate traders at the exact same time. The MBO data would not show one trade message of volume ten and price 400 U.S. cents, but three distinct trade messages, one for each trader – e.g., of volumes three, four and three. We average the midpoint price for all trade prices to diminish this effect.

45 Similarly, the average midpoint price was 523 U.S. cents for the wheat futures market and 896 U.S. cents, 302 dollars, 30 U.S. cents and 108 U.S. cents for, respectively, the soybean, soybean meal, soybean oil and live cattle futures markets.

46 The respective contract units (underlying values) for the wheat, soybean, soybean meal, soybean oil and live cattle futures markets are 5000 bushels ($26,150), 5000 bushels ($44,800), 100 short tons ($30,200), 60,000 pounds ($18,000) and 40,000 pounds ($43,200).

47 Similar to why we do not use the average trade price, we do not use the average trade volume. We use the average volume that was added to the LOB per order, assuming this is the volume traders actually wanted to trade and hence coming closer to the notional value they wish to trade.

of detected spoof orders (104,200), their effect on liquidity costs is among the least of all studied markets. For example, the lowest increase in liquidity costs of all markets occurs on the ask side of the soybean market when bid spoof orders are used (0.00004%). Conversely, despite the live cattle market having the lowest frequency of spoof orders (4,080), their impact is the highest of all studied markets. Annual liquidity costs can be calculated by multiplying the number of bid and ask spoof orders by the liquidity costs for trading a single contract in Table 5.5. For example, the soybean futures market had 51,141 ask spoof orders. Assuming one contract is traded after each ask spoof event, this results in total liquidity costs of $2,045.64 on the bid side and $1,022.82 on the ask side.

5.4.3 Controlling for Irregular Timestamps

As mentioned in section 5.3, we added the time-difference (in seconds) between two consecutive messages as a variable to control for the irregularity of the data. The coefficients for this variable *dTimestamp* can be found in Appendix 5.G. From the 120 reported coefficients, 106 coefficients indicate a negative relationship between *dTimestamp* and the respective liquidity measure. This variable can also be used as a proxy for reaction time by traders: an increase (decrease) in time between two messages indicates a slower (faster) reaction time by traders. Hence, most coefficients show a negative relationship between traders' response rates and liquidity. However, we cannot draw any meaningful conclusions since we did not differentiate between different responses, that is message type (i.e., add order, cancel order, modify order, trade) or time period (i.e., before, during or after spoofing) for this variable. Moreover, since we only studied time windows containing spoofing, we do not know whether this result holds true under regular market circumstances. Future research is needed to delve deeper into this issue.

5.5 DISCUSSION AND CONCLUSION

This paper studies the frequency of spoofing and its impact on liquidity dimensions in agricultural futures markets. Specifically, we study 1) the frequency of the single spoofing strategy in the corn, wheat, soybean, soybean meal, soybean oil and live cattle markets; and 2) using panel data regressions, how liquidity was affected in these markets before, during and after spoofing. Between July 2019 and June 2020, we identified the least (4,080) spoof orders in the CME Group live cattle futures market and the most (104,200) spoof orders in the CME Group soybean futures market. While we found no consistency in how liquidity dimensions respond to spoofing *across* markets, liquidity dimensions *within* a single futures market respond fairly consistently. Liquidity in the corn, soybean, soybean meal and soybean oil futures markets generally worsens after spoofing, while liquidity in the wheat and live cattle markets improves after spoofing. Hence, surprisingly, small agricultural markets that face hedging challenges might benefit from spoofing. Moreover, results seem to suggest an inverse relationship between liquidity costs and frequency of identified spoof orders. That is, the impact on relative liquidity costs was highest (lowest) in the live cattle (soybean) futures market, which had the lowest (highest) spoof order frequency. Future research is needed to confirm the existence of this inverse relationship.

Our findings provide a nuanced perspective on the impact of spoofing on liquidity, as it suggests that liquidity improves in certain markets after spoofing. This observation highlights the complex dynamics between market manipulation and liquidity. Moreover, it adds a potentially controversial result to the dialogue on the role of market manipulation in markets. Although future research is needed into why liquidity dimensions improve after spoofing in the wheat and live cattle markets, we propose that this might be related to the fact that these markets are generally less liquid than the corn, soybean, soybean meal and soybean oil markets. Thus, there might be a 'general liquidity threshold' that determines how markets respond to spoof orders, above which markets would experience a decrease and below which markets would experience an increase in liquidity after spoofing. Moreover, markets responding similarly to spoofing – be it through increasing or decreasing liquidity – might have common characteristics that determine this response. Another avenue worth exploring is to study how markets are perceived in terms of fairness and market manipulation. From the six markets in this study, the wheat and live cattle markets have the least identified spoof orders. Market participants might perceive these markets as more trustworthy and subject to little market manipulation. Hence, when spoof orders do occur, these might be perceived as genuine orders and thus attract more market participants. This could also explain why liquidity costs in the live cattle (soybean) market – the market with the least (most) identified spoof orders – are impacted the most (least) after spoofing, compared to the other markets under study.

Our findings concerning the *dTimestamp* variable contribute to the debate about the use of irregular data in timeseries regressions. We controlled for this irregularity by including the *dTimestamp* variable, which captures the time between observations. Indeed, the results showed that this controlling was necessary, given that the unequal spacing provided for through this variable contributes to explaining the liquidity measures. Future methodological research is necessary to deal with the impact of using irregular data.

This study has several limitations: firstly, we cannot know with absolute certainty that the identified orders are spoof orders since only regulatory agencies or courts can issue this verdict. Moreover, although our analysis focuses on agricultural futures markets, the levels of the spoofing criteria are based on 28 CFTC examples mostly representing non-agricultural futures markets. In addition, the criteria for identifying spoof orders do not account for any modifications or partial executions, so the number of spoof orders identified as being part of the single spoofing strategy may be an underestimation. Conversely, the use of the single spoofing strategy might also be overestimated, since it can also include other spoofing strategies: further inspection of the identified spoof orders also revealed patterns similar to layered spoofing. For example, the addition of 36 bid orders of volume eight at five distinct price levels within three seconds. This reveals another limitation of our study: the overlapping time windows in our analysis. Around each spoof order, a time window of sixty seconds is extracted, so when spoof orders occur in rapid succession, these time windows will overlap. Hence, some time windows might be overrepresented in our analysis. In addition, the panel data regressions use a fixed number of lags for each market, which might not be the optimal number. More robustness checks are needed, for example with the parameters of the spoof order criteria and the dollar value used for the APM calculation. Finally, our data is irregularly spaced and, although we have attempted to control for this by including a timestamp variable in the regressions, more research is needed to test how this affects results.

Our results have implications for many financial stakeholders. First, the characteristics we use to identify the single spoofing strategy can be used by market participants and regulators to monitor the market, as well as for surveillance purposes. As discussed more extensively below, market analysts can use the characteristics to monitor markets and aid the decision-making process in which market(s) (not) to trade. Companies can use the characteristics for internal compliance purposes, monitoring their own market activity to ensure they are not engaged in single spoofing. Regulators can enhance their surveillance systems by developing alerts based on the spoofing characteristics used in this paper.

Second, the results regarding the impact of spoofing on liquidity dimensions have different implications for different types of market participants. In the discussion below, we focus exclusively on the impact of spoofing on liquidity, omitting its impact on transaction prices. The **tightness** dimension is important for market participants who capitalize on small price movements within short time frames. Particularly market participants such as scalpers, arbitrageurs, and high-frequency traders (HFT) need to keep a close eye on the tightness dimension after spoofing, since they utilize the bid-ask spread, find arbitrage opportunities in small price discrepancies between markets and operate on thin margins (Menkveld, 2013; Shah & Brorsen, 2011). In markets where the tightness dimension worsens (improves) due to spoofing, a wider (narrower) spread means higher (lower) transaction costs and worse (better) opportunities for profit on small price differences. Hence, after spoofing, these market participants may find trading opportunities in the live cattle market and abstain from trading in the corn, soybean, soybean meal and soybean oil markets. The opposite may be true for arbitrageurs, such as spread traders, who seek to take advantage of widening spreads (Wang et al., 2014).

The **depth** dimension is important for market participants that trade large volumes, such as institutional investors and hedgers (Pennings & Meulenberg, 1997a, 1997b). These market participants wish to trade large quantities without running through the LOB – thereby incurring additional liquidity costs – and impacting the price significantly. These large-volume traders need sufficient depth to absorb their trades without excessive market impact and to help them conceal their strategy, for example when they need to roll over their position, by making their orders less noticeable. Spoofing can alter trading decisions based on available liquidity, and the results show general and specific scenarios in which large-volume traders can benefit or suffer from spoofing. In general, hedgers and institutional investors, for example, benefit from increased depth after spoofing in the wheat and live cattle markets and suffer from decreased depth in the corn, soybean, soybean meal and soybean oil markets. In a more specific scenario, a farmer might benefit from the increased bid depth by selling corn futures contracts after spoofing occurred on the bid side.

The effects of spoofing on the **resiliency** dimension are particularly interesting for long-term investors who benefit from stable trading environments, such as institutional investors, long-term speculators and hedgers. For example, hedgers need markets to recover quickly from price shocks or market manipulation to maintain their hedging effectiveness and, similarly, portfolio managers need resilient markets to preserve the value and predictability of their investments over time. Therefore, these market participants can value more resilient markets over less resilient ones. Resilience in the live cattle market improves after spoofing, while it deteriorates in the soybean, soybean meal and soybean oil markets. Resiliency in the corn and wheat market varies depending on the side of the LOB and on whether bid or ask spoof orders are used.

Besides market participants, liquidity dimensions also influence other financial market stakeholders. Market analysts, for example, use them to predict market trends and stability. Hence, changes in liquidity due to spoofing may directly affect trading strategies. Moreover, regulators and exchanges monitor liquidity to, among others, protect traders and maintain market integrity. They can use the insights from this study for market surveillance – for example, by monitoring more closely markets that are more prone to or more affected by spoofing than other markets – or for modifying the market design to discourage spoofing.

Future research could extend our analysis by including more spoofing strategies, such as layering or flipping, and allowing for order modification and partial execution. Spoof orders do not necessarily have to be cancelled immediately or quickly: they can be implicitly cancelled by first modifying them to deeper levels in the LOB – to avoid the risk of execution – and cancelling them at a later time (Verhulst & Pennings, 2023). This study only focused on spoof orders that were successful in not getting executed. However, future research could include spoof orders that were partially executed. Moreover, future research could quantify both spoof orders and genuine orders, thus linking potential genuine orders to spoof orders in anonymized data. Our analysis of the effects of spoofing on liquidity could be extended by increasing the time window around spoofing and studying when liquidity levels after spoofing revert to the same levels as before spoofing. In addition, we used the *dTimestamp* variable to control for irregular data. However, since it can also serve as a proxy for 'reaction time' of the market, this variable could yield interesting research questions when differentiating between the types of messages, including questions such as: how fast are orders added, cancelled or modified; or how fast are orders added, cancelled or modified during spoofing versus in regular market circumstances? Moreover, methodological research is needed to address the impact of irregular data on models and to develop methods for coping with this. Finally, future research could study 1) the presence of a general liquidity threshold or common market characteristics, that determine(s) how markets respond to spoof orders; 2) the existence of an inverse relationship between spoofing frequency and liquidity costs; and 3) the perception of markets in terms of fairness/market manipulation and whether this affects how markets respond to spoofing.

Chapter 6

General conclusions and discussions

Identifying and analyzing market manipulation has proven to be difficult, as academics are faced with numerous financial big data and market manipulation research challenges (Chapter 1). The goal of this dissertation is to effectively identify and analyze market manipulation in a high-frequency data context, by applying particle physics tools to financial market data. We aimed to answer the following main research question: *'How can market manipulation in a high-frequency context be identified and analyzed?'* Four sub-research questions (RQ) were formulated to answer the main research question. Chapter 2 introduced a new visualization tool rooted in particle physics, to make high-frequency trading and high-paced events more visible utilizing all information of the data (RQ1). Chapter 3 applied this new visualization tool from Chapter 2 to a well-known spoofing case, thereby making spoofing characteristics and motivations visible (RQ2). Chapter 4 disentangled spoofing by developing a conceptual framework consisting of spoofing characteristics, based on academic literature, legislation, rulings and expert knowledge from the industry (RQ3). Using specific characteristics from the conceptual framework of Chapter 4, Chapter 5 subsequently identified spoofing in agricultural futures markets, also assessing its frequency and impact on liquidity in these markets (RQ4).

This chapter first answers the sub- and main research questions using the research findings of Chapter 2 to 5. Theoretical, methodological and practical contributions are considered, followed by a discussion section, a note on the limitations, suggestions for future research and a closing section.

6.1 INTEGRATING RESEARCH FINDINGS WITH RESEARCH QUESTIONS

RQ1: How can we improve understanding of high-frequency markets and developments therein?

We can improve our understanding of high-frequency markets and developments therein by using message-based data, visualizations and different disciplinary perspectives. First, in Chapter 2 we found that using all the information embedded in messages – rather than taking aggregated snapshots of the data – makes high-frequency trading and information-dense events more apparent. The LOB is visualized with each activity in the market, regardless of whether these activities occurred simultaneously. This allows for better observation and inspection of time windows containing considerable amounts of market activity and changes occurring in a split-second, for example, market activity around the announcement of an USDA report (as demonstrated in Chapter 2) or around market manipulation (as demonstrated in Chapter 3). The details and nuances of market behavior would go unnoticed if snapshots of the data were used. The difference between visualizing messages and snapshots is not simply a matter of zooming: when creating snapshots, messages are compressed into seconds, thus amounting to a loss of information. Not only is the use of messages beneficial for visualization purposes, Chapter 5 also shows that the irregular spacing of messages statistically significantly impacts the results, as demonstrated for liquidity. This means that the time difference between two consecutive messages contains information explaining liquidity. This information embedded in messages might be overlooked when creating snapshots of the data, stressing the importance of using messages to better understand high-frequency markets and developments therein. Second, the new visualization methodology rooted in particle physics improves our understanding of high-frequency markets. Chapter 2 demonstrated that the message-based visualization is effective in visualizing market activity in the LOB, together with additional variables characterizing the market. The visualization capitalizes on histograms by binning the axes and plotting messages without being dependent on the regular intervals of time. Individual trading actions are put in perspective of their market context, and even between market contexts due to the ability to visualize multiple high-frequency markets in conjunction. The visualization methodology combines multiple different data sources – e.g., LOB and trade data – and its customizability allows for studying an extensive number of research topics to enhance the understanding of high-frequency markets. For example, visualizing the LOB along with the number of added, modified and deleted orders in the closing-period on a settlement day. Indeed, in Chapter 3 we found that it is also effective in visualizing market manipulation – as manipulation appears as a clearly visible hotspot – as well as the variables characterizing market manipulation. This wide application of the visualization methodology makes it useful for many financial market stakeholders interested in understanding high-frequency market data, as it goes beyond traditional methods. Third, Chapter 4 demonstrated that market behavior in high-frequency markets can be better understood from multiple disciplinary perspectives. For example, a legal perspective can explain market participants changing their behavior to comply with the law; an economic perspective can explain why manipulative traders participate in illegal practices, or show the effects of manipulative behavior on the market; and a market microstructure perspective can explain the order volumes and price levels chosen by participants. Thus, delineating a single topic from various perspectives helps in better understanding market behavior in high-frequency markets.

RQ2: How can we improve understanding of market manipulation in a high-frequency context?

Similar to RQ1, our understanding of market manipulation in a high-frequency context can be improved through the use of message-based data and visualizations, and by applying different perspectives. Messages ensure all information is considered, visualizations show all individual actions by market manipulators in the context of the full market and differing perspectives shed light on how, why and when participants manipulate the market. In Chapter 3 we found that, rather than sifting through raw data, traders' actions can be directly observed, followed and analyzed within the broader market context using the message-based visualization methodology from Chapter 2. The visualizations reveal spoofing behavior, market participants' responses to spoofing and they can visualize characteristics indicative of spoofing. The visualizations also helped us find an alternative motivation for spoofing: apart from moving the price, spoofing is also used to attract additional liquidity to the market. This methodology simplifies the process of identifying and understanding manipulative behaviors occurring within milliseconds – actions that may not be noticeable through traditional financial analysis tools. Our understanding of market manipulation can further be improved by statistically testing what happens in the market before, during and after market manipulation. Illustrated for liquidity, Chapters 3 and 5 statistically tested how liquidity is impacted by spoofing. We found that, contrary to what is often thought, liquidity improves in several markets after spoofing. Complementing the visualizations and statistical analysis, the conceptual framework developed in Chapter 4 further enhances our understanding of market manipulation in two ways. First, the methodology used to produce the conceptual framework fully delineates spoofing into dimensions and attributes. Second, the conceptual framework was applied to 204 public spoofing cases. Combined, the insights from the conceptual framework in Chapter 4, the visualizations from Chapter 3 and the statistical analysis from Chapters 3 and 5, provide a comprehensive view on spoofing, bringing us closer to understanding the market dynamics around market manipulation and giving insight into actual market behavior – be it with benign or malicious intent.

RQ3: What are the characteristics of the market manipulation practice of 'spoofing' in a high-frequency context?

In Chapter 4 we delineated spoofing strategies and developed a conceptual framework containing spoofing dimensions and attributes to analyze spoofing cases, thereby characterizing spoofing. Spoofing strategies consist of three elements: 1) the action, or the type of spoofing used; 2) the reaction, or the spoofer's desired impact on the market; and 3) the goal, or the objective the spoofer wishes to achieve. Some spoofing strategies only use the action and goal elements, for example flashing single spoof orders (action) to test market depth at certain LOB levels (goal). Other spoofing strategies use all three elements, for example using layered spoofing (action) to create an imbalance in the market (reaction) in order to buy or sell at a better price (goal). The conceptual framework was developed by gathering spoofing characteristics from academic literature, legislation, rulings and industry knowledge. These characteristics (attributes) were then grouped by a common theme into dimensions of the conceptual framework. For example, the dimension 'spoof order' consists, among others, of the attributes 'size (volume)' and 'cancellation time (seconds) after placement'; and the dimension 'spoofing: general' contains dimensions such as 'spoofing type' and 'algorithmic or manual'. Six out of nine dimensions were directly related to the spoofing strategy – rather than, for example, describing spoofing case information – which totaled to 69 spoofing attributes. By analyzing and integrating diverse sources and perspectives, we systematically captured all relevant spoofing information and characteristics. Lawmakers, regulators, exchanges and market participants can evaluate trading behavior and strategies against this concept of spoofing. In Chapter 4, we applied the conceptual framework to 204 public spoofing cases and discussed numerous characteristics which highlight the nuances and complexities of spoofing. In Chapter 5, we used the conceptual framework to gain insight in 1) the frequency of spoof orders; and 2) the impact of spoof orders on liquidity. We selected key spoofing characteristics from the conceptual framework to identify spoof orders in agricultural futures markets, which proved to be promising. One of these spoofing characteristics related to the visualizations of the LOB in Chapter 3, which showed that spoofing often presents itself as a hotspot in volume. These hotspots were transformed into the 'LOB state difference' variable in Chapter 5, characterizing spoof orders as those with a relatively large impact on LOB volume. Moreover, several attributes of the conceptual framework had been previously visualized in Chapter 3, demonstrating that visually depicting these characteristics can offer guidance in effectively and quickly identifying large deviations that may indicate manipulative behavior.

RQ4: What is the frequency and impact on liquidity of the market manipulation practice of 'spoofing' in high-frequency markets?

In Chapter 5, we found that spoof orders comprised approximately 0.10% of all added limit orders in the corn futures market, and 0.06% in the wheat, 0.17% in the soybean, 0.07% in the soybean meal, 0.06% in the soybean oil and 0.02% in the live cattle futures markets at CME Group between July 2019 and June 2020. Its economic impact was measured on three liquidity dimensions and in terms of liquidity costs. While we found no consistent impact of spoof orders on liquidity dimensions across markets, liquidity dimensions within a single market responded similarly. Generally, liquidity worsens after spoofing in the corn, soybean, soybean meal and soybean oil markets, whereas liquidity improves after spoofing in the wheat and live cattle markets. The latter was also reported in Chapter 3, as anecdotal evidence of the JPMorgan spoofing case indicated liquidity had improved after spoofing. Moreover, results suggested an inverse relationship between spoofing frequency and economic impact in terms of liquidity costs: a lower spoofing frequency seemed to be associated with higher absolute liquidity costs and vice versa – regardless of whether the spoofing has a positive or negative impact on liquidity costs. The economic impact was demonstrated in Chapter 3 for the JPMorgan spoof orders, by showing how price, volume, cancellations, trade volume and liquidity costs behaved around these spoof orders. Chapter 4 discussed economic impact in terms of LOB (im)balance and monetary aspects, such as market losses and the amount of harm inflicted on the market.

Main RQ: How can market manipulation in a high-frequency context be identified and analyzed?

Data in a high-frequency context demands more advanced tools and methodologies to effectively manage and analyze it. Market manipulation in a high-frequency context can be better identified and analyzed by taking a holistic approach that integrates the areas of managing financial big data and studying market manipulation. This dissertation adopts this type of approach, providing a comprehensive framework for addressing the challenges outlined in Figure 1.1. Specifically, we contribute to the main research question by providing tools and methodologies rooted in particle physics. We found in Chapters 2, 3 and 5 that the ROOT framework proves effective for storing, processing, visualizing and analyzing high-frequency data, thus addressing the financial big data challenges. This is the essential first step before any market manipulation can be identified or analyzed. Next, this dissertation covers all market manipulation research aspects from start to finish, providing methodologies to delineate, characterize, identify, visualize and analyze market manipulation. In doing so, it provides a comprehensive framework for identifying and analyzing market manipulation in a high-frequency context.

6.2 THEORETICAL AND PRACTICAL CONTRIBUTIONS

This dissertation has theoretical, methodological and practical contributions, most of which relate to the financial big data and market manipulation research challenges depicted in Figure 1.1.

6.2.1 Theoretical and methodological contributions

The literature and research on spoofing, particularly in futures markets, is scarce. No consistent definitions of spoofing strategies are available, and the impact of spoofing on markets is understudied. This dissertation contributes to the spoofing literature by describing a theory of spoofing and demonstrating its impact on liquidity. In Chapter 4, we defined the elements of spoofing strategies and provided a conceptual framework to disentangle spoofing. Whereas previous literature shows that spoofing is harmful for markets, in Chapter 5 we demonstrated that in some cases spoofing benefits the market in terms of liquidity, and in Chapter 3 we showed that attracting liquidity can also be a motivation for spoofing. These elements contribute to the theoretical understanding of spoofing. In addition, Chapter 4 contributes by providing a dataset containing all public spoofing cases up until June 2022, to overcome the limitation of existing empirical studies.

We make a methodological contribution to market manipulation research by providing methodologies for all its constituent aspects: from collecting public information and delineating its implications, to characterizing and identifying manipulative behaviors, to visualizing these behaviors and ultimately assessing their impact on the market. The results of applying these methodologies contribute empirically to spoofing literature. While the application focuses on spoofing, the methodologies used are applicable to other market manipulation types and different trading environments. The research further contributes methodologically by providing a novel visualization method derived from particle physics (Chapter 2) and by adding to the empirical discussion whether and why research should use messages or snapshots of the data. Finally, this dissertation contributes through its interdisciplinary nature of applying particle physics methodologies and tools to high-frequency market data.

6.2.2 Practical contributions

The visualization tool presented in Chapters 2 and 3, together with the identification and analysis procedure for market manipulation (Chapter 5), are elements for the development of a detection tool. This would also require the methodology and conceptual framework for delineating market manipulation from Chapter 4, to fully understand all elements and characteristics of certain types of market manipulation and shed light on these practices from multiple perspectives. Hence, the research in this dissertation practically contributes to the development of a market manipulation detection tool for market surveillance purposes.

The financial big data challenges outlined in Figure 1.1 are addressed by a comprehensive approach that spans all critical aspects, from data storage and processing to visualization and analysis. The main data analysis framework that was used – ROOT – is open source, making this research accessible to many scientific and industry stakeholders. Hence, this dissertation has laid fundamental groundwork required to manage and research high-frequency market data. Moreover, it offers guidance for the process of researching market manipulation, which benefits many financial market stakeholders. Regulatory agencies can use the conceptual framework in Chapter 4 to select specific characteristics to study or identify spoofing. In turn, the identified spoof orders can then be visualized (Chapter 2 and 3) to better understand how markets respond to these events or to establish any spillover effects between markets. For example, after Chapter 3 was published, the CFTC published the first spoofing case with visualizations, underlining the need and usefulness of visualization tools in market surveillance. Statistical tests can subsequently validate the visualized market response (Chapter 5), thereby providing scientifically backed, robust evidence. Market surveillance systems can be improved, and are already being improved, using similar approaches to those proposed in this dissertation. Members of the International Expert Group on Market Surveillance (IMS Group) are currently examining the implementation of the metrics and tools developed in this dissertation. Similarly, scientists can better research market manipulation, as they can use the methodologies proposed here to identify and study the impact of other manipulative practices. Market participants benefit through a deeper understanding of spoofing in high-frequency markets, enabling them to enhance their internal monitoring and compliance. The detailed overview of spoofing characteristics can help market participants to better identify when spoofing is occurring, empowering them to make informed decisions about trading under those conditions.

6.3 DISCUSSION

This section provides a critical reflection on the methodologies proposed in this dissertation and puts the dissertation into a broader perspective.

Practical applicability. The dissertation has provided several elements which can be used to develop a detection tool for market surveillance purposes. One of these elements is the new visualization methodology, with Chapters 2 and 3 highlighting the effectiveness of visualization. However, there are two major drawbacks that may raise questions about whether this methodology is actually practically applicable in its current state: 1) ROOT was not written for financials nor is it taught in any textbooks or classes, which can create high barriers to using it; and 2) the generated visualizations use historic data and are not (yet) compatible with real-time data. The latter issue can be particularly problematic for exchanges, if they wish to monitor markets in real time. Moreover, the visualizations can benefit from a more user-friendly layout, such as a dashboard, and from compatibility with more data sources. For example, linking tweet datasets or trader-specific information like open interest per trader to the LOB. Statistical analyses such as the ones in Chapters 3 and 5 can aid regulators and exchanges in differentiating between statistically different market behavior, which in turn helps to identify manipulative practices. Developments in the market are following each other rapidly, but the identification and analysis of manipulative behavior can be achieved with relatively straightforward tools and methodologies.

Proving intent. The crucial element in the U.S. legal definition of spoofing is the word 'intent'. Studying the 204 spoofing cases – particularly those from the CFTC due to their detailed nature – suggested an evolution in the type of evidence used against spoofing. Whereas early spoofing cases often relied on supporting evidence such as chat logs and phone calls, more recent cases tend to be more quantitative and data driven. This raises the question of whether – and how – traders' intent can be inferred from hard data alone. Although publicly reported evidence of some spoofing cases is based on only hard data, I would argue that the approach to gathering evidence needs to incorporate more statistical methods to strengthen legal arguments. There is no need for regulatory agencies to rush onto the AI bandwagon, since simple statistics can already be very effective. Distributions, for example, are a simple and effective way to incorporate a more statistical-based approach in gathering evidence. A case in point is the LOB state distribution used in Chapter 5, a simple but effective measure for identifying orders with a large impact on the market. A different example of using distributions is comparing the trading behavior of a single trader with the distribution of the market. This might lead to conclusions such as 'Trader A's cancellation time is statistically significantly different from the market average, falling within the 0.01% tail of the market distribution'. Such highly divergent behavior by a trader compared to the market might indicate a different intent. The use of distributions and statistical tests will strengthen legal arguments about this differing intent, making them more solidly based on science. It is important to note that these distributions should not be considered in isolation but combined with other characteristics indicative of market manipulation. As such, visualizations like those presented in the paper offer a more intuitive understanding of what is happening in the market. This is especially true and helpful when a spoofing case is brought to court, where people with non-trading backgrounds must decide what is and what is not genuine market behavior.

Legitimate vs. illegitimate behavior. While Chapters 4 and 5 help to define and identify spoofing, differentiating between legitimate and illegitimate behavior remains a complex challenge, especially since some legitimate trading behavior falls under the definition of spoofing. Leuchtkafer (2015), for example, raises the question at what point market making becomes spoofing and illustrates this with an example of an HFT market maker posting on eight exchanges and cancelling on the other seven when any one of their orders trades. Hence, this HFT market maker fully intended to cancel seven out of eight orders. While this is common practice in HFT market making and not perceived as spoofing, available liquidity thus gets misjudged as they are "offer[ing] more liquidity than they're prepared to trade in one go" (Leuchtkafer, 2015, p.6). Market makers are an important category of traders, accounting for most of the market activity. Due to the nature of our data, we have not been able to consider this difference between spoofing and HFT market making, either in definitions or in the identification of spoofing. Our data does not include labels that categorize market participants, which would be needed to differentiate between market makers and other types of market participants. Still, this dissertation can help to resolve this issue. For example, the conceptual framework from Chapter 4 can offer guidance in making the case why this market-making behavior should (not) be labeled as spoofing or manipulative behavior. Furthermore, the discussion by Leuchtkafer raises an important issue with market manipulation: it is not only about the *intention* of the trader, but also about the *misleading impact* on the market. This is also evident from the analyzed spoofing cases in Chapter 4, most of which report a combination of intention and misleading impact on market participants. Spoofing and the aforementioned HFT market-making practice both mislead market participants by suggesting more volume than is actually the case. In the event of spoofing, the spoof orders are intentionally used to mislead market participants and make them behave in ways they would otherwise not have behaved. So, is this also the case for the HFT market-making practice described above? Perhaps. Or do we have to ask a different question: is this behavior induced by the market's microstructure? Market makers are, after all, incentivized to add liquidity to the market resulting in, among others, the spoofing-like behavior described above. The market microstructure is a critical, yet poorly studied factor that can encourage or discourage market manipulation and requires more research. The sections below highlight two elements of the market microstructure: iceberg orders and trading front ends.

Iceberg orders. Following the line of reasoning that liquidity is misjudged due to HFT market making (Leuchtkafer, 2015), iceberg orders can also be viewed as misleading. Iceberg orders are orders that display only a fraction of the total order while the rest of the volume is hidden for other market participants (Buti & Rindi, 2013; Shang et al., 2021). They play an important role for traders wishing to trade large volumes without incurring additional liquidity costs, such as institutional investors and hedgers. However, these orders can cause market participants to misjudge liquidity, since *less* liquidity is being offered than traders are actually willing to trade in one go. Does this mean that iceberg orders should be considered misleading? Do iceberg orders make traders act in ways they would otherwise not have? If so, are iceberg orders desirable? Iceberg orders can also hide the market manipulator's true buy or sell interest. Functionalities like these, which are implemented with good intentions, should be carefully considered since they can also be abused and exploited. Careful evaluation is necessary to determine whether the benefits outweigh the costs in this challenging trade-off.

Trading front ends. Besides the functionalities that the trading platforms of exchanges provide, market participants can use trading platforms by independent software vendors. Previous spoofing cases have demonstrated that these independent trading platforms can provide features that can be exploited for market manipulation, even when these features are sometimes implemented exactly to prevent market manipulation. Two examples are the SuperGUI system in the Krishna Mohan spoofing case (CFTC, 2018d) and the avoid-orders-that-cross functionality in the Igor Oystacher spoofing case (CFTC, 2015a). This raises the question if the added features of these front ends are truly necessary, or if they facilitate market manipulation. More research is needed into the impact of these independent trading front ends on market manipulation.

Who does it benefit? The main purposes of futures markets are to facilitate risk management through hedging and to enhance price discovery. The discussion about the market microstructure – including incentives, market design and trading front ends – raises the question to what extent additional features, faster trading, etc. enhance or hinder the purposes of futures markets. For example, does the HFT arms race benefit futures markets? Do price limits and different matching algorithms – or the market design in general – benefit the price discovery process of futures markets? Are all these features and functionalities necessary and beneficial for the functioning of futures markets, or do they make trading overly complex? Insights and solutions can also be drawn from other types of markets and industries. Consider, for example, the Dutch fruit and vegetable market, where one of the trading mechanisms is the online auction clock. For each product, the starting price is set at a specified amount, which decreases stepwise until a buyer presses a button to stop the price. The buyer then purchases the product at the price at which they stopped the clock. While the trading system itself might not be of interest to financial markets, the concept underlying the auction clock is an interesting one: it ensures that the earliest buyer secures the product, independent of any network latency. Behind the scenes, the auction system logs the time from when the price begins to decline to when a buyer presses the button. The buyer who responds the fastest gets the product. This also means that sometimes the price on the clock "bounces back up", when the system records that one buyer stopped the price but then receives new information that another buyer was even faster. Thus, the concept relies on buyers' actual reaction times, rather than their proximity to the server or levels of latency. Closely related to this is the implementation of CERN's White Rabbit (CERN, 2024) at Deutsche Börse, which synchronizes the clocks of traders to the same time source used by Deutsche Börse with a sub-nanosecond accuracy (Deutsche Börse Group, 2024; Eurex, 2019). This ensures accurate time synchronization across their trading platforms and accurate market activity timestamps. A natural next step could be to match trades as described above, relying on reaction time rather than proximity to the server and minimal latency. These kinds of differing mechanisms can provide alternative perspectives when designing markets.

Uniting market surveillance. A shift in market surveillance is necessary for regulatory agencies to keep up with the rapidly evolving high-frequency markets. Particularly in Europe, where relatively few spoofing cases have been reported and markets are fragmented – and, hence, many regulatory agencies exist, facing the same challenges regarding financial market data and market surveillance. Recent efforts to unify Europe's fragmented capital markets through central supervision have proven unsuccessful (Tamma et al., 2024). However, a feasible middle ground might be to create a strong collaboration. Rather than trying to solve the same challenges individually, regulatory agencies across the world need to collaborate more in their market surveillance efforts. Alongside pooling resources and coordinating responses, the sharing of information, best practices and data would strengthen financial markets and benefit all stakeholders. For example, both regulators and exchanges would benefit from 1) sharing market surveillance experiences; 2) creating common ground; and 3) gaining more clarity on how to interpret the law and identify market manipulation. Moreover, it would also allow for better cross-exchange market surveillance. This clarity would benefit market participants by providing them with a better understanding of what is considered legitimate and illegitimate trading behavior. The conceptual framework in Chapter 4 could help to create a common understanding of and language for market manipulation practices. Once regulatory agencies develop this common language, countries can conduct business more quickly, both individually and collectively, creating an equal playing field for all. A worldwide market surveillance collaboration would transform the global market surveillance landscape, ensuring fairer and more resilient financial markets.

6.4 LIMITATIONS AND FUTURE RESEARCH

This dissertation has several general limitations besides those outlined for each specific study. First, its main research question is formulated rather broadly and has no straightforward answer, since different methodologies produce different answers. Its focus is on spoofing as a type of market manipulation, and it applies particle physics methodologies and tools to answer the main research question. However, spoofing is just one of many types of market manipulations and other methodologies can be applied as well, such as a qualitative or a legal approach. Moreover, it might be concluded from the main research question that the goal is to develop a detection tool to identify and analyze market manipulation in a high-frequency context. However, we do not provide such a detection tool, though the separate studies can be considered building blocks towards such a tool. Second, our attempts to include the legal perspective have only scratched the surface. While Chapter 4 includes legal definitions and rulings from regulatory agencies, it does not delve any deeper into the legal complexities and court decisions involved. Future research could seek to combine the legal and economic components and integrate, for example, judges' decision models. Third, some of the results in this dissertation are ambiguous. Though Chapter 5 demonstrates that liquidity improves after spoofing in some agricultural futures markets, this does not mean that spoofing is 'good' for these markets. Then, there are the (inevitable) omissions: Chapter 3 briefly discussed the impact of spoofing on the price but does not include other factors that could also be relevant, such as its impact on price discovery and trust in the market. Also, we focus on liquidity (costs), but other (macro-economic) costs are not considered. For example, the impact of spoofing on the hedging effectiveness of futures markets, which in turn can affect the capital costs and the viability of organizations. Other methodologies than the ones rooted in particle physics could be more suitable for these types of studies.

Future research can apply the methodologies from this dissertation – i.e., the delineation, characterization, identification, visualization and analysis of spoofing – to other types of market manipulation, other types of markets (e.g., stock and energy markets) and different countries. Research can be extended by studying other types of market impact and economic costs, such as the impact of market manipulation on the perceived trustworthiness of markets or on the final price consumers pay. Moreover, Chapter 5 demonstrated that the irregular spacing between messages significantly impacts results. More methodological research is needed on how to address this irregular spacing and to evaluate if it is relevant for all types of research questions. Also, the visualizations in Chapter 2 and 3 can be extended to include more variables indicative of market manipulation. Given that the spoof orders produce a sharp color contrast in the visualizations (Chapter 3), which was expressed in the LOB state variable (Chapter 5), one might ask whether other elements can also be expressed in a similar manner. If so, are these elements worth visualizing at all, or is the visualization only beneficial as a support tool for inspecting market manipulation?

Future research can also address the discussion points raised in section 6.3 and the limitations described above. In short, this means addressing the issues of 1) the evidence needed to prove market manipulation; 2) the difference between (HFT) market making and manipulative practices; 3) market manipulation in the market microstructure context (e.g., iceberg orders, trading front ends and market design); 4) achieving a deeper integration between the legal and economic perspective, or other perspectives; and 5) the application of methodologies and tools from other disciplines to financial market big data.

6.5 CLOSING

While applying particle physics tools to financial market data may not be an obvious approach, we have proven that it is a highly effective one. As the volume and complexity of financial markets continue to grow, the integration of diverse fields becomes increasingly important. This applies not only to financial markets and big data, but also to other industries and issues involving smaller data. Scientists and industry stakeholders can benefit from stepping outside their comfort zones and drawing lessons from other disciplines. This requires an open mindset and a willingness to abandon the 'we have always done it this way' mentality, which is essential in a rapidly evolving world. This dissertation illustrated the value of integrating highly diverse scientific fields to address complex challenges, in hopes of inspiring future research to do the same.

References

Aidov, A., & Daigler, R. T. (2015). Depth Characteristics for the Electronic Futures Limit Order Book. *The Journal of Futures Markets*, *35*(6), 542–560. https://doi.org/https://doi.org/10.1002/fut.21706

Aldridge, I., & Krawciw, S. (2017). *Real-Time Risk: What Investors Should Know About FinTech, High-Frequency Trading, and Flash Crashes*. John Wiley & Sons. https://doi.org/10.1002/9781119319030

Antcheva, I., Ballintijn, M., Bellenot, B., Biskup, M., Brun, R., Buncic, N., Canal, Ph., Casadei, D., Couet, O., Fine, V., Franco, L., Ganis, G., Gheata, A., Maline, D. G., Goto, M., Iwaszkiewicz, J., Kreshuk, A., Segura, D. M., Maunder, R., … Tadel, M. (2009). ROOT — A C++ framework for petabyte data storage, statistical analysis and visualization. *Computer Physics Communications*, *180*(12), 2499–2512. https://doi.org/10.1016/j.cpc.2009.08.005

Arzandeh, M., & Frank, J. (2019). Price Discovery in Agricultural Futures Markets: Should We Look beyond the Best Bid-Ask Spread? *American Journal of Agricultural Economics*, *101*(5), 1482–1498. https://doi.org/10.1093/ajae/aaz001

Baruch, S., Panayides, M., & Venkataraman, K. (2017). Informed trading and price discovery before corporate events. *Journal of Financial Economics*, *125*(3), 561–588. https://doi.org/10.1016/j.jfineco.2017.05.008

Battalio, R., Corwin, S. A., & Jennings, R. (2016). Can Brokers Have It All? On the Relation between Make-Take Fees and Limit Order Execution Quality. *The Journal of Finance*, *71*(5), 2193–2238. https://doi.org/10.1111/jofi.12422

Bayraktar, E., & Munk, A. (2018). Mini-Flash Crashes, Model Risk, and Optimal Execution. *Market Microstructure and Liquidity*, *04*(01n02), 1850010. https://doi.org/10.1142/S2382626618500107

Bhattacharya, U., Kuo, W.-Y., Lin, T.-C., & Zhao, J. (2018). Do Superstitious Traders Lose Money? *Management Science*, *64*(8), 3772–3791. https://doi.org/10.1287/mnsc.2016.2701

Biais, B., Bisière, C., & Spatt, C. (2010). Imperfect Competition in Financial Markets: An Empirical Study of Island and Nasdaq. *Management Science*, *56*(12), 2237–2250. https://doi.org/10.1287/mnsc.1100.1243

Bongiovanni, S., Borkovec, M., & Sinclair, R. D. (2006). Let's Play Hide-and-Seek. *The Journal of Trading*, *1*(3), 34–46. https://doi.org/10.3905/jot.2006.644087

Brogaard, J. (2010). *High frequency trading and its impact on market quality*.

Brogaard, J., & Garriott, C. (2019). High-Frequency Trading Competition. *Journal of Financial and Quantitative Analysis*, *54*(4), 1469–1497. https://doi.org/10.1017/S0022109018001175

Brogaard, J., Hendershott, T., & Riordan, R. (2014). High-Frequency Trading and Price Discovery. *Review of Financial Studies*, *27*(8), 2267–2306. https://doi.org/10.1093/rfs/hhu032

Brogaard, J., Hendershott, T., & Riordan, R. (2019). Price Discovery without Trading: Evidence from Limit Orders. *The Journal of Finance*, *74*(4), 1621–1658. https://doi.org/10.1111/jofi.12769

Brogaard, J., Li, D., & Yang, J. (2022). Does High Frequency Market Manipulation Harm Market Quality? *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4280120

Brolley, M. (2020). Price Improvement and Execution Risk in Lit and Dark Markets. *Management Science*, *66*(2), 863–886. https://doi.org/10.1287/mnsc.2018.3204

Brun, R., & Rademakers, F. (1997). ROOT - an object oriented data analysis framework. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, *389*(1–2), 81–86.

Buti, S., & Rindi, B. (2013). Undisclosed orders and optimal submission strategies in a limit order market. *Journal of Financial Economics*, *109*(3), 797–812. https://doi.org/10.1016/j.jfineco.2013.04.002

Buti, S., Rindi, B., & Werner, I. M. (2017). Dark pool trading strategies, market quality and welfare. *Journal of Financial Economics*, *124*(2), 244–265. https://doi.org/10.1016/j.jfineco.2016.02.002

Canellos, G. S., Rangwala, T. S., Walfish, D. R., Jou, J. K., & Palladino, S. L. (2016). *The Law Surrounding Spoofing in the Derivatives and Securities Markets*. https://www.milbank.com/a/web/24678/Spoofing-in-the-Derivatives-and-Securities-Markets-June-22016.pdf

Cao, C., Hansch, O., & Wang, X. (2009). The information content of an open limit-order book. *Journal of Futures Markets*, *29*(1), 16–41. https://doi.org/10.1002/fut.20334

Cao, Y., Li, Y., Coleman, S., Belatreche, A., & McGinnity, T. M. (2014). Detecting price manipulation in the financial market. *2014 IEEE Conference on Computational Intelligence for Financial Engineering & Economics (CIFEr)*, 77–84. https://doi.org/10.1109/CIFEr.2014.6924057

Cao, Y., Li, Y., Coleman, S., Belatreche, A., & McGinnity, T. M. (2015). Adaptive Hidden Markov Model With Anomaly States for Price Manipulation Detection. *IEEE Transactions on Neural Networks and Learning Systems*, *26*(2), 318–330. https://doi.org/10.1109/TNNLS.2014.2315042

Capponi, A., Menkveld, A. J., & Zhang, H. (2019). Large Orders in Small Markets: On Optimal Execution with Endogenous Liquidity Supply. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3326313

Cartea, Á., Jaimungal, S., & Wang, Y. (2020). Spoofing and Price Manipulation in Order-Driven Markets. *Applied Mathematical Finance*, *27*(1–2), 67–98. https://doi.org/10.1080/1350486X.2020.1726783

CBOT. (2022). *CBOT Rulebook: Chapter 5*. https://www.cmegroup.com/content/dam/cmegroup/rulebook/CBOT/I/5.pdf

Cenesizoglu, T., & Grass, G. (2018). Bid- and ask-side liquidity in the NYSE limit order book. *Journal of Financial Markets*, *38*, 14–38. https://doi.org/10.1016/j.finmar.2017.10.002

CERN. (2018a). *About ROOT*. https://root.cern.ch/about-root

CERN. (2018b). *ROOT*. https://root.cern

CERN. (2020a). *Physics*. https://home.cern/science/physics

CERN. (2020b). *The Large Hadron Collider*. https://home.cern/science/accelerators/large-hadron-collider

CERN. (2020c). *The Standard Model*. https://home.cern/science/physics/standard-model

CERN. (2021a). *RooFit*. https://root.cern.ch/doc/master/group__Roofitmain.html

CERN. (2021b). *RooStats*. https://root.cern.ch/doc/v606/group__Roostats.html

CERN. (2021c). *TMVA*. https://root.cern.ch/tmva

CERN. (2023a). *Physics*. https://home.cern/science/physics

CERN. (2023b). *The SWAN Service*. https://swan.web.cern.ch/swan/

CERN. (2024). *The White Rabbit Project*. https://white-rabbit.web.cern.ch/

CFTC. (2015a). *Complaint for Injunctive and Other Equitable Relief and Civil Monetary Penalties Under the Commodity Exchange Act, Case Number: 15-cv-09196*. https://www.cftc.gov/sites/default/files/idc/groups/public/@lrenforcementactions/documents/legalpleading/enfigorcomplnt101915.pdf

CFTC. (2015b). *Complaint for Injunctive Relief, Civil Monetary Penalties, and Other Equitable Relief, Case Number: 15-cv-03398*. https://www.cftc.gov/sites/default/files/idc/groups/public/@lrenforcementactions/documents/legalpleading/enfsaraocomplaint041715.pdf

CFTC. (2017a). *Order Instituting Proceedings Pursuant to Section 6(c) and (d) of the Commodity Exchange Act, Making Findings, and Imposing Remedial Sanctions, Case Number: 17-06*. https://www.cftc.gov/sites/default/files/idc/groups/public/@lrenforcementactions/documents/legalpleading/enfcitigroupglobalorder011917.pdf

CFTC. (2017b). *Order Instituting Proceedings Pursuant to Section 6(c) and (d) of the Commodity Exchange Act, Making Findings, and Imposing Remedial Sanctions, Case Number: 17-14*. https://www.cftc.gov/sites/default/files/idc/groups/public/@lrenforcementactions/documents/legalpleading/enfdavidlieworder060217.pdf

CFTC. (2018a). *Complaint for Injunctive Relief, Civil Monetary Penalties, and Other Equitable Relief, Case Number: 18-cv-00158*. https://www.cftc.gov/sites/default/files/idc/groups/public/@lrenforcementactions/documents/legalpleading/enfandreflotroncomplaint012918.pdf

CFTC. (2018b). *Complaint for Injunctive Relief, Civil Monetary Penalties, and Other Equitable Relief, Case Number: 18-cv-00603*. https://www.cftc.gov/sites/default/files/idc/groups/public/@lrenforcementactions/documents/legalpleading/enfvorleychanucomplaint012618.pdf

CFTC. (2018c). *Complaint for Injunctive Relief, Civil Monetary Penalties, and Other Equitable Relief, Case Number: 18-cv-00619*. https://www.cftc.gov/sites/default/files/idc/groups/public/@lrenforcementactions/documents/legalpleading/enfthakkarcomplaint012818.pdf

CFTC. (2018d). *Complaint for Injunctive Relief, Civil Monetary Penalties, and Other Relief, Case Number: 18-cv-00260*. https://www.cftc.gov/sites/default/files/idc/groups/public/@lrenforcementactions/documents/legalpleading/enfmohancomplaint012818.pdf

CFTC. (2018e). *Complaint for Injunctive Relief, Civil Monetary Penalties, and Other Relief, Case Number: 18-cv-00620*. https://www.cftc.gov/sites/default/files/idc/groups/public/@lrenforcementactions/documents/legalpleading/enfjiongshengorder012918.pdf

R

CFTC. (2018f). *Order Instituting Proceedings Pursuant to Section 6(c) and (d) of the Commodity Exchange Act, Making Findings, and Imposing Remedial Sanctions, Case Number: 18-06*. https://www.cftc.gov/sites/default/files/idc/groups/public/@lrenforcementactions/documents/legalpleading/enfdeutschebankagorder012918.pdf

CFTC. (2018g). *Order Instituting Proceedings Pursuant to Section 6(c) and (d) of the Commodity Exchange Act, Making Findings, and Imposing Remedial Sanctions, Case Number: 18-07*. https://www.cftc.gov/sites/default/files/idc/groups/public/@lrenforcementactions/documents/legalpleading/enfusbagorder012918.pdf

CFTC. (2018h). *Order Instituting Proceedings Pursuant to Section 6(c) and (d) of the Commodity Exchange Act, Making Findings, and Imposing Remedial Sanctions, Case Number: 19-01*. https://www.cftc.gov/sites/default/files/2018-10/enfkamaldeepdandhiorder101118.pdf

CFTC. (2019a). *Complaint for Injunctive Relief, Civil Monetary Penalties, and Other Equitable Relief, Case Number: 19-cv-6163*. https://www.cftc.gov/media/2506/enfmichaelgreggcomplaint091619/download

CFTC. (2019b). *Order Instituting Proceedings Pursuant to Section 6(c) and (d) of the Commodity Exchange Act, Making Findings, and Imposing Remedial Sanctions, Case Number: 19-07*. https://www.cftc.gov/media/2141/enfmerrilllynchorder062519/download

CFTC. (2019c). *ORDER INSTITUTING PROCEEDINGS PURSUANT TO SECTION 6(c) AND (d) OF THE COMMODITY EXCHANGE ACT, MAKING FINDINGS, AND IMPOSING REMEDIAL SANCTIONS, Case Number: 19-26*. https://www.cftc.gov/media/2516/enfchristiantrunzorder091619/download

CFTC. (2019d). *Order Instituting Proceedings Pursuant to Section 6(c) and (d) of the Commodity Exchange Act, Making Findings, and Imposing Remedial Sanctions, Case Number: 19-31*. https://www.cftc.gov/media/2601/enfigorchernomzavorder093019/download

CFTC. (2019e). *Order Instituting Proceedings Pursuant to Section 6(c) and (d) of the Commodity Exchange Act, Making Findings, and Imposing Remedial Sanctions, Case Number: 19-46*. https://www.cftc.gov/media/2681/enfmitsubishiorder093019/download

CFTC. (2019f). *Order Instituting Proceedings Pursuant to Section 6(c) and (d) of the Commodity Exchange Act, Making Findings, and Imposing Remedial Sanctions, Case Number: 20-06*. https://www.cftc.gov/media/2986/enftowerresearchorder110619/download

CFTC. (2020a). *CFTC Orders JPMorgan to Pay Record $920 Million for Spoofing and Manipulation*. https://www.cftc.gov/PressRoom/PressReleases/8260-20

CFTC. (2020b). *Complaint for Injunctive Relief, Civil Monetary Penalties, and Other Equitable Relief Case Number: 20-cv-5777*. https://www.cftc.gov/media/4831/enfromancomplaint092920/download

CFTC. (2020c). *Order Instituting Proceedings Pursuant to Section 6(c) and (d) of the Commodity Exchange Act, Making Findings, and Imposing Remedial Sanctions, Case Number 20-17*. https://www.cftc.gov/media/4026/enfdeutschebanksecuritiesorder061820/download

CFTC. (2020d). *Order Instituting Proceedings Pursuant to Section 6(c) and (d) of the Commodity Exchange Act, Making Findings, and Imposing Remedial Sanctions, Case Number: 20-27*. https://www.cftc.gov/media/4411/enfbankofnovascotiaspoofingorder081920/download

CFTC. (2020e). *Order Instituting Proceedings Pursuant to Section 6(c) and (d) of the Commodity Exchange Act, Making Findings, and Imposing Remedial Sanctions, Case Number: 20-69*. https://www.cftc.gov/media/4826/enfjpmorganchaseorder092920/download

CFTC. (2022). *Complaint for Injunctive Relief, Civil Monetary Penalties, and Other Relief, Case Number: 22-cv-01925*. https://www.cftc.gov/media/7166/enfdavidskuddercomplaint041422/download

CFTC-SEC. (2010). *Findings Regarding The Market Events of May 6, 2010*.

Chakrabarty, B., & Shaw, K. W. (2008). Hidden Liquidity: Order Exposure Strategies Around Earnings Announcements. *Journal of Business Finance & Accounting*, *35*(9–10), 1220–1244. https://doi.org/10.1111/j.1468-5957.2008.02111.x

Chen, J., & Hsieh, P.-F. (2023). Price Impact from Order Revision: Evidence from Order Spoofing. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4304143

Chen, N., Kou, S., & Wang, C. (2018). A Partitioning Algorithm for Markov Decision Processes with Applications to Market Microstructure. *Management Science*, *64*(2), 784–803. https://doi.org/10.1287/mnsc.2016.2639

Chen, T., Gao, Z., He, J., Jiang, W., & Xiong, W. (2019). Daily price limits and destructive market behavior. *Journal of Econometrics*, *208*(1), 249–264. https://doi.org/10.1016/j.jeconom.2018.09.014

Chordia, T., Hu, J., Subrahmanyam, A., & Tong, Q. (2019). Order Flow Volatility and Equity Costs of Capital. *Management Science*, *65*(4), 1520–1551. https://doi.org/10.1287/mnsc.2017.2848

CME. (2022). *CME Rulebook: Chapter 5*. https://www.cmegroup.com/content/dam/cmegroup/rulebook/CME/I/5/5.pdf

CME Group. (2013a). *Notice of Disciplinary Action CBOT 11-8581-BC*. https://www.cmegroup.com/tools-information/lookups/advisories/disciplinary/CBOT-11-8581-BC-PANTHER-ENERGY-TRADING.html

CME Group. (2013b). *Notice of Disciplinary Action CME 11-8581-BC*. https://www.cmegroup.com/tools-information/lookups/advisories/disciplinary/CME-11-8581-BC-MICHAEL-COSCIA.html

CME Group. (2014). *CFTC Regulation 40.6(a) Certification. Adoption of Rule 575 ("Disruptive Practices Prohibited") and Issuance of CME Group Market Regulation Advisory Notice RA1405-5. CME Submission No. 14-367 (1 of 4)*. https://www.cftc.gov/sites/default/files/filings/orgrules/14/08/rule082814cmedcm001.pdf

CME Group. (2017a). *Notice of Disciplinary Action CBOT-15-0238-BC*. https://www.cmegroup.com/notices/disciplinary/2017/12/cbot-15-0238-bc-marc-christopher-michelotti.html

CME Group. (2017b). *Notice of Disciplinary Action CBOT-16-0416-BC*. https://www.cmegroup.com/notices/disciplinary/2017/09/cbot-16-0416-bc-andrew-anaszewicz.html

CME Group. (2018). *Notice of Disciplinary Action NYMEX 16-0504-BC*. https://www.cmegroup.com/notices/disciplinary/2018/02/NYMEX-16-0504-BC-PETER-C-MILLER.html

CME Group. (2019). *Notice of Disciplinary Action CME 15-0334-BC*. https://www.cmegroup.com/notices/disciplinary/2019/09/CME-15-0334-BC-GREGORY-KOBUS.html

CME Group. (2020a). *Develop to CME Globex*. https://www.cmegroup.com/globex/develop-to-cme-globex.html

CME Group. (2020b). *MDP 3.0 - Market by Price - Multiple Depth Book*. https://www.cmegroup.com/confluence/display/EPICSANDBOX/MDP+3.0+-+Market+by+Price+-+Multiple+Depth+Book

Colliard, J.-E., & Hoffmann, P. (2017). Financial Transaction Taxes, Market Composition, and Liquidity. *The Journal of Finance*, *72*(6), 2685–2716. https://doi.org/10.1111/jofi.12510

Comerton-Forde, C., Malinova, K., & Park, A. (2018). Regulating dark trading: Order flow segmentation and market quality. *Journal of Financial Economics*, *130*(2), 347–366. https://doi.org/10.1016/j.jfineco.2018.07.002

Cont, R., Kukanov, A., & Stoikov, S. (2014). The Price Impact of Order Book Events. *Journal of Financial Econometrics*, *12*(1), 47–88. https://doi.org/10.1093/jjfinec/nbt003

Coppler, C. (2015). The Anti-Spoofing Statue: Vague as Applied to the Hypothetically Legitimate Trader. *American University Business Law Review*, *5*, 261–290.

Couleau, A., Serra, T., & Garcia, P. (2020). Are Corn Futures Prices Getting "Jumpy"? *American Journal of Agricultural Economics*, *102*(2), 569–588. https://doi.org/10.1002/ajae.12030

Dacorogna, M. M., Gençay, R., Müller, U. A., Olsen, R. B., & Pictet, O. V. (2001). *An Introduction to High-Frequency Finance*. Elsevier. https://doi.org/10.1016/B978-0-12-279671-5.X5000-X

Dalko, V., Michael, B., & Wang, M. (2020). Spoofing: effective market power building through perception alignment. *Studies in Economics and Finance*, *37*(3), 497–511. https://doi.org/10.1108/SEF-09-2019-0346

Dalko, V., & Wang, M. H. (2018). How Effective are the Order-to-Trade Ratio and Resting Time Regulations? *Journal of Financial Regulation*, *4*(2), 321–325. https://doi.org/10.1093/jfr/fjy007

Dalko, V., & Wang, M. H. (2020a). High-frequency trading: Order-based innovation or manipulation? *Journal of Banking Regulation*, *21*(4), 289–298. https://doi.org/10.1057/s41261-019-00115-y

Dalko, V., & Wang, M. H. (2020b). Volume Limit: An Effective Response to the India Flash Crash? *Journal of Financial Regulation*, *5*(2), 249–255. https://doi.org/10.1093/jfr/fjz006

de Boer, T. A. P., Gardebroek, C., Pennings, J. M. E., & Trujillo-Barrera, A. (2022). Intraday liquidity in soybean complex futures markets. *Journal of Futures Markets*, *42*(7), 1189–1211. https://doi.org/10.1002/fut.22325

Deb, S. S., Kalev, P. S., & Marisetty, V. B. (2010). Are price limits really bad for equity markets? *Journal of Banking & Finance*, *34*(10), 2462–2471. https://doi.org/10.1016/j.jbankfin.2010.04.001

Debie, P., Gardebroek, C., Hageboeck, S., Leeuwen, P., Moneta, L., Naumann, A., Pennings, J. M. E., Trujillo-Barrera, A. A., & Verhulst, M. E. (2022). Unravelling the JPMorgan spoofing case using particle physics visualization methods. *European Financial Management*. https://doi.org/10.1111/eufm.12353

Deutsche Börse Group. (2024). *Time services*. https://www.deutsche-boerse.com/dbg-en/products-services/ps-technology/ps-connectivity-services/ps-connectivity-services-time-services

Do, B. L., & Putnins, T. J. (2023). Detecting Layering and Spoofing in Markets. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4525036

DOJ. (2022). *Ex-Wall Street Trader Convicted of Fraud in Precious Metals Spoofing Scheme*. https://www.justice.gov/opa/pr/ex-wall-street-trader-convicted-fraud-precious-metals-spoofing-scheme

Du, S., & Zhu, H. (2017). What is the Optimal Trading Frequency in Financial Markets? *The Review of Economic Studies*, *84*(4), 1606–1651. https://doi.org/10.1093/restud/rdx006

Dugast, J. (2018). Unscheduled News and Market Dynamics. *The Journal of Finance*, *73*(6), 2537–2586. https://doi.org/10.1111/jofi.12717

Easley, D., & O'Hara, M. (1987). Price, trade size, and information in securities markets. *Journal of Financial Economics*, *19*(1), 69–90. https://doi.org/https://doi.org/10.1016/0304-405X(87)90029-8

Engle, R. F., & Russell, J. R. (1998). Autoregressive Conditional Duration: A New Model for Irregularly Spaced Transaction Data. *Econometrica*, *66*(5), 1127–1162. https://doi.org/10.2307/2999632

Erenburg, G., & Lasser, D. (2009). Electronic limit order book and order submission choice around macroeconomic news. *Review of Financial Economics*, *18*(4), 172–182. https://doi.org/https://doi.org/10.1016/j.rfe.2009.06.001

Eurex. (2019). *High Precision Time (White Rabbit) Pilot*. Eurex. https://www.eurex.com/ex-en/support/initiatives/archive/high-precision-time-white-rabbit-pilot

European Commission. (2015). *Commission Delegated Regulation (EU) 2016/522 of 17 December 2015*. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32016R0522&qid=1658412857818

Fabozzi, F. J., Focardi, S. M., & Jonas, C. (2011). High-frequency trading: methodologies and market impact. *Review of Futures Markets*, *9*(Special Issue), 7–38.

FINRA. (2022). *Security Futures*. https://www.finra.org/rules-guidance/key-topics/security-futures

FIXtrading. (2020). *Financial Information eXchange (FIX) Protocol*. https://www.fixtrading.org/what-is-fix/

Foucault, T., Moinas, S., & Theissen, E. (2007). Does Anonymity Matter in Electronic Limit Order Markets? *Review of Financial Studies*, *20*(5), 1707–1747. https://doi.org/10.1093/rfs/hhm027

Fox, M. B., Glosten, L. R., & Guan, S. S. (2021). Spoofing and its Regulation. *Columbia Business Law Review*, *1244*.

Gai, J., Choi, D. J., O'Neal, D., Ye, M., & Sinkovits, R. S. (2014). Fast construction of nanosecond level snapshots of financial markets. *Concurrency and Computation: Practice and Experience*, *26*(13), 2149–2156. https://doi.org/10.1002/cpe.3220

Golub, A., Keane, J., & Poon, S.-H. (2012). High Frequency Trading and Mini Flash Crashes. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2182097

Gomber, P., & Schweickert, U. (2002). The Market Impact - Liquidity Measure in Electronic Securities Trading. *Die Bank*, *7*. https://www.xetra.com/resource/blob/5480/784401adc4e0e19bf32e37edcb45926d/data/The-Market-Impact-Liquidity-Measure-in-Electronic-Securities-Trading.pdf

Gomber, P., Schweickert, U., & Theissen, E. (2015). Liquidity Dynamics in an Electronic Open Limit Order Book: an Event Study Approach. *European Financial Management*, *21*(1), 52–78. https://doi.org/10.1111/j.1468-036X.2013.12006.x

Hachmeister, A. (2007). *Informed Traders as Liquidity Providers: Evidence from the German Equity Market*. Deutscher Universitätsverlag Wiesbaden. https://doi.org/10.1007/978-3-8350-9577-9

Hasbrouck, J. (2017). Securities Trading: Principles and Procedures. *Manuscript, Version 12*, 1–195.

Hasbrouck, J. (2018). High-Frequency Quoting: Short-Term Volatility in Bids and Offers. *Journal of Financial and Quantitative Analysis*, *53*(2), 613–641. https://doi.org/10.1017/S0022109017001053

Hasbrouck, J. (2021). Securities Trading: Principles and Procedures. *Manuscript, Version 12c*, 1–217. https://pages.stern.nyu.edu/~jhasbrou/STPP/drafts/STPPms12c.pdf

Hautsch, N., & Horvath, A. (2019). How effective are trading pauses? *Journal of Financial Economics*, *131*(2), 378–403. https://doi.org/10.1016/j.jfineco.2017.12.011

He, X., Serra, T., & Garcia, P. (2021). Resilience in "Flash Events" in the Corn and Lean Hog Futures Markets. *American Journal of Agricultural Economics*, *103*(2), 743–764. https://doi.org/10.1111/ajae.12146

Hirsch, M., Cook, D., Lajbcygier, P., & Hyndman, R. (2019). Revealing High-Frequency Trading Provision of Liquidity with Visualization. *Proceedings of the 2nd International Conference on Software Engineering and Information Management*, 157–165.

ICE Futures U.S. (2021a). *Disciplinary Notice, Case Number: 2019-029*. https://www.theice.com/publicdocs/futures_us/disciplinary_notices/ICE_Futures_US_Doug_Moore_20210831.pdf

ICE Futures U.S. (2021b). *Disciplinary Notice, Case Number: 2019-043*. https://www.theice.com/publicdocs/futures_us/disciplinary_notices/ICE_Futures_US_Wai_Yui_Tam_20211105.pdf

ICE Futures U.S. (2022). *ICE Futures U.S.®, Inc. TRADING RULES*. https://www.theice.com/publicdocs/rulebooks/futures_us/4_Trading.pdf

Ito, T., & Yamada, M. (2018). Did the reform fix the London fix problem? *Journal of International Money and Finance*, *80*, 75–95. https://doi.org/10.1016/j.jimonfin.2017.10.004

Kahraman, B., & Tookes, H. E. (2017). Trader Leverage and Liquidity. *The Journal of Finance*, *72*(4), 1567–1610. https://doi.org/10.1111/jofi.12507

Kandel, E., Rindi, B., & Bosetti, L. (2012). The effect of a closing call auction on market quality and trading strategies. *Journal of Financial Intermediation*, *21*(1), 23–49. https://doi.org/10.1016/j.jfi.2011.03.002

Kerr, J., Sadka, G., & Sadka, R. (2019). Illiquidity and Price Informativeness. *Management Science*, *66*(1), 334–351. https://doi.org/https://doi.org/10.1287/mnsc.2018.3154

Kim, K. A., & Park, J. (2010). Why Do Price Limits Exist in Stock Markets? A Manipulation-Based Explanation. *European Financial Management*, *16*(2), 296–318. https://doi.org/10.1111/j.1468-036X.2008.00456.x

Kirilenko, A., Kyle, A. S., Samadi, M., & Tuzun, T. (2017). The Flash Crash: High-Frequency Trading in an Electronic Market. *The Journal of Finance*, *72*(3), 967–998. https://doi.org/10.1111/jofi.12498

Kyle, A. S. (1985). Continuous Auctions and Insider Trading. *Econometrica: Journal of the Econometric Society*, *53*(6), 1315–1335. https://doi.org/10.2307/1913210

Leangarun, T., Tangamchit, P., & Thajchayapong, S. (2016). Stock price manipulation detection based on mathematical models. *International Journal of Trade, Economics and Finance*, *7*(3), 81–88. https://doi.org/10.1109/ICACI.2016.7449848

Lee, E. J., Eom, K. S., & Park, K. S. (2013). Microstructure-based manipulation: Strategic behavior and performance of spoofing traders. *Journal of Financial Markets*, *16*(2), 227–252. https://doi.org/10.1016/j.finmar.2012.05.004

Leonard, G., Cao, Y., Haas, M., & Mocek, G. (2020). The legal and economic implications from recent UK spoofing cases. *Journal of Financial Compliance*, *4*(2), 179–193.

Leuchtkafer, R. T. (2015). *[Letter from R.T. Leuchtkafer to the Securities and Exchange Commission, 2015]*. https://www.sec.gov/comments/sr-bats-2015-57/bats201557-3.pdf

Li, H., Polukarov, M., & Ventre, C. (2023). Detecting Financial Market Manipulation with Statistical Physics Tools. *4th ACM International Conference on AI in Finance*, 1–1. https://doi.org/10.1145/3604237.3626871

Li, H., & Yang, S. Y. (2022). Impact of False Information from Spoofing Strategies: An ABM Model of Market Dynamics. *2022 IEEE Symposium on Computational Intelligence for Financial Engineering and Economics (CIFEr)*, 1–10. https://doi.org/10.1109/CIFEr52523.2022.9776070

Li, S., Wang, X., & Ye, M. (2019). *Who Provides Liquidity, and When?* (25972). https://doi.org/10.3386/w25972

Lin, T. C. W. (2017). The New Market Manipulation. *Emory Law Journal*, *66*, 1253–1314.

Linton, O., & Mahmoodzadeh, S. (2018). Implications of High-Frequency Trading for Security Markets. *Annual Review of Economics*, *10*(1), 237–259. https://doi.org/10.1146/annurev-economics-063016-104407

LLVM. (2021). *Clang: a C language family fronted for LLVM*. https://clang.llvm.org/

MacKenzie, D. (2021). *Trading at the Speed of Light*. Princeton University Press. https://doi.org/10.1515/9780691217796

MacKenzie, D. (2022). Spoofing: Law, materiality and boundary work in futures trading. *Economy and Society*, *51*(1), 1–22. https://doi.org/10.1080/03085147.2022.1987753

Mark, G. (2019). Spoofing and Layering. *The Journal of Corporation Law*, *45*(2), 101–169.

Martínez Miranda, E., Phelps, S., & Howard, M. J. (2019). Order flow dynamics for prediction of order cancelation and applications to detect market manipulation. *High Frequency*, *2*(1), 4–36. https://doi.org/10.1002/hf2.10026

Martínez-Miranda, E., McBurney, P., & Howard, M. J. W. (2016). Learning unfair trading: A market manipulation analysis from the reinforcement learning perspective. *2016 IEEE Conference on Evolving and Adaptive Intelligent Systems (EAIS)*, 103–109. https://doi.org/10.1109/EAIS.2016.7502499

Mendonça, L., & De Genaro, A. (2020). Detection and analysis of occurrences of spoofing in the Brazilian capital market. *Journal of Financial Regulation and Compliance*, *28*(3), 369–408. https://doi.org/10.1108/JFRC-07-2019-0092

Menkveld, A. J. (2013). High frequency trading and the new market makers. *Journal of Financial Markets*, *16*(4), 712–740. https://doi.org/10.1016/j.finmar.2013.06.006

Menkveld, A. J., & Yueshen, B. Z. (2019). The Flash Crash: A Cautionary Tale About Highly Fragmented Markets. *Management Science*, *65*(10), 4470–4488. https://doi.org/https://doi.org/10.1287/mnsc.2018.3040

Menkveld, A. J., & Zoican, M. A. (2017). Need for Speed? Exchange Latency and Liquidity. *The Review of Financial Studies*, *30*(4), 1188–1228. https://doi.org/10.1093/rfs/hhx006

Michaels, D. (2020, September 29). JPMorgan Paying $920 Million to Resolve Market Manipulation Probes. *Wall Street Journal*.

Montgomery, J. D. (2016). Spoofing, Market Manipulation, and the Limit-Order Book. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2780579

Nemko, M. (2016). Irrationality: The Eminents interview with Nobel Prize winner Robert Shiller. *Psychology Today*. https://www.psychologytoday.com/intl/blog/how-do-life/201610/irrationality

Neurensic. (2016). *Spoofing Similarity Model*. https://neurensic.com/spoofing-similarity-model/

NYMEX. (2022). *NYMEX Rulebook: Chapter 5*. https://www.cmegroup.com/content/dam/cmegroup/rulebook/NYMEX/1/5.pdf

Paddrik, M. E., Haynes, R., Todd, A. E., Scherer, W. T., & Beling, P. A. (2016). Visual Analysis to support regulators in electronic order book markets. *Environment Systems and Decisions*, *36*(2), 167–182. https://doi.org/https://doi.org/10.1007/s10669-016-9597-2

Park, Y.-H., & Abruzzo, N. (2016). An Empirical Analysis of Futures Margin Changes: Determinants and Policy Implications. *Journal of Financial Services Research*, *49*(1), 65–100. https://doi.org/10.1007/s10693-014-0212-8

Pennings, J. M. E., & Meulenberg, M. T. G. (1997a). Hedging efficiency: a futures exchange management approach. *Journal of Futures Markets*, *17*(5), 599–615.

Pennings, J. M. E., & Meulenberg, M. T. G. (1997b). The hedging performance in new agricultural futures markets: A note. *Agribusiness*, *13*(3), 295–300. https://doi.org/10.1002/(SICI)1520-6297(199705/06)13:3<295::AID-AGR4>3.0.CO;2-W

Pennings, J. M. E., Verhulst, M. E., Debie, P., Giltjes, M., Vletter-van Dort, H. M., van Dijck, G., Tekinerdogan, B., Gardebroek, C., Heijmans, R. H., Aerts, T. G., Broeders, D. W. G. A., Kiliç, M., Knook, J., Akerboom, S., Israel, M., Duivenvoorden, S., van Marle, E., Kennedy, N., Vos, K., … Ziogas, N. (2020). *Solving regulatory stalemates: developing a methodology to identify illegitimate behaviour using particle physics*. NWO Research proposal NWA.1334.19.006, NWA L2 – Theme 2019 Innovation of supervision, pages 1- 31.

Peress, J., & Schmidt, D. (2020). Glued to the TV: Distracted noise traders and stock market liquidity. *The Journal of Finance*, *75*(2), 1083–1133. https://doi.org/https://doi.org/10.1111/jofi.12863

Polansek, T. (2015). *High-frequency trader convicted in first U.S. spoofing case*. Reuters. https://www.reuters.com/article/us-court-spoofing-verdict-idUSKCN0SS2QQ20151104

Putniņš, T. J. (2012). Market Manipulation: A Survey. *Journal of Economic Surveys*, *26*(5), 952–967. https://doi.org/10.1111/j.1467-6419.2011.00692.x

Putniņš, T. J. (2020). An overview of market manipulation. *Corruption and Fraud in Financial Markets: Malpractice, Misconduct and Manipulation*, *13*.

Ranaldo, A. (2004). Order aggressiveness in limit order book markets. *Journal of Financial Markets*, *7*(1), 53–74. https://doi.org/10.1016/S1386-4181(02)00069-1

Rösch, C. G., & Kaserer, C. (2013). Market liquidity in the financial crisis: The role of liquidity commonality and flight-to-quality. *Journal of Banking & Finance*, *37*(7), 2284–2302. https://doi.org/10.1016/j.jbankfin.2013.01.009

Sahu, R. (2022). A Look into 'Market Manipulation' With a Focus on Pump and Dump Schemes and Initial Considerations for Responding to a Regulatory Query. *Business Law Review*, *43*(Issue 3), 114–121. https://doi.org/10.54648/BULA2022016

Sanders, J. I. (2016). Spoofing: A Proposal for Normalizing Divergent Securities and Commodities Futures Regimes. *Wake Forest Law Review*, *51*, 517–542.

Sar, M. (2017). Dodd-Frank and the spoofing prohibition in commodities markets. *Fordham Journal of Corporate & Financial Law*, *22*, 383–416.

Schoenberg, T., & Robinson, M. (2020, September 29). JPMorgan Admits Spoofing by 15 Traders, Two Desks in Record Deal. *Bloomberg*.

SEC. (2021). *SEC Charges North Carolina Day Trader for Manipulative Trading (Administrative Proceeding File No. 3-20599)*. https://www.sec.gov/enforce/33-10989-s

Sensoy, A. (2019). Commonality in ask-side vs. bid-side liquidity. *Finance Research Letters*, *28*, 198–207. https://doi.org/10.1016/j.frl.2018.04.020

Shah, S., & Brorsen, W. (2011). Electronic vs. Open Outcry: Side-by-Side Trading of KCBT Wheat Futures. *Journal of Agricultural and Resource Economics*, *36*(1), 48–62.

Shang, Q., Serra, T., Garcia, P., & Mallory, M. (2021). Looking under the surface: An analysis of iceberg orders in the U.S. agricultural futures markets. *Agricultural Economics*, *52*(4), 679–699. https://doi.org/10.1111/agec.12642

Shkilko, A., & Sokolov, K. (2020). Every Cloud Has a Silver Lining: Fast Trading, Microwave Connectivity, and Trading Costs. *The Journal of Finance*, *75*(6), 2899–2927. https://doi.org/10.1111/jofi.12969

Sinkovits, R., Feng, T., & Ye, M. (2014). Fast, Low-Memory Algorithm for Construction of Nanosecond Level Snapshots of Financial Markets. *Proceedings of the 2014 Annual Conference on Extreme Science and Engineering Discovery Environment - XSEDE '14*, 1–5. https://doi.org/10.1145/2616498.2616501

Smith, T. (2023). *An exabyte of disk storage at CERN*. https://home.web.cern.ch/news/news/computing/exabyte-disk-storage-cern

Stenfors, A., Dilshani, K., Guo, A., & Mere, P. (2023). A Model to Quantify the Risk of Cross-Product Manipulation: Evidence from the European Government Bond Futures Market. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4546523

Stenfors, A., Doraghi, M., Soviany, C., Susai, M., & Vakili, K. (2023). Cross-market spoofing. *Journal of International Financial Markets, Institutions and Money*, *83*, 101735. https://doi.org/10.1016/j.intfin.2023.101735

Stenfors, A., & Susai, M. (2021). Spoofing and pinging in foreign exchange markets. *Journal of International Financial Markets, Institutions and Money*, *70*, 101278. https://doi.org/10.1016/j.intfin.2020.101278

Tamma, P., Foy, H., & Hancock, A. (2024). Majority of EU states object to capital markets reform push. *Financial Times*. https://www.ft.com/content/6164fb0d-634f-444b-b7e5-069c590f24ca

Tao, X., Day, A., Ling, L., & Drapeau, S. (2022). On detecting spoofing strategies in high-frequency trading. *Quantitative Finance*, *22*(8), 1405–1425. https://doi.org/10.1080/14697688.2022.2059390

Tejedor, E., & Kothuri, P. (2018). *CERN's Platform for Data Analysis with Spark*. https://www.slideshare.net/databricks/cerns-next-generation-data-analysis-platform-with-apache-spark-with-enric-tejedor

Trebbi, F., & Xiao, K. (2019). Regulation and Market Liquidity. *Management Science*, *65*(5), 1949–1968. https://doi.org/https://doi.org/10.1287/mnsc.2017.2876

Tripathi, A., Vipul, V., & Dixit, A. (2020). Limit order books: a systematic review of literature. *Qualitative Research in Financial Markets*, *12*(4), 505–541. https://doi.org/10.1108/QRFM-07-2019-0080

United States. (2010). *Dodd-Frank Wall Street Reform and Consumer Protection Act: Conference Report (to Accompany HR 4173)*.

United States v. Coscia. (2017). *866 F.3d 782 (7th Cir. 2017)*. https://casetext.com/case/united-states-v-coscia-4

United States v. Williams. (2008). *553 U.S. 285*. https://supreme.justia.com/cases/federal/us/553/06-694/index.pdf

U.S. Code. (2023a). *7 USC 6c: Prohibited transactions*. https://uscode.house.gov/browse/prelim@title7/chapter1&edition=prelim

U.S. Code. (2023b). *18 USC 1343: Fraud by wire, radio, or television*. https://uscode.house.gov/browse/prelim@title18/part1/chapter63&edition=prelim

U.S. Congress. (1933). *Securities Act of 1933*. https://www.govinfo.gov/content/pkg/COMPS-1884/pdf/COMPS-1884.pdf

U.S. Congress. (1934). *Securities Exchange Act of 1934*. https://www.govinfo.gov/content/pkg/COMPS-1885/pdf/COMPS-1885.pdf

Verhulst, M. E., Debie, P., Hageboeck, S., Pennings, J. M. E., Gardebroek, C., Naumann, A., van Leeuwen, P., Trujillo-Barrera, A. A., & Moneta, L. (2021). When Two Worlds Collide: Using Particle Physics Tools to Visualize the Limit Order Book. *Journal of Futures Markets*, *41*(11), 1715–1734.

Verhulst, M. E., & Pennings, J. M. E. (2023). Spoofing in U.S. Futures Markets: An Interdisciplinary Approach. *[Working Paper]*.

Wang, X., Garcia, P., & Irwin, S. H. (2014). The Behavior of Bid-Ask Spreads in the Electronically-Traded Corn Futures Market. *American Journal of Agricultural Economics*, *96*(2), 557–577.

Wang, X., Hoang, C., Vorobeychik, Y., & Wellman, M. P. (2021). Spoofing the Limit Order Book: A Strategic Agent-Based Analysis. *Games*, *12*(2), 46. https://doi.org/10.3390/g12020046

Wang, X., Hoang, C., & Wellman, M. P. (2020). Learning-based trading strategies in the face of market manipulation. *Proceedings of the First ACM International Conference on AI in Finance*, 1–8. https://doi.org/10.1145/3383455.3422568

Wang, X., Vorobeychik, Y., & Wellman, M. P. (2018). A Cloaking Mechanism to Mitigate Market Manipulation. *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence*, 541–547. https://doi.org/10.24963/ijcai.2018/75

Wang, X., & Wellman, M. P. (2020). Market Manipulation: An Adversarial Learning Framework for Detection and Evasion. *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence*, 4626–4632. https://doi.org/10.24963/ijcai.2020/638

Wang, Y.-Y. (2015). Strategic spoofing order trading by different types of investors in the futures markets. In *Wall Street Journal*.

Williams, B., & Skrzypacz, A. (2020). Spoofing in Equilibrium. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3742327

Yang, S., Paddrik, M., Hayes, R., Todd, A., Kirilenko, A., Beling, P., & Scherer, W. (2012). Behavior based learning in identifying High Frequency Trading strategies. *2012 IEEE Conference on Computational Intelligence for Financial Engineering & Economics (CIFEr)*, 1–8. https://doi.org/10.1109/CIFEr.2012.6327783

Yao, C., & Ye, M. (2018). Why Trading Speed Matters: A Tale of Queue Rationing under Price Controls. *The Review of Financial Studies*, *31*(6), 2157–2183. https://doi.org/10.1093/rfs/hhy002

Zhai, J., Cao, Y., & Ding, X. (2018). Data analytic approach for manipulation detection in stock market. *Review of Quantitative Finance and Accounting*, *50*(3), 897–932. https://doi.org/10.1007/s11156-017-0650-0

Zhai, J., Cao, Y., Yao, Y., Ding, X., & Li, Y. (2017). Computational intelligent hybrid model for detecting disruptive trading activity. *Decision Support Systems*, *93*, 26–41. https://doi.org/10.1016/j.dss.2016.09.003

Appendices

Appendix 2.A

A LOB is a system for mapping all available demand for and supply of securities and financial instruments (futures and options) at a specific time. Table A.I illustrates what a LOB looks like.

Table 2.A.1 | Example of a LOB

	Level	Price	Volume
Ask (sell)	5	€55.00	2
	4	€54.60	10
	3	€53.50	2
	2	€53.00	4
	1	€50.00	9
Bid (buy)	1	€49.00	12
	2	€48.00	4
	3	€47.50	8
	4	€46.00	3
	5	€40.00	1

The top five rows are the first five levels on the ask side of the LOB; this is where orders rest from traders who wish to sell (supply). They are mapped in ascending order, with the first level indicating the lowest price at which a trader is willing to sell their security/instrument. The bottom five rows are the first five levels on the bid side of the LOB – i.e., where orders rest from traders wishing to buy (demand). The price levels are in descending order, with the first level indicating the highest price a trader is willing to pay for a security/instrument. Each price level has a volume associated with it. This is the aggregated volume for all traders wishing to buy/sell at that specific price level. Limit orders rest in the LOB until they are matched or cancelled. Market orders take the current market price (Arzandeh & Frank, 2019). Thus, a trader entering a sell (buy) market order will receive a price of €49.00 (€50.00) in the example presented in Table 2.A.1. The best bid and ask prices are the prices at the first level of the LOB, i.e., the highest bid and lowest ask price. The difference between the best ask and bid price is called the bid-ask spread; in Table 2.A.1, this spread is €1.00 (€50.00 – €49.00).

Appendix 2.B

Particle physics studies the fundamental constituents of matter (CERN, 2020a). One of its most prominent institutions is the European Organization for Nuclear Research (CERN). At CERN, physicists conduct research using the Large Hadron Collider (LHC) (CERN, 2020b). Although the LHC and LOB are different in nature, they show similarities in terms of the data they generate and the analysis techniques they require. LHC data is highly granular, as particle properties are measured by hundreds of thousands of sensors, producing a detailed stream of values. LHC software reconstructs the particles' physical properties from these measured values. Almost everything at the LHC is obeying distributions: the initial collision (what collides and how?), the initial collision products (how did the particles react?), the properties of the particles flying through the detector (how did the initial particles transform?) and the sensors' measurement uncertainties. Due to these distributions, a single measurement is of very limited value: the wealth of the data is only accessible through statistical analyses. These analyses are performed on distributions of physical properties and their correlations, measuring the significance of how well models of the fundamental laws of physics (e.g., the Standard Model[48] of particle physics) describe the collision products and looking for deviations from these models.

Much of this also applies to financial data. The market microstructure and the strive of market participants to make a profit or manage their risks define boundary conditions for trading actions, producing causality similar to that described in the laws of physics. While each single action might appear to follow a random distribution, behavioral patterns emerge from statistical analysis of the ensemble of actions. Here, measurements of the LHC correspond to actions and indicators of the market.

A notable difference, however, is the time dependence of events. While particle collisions are stochastically independent from each other, i.e., any previous collision has no effect on a subsequent collision, financial events exhibit a high degree of (temporal) correlation. Nevertheless, some of the tools and many of the approaches are applicable to financial data, analysis and visualization, opening up new opportunities for gaining knowledge and understanding from financial data.

48 The Standard Model of particle physics is a theory that explains how the basic building blocks of the universe (fundamental particles) are related to fundamental forces. It is a well-tested theory, as it has successfully explained almost all experimental results and precisely predicted phenomena (CERN, 2020c).

Appendix 2.C

MDP data provides the market messages required to recreate the LOB with millisecond precision. Each file contains all contracts of the same futures contract, ordered by message number and time of arrival. A contract's data is spread across multiple files, with adjacent time windows within each file. Each file contains three types of messages. Two types of messages are used for initialization and metadata (i.e., the "definition message" and the "security status message" respectively), and one is used for the incremental LOB updates. The definition and security status messages only exist once per contract per file. They contain information such as ID, name, expiration date, order book depth and tick size. The message that is used for incremental LOB updates consists of many different types of submessages. The main six submessages are messages to update the order book, i.e., insert a new bid or ask level, change existing bid or ask levels and delete existing bid or ask levels. Furthermore, there is one submessage that indicates when a trade took place and ten more submessages containing statistics such as opening price and settlement price (CME Group, 2020b).

To recreate the LOB, all messages are read and orders are processed in an iterative process. First, an empty order book of depth ten is initiated. Next, ten insertion submessages are initiated for the bid side and another ten for the ask side. Together, these twenty messages generate the first state of the LOB. Submessages that record changes in volume (at a certain level) replace the value of the quantity at the relevant price level entirely, meaning that these submessages contain the new actual quantity, rather than the volume that needs to be added or subtracted from the current quantity. If the volume at a certain level reaches zero, this level is deleted from the LOB with the delete-level submessage. If new volume is added to a previously non-existent level (zero volume), an insert-level submessage is sent (CME Group, 2020b).

Appendix 3.A

3.A.1 TRADITIONAL SPOOFING: 10-YEAR T-NOTE MARCH 2010 CONTRACT

The 10-Year T-Note March 2010 contract is visualized on February 4, 2010 from 13:27:20 to 13:30:05.

Table 3.A.1 | Spoofing actions on February 4, 2010 in the 10-Year T-Note March 2010 futures market

Time	Order type	LOB side	Action	Price (points)	Volume
13:27:27.281	Genuine order	Bid	Add	118.265625	10
13:30:00.545	Spoof order	Ask	Add	118.28125	1000
13:30:00.677	*Complete genuine order executed*				
13:30:01.469	Spoof order	Ask	Cancel	118.28125	1000

Note: This table presents the various spoofing actions JPM took on February 4, 2010 in the 10-Year T-Note March 2010 futures market. Per spoof action, the table reports the timestamp (*Time*), whether it concerned a genuine or spoof order (*Order type*), the LOB side the spoof action occurred on (*LOB side*), whether the order from the spoof action was added or cancelled (*Action*), the price level affected by the spoof action (*Price (points)*) and the volume related to the spoof action (*Volume*).

Table 3.A.2 | LOB state one millisecond before placement of the genuine order from the 10-Year T-Note March 2010 spoof

Bid volume	Bid price (points)	Level	Ask price (points)	Ask volume
1163	118.266	1	118.281	780
2788	118.250	2	118.297	1946
2551	118.234	3	118.312	2177
2113	118.219	4	118.328	2225
2144	118.203	5	118.344	1955
1999	118.188	6	118.359	2031
1673	118.172	7	118.375	2612
2291	118.156	8	118.391	1950
1365	118.141	9	118.406	3248
2758	118.125	10	118.422	1840

Note: This table reports the state of the LOB one millisecond before the genuine order from the 10-Year T-Note March 2010 spoof was added. It shows the prices and volumes of each level on the bid and ask side.

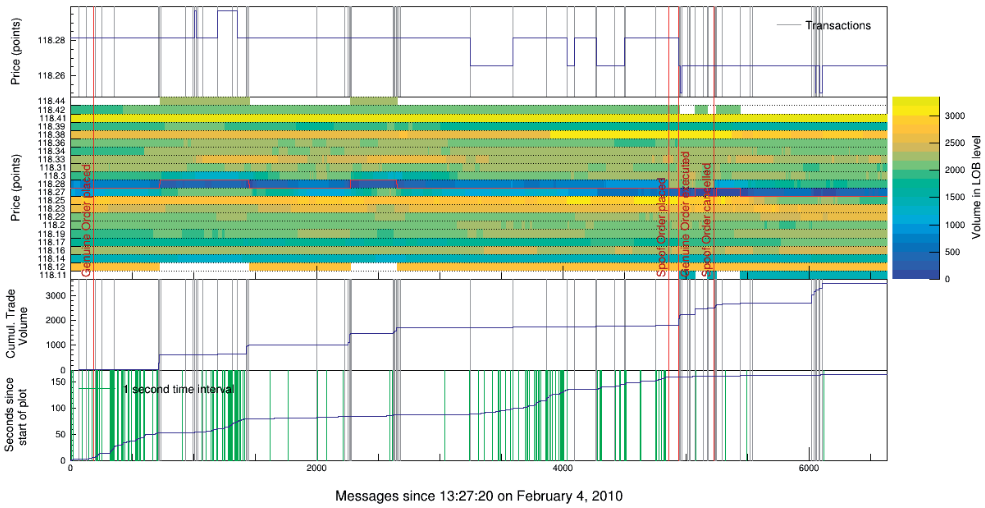


Figure 3.A.1 | 10-Year T-Note March 2010 LOB and trade behavior around the spoof of February 4, 2010. This figure visualizes the LOB and trade behavior around the spoof of February 4, 2010 in the 10-Year T-Note March 2010 futures market. The *first* panel shows the price of the last trade that took place (blue line) and when a trade took place (gray line). The *second* panel shows the volumes at the individual bid and ask levels between prices of 118.11 and 118.44 points. Each unit on the x-axis is one message. The y-axis represents the price of the 10-Year T-Note in points. The color represents the volume at each price level of the LOB for each message. The scale ranges from blue to yellow, with the color becoming a brighter yellow as volume increases at that price level. The red line is the midpoint. The *third* panel shows the cumulative trade volume per second. The fourth panel shows how much time passes between messages reported by the exchange. A steeper (flatter) blue line signals a lower (higher) rate of messages, given that a steeper (flatter) line signals more (less) time progression. The red vertical lines signal when the JPM spoofing activities took place, from left to right: when the genuine order was placed, when the spoof order of 1000 contracts was placed, when the genuine order was executed and when the spoof order was cancelled.

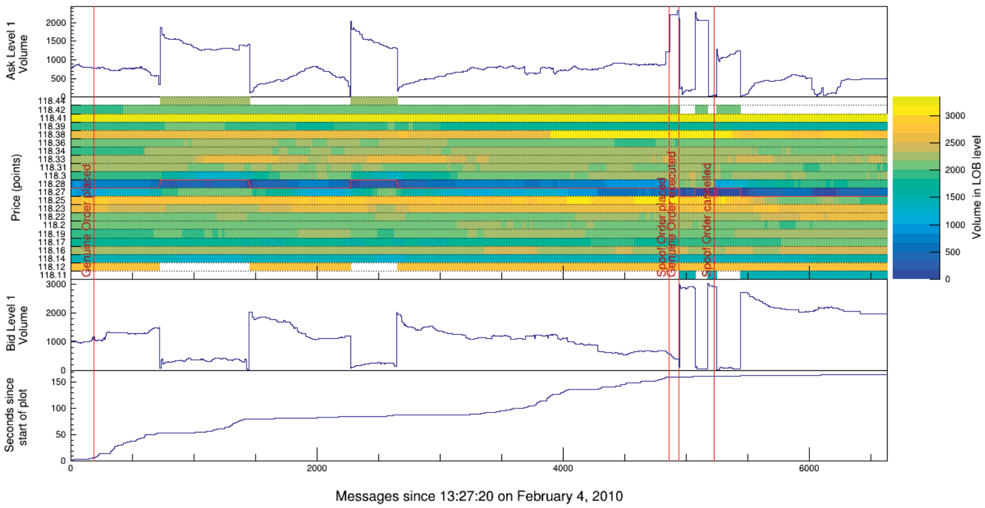


Figure 3.A.2 | 10-Year T-Note March 2010 first-level volume behavior around the spoof of February 4, 2010. This figure visualizes first-level bid and ask volume behavior around the spoof of February 4, 2010 in the 10-Year T-Note March 2010 futures market. The *first* panel shows the volume of the best ask level. The *second* panel shows the volumes at the individual bid and ask levels between prices of 118.11 and 118.44 points. Each unit on the x-axis is one message. The y-axis represents the price of the 10-Year T-Note in points. The color represents the volume at each price level of the LOB for each message. The scale ranges from blue to yellow, with the color becoming a brighter yellow as volume increases at that price level. The red line is the midpoint. The *third* panel shows the volume of the best bid level. The *fourth* panel shows how much time passes between messages reported by the exchange. A steeper (flatter) blue line signals a lower (higher) rate of messages, given that a steeper (flatter) line signals more (less) time progression. The red vertical lines signal when the JPM spoofing activities took place, from left to right: when the genuine order was placed, when the spoof order of 1000 contracts was placed, when the genuine order was executed and when the spoof order was cancelled.

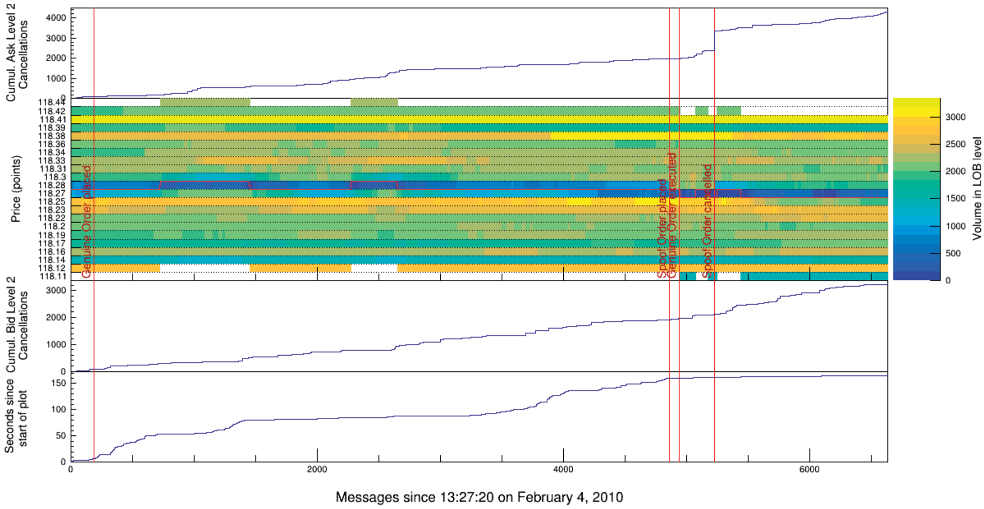


Figure 3.A.3 | 10-Year T-Note March 2010 second-level cancellation behavior around the spoof of February 4, 2010. This figure visualizes cumulative second-level bid and ask cancellation volume around the spoof of February 4, 2010 in the 10-Year T-Note March 2010 futures market. The *first* panel shows the cumulative volume of cancellations of the second ask level. The *second* panel shows the volumes at the individual bid and ask levels between prices of 118.11 and 118.44 points. Each unit on the x-axis is one message. The y-axis represents the price of the 10-Year T-Note in points. The color represents the volume at each price level of the LOB for each message. The scale ranges from blue to yellow, with the color becoming a brighter yellow as volume increases at that price level. The red line is the midpoint. The *third* panel shows the cumulative volume of cancellations of the second bid level. The *fourth* panel shows how much time passes between messages reported by the exchange. A steeper (flatter) blue line signals a lower (higher) rate of messages, given that a steeper (flatter) line signals more (less) time progression. The red vertical lines signal when the JPM spoofing activities took place, from left to right: when the genuine order was placed, when the spoof order of 1000 contracts was placed, when the genuine order was executed and when the spoof order was cancelled.

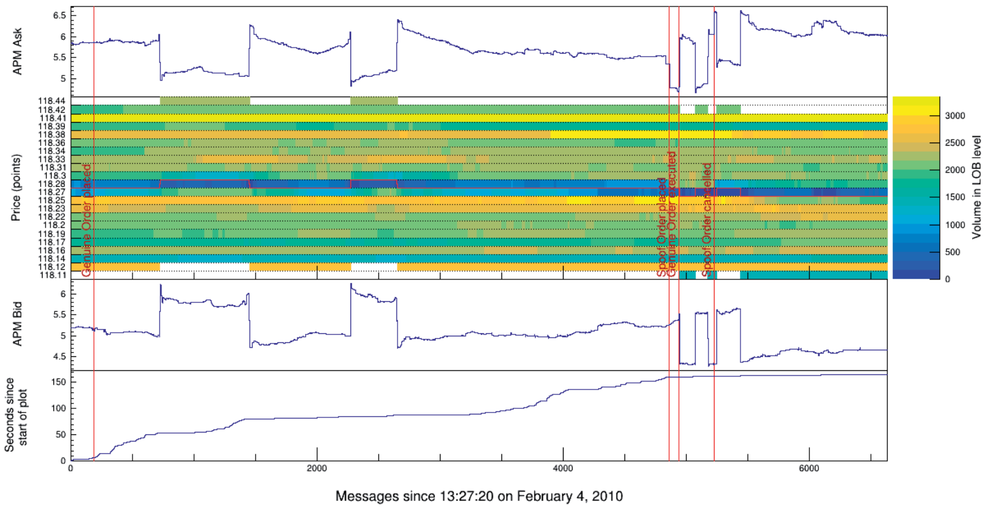


Figure 3.A.4 | 10-Year T-Note March 2010 bid and ask APM behavior around the spoof of February 4, 2010. This figure visualizes bid and ask liquidity costs (APM) behavior around the spoof of February 4, 2010 in the 10-Year T-Note March 2010 futures market. The *first* panel shows the APM of the ask side. APM measures the liquidity costs (in basis points) of a trader who wants to buy or sell a specific dollar value by submitting market orders. The *second* panel shows the volumes at the individual bid and ask levels between prices of 118.11 and 118.44 points. Each unit on the x-axis is one message. The y-axis represents the price of the 10-Year T-Note in points. The color represents the volume in each price level of the LOB for each message. The scale ranges from blue to yellow, with the color becoming a brighter yellow as volume increases at that price level. The red line is the midpoint. The *third* panel shows the APM for the bid side. The *fourth* panel shows how much time passes between messages reported by the exchange. A steeper (flatter) blue line signals a lower (higher) rate of messages, given that a steeper (flatter) line signals more (less) time progression. The red vertical lines signal when the JPM spoofing activities took place, from left to right: when the genuine order was placed, when the spoof order of 1000 contracts was placed, when the genuine order was executed and when the spoof order was cancelled.

Table 3.A.3 | Mean ask liquidity costs (bps) around the 10-Year T-Note March 2010 spoof for different time windows

Time window	Before vs. during	During vs. after	Before vs. after
Spoof duration	5.525 = 5.436	5.436 < 5.855***	5.525 < 5.855***
10 seconds	5.535 = 5.436	5.436 < 5.952***	5.535 < 5.952***
30 seconds	5.545 = 5.436	5.436 < 5.761***	5.545 < 5.761***
1 minute	5.595 > 5.436*	5.436 = 5.467	5.595 > 5.467***
5 minutes	5.712 > 5.436***	5.436 < 5.655***	5.712 > 5.655***

Note: The table reports the mean liquidity costs (bps, measured by APM) around the spoof in the 10-Year T-Note March 2010 market for different periods and various time windows. *before* represents the time up until the spoof order was added; *during* the period from when the spoof order was added until it was cancelled; and *after* the time following the cancellation of the spoof order. Five different time windows are used, the *Spoof duration* time window being 0.910 seconds. A lower APM indicates that liquidity costs are low and, hence, liquidity is high. Welch's *t*-tests were used to test for mean differences between the periods. Significance at the 0.1%, 1% and 5% (two-tailed) levels is indicated by ***, ** and *, respectively.

3.A.2 TRADITIONAL SPOOFING WITH ICEBERG ORDERS: SILVER MARCH 2014 CONTRACT

Figures are visualized for the Silver March 2014 contract from 01:59:20 to 01:59:30 on December 10, 2013.

Table 3.A.4 | Spoofing actions on December 10, 2013 in the Silver March 2014 futures market

Time	Order type	LOB side	Action	Price	Volume
01:59:22.386	Genuine order	Ask	Add	$19.97	5 displayed 15 hidden
01:59:26.901	Spoof order	Bid	Add	$19.96	100
01:59:26.902	*Beginning execution complete genuine order*				
01:59:27.729	Spoof order	Bid	Cancel	$19.96	100

Note: This table presents the various spoofing actions JPM took on December 10, 2013 in the Silver March 2014 futures market. Per spoof action, the table reports the timestamp (*Time*), whether it concerned a genuine or spoof order (*Order type*), the LOB side the spoof action occurred on (*LOB side*), whether the order from the spoof action was added or cancelled (*Action*), the price level affected by the spoof action (*Price*) and the volume related to the spoof action (*Volume*).

Table 3.A.5 | LOB state one millisecond before placement of the genuine order from the Silver March 2014 spoof

Bid volume	Bid price	Level	Ask price	Ask volume
2	$19.965	1	$19.970	2
6	$19.960	2	$19.975	11
10	$19.955	3	$19.980	10
17	$19.950	4	$19.985	10
13	$19.945	5	$19.990	10
11	$19.940	6	$19.995	8
11	$19.935	7	$20.000	29
13	$19.930	8	$20.005	5
25	$19.925	9	$20.010	13
11	$19.920	10	$20.015	7

Note: This table reports the state of the LOB one millisecond before the genuine order from the Silver March 2014 spoof was added. It shows the prices and volumes of each level on the bid and ask side.

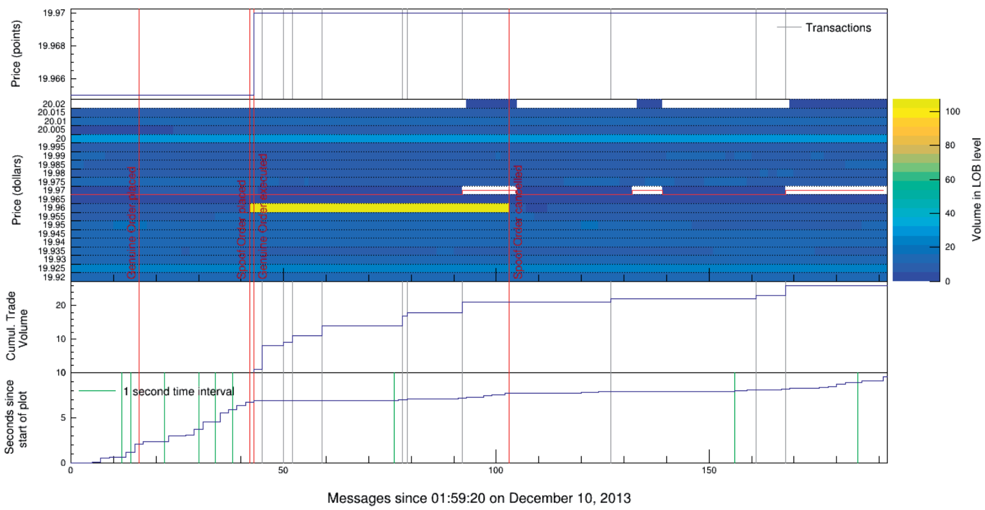


Figure 3.A.5 | Silver March 2014 LOB and trade behavior around the spoof of December 10, 2013. This figure visualizes the LOB and trade behavior around the spoof of December 10, 2013 in the Silver March 2014 futures market. The *first* panel shows the price of the last trade that took place (blue line) and when a trade took place (gray line). The *second* panel shows the volumes at the individual bid and ask levels between prices of $19.92 and $20.02. Each unit on the x-axis is one message. The y-axis represents the price of Silver in dollars. The color represents the volume at each price level of the LOB for each message. The scale ranges from blue to yellow, with the color becoming a brighter yellow as volume increases at that price level. The red line is the midpoint. The *third* panel shows the cumulative trade volume per second. The fourth panel shows how much time passes between messages reported by the exchange. A steeper (flatter) blue line signals a lower (higher) rate of messages, given that a steeper (flatter) line signals more (less) time progression. The red vertical lines signal when the JPM spoofing activities took place, from left to right: when the genuine iceberg order was placed, when the spoof order of 100 contracts was placed, when the first contract of the genuine iceberg order was executed and when the spoof order was cancelled.

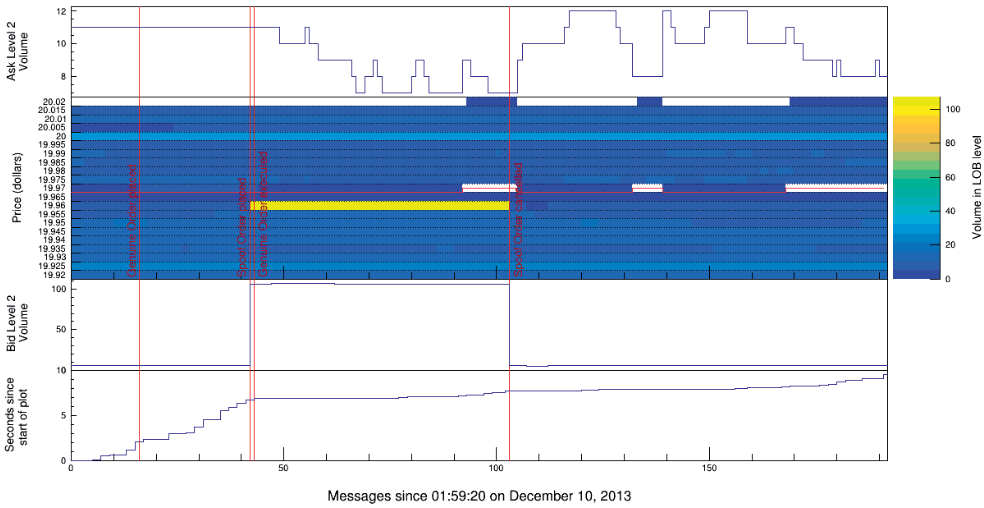


Figure 3.A.6 | Silver March 2014 second-level volume behavior around the spoof of December 10, 2013. This figure visualizes second-level bid and ask volume behavior around the spoof of December 10, 2013 in the Silver March 2014 futures market. The *first* panel shows the volume of the second ask level. The *second* panel shows the volumes at the individual bid and ask levels between prices of \$19.92 and \$20.02. Each unit on the x-axis is one message. The y-axis represents the price of Silver in dollars. The color represents the volume at each price level of the LOB for each message. The scale ranges from blue to yellow, with the color becoming a brighter yellow as volume increases at that price level. The red line is the midpoint. The *third* panel shows the volume of the second bid level. The *fourth* panel shows how much time passes between messages reported by the exchange. A steeper (flatter) blue line signals a lower (higher) rate of messages, given that a steeper (flatter) line signals more (less) time progression. The red vertical lines signal when the JPM spoofing activities took place, from left to right: when the genuine iceberg order was placed, when the spoof order of 100 contracts was placed, when the first contract of the genuine order was executed and when the spoof order was cancelled.

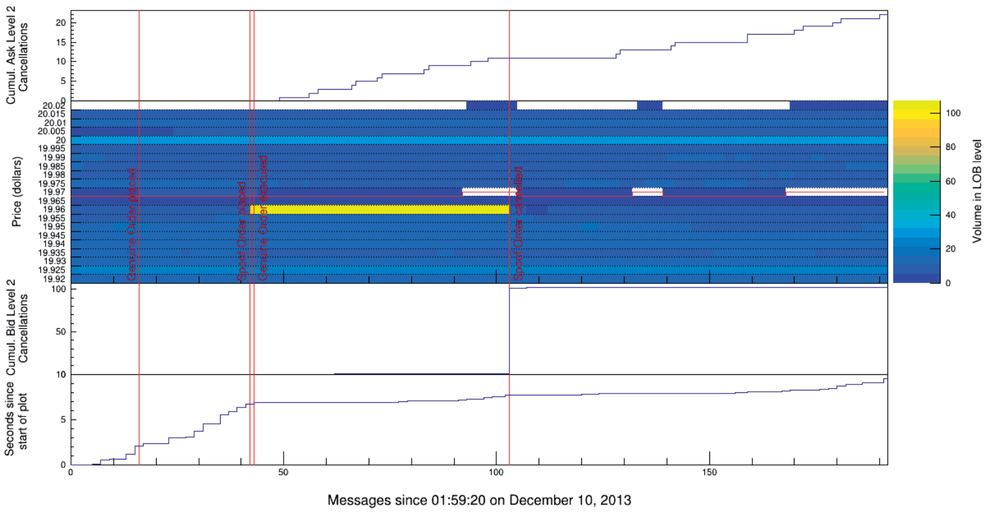


Figure 3.A.7 | Silver March 2014 second-level cancellation behavior around the spoof of December 10, 2013. This figure visualizes cumulative second-level bid and ask cancellation volume around the spoof of December 10, 2013 in the Silver March 2014 futures market. The *first* panel shows the cumulative volume of cancellations of the second ask level. The *second* panel shows the volumes at the individual bid and ask levels between prices of $19.92 and $20.02. Each unit on the x-axis is one message. The y-axis represents the price of Silver in dollars. The color represents the volume at each price level of the LOB for each message. The scale ranges from blue to yellow, with the color becoming a brighter yellow as volume increases at that price level. The red line is the midpoint. The *third* panel shows the cumulative volume of cancellations of the second bid level. The *fourth* panel shows how much time passes between messages reported by the exchange. A steeper (flatter) blue line signals a lower (higher) rate of messages, given that a steeper (flatter) line signals more (less) time progression. The red vertical lines signal when the JPM spoofing activities took place, from left to right: when the genuine iceberg order was placed, when the spoof order of 100 contracts was placed, when the first contract of the genuine order was executed and when the spoof order was cancelled.

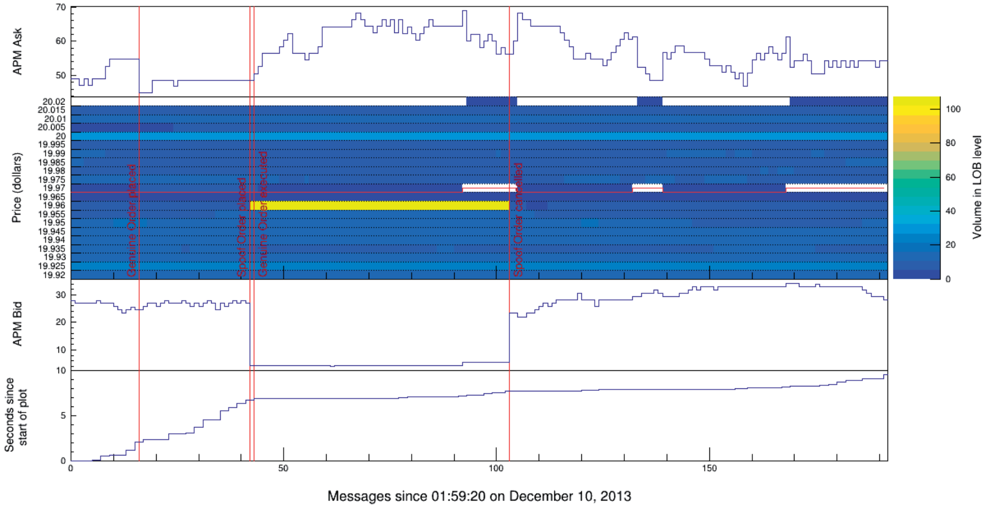


Figure 3.A.8 | Silver March 2014 bid and ask APM behavior around the spoof of December 10, 2013. This figure visualizes bid and ask liquidity costs (APM) behavior around the spoof of December 10, 2013 in the Silver March 2014 futures market. The *first* panel shows the APM of the ask side. APM measures the liquidity costs (in basis points) of a trader who wants to buy or sell a specific dollar value by submitting market orders. The *second* panel shows the volumes at the individual bid and ask levels between prices of $19.92 and $20.02. Each unit on the x-axis is one message. The y-axis represents the price of Silver in dollars. The color represents the volume in each price level of the LOB for each message. The scale ranges from blue to yellow, with the color becoming a brighter yellow as volume increases at that price level. The red line is the midpoint. The *third* panel shows the APM for the bid side. The *fourth* panel shows how much time passes between messages reported by the exchange. A steeper (flatter) blue line signals a lower (higher) rate of messages, given that a steeper (flatter) line signals more (less) time progression. The red vertical lines signal when the JPM spoofing activities took place, from left to right: when the genuine iceberg order was placed, when the spoof order of 100 contracts was placed, when the first contract of the genuine iceberg order was executed and when the spoof order was cancelled.

Table 3.A.6 | Mean bid liquidity costs (bps) around the Silver March 2014 spoof for different time windows

Time window	Before vs. during	During vs. after	Before vs. after
Spoof duration	26.876 > 5.006***	5.006 < 31.857***	26.876 < 31.857***
10 seconds	26.195 > 5.006***	5.006 < 28.973***	26.195 < 28.973***
30 seconds	18.250 > 5.006***	5.006 < 31.620***	18.250 < 31.620***
1 minute	15.413 > 5.006***	5.006 < 42.309***	15.413 < 42.309***
5 minutes	23.472 > 5.006***	5.006 < 39.346***	23.472 < 39.346***

Note: The table reports the mean liquidity costs (bps, measured by APM) around the spoof in the Silver March 2014 market for different periods and various time windows. *Before* represents the time up until the spoof order was added; *during* the period from when the spoof order was added until it was cancelled; and *after* the time following the cancellation of the spoof order. Five different time windows are used, the *Spoof duration* time window being 0.820 seconds. A lower APM indicates that liquidity costs are low and, hence, liquidity is high. Welch's *t*-tests were used to test for mean differences between the periods. Significance at the 0.1%, 1% and 5% (two-tailed) levels is indicated by ***, ** and *, respectively.

3.A.3.1 LAYERED SPOOFING: SILVER MARCH 2012 CONTRACT

The Silver March 2012 spoofing is visualized from 11:59:30 to 11:59:45 on December 12, 2011.

Table 3.A.7 | Spoofing actions on December 12, 2011 in the Silver March 2012 futures market

Time	Order type	LOB side	Action	Price	Volume
11:59:36.672	Genuine order	Ask	Add	$31.085	5
11:59:39.171	Spoof layer 1	Bid	Add	$31.075	10
11:59:39.369	Spoof layer 2	Bid	Add	$31.080	10
11:59:39.523	Spoof layer 3	Bid	Add	$31.080	10
11:59:39.687	Spoof layer 4	Bid	Add	$31.080	10
11:59:39.698	*Complete genuine order executed*				
11:59:39.837	Spoof layer 5	Bid	Add	$31.080	10
11:59:40.335	Spoof layer 2	Bid	Cancel	$31.080	10
11:59:40.337	Spoof layer 3-5	Bid	Cancel	$31.080	30
11:59:40.663	Spoof layer 1	Bid	Cancel	$31.075	10

Note: This table presents the various spoofing actions JPM took on December 12, 2011 in the Silver March 2012 futures market. Per spoof action, the table reports the timestamp (*Time*), whether it concerned a genuine or spoof order (*Order type*), the LOB side the spoof action occurred on (*LOB side*), whether the order from the spoof action was added or cancelled (*Action*), the price level affected by the spoof action (*Price*) and the volume related to the spoof action (*Volume*).

Table 3.A.8 | LOB state one millisecond before placement of the genuine order from the Silver March 2012 spoof

Bid volume	Bid price	Level	Ask price	Ask volume
1	$31.075	1	$31.085	3
6	$31.070	2	$31.090	2
8	$31.065	3	$31.095	3
10	$31.060	4	$31.100	13
7	$31.055	5	$31.105	4
8	$31.050	6	$31.110	9
6	$31.045	7	$31.115	5
9	$31.040	8	$31.120	8
6	$31.035	9	$31.125	7
17	$31.030	10	$31.130	7

Note: This table reports the state of the LOB one millisecond before the genuine order from the Silver March 2012 spoof was added. It shows the prices and volumes of each level on the bid and ask side.

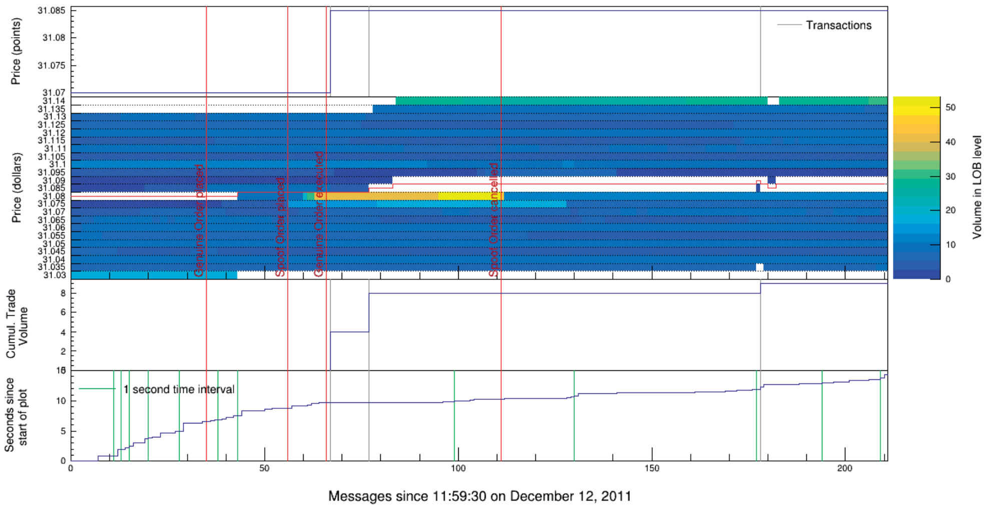


Figure 3.A.9 | Silver March 2012 LOB and trade behavior around the spoof of December 12, 2011. This figure visualizes the LOB and trade behavior around the spoof of December 12, 2011 in the Silver March 2012 futures market. The *first* panel shows the price of the last trade that took place (blue line) and when a trade took place (gray line). The *second* panel shows the volumes at the individual bid and ask levels between prices of $31.03 and $31.14. Each unit on the x-axis is one message. The y-axis represents the price of Silver in dollars. The color represents the volume at each price level of the LOB for each message. The scale ranges from blue to yellow, with the color becoming a brighter yellow as volume increases at that price level. The red line is the midpoint. The *third* panel shows the cumulative trade volume per second. The fourth panel shows how much time passes between messages reported by the exchange. A steeper (flatter) blue line signals a lower (higher) rate of messages, given that a steeper (flatter) line signals more (less) time progression. The red vertical lines signal when the JPM spoofing activities took place, from left to right: when the genuine order was placed, when the first spoof order of 10 contracts was placed, when the genuine order was executed and when the first spoof order was cancelled.

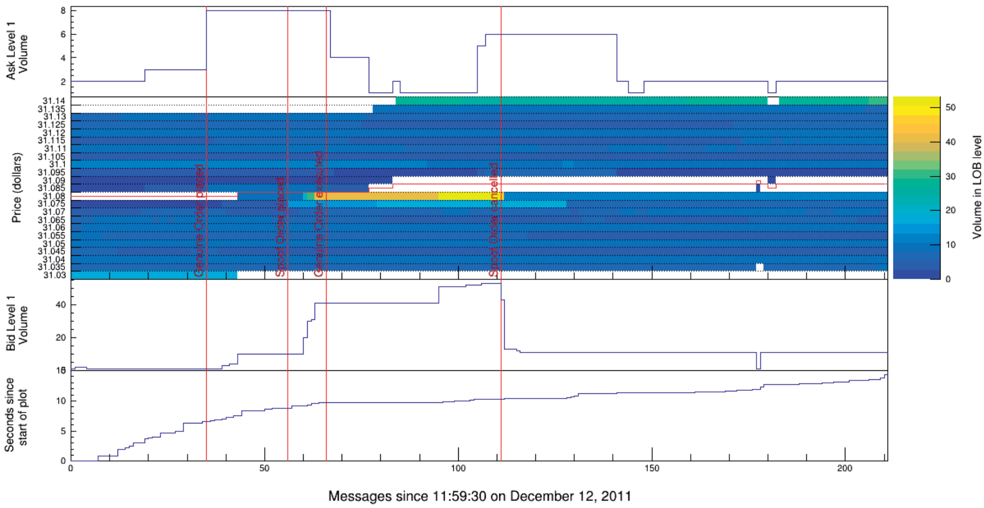


Figure 3.A.10 | Silver March 2012 first-level volume behavior around the spoof of December 12, 2011. This figure visualizes first-level bid and ask volume behavior around the spoof of December 12, 2011 in the Silver March 2012 futures market. The *first* panel shows the volume of the best ask level. The *second* panel shows the volumes at the individual bid and ask levels between prices of $31.03 and $31.14. Each unit on the x-axis is one message. The y-axis represents the price of Silver in dollars. The color represents the volume at each price level of the LOB for each message. The scale ranges from blue to yellow, with the color becoming a brighter yellow as volume increases at that price level. The red line is the midpoint. The *third* panel shows the volume of the best bid level. The *fourth* panel shows how much time passes between messages reported by the exchange. A steeper (flatter) blue line signals a lower (higher) rate of messages, given that a steeper (flatter) line signals more (less) time progression. The red vertical lines signal when the JPM spoofing activities took place, from left to right: when the genuine order was placed, when the first spoof order of 10 contracts was placed, when the genuine order was executed and when the first spoof order was cancelled.

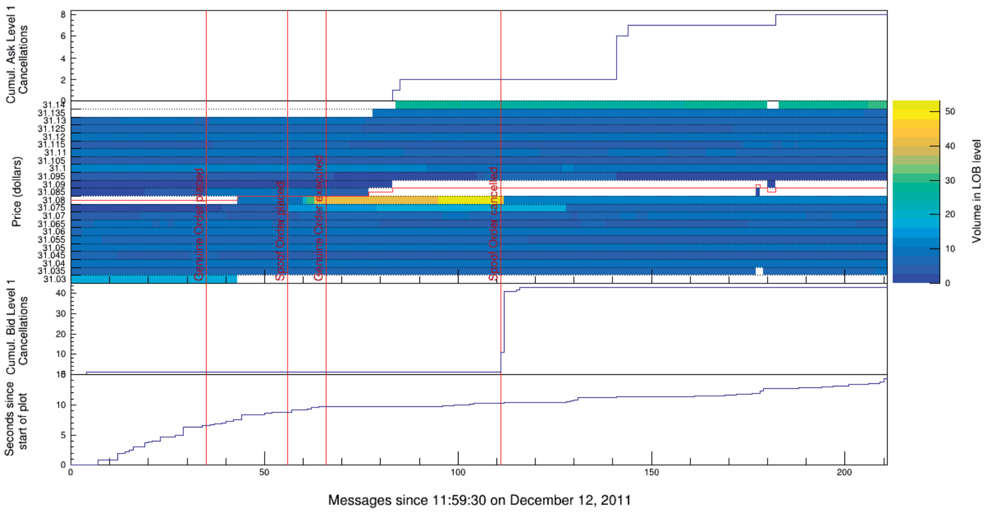


Figure 3.A.11 | Silver March 2012 first-level cancellation behavior around the spoof of December 12, 2011. This figure visualizes cumulative first-level bid and ask cancellation volume around the spoof of December 12, 2011 in the Silver March 2012 futures market. The *first* panel shows the cumulative volume of cancellations of the best ask level. The *second* panel shows the volumes at the individual bid and ask levels between prices of $31.03 and $31.14. Each unit on the x-axis is one message. The y-axis represents the price of Silver in dollars. The color represents the volume at each price level of the LOB for each message. The scale ranges from blue to yellow, with the color becoming a brighter yellow as volume increases at that price level. The red line is the midpoint. The *third* panel shows the cumulative volume of cancellations of the best bid level. The *fourth* panel shows how much time passes between messages reported by the exchange. A steeper (flatter) blue line signals a lower (higher) rate of messages, given that a steeper (flatter) line signals more (less) time progression. The red vertical lines signal when the JPM spoofing activities took place, from left to right: when the genuine order was placed, when the first spoof order of 10 contracts was placed, when the genuine order was executed and when the first spoof order was cancelled.

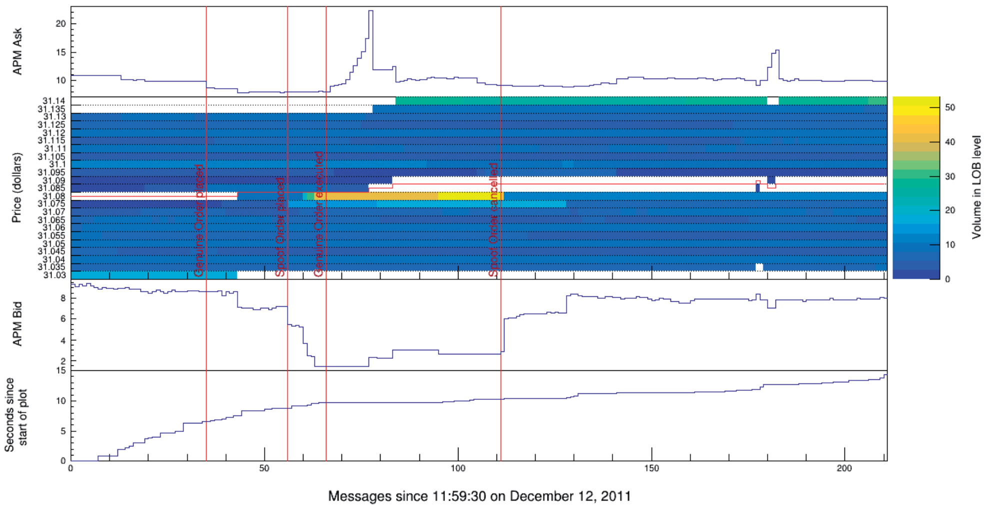


Figure 3.A.12 | Silver March 2012 bid and ask APM behavior around the spoof of December 12, 2011. This figure visualizes bid and ask liquidity costs (APM) behavior around the spoof of December 12, 2011 in the Silver March 2012 futures market. The *first* panel shows the APM of the ask side. APM measures the liquidity costs (in basis points) of a trader who wants to buy or sell a specific dollar value by submitting market orders. The *second* panel shows the volumes at the individual bid and ask levels between prices of $31.03 and $31.14. Each unit on the x-axis is one message. The y-axis represents the price of Silver in dollars. The color represents the volume in each price level of the LOB for each message. The scale ranges from blue to yellow, with the color becoming a brighter yellow as volume increases at that price level. The red line is the midpoint. The *third* panel shows the APM for the bid side. The *fourth* panel shows how much time passes between messages reported by the exchange. A steeper (flatter) blue line signals a lower (higher) rate of messages, given that a steeper (flatter) line signals more (less) time progression. The red vertical lines signal when the JPM spoofing activities took place, from left to right: when the genuine order was placed, when the first spoof order of 10 contracts was placed, when the genuine order was executed and when the first spoof order was cancelled.

Table 3.A.9 | Mean bid liquidity costs (bps) around the Silver March 2012 spoof for different time windows

Time window	Before vs. during	During vs. after	Before vs. after
Spoof duration	7.737 > 3.939***	3.939 < 8.057***	7.737 < 8.057***
10 seconds	8.715 > 3.939***	3.939 < 7.957***	8.715 > 7.957***
30 seconds	9.430 > 3.939***	3.939 < 8.612***	9.430 > 8.612***
1 minute	9.316 > 3.939***	3.939 < 8.555***	9.316 > 8.555***
5 minutes	8.208 < 3.939 ***	3.939 < 9.056***	8.208 < 9.056***

Note: The table reports the mean liquidity costs (bps, measured by APM) around the spoof in the Silver March 2012 market for different periods and various time windows. *Before* represents the time up until the spoof order was added; *during* the period from when the spoof order was added until it was cancelled; and *after* the time following the cancellation of the spoof order. Five different time windows are used, the *Spoof duration* time window being 1.47 seconds. A lower APM indicates that liquidity costs are low and, hence, liquidity is high. Welch's *t*-tests were used to test for mean differences between the periods. Significance at the 0.1%, 1% and 5% (two-tailed) levels is indicated by ***, ** and *, respectively.

3.A.3.2 LAYERED SPOOFING: SILVER MAY 2014 CONTRACT

Spoofing in the Silver May 2014 contract is visualized on March 5, 2014 from 08:18:35 to 08:18:50.

Table 3.A.10 | Spoofing actions on March 5, 2014 in the Silver May 2014 futures market

Time	Order type	LOB side	Action	Price	Volume
08:18:39.699	Genuine order	Ask	Add	$21.275	2
08:18:40.433	Spoof layer 1	Bid	Add	$21.255	2
08:18:40.587	Spoof layer 2	Bid	Add	$21.260	2
08:18:40.707	Spoof layer 3	Bid	Add	$21.265	2
08:18:40.819	Spoof layer 4	Bid	Add	$21.265	2
08:18:40.939	Spoof layer 5	Bid	Add	$21.265	2
08:18:41.058	Spoof layer 6	Bid	Add	$21.265	2
08:18:41.194	Spoof layer 7	Bid	Add	$21.270	2
08:18:41.322	Spoof layer 8	Bid	Add	$21.270	2
08:18:41.459	Spoof layer 9	Bid	Add	$21.270	2
08:18:41.587	Spoof layer 10	Bid	Add	$21.270	2
08:18:41.595	*Complete genuine order executed*				
08:18:42.188	Spoof layer 7-10	Bid	Cancel	$21.270	8
08:18:42.332	Spoof layer 3-6	Bid	Cancel	$21.265	8
08:18:42.450	Spoof layer 2	Bid	Cancel	$21.260	2
08:18:42.572	Spoof layer 1	Bid	Cancel	$21.255	2

Note: This table presents the various spoofing actions JPM took on March 5, 2014 in the Silver May 2014 futures market. Per spoof action, the table reports the timestamp (*Time*), whether it concerned a genuine or spoof order (*Order type*), the LOB side the spoof action occurred on (*LOB side*), whether the order from the spoof action was added or cancelled (*Action*), the price level affected by the spoof action (*Price*) and the volume related to the spoof action (*Volume*).

Table 3.A.11 | LOB state one millisecond before placement of the genuine order from the Silver May 2014 spoof

Bid volume	Bid price	Level	Ask price	Ask volume
8	$21.270	1	$21.275	10
15	$21.265	2	$21.280	16
21	$21.260	3	$21.285	14
21	$21.255	4	$21.290	18
22	$21.250	5	$21.295	16
19	$21.245	6	$21.300	20
16	$21.240	7	$21.305	10
8	$21.235	8	$21.310	17
26	$21.230	9	$21.315	7
11	$21.225	10	$21.320	22

Note: This table reports the state of the LOB one millisecond before the genuine order from the Silver May 2014 spoof was added. It shows the prices and volumes of each level on the bid and ask side.

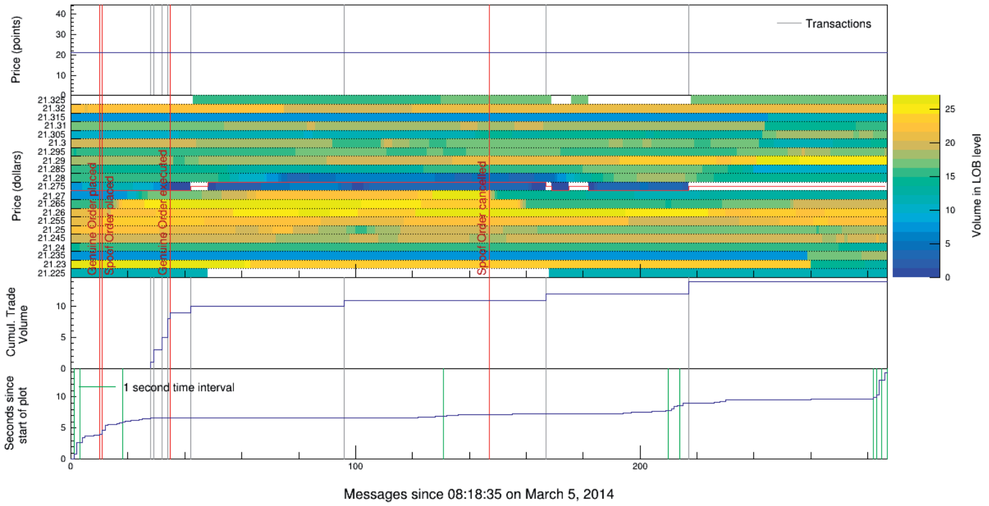


Figure 3.A.13 | Silver May 2014 LOB and trade behavior around the spoof of March 5, 2014. This figure visualizes the LOB and trade behavior around the spoof of March 5, 2014 in the Silver May 2014 futures market. The *first* panel shows the price of the last trade that took place (blue line) and when a trade took place (gray line). The *second* panel shows the volumes at the individual bid and ask levels between prices of $21.225 and $21.325. Each unit on the x-axis is one message. The y-axis represents the price of Silver in dollars. The color represents the volume at each price level of the LOB for each message. The scale ranges from blue to yellow, with the color becoming a brighter yellow as volume increases at that price level. The red line is the midpoint. The *third* panel shows the cumulative trade volume per second. The fourth panel shows how much time passes between messages reported by the exchange. A steeper (flatter) blue line signals a lower (higher) rate of messages, given that a steeper (flatter) line signals more (less) time progression. The red vertical lines signal when the JPM spoofing activities took place, from left to right: when the genuine order was placed, when the first spoof order of 2 contracts was placed, when the genuine order was executed and when the first spoof order was cancelled.

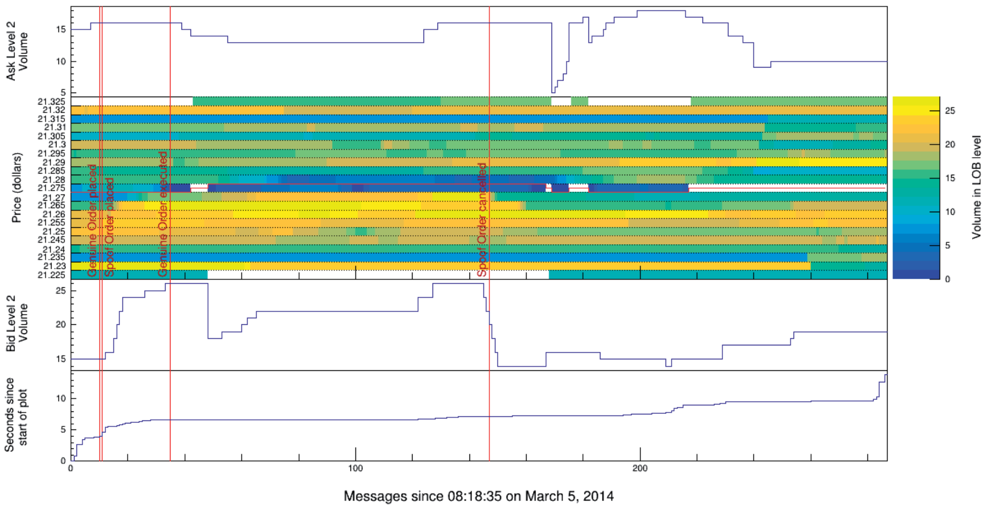


Figure 3.A.14 | Silver May 2014 second-level volume behavior around the spoof of March 5, 2014. This figure visualizes second-level bid and ask volume behavior around the spoof of March 5, 2014 in the Silver May 2014 futures market. The *first* panel shows the volume of the second ask level. The *second* panel shows the volumes at the individual bid and ask levels between prices of $21.225 and $21.325. Each unit on the x-axis is one message. The y-axis represents the price of Silver in dollars. The color represents the volume at each price level of the LOB for each message. The scale ranges from blue to yellow, with the color becoming a brighter yellow as volume increases at that price level. The red line is the midpoint. The *third* panel shows the volume of the second bid level. The *fourth* panel shows how much time passes between messages reported by the exchange. A steeper (flatter) blue line signals a lower (higher) rate of messages, given that a steeper (flatter) line signals more (less) time progression. The red vertical lines signal when the JPM spoofing activities took place, from left to right: when the genuine order was placed, when the first spoof order of 2 contracts was placed, when the genuine order was executed and when the first spoof order was cancelled.

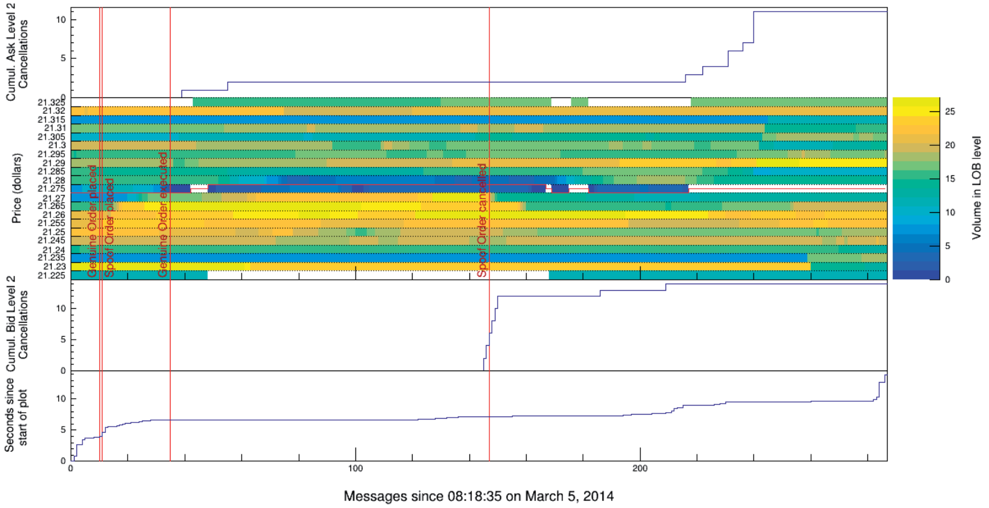


Figure 3.A.15 | Silver May 2014 second-level cancellation behavior around the spoof of March 5, 2014. This figure visualizes cumulative second-level bid and ask cancellation volume around the spoof of March 5, 2014 in the Silver May 2014 futures market. The *first* panel shows the cumulative volume of cancellations of the second ask level. The *second* panel shows the volumes at the individual bid and ask levels between prices of $21.225 and $21.325. Each unit on the x-axis is one message. The y-axis represents the price of the Silver in dollars. The color represents the volume at each price level of the LOB for each message. The scale ranges from blue to yellow, with the color becoming a brighter yellow as volume increases at that price level. The red line is the midpoint. The *third* panel shows the cumulative volume of cancellations of the second bid level. The *fourth* panel shows how much time passes between messages reported by the exchange. A steeper (flatter) blue line signals a lower (higher) rate of messages, given that a steeper (flatter) line signals more (less) time progression. The red vertical lines signal when the JPM spoofing activities took place, from left to right: when the genuine order was placed, when the first spoof order of 2 contracts was placed, when the genuine order was executed and when the first spoof order was cancelled.

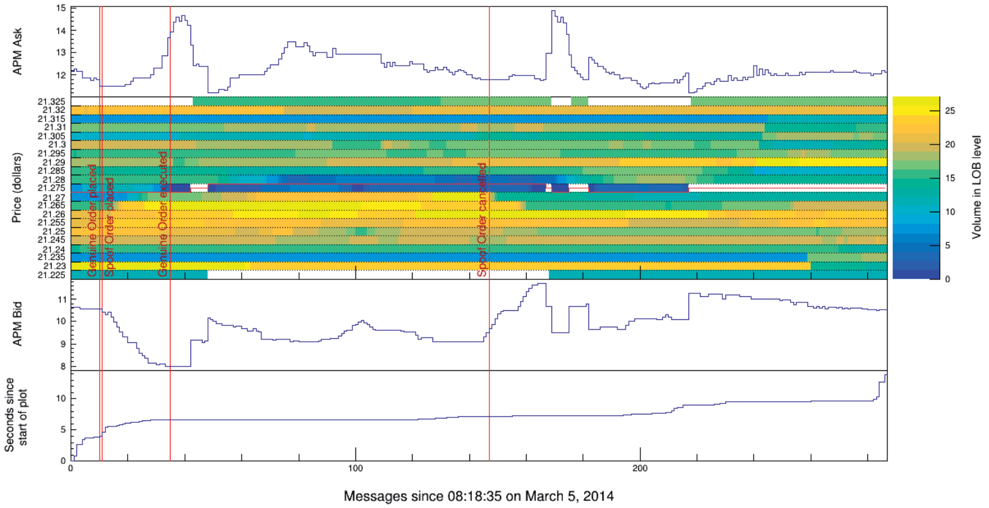


Figure 3.A.16 | Silver May 2014 bid and ask APM behavior around the spoof of March 5, 2014. This figure visualizes bid and ask liquidity costs (APM) behavior around the spoof of March 5, 2014 in the Silver May 2014 futures market. The *first* panel shows the APM of the ask side. APM measures the liquidity costs (in basis points) of a trader who wants to buy or sell a specific dollar value by submitting market orders. The *second* panel shows the volumes at the individual bid and ask levels between prices of $21.225 and $21.325. Each unit on the x-axis is one message. The y-axis represents the price of the Silver in dollars. The color represents the volume in each price level of the LOB for each message. The scale ranges from blue to yellow, with the color becoming a brighter yellow as volume increases at that price level. The red line is the midpoint. The *third* panel shows the APM for the bid side. The *fourth* panel shows how much time passes between messages reported by the exchange. A steeper (flatter) blue line signals a lower (higher) rate of messages, given that a steeper (flatter) line signals more (less) time progression. The red vertical lines signal when the JPM spoofing activities took place, from left to right: when the genuine order was placed, when the first spoof order of 2 contracts was placed, when the genuine order was executed and when the first spoof order was cancelled.

Table 3.A.12 | Mean bid liquidity costs (bps) around the Silver May 2014 spoof for different time windows

Time window	Before vs. during	During vs. after	Before vs. after
Spoof duration	10.558 > 9.518***	9.518 < 10.445***	10.588 > 10.445***
10 seconds	10.304 > 9.518***	9.518 < 10.513***	10.304 < 10.513***
30 seconds	10.086 > 9.518***	9.518 < 10.907***	10.086 < 10.907***
1 minute	10.122 > 9.518***	9.518 < 10.650***	10.122 < 10.650***
5 minutes	10.686 > 9.518***	9.518 < 10.857***	10.686 < 10.857***

Note: The table reports the mean liquidity costs (bps, measured by APM) around the spoof in the Silver May 2014 market for different periods and various time windows. *Before* represents the time up until the spoof order was added; *during* the period from when the spoof order was added until it was cancelled; and a*fter* the time following the cancellation of the spoof order. Five different time windows are used, the *Spoof duration* time window being 2.13 seconds. A lower APM indicates that liquidity costs are low and, hence, liquidity is high. Welch's *t*-tests were used to test for mean differences between the periods. Significance at the 0.1%, 1% and 5% (two-tailed) levels is indicated by ***, ** and *, respectively.

3.A.3.3 LAYERED SPOOFING: GOLD APRIL 2014 CONTRACT

The Gold April 2014 contract is visualized from 08:02:10 to 08:02:30 on March 3, 2014.

Table 3.A.1 | Spoofing actions on March 3, 2014 in the Gold April 2014 futures market

Time	Order type	LOB side	Action	Price	Volume
08:02:17.998	Genuine order	Ask	Add	$1348.2	5
08:02:19.360	Spoof layer 1	Bid	Add	$1347.9	5
08:02:19.784	Spoof layer 2	Bid	Add	$1348.0	5
08:02:20.152	Spoof layer 3	Bid	Add	$1348.1	5
08:02:20.560	Spoof layer 4	Bid	Add	$1348.1	5
08:02:21.071	Spoof layer 5	Bid	Add	$1348.1	5
08:02:21.529	Spoof layer 6	Bid	Add	$1348.1	5
08:02:21.530			*Complete genuine order executed*		
08:02:22.257	Spoof layer 3-6	Bid	Cancel	$1348.1	20
08:02:24.560	Spoof layer 2	Bid	Cancel	$1348.0	5
08:02:24.920	Spoof layer 1	Bid	Cancel	$1347.9	5

Note: This table presents the various spoofing actions JPM took on March 3, 2014 in the Gold April 2014 futures market. Per spoof action, the table reports the timestamp (*Time*), whether it concerned a genuine or spoof order (*Order type*), the LOB side the spoof action occurred on (*LOB side*), whether the order from the spoof action was added or cancelled (*Action*), the price level affected by the spoof action (*Price*) and the volume related to the spoof action (*Volume*).

Table 3.A.14 | LOB state one millisecond before placement of the genuine order from the Gold April 2014 spoof

Bid volume	Bid price	Level	Ask price	Ask volume
7	$1348.1	1	$1348.2	4
16	$1348.0	2	$1348.3	17
16	$1347.9	3	$1348.4	17
23	$1347.8	4	$1348.5	16
15	$1347.7	5	$1348.6	14
44	$1347.6	6	$1348.7	12
13	$1347.5	7	$1348.8	12
9	$1347.4	8	$1348.9	11
12	$1347.3	9	$1349.0	31
20	$1347.2	10	$1349.1	11

Note: This table reports the state of the LOB one millisecond before the genuine order from the Gold April 2014 spoof was added. It shows the prices and volumes of each level on the bid and ask side.

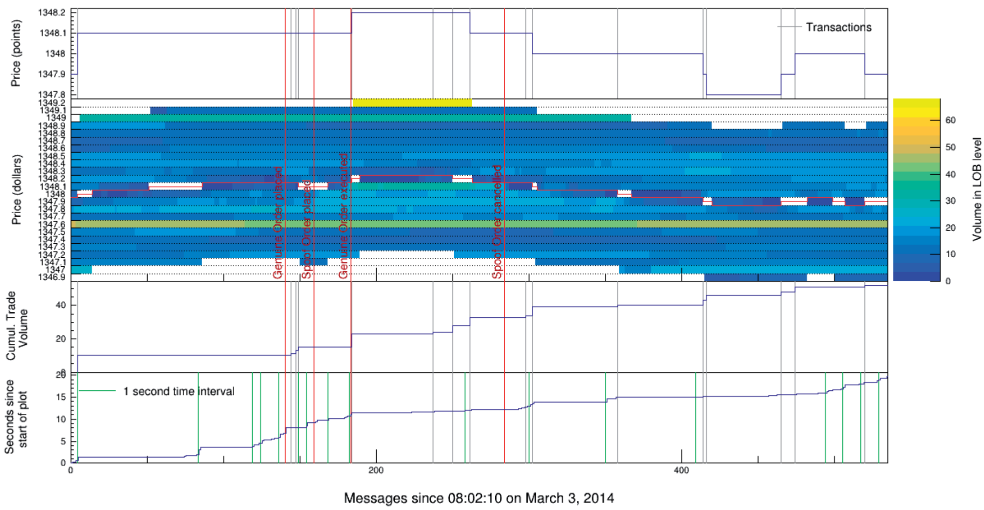


Figure 3.A.17 | Gold April 2014 LOB and trade behavior around the spoof of March 3, 2014. This figure visualizes the LOB and trade behavior around the spoof of March 3, 2014 in the Gold April 2014 futures market. The *first* panel shows the price of the last trade that took place (blue line) and when a trade took place (gray line). The *second* panel shows the volumes at the individual bid and ask levels between prices of $1346.9 and $1349.2. Each unit on the x-axis is one message. The y-axis represents the price of Gold in dollars. The color represents the volume at each price level of the LOB for each message. The scale ranges from blue to yellow, with the color becoming a brighter yellow as volume increases at that price level. The red line is the midpoint. The *third* panel shows the cumulative trade volume per second. The fourth panel shows how much time passes between messages reported by the exchange. A steeper (flatter) blue line signals a lower (higher) rate of messages, given that a steeper (flatter) line signals more (less) time progression. The red vertical lines signal when the JPM spoofing activities took place, from left to right: when the genuine order was placed, when the first spoof order of 5 contracts was placed, when the genuine order was executed and when the first spoof order was cancelled.

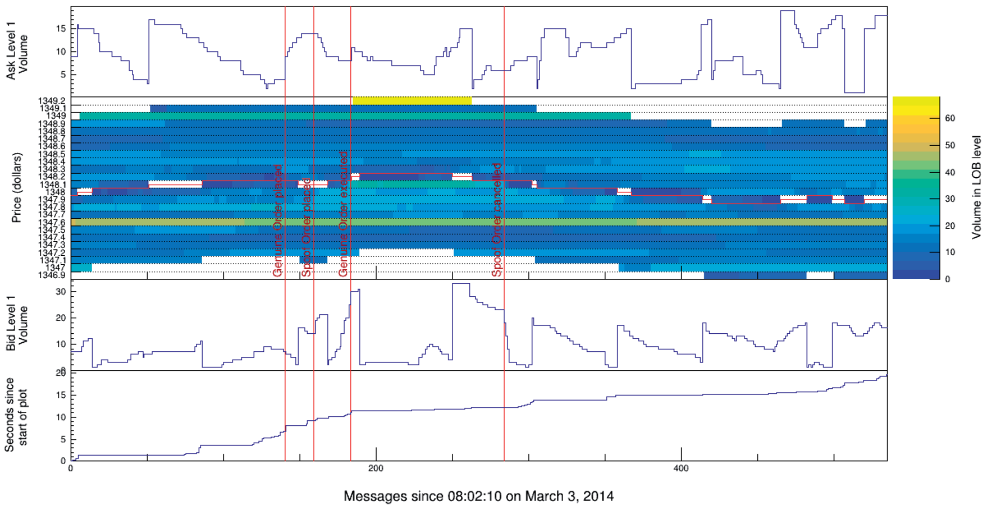


Figure 3.A.18 | Gold April 2014 first-level volume behavior around the spoof of March 3, 2014. This figure visualizes first-level bid and ask volume behavior around the spoof of March 3, 2014 in the Gold April 2014 futures market. The *first* panel shows the volume of the best ask level. The *second* panel shows the volumes at the individual bid and ask levels between prices of $1346.9 and $1349.2. Each unit on the x-axis is one message. The y-axis represents the price of Gold in dollars. The color represents the volume at each price level of the LOB for each message. The scale ranges from blue to yellow, with the color becoming a brighter yellow as volume increases at that price level. The red line is the midpoint. The *third* panel shows the volume of the best bid level. The *fourth* panel shows how much time passes between messages reported by the exchange. A steeper (flatter) blue line signals a lower (higher) rate of messages, given that a steeper (flatter) line signals more (less) time progression. The red vertical lines signal when the JPM spoofing activities took place, from left to right: when the genuine order was placed, when the first spoof order of 5 contracts was placed, when the genuine order was executed and when the first spoof order was cancelled.

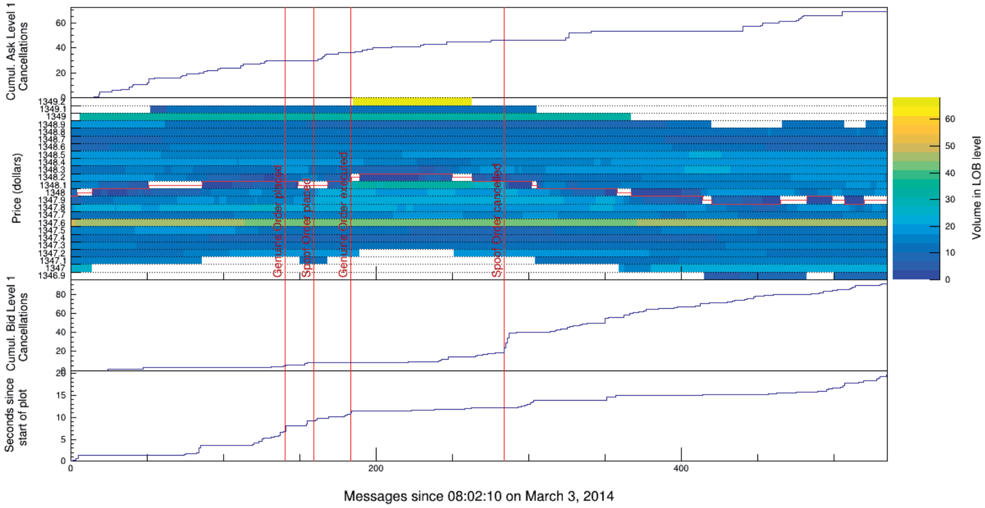


Figure 3.A.19 | Gold April 2014 first-level cancellation behavior around the spoof of March 3, 2014. This figure visualizes cumulative first-level bid and ask cancellation volume around the spoof of March 3, 2014 in the Gold April 2014 futures market. The *first* panel shows the cumulative volume of cancellations of the best ask level. The *second* panel shows the volumes at the individual bid and ask levels between prices of $1346.9 and $1349.2. Each unit on the x-axis is one message. The y-axis represents the price of Gold in dollars. The color represents the volume at each price level of the LOB for each message. The scale ranges from blue to yellow, with the color becoming a brighter yellow as volume increases at that price level. The red line is the midpoint. The *third* panel shows the cumulative volume of cancellations of the best bid level. The *fourth* panel shows how much time passes between messages reported by the exchange. A steeper (flatter) blue line signals a lower (higher) rate of messages, given that a steeper (flatter) line signals more (less) time progression. The red vertical lines signal when the JPM spoofing activities took place, from left to right: when the genuine order was placed, when the first spoof order of 5 contracts was placed, when the genuine order was executed and when the first spoof order was cancelled.

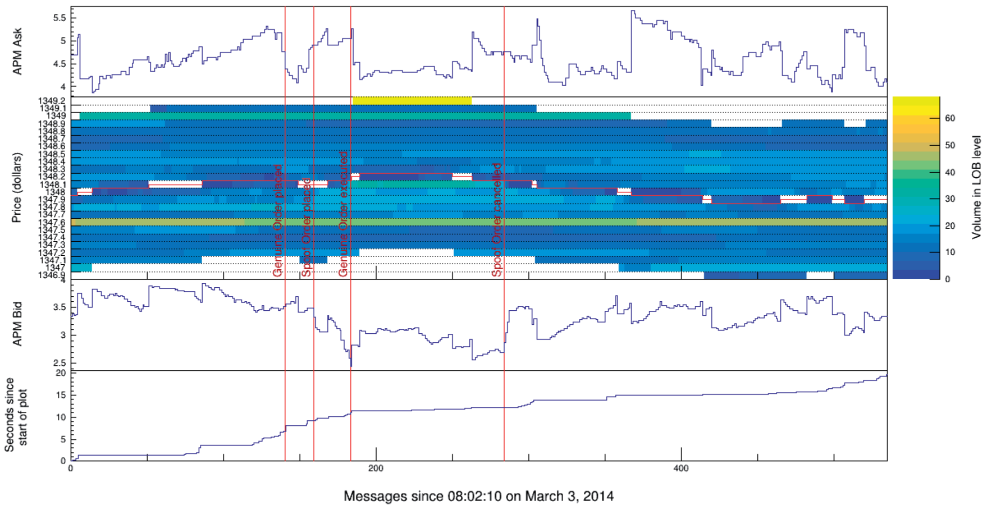


Figure 3.A.20 | Gold April 2014 bid and ask APM behavior around the spoof of March 3, 2014. This figure visualizes bid and ask liquidity costs (APM) behavior around the spoof of March 3, 2014 in the Gold April 2014 futures market. The *first* panel shows the APM of the ask side. APM measures the liquidity costs (in basis points) of a trader who wants to buy or sell a specific dollar value by submitting market orders. The *second* panel shows the volumes at the individual bid and ask levels between prices of $1346.9 and $1349.2. Each unit on the x-axis is one message. The y-axis represents the price of Gold in dollars. The color represents the volume in each price level of the LOB for each message. The scale ranges from blue to yellow, with the color becoming a brighter yellow as volume increases at that price level. The red line is the midpoint. The *third* panel shows the APM for the bid side. The *fourth* panel shows how much time passes between messages reported by the exchange. A steeper (flatter) blue line signals a lower (higher) rate of messages, given that a steeper (flatter) line signals more (less) time progression. The red vertical lines signal when the JPM spoofing activities took place, from left to right: when the genuine order was placed, when the first spoof order of 5 contracts was placed, when the genuine order was executed and when the first spoof order was cancelled.

Table 3.A.15 | Mean bid liquidity costs (bps) around the Gold April 2014 spoof for different time windows

Time window	Before vs. during	During vs. after	Before vs. after
Spoof duration	3.475 > 3.161***	3.161 < 3.382***	3.475 > 3.382***
10 seconds	3.508 > 3.161***	3.161 < 3.356***	3.508 > 3.356***
30 seconds	3.368 > 3.161***	3.161 < 3.283***	3.368 > 3.283***
1 minute	3.280 > 3.161***	3.161 < 3.276***	3.280 = 3.276
5 minutes	3.540 > 3.161***	3.161 < 3.878***	3.540 < 3.878***

Note: The table reports the mean liquidity costs (bps, measured by APM) around the spoof in the Gold April 2014 market for different periods and various time windows. *Before* represents the time up until the spoof order was added; *during* the period from when the spoof order was added until it was cancelled; and *after* the time following the cancellation of the spoof order. Five different time windows are used, the *Spoof duration* time window being 5.56 seconds. A lower APM indicates that liquidity costs are low and, hence, liquidity is high. Welch's *t*-tests were used to test for mean differences between the periods. Significance at the 0.1%, 1% and 5% (two-tailed) levels is indicated by ***, ** and *, respectively.

Appendix 4.A

Five spoofing types are explained, including examples. The examples are simplified and, unless specified otherwise, it is assumed that 1) the spoofer wishes to establish a false sense of supply and demand in the LOB; and 2) the spoofer's goal is to move the price. Note that, while the spoofing types discussed in this appendix have been identified by regulators such as the CFTC and other market stakeholders, this does not mean that these are the only types of spoofing. Other spoofing types may exist that have not yet been identified or named.

4.A.1 SINGLE SPOOFING

We define "single spoofing" as the simplest, most basic type of spoofing, using a single spoof order. In this event, the spoofer places a relatively small order on one side of the LOB intended for execution (the genuine order), followed by a relatively large order on the opposite side of the LOB not intended to be executed (the spoof order) (Debie et al., 2022). The spoof order is used to create a false sense of supply and demand (i.e., market depth) in the LOB and induce the market into the direction of the genuine order (Dalko & Wang, 2020a). Shortly after placement, the spoof order is cancelled or modified deeper into the LOB and then cancelled, either because the genuine order has been executed or because the risk of the spoof order being executed becomes too high. After cancellation, the false supply/demand created is gone and the spoofer succeeded in buying or selling at a better price, sooner or in larger quantities because of the seemingly increased liquidity (CFTC, 2019c; Dalko & Wang, 2018; Debie et al., 2022).

Table 4.A.1 shows a hypothetical example to illustrate single spoofing. The example uses a hypothetical market with a LOB consisting of three levels on the bid and ask sides. Actions by the spoofer are presented in bold and are underlined. Panel A shows the initial state of the market, which has a best bid price of $49.00 and a best ask price of $50.00. The spoofer wishes to buy one contract for $48.00, which is a lower price than the current best bid price. Hence, they add a genuine order at the second bid level. Next, the spoofer wishes to push the price towards their genuine order. They add a single, relatively large spoof order on the ask side to create the impression of increased sell interest (Panel B). Due to the (seemingly) greater sell interest, other market participants expect the price to fall and are induced to trade. They adjust their positions or start selling at once – before the expected price drop –

Table 4.A.1 | Hypothetical example of single spoofing

A) Genuine order is added			
	Level	Price ($)	Volume
Ask	3	53.50	2
	2	53.00	4
	1	50.00	9
Bid	1	49.00	7
	2	48.00	4 **+1**
	3	47.50	12

B) Single spoof order is added			
	Level	Price ($)	Volume
Ask	3	53.50	2
	2	53.00	4 **+50**
	1	50.00	9
Bid	1	49.00	7
	2	48.00	**5**
	3	47.50	12

C) Genuine order is executed			
	Level	Price ($)	Volume
Ask	3	53.50	2
	2	53.00	**54**
	1	50.00	9
Bid	1	47.50	10
	2	47.00	2
	3	46.50	7

D) Spoof order is cancelled			
	Level	Price ($)	Volume
Ask	3	53.50	2
	2	53.00	4
	1	50.00	9
Bid	1	47.50	10
	2	47.00	2
	3	46.50	7

Note: This table shows an example of the 'single spoofing' type. Panels A through D show a limit order book (LOB) of a hypothetical market with three levels on the bid and ask sides. Actions by the spoofer are presented in bold and are underlined. Panel A shows the state of the LOB before a spoof order is added, and the spoofer adds a genuine order to buy one contract at price level $48.00. A spoof order of volume 50 is added on the ask side at price level $53.00 in Panel B. The market responds to the spoof order in Panel C, and the genuine order is executed. The spoof order of volume 50 is cancelled, and the new state of the LOB is displayed in Panel D.

while the price is still high. This often occurs in herd behavior (Dalko & Wang, 2018). In the example, Panel C shows that, due to the spoof order, market participants traded fourteen contracts on the bid side: seven contracts for $49.00, five contracts for $48.00 and two contracts for $47.50. The spoofer successfully executed their genuine order, buying one contract for $47.50. The last step for the spoofer is to cancel the spoof order, resulting in the disappearance of the LOB imbalance created by the spoofer themself (Panel D).

Linking the above example to Figure 4.1, the single spoofing (action) is used to create a false sense of supply (reaction), in order to buy at a lower price (goal). Before the spoof order was added, the total bid side consisted of 23 contracts and the total ask side of 15 contracts, so the ratio of bid to ask orders was 1.53:1. When the spoof order was added, the ratio of bid to ask orders changed to 0.37:1, the total bid side now consisting of 24 contracts and the ask side of 65 contracts. Thus, before the spoof order was added, the LOB was already imbalanced, showing more bid orders. When the spoof order was added, however, the imbalance grew significantly in the inverse direction, exhibiting more ask than bid orders. Moreover, using the figures from the example, the spoofer succeeded in buying one futures contract for $48.00, while the market price was $49.00, i.e., $1 cheaper than without spoofing, excluding transaction costs. Although the prices in the example do not represent real market prices, the underlying values of the WTI Crude Oil futures market are used to give an impression of the underlying value of a $1 price move, where a profit of $1 per futures contract would represent an underlying value of $1000. In turn, the victim of such spoofing would lose $1 per futures contract, i.e., an underlying value of $1000 WTI Crude Oil futures market, as futures markets are a zero-sum game.

4.A.2 LAYERED SPOOFING OR LAYERING

Layered spoofing, or layering, is similar to single spoofing but uses multiple spoof orders rather than a single spoof order. The CFTC defines these layered spoof orders as *"orders with gradually increasing or decreasing prices"* (CFTC, 2019e). However, spoofers can also place multiple spoof orders at the same price level. Hence, in this paper, we define layered spoofing as spoofing through multiple spoof orders, regardless of whether these are being placed at different price levels or at a single price level. Depending on the exchange, the

Table 4.A.2 | Hypothetical example of layered spoofing

A) Genuine order is added

	Level	Price ($)	Volume
Ask	3	53.50	2
	2	53.00	4
	1	50.00	9
Bid	1	49.00	7
	2	48.00	4 +1
	3	47.50	12

B) Multiple spoof orders are added

	Level	Price ($)	Volume
Ask	3	53.50	2 +10
	2	53.00	4 +10
	1	50.00	9 +10
Bid	1	49.00	7
	2	48.00	5
	3	47.50	12

C) Genuine order is executed

	Level	Price ($)	Volume
Ask	3	53.50	12
	2	53.00	14
	1	50.00	19
Bid	1	47.50	10
	2	47.00	2
	3	46.50	7

D) Spoof orders are cancelled

	Level	Price ($)	Volume
Ask	3	53.50	2
	2	53.00	4
	1	50.00	9
Bid	1	47.50	10
	2	47.00	2
	3	46.50	7

Note: This table shows an example of 'layered spoofing'. Panels A through D show a limit order book (LOB) of a hypothetical market with three levels on the bid and ask sides. Actions by the spoofer are presented in bold and are underlined. Panel A shows the state of the LOB before a spoof order is added, and the spoofer adds a genuine order to buy one contract at price level $48.00. Three spoof orders, each with a volume of ten contracts, are added on the ask side at the respective price levels of $50.00, $53.00 and $53.50 in Panel B. The market responds to the spoof orders in Panel C, and the genuine order is executed. The three spoof orders are cancelled, and the new state of the LOB is displayed in Panel D.

advantage of using multiple spoof orders is that other market participants are more likely to perceive these orders as coming from different sources. Spoofers can also layer their spoof orders at multiple price levels, after which they modify the spoof orders to the same price level. This is called “collapsing of layers” and can be used to circumvent individual order-size limits (Neurensic, 2016).

Table 4.A.2 illustrates layered spoofing through the same hypothetical example as discussed in Table 4.A.1. The only difference is in Panel B: rather than using a single spoof order of volume 50 (Table 4.A.1), the spoofer uses three layered spoof orders. Each spoof order is of volume 10 and placed at the first through third ask levels to create an increase in selling pressure and a LOB imbalance. Referring to the elements of Figure 4.1, the layered spoof (action) is used to create a false sense of supply (reaction) in order to buy at a lower price (goal). Recall that the ratio of bid to ask orders was 1.53:1 before the spoof orders were added. Due to the spoof layers, the ask side volume increases by 30 contracts to a total of 45 contracts, and the bid to ask ratio drops to 0.53:1. Hence, while there were more bids than asks before the spoofing started, the placement of the layered spoof orders inverts the imbalance. As with the single spoofing example, the spoofer was able to buy for $48.00, while the market price was $49.00 and, likewise, the gains and underlying values remain unchanged.

4.A.3 FLIPPING

In the event of flipping, the spoofer places a spoof order on one side of the market and switches (“flips”) their position to the other side of the market (CFTC, 2015a). The spoofer places a spoof order at or near the best price level on one side of the market to attract more volume at these levels. Next, the spoofer cancels their spoof order and simultaneously adds an aggressive genuine order on the other side of the market for the same price as the cancelled spoof order, to be executed against the remaining volume at that price level (CFTC, 2015a; MacKenzie, 2022; Sar, 2017). The order can be flipped within a matter of 0.005 seconds (Sar, 2017). This can be done by adding and removing orders or by exploiting the trading functionalities provided by, for example, independent vendors. A case in point is Igor Oystacher (CFTC, 2015a), who exploited the trading function “avoid orders that cross” to execute a flipping strategy.

Table 4.A.3 shows a hypothetical example of flipping, using the same LOB as in previous examples. The spoofer wants to buy 20 contracts for $50.00, but as Panel A shows, the available sell volume for $50.00 is only nine contracts. If the spoofer were to buy 20 contracts in this state of the market, they would incur liquidity costs as they would have to pay more than $50.00 to buy 20 contracts. Hence, the spoofer adds a spoof order on the ask side for 20

Table 4.A.3 | Hypothetical example of flipping

A) Spoof order is added

	Level	Price ($)	Volume
Ask	3	53.50	2
	2	53.00	4
	1	50.00	9 +20
Bid	1	49.00	7
	2	48.00	4
	3	47.50	12

B) Other traders add volume to first ask level

	Level	Price ($)	Volume
Ask	3	53.50	2
	2	53.00	4
	1	50.00	50
Bid	1	49.00	7
	2	48.00	4
	3	47.50	12

C) Spoof order is flipped and add genuine buy market order

	Level	Price ($)	Volume
Ask	3	53.50	2
	2	53.00	4
	1	50.00	30
Market order: buy 20 contracts for 50.00			
Bid	1	49.00	7
	2	48.00	4
	3	47.50	12

D) State LOB after market order is executed

	Level	Price ($)	Volume
Ask	3	53.50	2
	2	53.00	4
	1	50.00	10
Bid	1	49.00	7
	2	48.00	4
	3	47.50	12

Note: This table shows an example of the 'flipping' spoofing. Panels A through D show a limit order book (LOB) of a hypothetical market with three levels on the bid and ask sides. Actions by the spoofer are presented in bold and are underlined. A spoof order of volume 20 is added in Panel A at the first ask level at price $50.00. Other market participants add volume to the first ask level – where the spoof order is resting – totaling to a volume of 50 contracts (Panel B). In Panel C, the spoofer cancels their spoof order and simultaneously adds a market order to buy 20 contracts for $50.00. The market order is executed, and the resulting state of the LOB is shown in Panel D.

contracts at price $50.00 (Panel A). By doing so, the spoofer attempts to attract more volume to this price level. Panel B shows that the spoofer was successful in attracting more volume since an additional 21 contracts were added by other market participants, totaling the volume at the first ask level to 50 contracts. The spoofer can now successfully buy 20 contracts for $50.00 by canceling their spoof order and almost simultaneously adding a buy market order for 20 contracts at price level $50.00, thus changing their position from sell to buy. Panel C demonstrates what happens to the LOB before the market order executes: the first ask level has a volume of 30 rather than 50 because the spoof order is cancelled. Panel D shows the final state of the LOB once the genuine market order executes against the resting ask volume: only ten contracts remain at the ask price level of $50.00.

Linking the flipping example to Figure 4.1, the flipping (action) is used to create a false sense of supply (reaction) to attract more volume to this price level, so as to buy a larger quantity without incurring additional liquidity costs (goal). The ask side volume at price level $50.00 increased by 222%, from 9 to 29 contracts. Total volume on the ask side increased from 15 to 35 contracts, i.e., by 133%. In this example, the spoof order attracted 21 more contracts to the price level of the spoof order, and the spoofer was able to buy 20 contracts for $50.00. If the spoofer had bought the 20 contracts via a market order before adding the spoof order, their order would have consumed all three price levels on the ask side. In this event, they could have bought nine contracts for $50.00, four contracts for $53.00, and two contracts for $53.50, and they would have had to forego the last five contracts due to a lack of ask volume. The total amount due by the spoofer for these 15 contracts would have been $769. Using the flipping strategy, however, allowed the spoofer to buy all twenty contracts at the lowest ask price available: now, the price of the same 15 contracts ($50.00 each) totals $750 – i.e., $19 cheaper than without spoofing, excluding transaction costs. Using the underlying values of the WTI Crude Oil futures market, this $19 difference translates into buying WTI Crude Oil for $19,000 less. Conversely, this means that the other market participants lost $19,000 in the WTI Crude Oil futures market.

4.A.4 SPREAD SQUEEZING

Spread squeezing is used in markets with a wide bid-ask spread. Spoof orders are not placed at existing LOB price levels but within the spread. In other words, buy spoof orders are placed above the best bid price and sell spoof orders are placed below the best ask price. Spoofers use this technique to entice other market participants to join or beat the new price levels. Once more volume is added to the new price levels, the spoofer trades genuine orders against this volume and cancels their spoof orders or vice versa (CME Group, 2017a; Neurensic, 2016). Note that, although spread squeezing and flipping seem very similar, there is an important distinction: in the flipping strategy, spoof orders are placed at or near the best bid/ask level, whereas spoof orders in the spread-squeeze strategy are placed at better levels, i.e., within the bid-ask spread.

Table 4.A.4 illustrates a hypothetical example of spread squeezing, using the same LOB as previous spoofing examples. Note that this example contains a single spoof order, but multiple spoof orders can also be placed using this technique. In this example, the spoofer wants to buy 10 contracts at a price of $49.50. Before the spoofer added their spoof order, the lowest ask price was $50.00. The spoofer adds a sell spoof order at price $49.50 with a volume of one contract, thereby creating a new best ask price (Panel A). Next, the spoofer waits until more volume is added to the newly established ask price. In this case, 19 contracts are added by other participants (Panel B). The spoofer then cancels their spoof order and adds a genuine market order to buy ten contracts for $49.50 each (Panel C). Panel D shows the final state of the market once the spoofer is finished. In this example, the spoof order is cancelled before adding a genuine order; otherwise, the genuine order would have traded against the spoof order. The spoof order can also be cancelled after the genuine order is added if more levels are added by market participants.

Table 4.A.4 | Hypothetical example of spread squeezing

	A) Spoof order is added to new level				**B) Other traders add volume to new best ask level**		
	Level	**Price ($)**	**Volume**		**Level**	**Price ($)**	**Volume**
Ask	3	53.00	4	**Ask**	3	53.00	4
	2	50.00	9		2	50.00	9
	1	49.50	+1		1	49.50	20
Bid	1	49.00	7	**Bid**	1	49.00	7
	2	48.00	4		2	48.00	4
	3	47.50	12		3	47.50	12

	C) Spoof order is cancelled and genuine order is added				**D) Market order is executed**		
	Level	**Price ($)**	**Volume**		**Level**	**Price ($)**	**Volume**
Ask	3	53.00	4	**Ask**	3	53.00	4
	2	50.00	9		2	50.00	9
	1	49.50	19		1	49.50	9
	Market order: buy 10 contracts for 49.50						
Bid	1	49.00	7	**Bid**	1	49.00	7
	2	48.00	4		2	48.00	4
	3	47.50	12		3	47.50	12

Note: This table shows an example of the 'spread-squeezing' spoofing. Panels A through D show a limit order book (LOB) of a hypothetical market with three levels on the bid and ask sides. Actions by the spoofer are presented in bold and are underlined. A spoof order of one contract is added in Panel A at a new best ask level at price $49.50. Other market participants add volume to this newly created first ask level, totaling to a volume of 20 contracts (Panel B). In Panel C, the spoofer cancels their spoof order and simultaneously adds a market order to buy 10 contracts for $49.50. The market order is executed, and the resulting state of the LOB is shown in Panel D.

Linking the above example to Figure 4.1, spread-squeezing (action) is used to tighten the bid-ask spread (reaction), so as to attract more market participants towards this tighter spread and buy at a lower price (goal). The lowest ask price was $50.00 before the spoof order and $49.50 after. Had the spoofer bought ten contracts without spoofing, they would have bought nine contracts for $50.00 and one contract for $53.00, totaling to $503 for ten contracts. However, the spoofer was able to buy ten contracts for $49.50 each, i.e., for $495 in total, due to the spread squeeze. This is $8 cheaper than without spoofing, a difference that would represent an underlying value of $8000 in the WTI Crude Oil futures market.

4.A.5 VACUUMING

Unlike the spoofing types mentioned above, in the vacuuming spoofing strategy, the spoofer adds their genuine order(s) and spoof order(s) on the same side of the market (that is, the LOB). Subsequently, the spoof orders are cancelled almost simultaneously – the "vac-

Table 4.A.5 | Hypothetical example of vacuuming

A) Spoof and genuine orders are added

	Level	Price ($)	Volume
Ask	3	53.50	2 +20
	2	53.00	4 +20
	1	50.00	9 +1
Bid	1	49.00	7
	2	48.00	4
	3	47.50	12

B) New state of the LOB

	Level	Price ($)	Volume
Ask	3	53.50	22
	2	53.00	24
	1	50.00	10
Bid	1	49.00	7
	2	48.00	4
	3	47.50	12

C) Spoof orders are cancelled

	Level	Price ($)	Volume
Ask	3	53.50	2
	2	53.00	4
	1	50.00	10
Bid	1	49.00	7
	2	48.00	4
	3	47.50	12

D) Genuine order is executed

	Level	Price ($)	Volume
Ask	3	54.00	5
	2	53.50	2
	1	53.00	4
Bid	1	49.00	7
	2	48.00	4
	3	47.50	12

Note: This table shows an example of 'vacuuming'. Panels A through D show a limit order book (LOB) of a hypothetical market with three levels on the bid and ask sides. Actions by the spoofer are presented in bold and are underlined. Panel A shows that the spoofer 1) adds a genuine order of one contract at the first ask level at price $50.00; and 2) adds two spoof orders of twenty contracts each at ask levels $53.00 and $53.50. The new state of the LOB is shown in Panel B. Next, in Panel C, the spoofer cancels their spoof orders simultaneously. Market participants respond to this significant change in sell interest and cross the spread, thereby executing the spoofer's genuine order (Panel D).

uum" – to signal a sudden and significant decline in interest, indicating a likely price change. By doing so, the spoofer tries to tempt other market participants to cross the spread and trade against their still resting genuine orders (CFTC, 2019d).

Table 4.A.5 illustrates how the vacuuming strategy is executed through a hypothetical example. The spoofer wishes to sell one contract for $50.00. They add one genuine ask order at price level $50.00 for one contract, and two spoof orders with a volume of 20 contracts each at price levels $53.00 and $53.50 (Panel A). Panel B shows the state of the LOB after these actions. Next, the spoofer cancels all spoof orders simultaneously (Panel C). Other market participants react to this significant change in sell interest and cross the spread: they buy ten contracts for $50.00 each, including the spoofer's genuine order (Panel D).

Linking the example to Figure 4.1, vacuuming (action) is used to create a false sense of supply and a significant change in sell interest (reaction), so as to induce more market participants to cross the spread (goal). The ask side had a total of 15 contracts before the spoof and genuine orders were added and a total of 56 contracts – an increase of 273% – afterwards. The ask side volume dropped by 40 contracts when all spoof orders were cancelled. More market participants were tempted to cross the spread because of the spoofing, resulting in a faster execution of the spoofer's genuine order. Hence, the spoofer's goal can be linked to the immediacy dimension of liquidity as they wish to trade more quickly (Hasbrouck, 2021). Had they been forced to postpone the trade, the market could have moved away from the genuine-order price level and the spoofer might have had to sell at a lower price than $50.00.

Appendix 4.B

Details on the spoofing examples mentioned in spoofing cases by the CFTC, CME Group and ICE, from January 2010 to June 2022

CFTC case number	Defendants (short name)	Market	Maturity month	Date	Timestamps (from – until; CT)
2008					
18-cv-00158	Andre Flotron	Gold	December	Aug 29, 2008	13:10:15+ ET
18-cv-00158	Andre Flotron	Gold	December	Oct 9, 2008	08:42:49 – 08:52:03 ET
2009					
20-69	JPMorgan	T-Bond	September 2009	Jul 20, 2009	07:47:13 – 07:47:23+
12-cv-8791	Eric Moncada, BES Capital & Serdika	Wheat	December 2009	Oct 27, 2009	09:36:00 – 09:38:01
12-cv-8791	Eric Moncada, BES Capital & Serdika	Wheat	December 2009	Oct 29, 2009	10:33:19 – 10:39:31
2010					
20-69	JPMorgan	10-Year T-Note	March 2010	Feb 4, 2010	13:27:27 – 13:30:02
15-cv-03398	Navinder Sarao	E-mini S&P 500	NA	May 4, 2010	11:00:36 – 11:12:39
15-cv-03398	Navinder Sarao	E-mini S&P 500	NA	May 5, 2010	11:22 – 12:30
15-cv-03398	Navinder Sarao	E-mini S&P 500	NA	May 6, 2010	09:20 and 11:17 – 13:40
2011					
18-cv-00603	James Vorley & Cedric Chanu	Gold	NA	Mar 16, 2011	06:03:21+
18-cv-00158	Andre Flotron	Gold	December	Sep 6, 2011	05:06:32 – 05:06:43 ET
20-69	JPMorgan	10-Year T-Note	December 2011	Sep 27, 2011	14:03:54 – 14:03:58+
20-69	JPMorgan	Silver	March 2012	Dec 12, 2011	11:59:36 – 11:59:41
2012					
18-cv-00603	James Vorley & Cedric Chanu	Gold	NA	Jan 6, 2012	01:59:04 – 01:59:25
18-cv-00603	James Vorley & Cedric Chanu	Gold	NA	Jun 6, 2012	01:26:57 – 01:27:15+
18-cv-00158	Andre Flotron	Gold	August	Jun 28, 2012	13:26:33 – 13:26:53 ET
15-cv-09196	Igor Oystacher & 3 Red Trading	Natural Gas	NA	Nov 30, 2012	08:02:34 – 08:02:36+

CFTC case number	Defendants (short name)	Market	Maturity month	Date	Timestamps (from – until; CT)
2013					
19-cv-06163	Michael Nowak & Gregg Smith	Silver	March 2013	Jan 25, 2013	07:55:28 – 07:55:30+
18-cv-00603	James Vorley & Cedric Chanu	Gold	NA	Jan 28, 2013	15:32:43 – 15:33:16
18-cv-00158	Andre Flotron	Silver	March	Feb 15, 2013	05:59:28 – 05:59:35 ET
18-cv-00619	Jitesh Thakkar & Edge Fin. Tech.	E-mini S&P 500	March 2013	Feb 22, 2013	10:17:07+
18-cv-00619	Jitesh Thakkar & Edge Fin. Tech.	E-mini S&P 500	March 2013	Feb 25, 2013	12:48:52+
18-cv-00620	Jiongsheng Zhao	E-mini S&P 500	June	Apr 15, 2013	01:11:16 – 01:11:35
19-cv-06163	Michael Nowak & Gregg Smith	Gold	June 2013	May 6, 2013	09:01:13 – 09:01:24
18-cv-00603	James Vorley & Cedric Chanu	Gold	June	May 23, 2013	01:05:45 – 01:07:58
19-cv-06163	Michael Nowak & Gregg Smith	Gold	June 2013	May 23, 2013	14:00:22 – 14:00:33
18-cv-00603	James Vorley & Cedric Chanu	Gold	August	Jun 2, 2013	19:52:13 – 19:55:15
18-cv-00158	Andre Flotron	Gold	August	Jul 1, 2013	10:11:06 – 10:11:56 ET
18-cv-00603	James Vorley & Cedric Chanu	Gold	NA	Jul 9, 2013	15:51:59 – 15:52:06
15-cv-03398	Navinder Sarao	E-mini S&P 500	NA	Sep 30, 2013	12:30:38 – 12:30:50
19-cv-06163	Michael Nowak & Gregg Smith	Gold	December 2013	Oct 7, 2013	07:34:45 – 07:34:52+
18-cv-00158	Andre Flotron	Gold	December	Oct 14, 2013	08:26:47 – 08:26:54 ET
18-cv-00158	Andre Flotron	Gold	December	Oct 17, 2013	11:13:35 – 11:15:50 ET
18-cv-00260	Krishna Mohan	E-mini Dow ($5)	December 2013	Nov 27, 2013	08:05:39 – 08:05:48
18-cv-00260	Krishna Mohan	E-mini Nasdaq-100	December 2013	Dec 2, 2013	03:01:50 – 03:02:12
18-cv-00260	Krishna Mohan	E-mini Nasdaq-100	December 2013	Dec 3, 2013	02:34:47 – 02:34:54
18-cv-00260	Krishna Mohan	E-mini Nasdaq-100	December 2013	Dec 9, 2013	05:27:41 – 05:27:48
20-69	JPMorgan	Silver	March 2014	Dec 10, 2013	01:59:22 – 01:59:28
19-01	Kamaldeep Gandhi	E-mini S&P 500	March 2014	Dec 16, 2013	04:45:56 – 04:46:14
2014					
15-cv-09196	Igor Oystacher & 3 Red Trading	E-mini S&P 500	NA	Jan 6, 2014	14:01:45 – 14:01:48+
20-69 and 19-cv-06163	JPMorgan and Michael Nowak & Gregg Smith	Gold	April 2014	Mar 3, 2014	08:02:17 – 08:02:23+
15-cv-03398	Navinder Sarao	E-mini S&P 500	NA	Mar 3, 2014	11:38:27 –11:38:33
20-69 and 19-16	JPMorgan and John Edmonds	Silver	May 2014	Mar 5, 2014	08:18:39 – 08:18:42+
18-cv-00620	Jiongsheng Zhao	E-mini S&P 500	March	Mar 5, 2014	22:44:12 – 22:44:41
19-01	Kamaldeep Gandhi	E-mini S&P 500	December 2014	Sep 17, 2014	10:09:14 – 10:09:17

CFTC case number	Defendants (short name)	Market	Maturity month	Date	Timestamps (from – until; CT)
2015					
15-cv-03497	Heet Khara & Nasim Salim	Silver	NA	Feb 18, 2015	04:02:17 – 04:02:28+
15-cv-03497	Heet Khara & Nasim Salim	Gold	April 2015	Mar 25, 2015	03:43:57 – 03:44:05
15-cv-03497	Heet Khara & Nasim Salim	Gold	June 2015	Apr 27, 2015	05:49:37 – 05:49:44+
19-cv-06163	Michael Nowak & Gregg Smith	Gold	August 2015	Jun 11, 2015	07:31:25 – 07:31:28
20-69	JPMorgan	Ultra T-Bond	September 2015	Jun 30, 2015	08:45:46 – 08:46:09
19-15	Corey Flaum	Gold	February 2016	Dec 31, 2015	11:39:10 – 11:40:35
2016					
18-cv-00620	Jiongsheng Zhao	E-mini S&P 500	June	Mar 21, 2016	07:31:28 – 07:31:35
20-69 and 19-26	JPMorgan and Christian Trunz	Platinum	July 2016	Jun 22, 2016	02:14:33 – 02:14:38
22-cv-01925	David Skudder, Global AG & Nesvick Trading Group	Soybean (futures & options)	NA	Jul 1, 2016	11:29:14 – 11:29:18
22-cv-01925	David Skudder, Global AG & Nesvick Trading Group	Soybean (futures & options)	NA	Jul 11, 2016	10:35:58 – 10:36:10
2017					
22-cv-01925	David Skudder, Global AG & Nesvick Trading Group	Soybean (futures)	NA	Aug 31, 2017	11:15:05 – 11:15:47
22-cv-01925	David Skudder, Global AG & Nesvick Trading Group	Soybean (futures)	NA	Dec 5, 2017	08:47:20 – 08:49:02
2018					
20-cv-05777	Roman Banovzay (Jr. & Sr.) & Bazur	Crude Oil	March 2018	Jan 25, 2018	06:31:40 – 06:31:45
20-cv-05777	Roman Banovzay (Jr. & Sr.) & Bazur	Crude Oil	March 2018	Feb 1, 2018	02:10:26 – 02:10:47
20-cv-05777	Roman Banovzay (Jr. & Sr.) & Bazur	Crude Oil	March 2018	Feb 12, 2018	03:02:13 – 03:02:30

Note: This table lists the spoofing cases that provide detailed examples in their descriptions. It lists the case number; shortened name of the defendants; the market that was spoofed, including maturity month; the date; and the beginning and ending timestamps of the spoofing examples. Detailed examples are included whenever the spoofing cases included specific timestamps. A "+" is added to every timestamp that lacks the exact time of when the spoofing ended. Some spoofing cases included examples without timestamps but containing information that could help derive the spoofing from the data; these examples are not included in this table but can be found in the following CFTC spoofing cases: Citigroup Global Markets Inc. (Case No. 17-06), Logista Advisors LLC (Case No. 17-29), James Vorley & Cedric Chanu (Case No. 18-cv-00603), Michael Franko (Case No. 18-35), Geneva Trading USA LLC (Case No. 18-37) and JPMorgan (Case No. 20-69).

Appendix 4.C

Verhulst, M. E. (2024). Online appendices for dissertation. [Data set]. *Zenodo*. https://doi.org/10.5281/zenodo.11472942

Appendix 5.A

Verhulst, M. E. (2024). Online appendices for dissertation. [Data set]. *Zenodo*. https://doi.org/10.5281/zenodo.11472942

Appendix 5.B

For illustration purposes, the LOB state approach is visualized for a known spoofing case, as the market data of this spoofing case were available to the authors through previous research. Specifically, we visualize a spoof order placed by a JPMorgan trader (CFTC, 2020e) in the Silver March 2014 futures market on December 10, 2013. The data consists of the CME Group's proprietary market-depth data set in Market Depth 3.0 format. Different from this format to MBO data is that each market-message provides an update to the LOB: for example, a price level being added, modified or removed. MBO data, on the other hand, contains messages specific to orders: for example, when an order is added, modified or removed. The LOB of the dataset for this example can be reconstructed up until level ten on each LOB side.

For each message[49] on December 10, 2013, we calculate the LOB state and calculate the difference between two consecutive messages to receive the LOB state differences. These differences are then plotted in a histogram in Figure 5.B.1. The two vertical lines represent the 1% and 99% boundaries of the histogram. The difference the JPMorgan spoof order causes in the LOB state is marked by the black circle on the right side of the figure.[50]

Figure 5.B.1 shows the spoof order by the JPMorgan trader as a clear outlier, as it falls well outside the 1% and 99% boundaries. Hence, combining the LOB-state difference approach with other spoofing criteria shows promising results for identifying spoof orders.

49 This is different from the approach in the paper, as here we calculate the LOB state for *all* messages, while in the paper we calculate this only for *added* orders.

50 It should be noted that with only ten LOB levels, large differences in the LOB state can be the result of the tenth level disappearing (or reappearing) in the LOB due to the addition (deletion) of a new (old) level. This effect is eliminated as much as possible in the paper by using fifty LOB levels on either side, rather than ten levels.

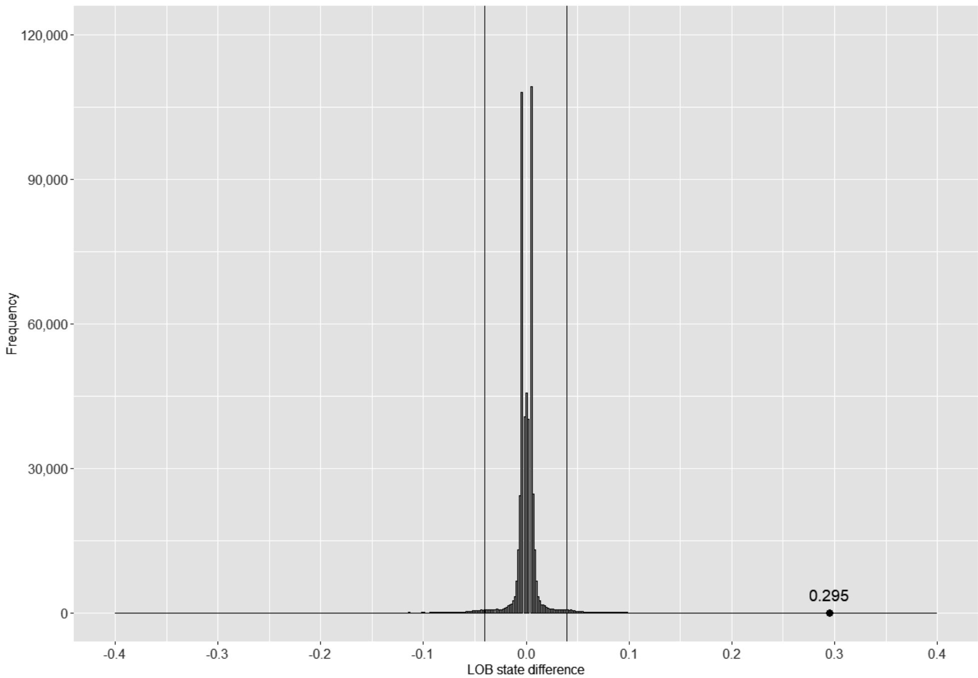


Figure 5.B.1 | LOB-state difference distribution on December 10, 2013 in the Silver March 2014 contract from CME Group. This figure shows the distribution in LOB-state differences on a day a JPMorgan trader placed a spoof order in the Silver March 2014 contract from CME Group. The two vertical lines represent the 1% and 99% boundaries of the distribution, and the black circle on the right of the figure represents the change the spoof order by JPMorgan caused in the LOB state.

Appendix 5.C

This paper defines a spoof order as an order that 1) is added to the LOB and cancelled within one second; 2) is not (partially) executed; 3) is not modified and then cancelled; 4) is placed on one of the top five LOB levels; 5) does not create a new best bid or ask level; and 6) causes a change in the LOB state that is larger than 98% of all other added orders in the market for that day. Criterion 6 depends on the following parameters: a) the type of message the LOB state change is calculated for. That is, if the LOB state change is calculated only for orders that are added to the LOB or for all market activity – adding, modifying or cancelling orders and transactions. b) Distribution time window; if distributions of the LOB state change are made per day, week or year. c) The number of levels that are used to calculate the LOB state, for example 10, 50 or 100 levels. d) The percentile that is used for the tails of the LOB-state change distributions. In the paper, we set the parameters to calculate the LOB state change for orders that are added to the LOB (parameter a); we make a distribution per day (parameter b); use 50 price levels per side (parameter c); and use the one percentile at both tails of the distribution (parameter d). For illustration purposes, Table 5.C.1 shows the number of spoof orders that are identified in the corn futures market of CME Group between July 2019 and June 2020, when parameters a) to d) from criterion 6 vary but criteria 1 to 5 are constant.

Table 5.C.1 | Variation in number of identified spoof orders in the corn futures market of CME Group between July 2019 and June 2020.

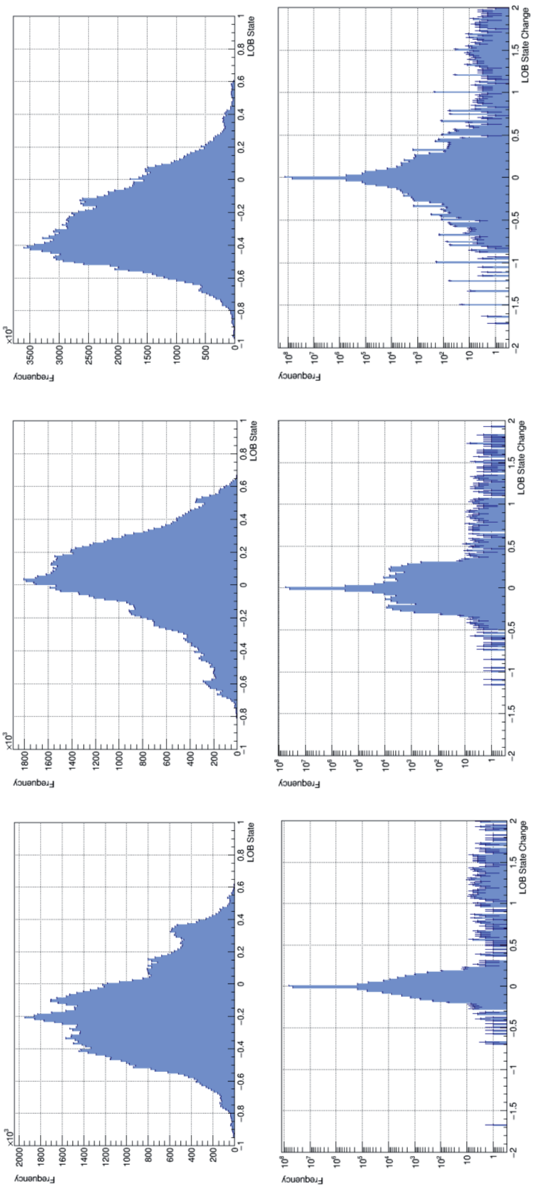
		Messages: all market activity			Messages: added orders		
		Distribution per			Distribution per		
Tail percentage	LOB levels	Day	Week	Year	Day	Week	Year
1%	10	60,342	60,490	61,042	42,940	43,536	47,099
	50	55,101	55,365	59,535	39,347	40,382	45,563
	100	51,533	52,374	59,930	37,432	38,592	44,791
0.10%	10	5790	5841	8162	4441	4661	7327
	50	5279	5400	7777	3749	3962	6755
	100	5104	5250	7728	3652	3826	6570
0.01%	10	902	925	1277	823	861	1230
	50	508	503	974	348	382	788
	100	513	525	904	349	374	720

Note: This table shows the variation in identified spoof orders in the corn futures market of CME Group between July 2019 and June 2020, when four parameters of spoofing criterion (6) differ and criterion (1) to (5) are constant. Spoofing criterion (6) states that order causes a change in the LOB state that is larger than 98% of all other added orders in the market for that day. The parameters that can vary are the type of message the limit order book (LOB) state change is calculated for (all market activity vs. added orders); if the distributions are made per day, week or year; the number of LOB levels that are used to calculate the LOB state; and the percentile that is used for the tails of the LOB-state change distributions. The LOB state is calculated by extracting the total ask volume from the total bid volume, and dividing this by the summation of the total bid and ask volume.

Appendix 5.D

Table 5.D.1 | LOB state of agricultural CME Group futures from July 2019 to June 2020.

A. Corn
B. Wheat
C. Soybean



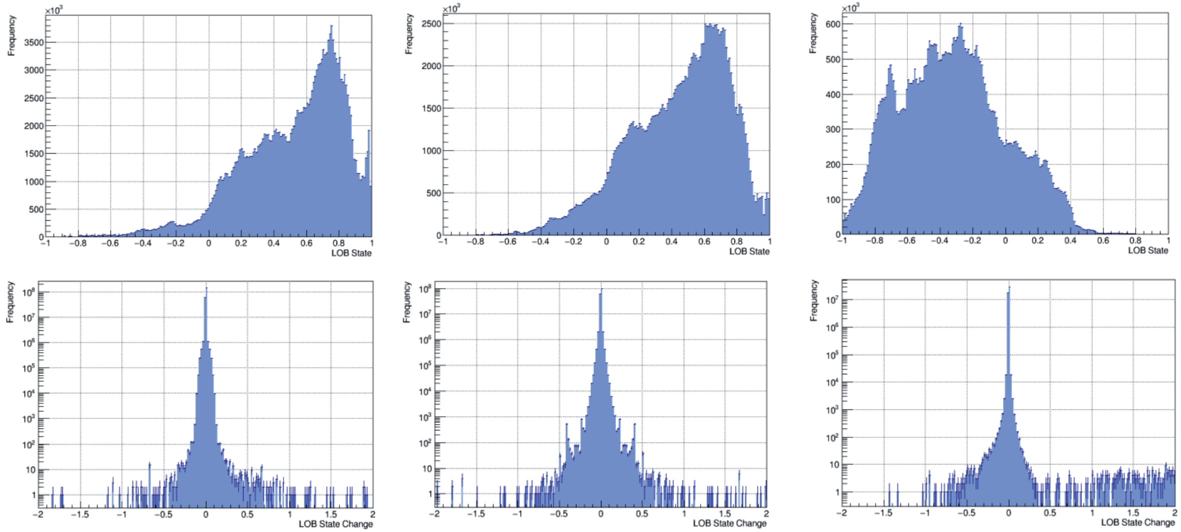
LOB state
LOB state change

Table 5.D.1 | Continued

D. Soybean meal E. Soybean oil F. Live cattle

LOB state

LOB state change



Note: This table shows the limit order book (LOB) state and LOB state change for the corn, wheat, soybean, soybean meal, soybean oil and live cattle futures markets. The LOB state is calculated for each message by extracting the total ask volume from the total bid volume, and dividing this by the summation of the total bid and ask volume. The LOB state change is calculated by the difference in LOB state between two consecutive messages. All message types are used for the distributions: added orders, canceled orders, modified orders and trades.

Appendix 5.E

Table 5.E.1 | Impact of spoofing on liquidity in the corn futures market from CME Group between July 2019 to June 2020.

Liquidity measure	Bid spoof orders			Ask spoof orders		
	During vs. before	**During vs. after**	**After vs. before**	**During vs. before**	**During vs. after**	**After vs. before**
Bid-ask spread	−0.0005***	0.0007***	0.0002***	2.234e-05	0.0003***	0.0003***
Bid depth $_{1–5}$	0.826***	−0.388***	0.438***	0.677***	−0.709***	−0.031
Bid depth $_{6–10}$	0.588***	−0.548***	0.040	−0.322***	0.239***	−0.083***
Ask depth $_{1–5}$	0.737***	−0.799***	−0.063	1.011***	−0.647***	0.364***
Ask depth $_{6–10}$	−0.305***	0.192***	−0.113***	0.563***	−0.598***	−0.035
Bid orders	−0.182***	0.157***	−0.025***	−0.021***	−0.014**	−0.035***
Ask orders	−0.008	−0.018**	−0.025***	−0.216***	0.179***	−0.036***
APM bid	−0.122***	0.103***	−0.019**	0.001	0.038**	0.039***
APM ask	−0.020***	0.037***	0.017***	−0.121***	0.118***	−0.003

* = 0.10; ** = 0.05; *** = 0.01.

Note: This table reports the coefficients and significance of how various liquidity measures behave around spoof orders, based on panel data regressions on 39,347 identified spoof orders – 19,293 bid spoof orders 20,054 ask spoof orders – in the corn futures market of CME Group between July 2019 to June 2020. *Before* represents 30 seconds before the spoof orders. *During* is the period from when the spoof order was added until it was cancelled; due to the research design this is a maximum of one second. *After* represents 30 seconds after the spoof orders were cancelled. In the column names, the period name after the 'vs.' is the reference period. Numbers are rounded to the third decimal, unless the fourth decimal is necessary for interpretation purposes.

Table 5.E.2 | Impact of spoofing on liquidity in the wheat futures market from CME Group between July 2019 to June 2020.

Liquidity measure	Bid spoof orders			Ask spoof orders		
	During vs. before	During vs. after	After vs. before	During vs. before	During vs. after	After vs. before
Bid-ask spread	0.0001	−0.0001	−4.005e-05	0.001***	−0.001**	8.158e-05
Bid depth $_{1–5}$	0.227***	−0.1442**	0.082***	0.032	0.092***	0.124***
Bid depth $_{6–10}$	−0.044**	0.0289	−0.015	−0.054**	0.027	−0.027*
Ask depth $_{1–5}$	−0.046**	0.137***	0.092***	0.285***	−0.182***	0.104***
Ask depth $_{6–10}$	0.007	−0.005	0.002	0.008	0.044**	0.051***
Bid orders	0.009***	0.009***	0.018***	−0.034***	0.047***	0.013***
Ask orders	−0.037***	0.053***	0.016***	0.042***	−0.018***	0.024***
APM bid	−0.006	−0.0005	−0.006***	0.054***	−0.057***	−0.003
APM ask	0.058***	−0.046***	0.012***	0.013	−0.010	0.003

* = 0.10; ** = 0.05; *** = 0.01.
Note: This table reports the coefficients and significance of how various liquidity measures behave around spoof orders, based on panel data regressions on 20,635 identified spoof orders – 10,884 bid spoof orders 9751 ask spoof orders – in the wheat futures market of CME Group between July 2019 to June 2020. *Before* represents 30 seconds before the spoof orders. *During* is the period from when the spoof order was added until it was cancelled; due to the research design this is a maximum of one second. *After* represents 30 seconds after the spoof orders were cancelled. In the column names, the period name after the 'vs.' is the reference period. Numbers are rounded to the third decimal, unless the fourth decimal is necessary for interpretation purposes.

Table 5.E.3 | Impact of spoofing on liquidity in the soybean futures market from CME Group between July 2019 to June 2020.

Liquidity measure	Bid spoof orders			Ask spoof orders		
	During vs. before	During vs. after	After vs. before	During vs. before	During vs. after	After vs. before
Bid-ask spread	0.0007***	−0.0007***	2.095e-05	0.0004***	−0.0003***	0.0001***
Bid depth $_{1–5}$	0.676***	−0.687***	−0.011	0.289***	−0.270***	0.019**
Bid depth $_{6–10}$	−9.04e-05	−0.006	−0.007	0.031*	−0.026	0.005
Ask depth $_{1–5}$	0.514***	−0.501***	0.013	0.782***	−0.760***	0.022**
Ask depth $_{6–10}$	0.087***	−0.113***	−0.026**	−0.064***	0.022	−0.042***
Bid orders	−0.101***	0.082***	−0.018***	0.023***	−0.027***	−0.004**
Ask orders	0.046***	−0.068***	−0.022***	−0.129***	0.101***	−0.028***
APM bid	−0.146***	0.146***	−0.0001	0.001	0.007	0.009*
APM ask	−0.041***	0.045***	0.004***	−0.059***	0.064***	0.005***

* = 0.10; ** = 0.05; *** = 0.01.
Note: This table reports the coefficients and significance of how various liquidity measures behave around spoof orders, based on panel data regressions on 104,200 identified spoof orders – 53,059 bid spoof orders 51,141 ask spoof orders – in the soybean futures market of CME Group between July 2019 to June 2020. *Before* represents 30 seconds before the spoof orders. *During* is the period from when the spoof order was added until it was cancelled; due to the research design this is a maximum of one second. *After* represents 30 seconds after the spoof orders were cancelled. In the column names, the period name after the 'vs.' is the reference period. Numbers are rounded to the third decimal, unless the fourth decimal is necessary for interpretation purposes.

Table 5.E.4 | Impact of spoofing on liquidity in the soybean meal futures market from CME Group between July 2019 to June 2020.

Liquidity measure	Bid spoof orders			Ask spoof orders		
	During vs. before	**During vs. after**	**After vs. before**	**During vs. before**	**During vs. after**	**After vs. before**
Bid-ask spread	0.0003***	−0.00005	0.0002***	0.0004***	−0.0002***	0.0002***
Bid depth $_{1–5}$	0.304***	−0.311***	−0.007	0.320***	−0.361***	−0.040***
Bid depth $_{6–10}$	−0.108***	0.035	−0.073***	−0.055**	−0.005	−0.060***
Ask depth $_{1–5}$	0.106***	−0.065***	0.042***	0.297***	−0.271***	0.027***
Ask depth $_{6–10}$	−1.266e-05	−0.008	−0.008	−0.010	−0.036***	−0.046***
Bid orders	−0.053***	0.023***	−0.031***	0.018***	−0.044***	−0.027***
Ask orders	−0.003	−0.011***	−0.014***	−0.035***	0.016***	−0.020***
APM bid	−0.118***	0.129***	0.011***	0.037***	−0.033***	0.004***
APM ask	0.036***	−0.031***	0.006**	−0.114***	0.122***	0.009**

* = 0.10; ** = 0.05; *** = 0.01.
Note: This table reports the coefficients and significance of how various liquidity measures behave around spoof orders, based on panel data regressions on 43,444 identified spoof orders – 20,461 bid spoof orders 22,983 ask spoof orders – in the soybean meal futures market of CME Group between July 2019 to June 2020. *Before* represents 30 seconds before the spoof orders. *During* is the period from when the spoof order was added until it was cancelled; due to the research design this is a maximum of one second. *After* represents 30 seconds after the spoof orders were cancelled. In the column names, the period name after the 'vs.' is the reference period. Numbers are rounded to the third decimal, unless the fourth decimal is necessary for interpretation purposes.

Table 5.E.5 | Impact of spoofing on liquidity in the soybean oil futures market from CME Group between July 2019 to June 2020.

Liquidity measure	Bid spoof orders			Ask spoof orders		
	During vs. before	During vs. after	After vs. before	During vs. before	During vs. after	After vs. before
Bid-ask spread	−4.806e-05***	0.00006***	9.082e-06***	−2.59e-05***	0.00003***	2.666e-06
Bid depth $_{1-5}$	0.332***	−0.39***	−0.055***	0.011	−0.016	−0.005
Bid depth $_{6-10}$	0.032	−0.199***	−0.168***	−0.064***	−0.047**	−0.111***
Ask depth $_{1-5}$	−0.046***	−0.010	−0.056***	0.169***	−0.174***	−0.006
Ask depth $_{6-10}$	−0.038***	−0.019***	−0.056***	−0.027***	−0.047***	−0.074***
Bid orders	−0.078***	0.025***	−0.053***	−0.060***	0.010***	−0.050***
Ask orders	−0.061***	0.014***	−0.048***	−0.059***	0.005*	−0.054***
APM bid	−0.031***	0.035***	0.005**	0.036***	−0.024***	0.012**
APM ask	0.055***	−0.048***	0.007***	−0.042***	0.060***	0.018***

* = 0.10; ** = 0.05; *** = 0.01.

Note: This table reports the coefficients and significance of how various liquidity measures behave around spoof orders, based on panel data regressions on 37,002 identified spoof orders – 18,387 bid spoof orders 18,615 ask spoof orders – in the soybean oil futures market of CME Group between July 2019 to June 2020. *Before* represents 30 seconds before the spoof orders. *During* is the period from when the spoof order was added until it was cancelled; due to the research design this is a maximum of one second. *After* represents 30 seconds after the spoof orders were cancelled. In the column names, the period name after the 'vs.' is the reference period. Numbers are rounded to the third decimal, unless the fourth decimal is necessary for interpretation purposes.

Table 5.E.6 | Impact of spoofing on liquidity in the live cattle futures market from CME Group between July 2019 to June 2020.

Liquidity measure	Bid spoof orders			Ask spoof orders		
	During vs. before	During vs. after	After vs. before	During vs. before	During vs. after	After vs. before
Bid-ask spread	−0.0003***	0.0002	−0.0001***	−0.0002***	0.000	−0.0002***
Bid depth $_{1-5}$	0.100***	−0.036	0.064***	−0.031	0.066***	0.035***
Bid depth $_{6-10}$	0.089***	−0.013	0.076***	0.099***	0.0003	0.100***
Ask depth $_{1-5}$	0.220***	−0.087**	0.133***	0.224***	−0.082***	0.143***
Ask depth $_{6-10}$	0.125***	0.059	0.184***	0.179***	−0.033	0.146***
Bid orders	0.012**	0.014**	0.025***	0.004	0.021***	0.024***
Ask orders	0.011*	0.028***	0.038***	0.013***	0.013***	0.027***
APM bid	−0.051	−0.007	−0.058***	0.203***	−0.240***	−0.037***
APM ask	0.044	−0.071*	−0.027**	−0.090***	0.071**	−0.020*

* = 0.10; ** = 0.05; *** = 0.01.

Note: This table reports the coefficients and significance of how various liquidity measures behave around spoof orders, based on panel data regressions on 4080 identified spoof orders – 1690 bid spoof orders 2390 ask spoof orders – in the live cattle futures market of CME Group between July 2019 to June 2020. *Before* represents 30 seconds before the spoof orders. *During* is the period from when the spoof order was added until it was cancelled; due to the research design this is a maximum of one second. *After* represents 30 seconds after the spoof orders were cancelled. In the column names, the period name after the 'vs.' is the reference period. Numbers are rounded to the third decimal, unless the fourth decimal is necessary for interpretation purposes.

Appendix 5.F

Table 5.F.1 | Panel data regressions of the impact of spoofing on liquidity in the corn futures market from CME Group between July 2019 to June 2020.

Variables	Bid-ask spread	Bid depth 1–5	Bid depth6–10	Ask depth 1–5	Ask depth 6–10	Bid orders	Ask orders	APM bid	APM ask
Bid spoof orders									
First lag	0.7299***	0.9613***	0.9567***	0.9447***	0.9340***	0.9508***	0.9321***	1.0042***	1.0172***
	(0.0055)	(0.0011)	(0.0021)	(0.0031)	(0.0051)	(0.0010)	(0.0067)	(0.0073)	(0.0047)
Second lag	0.1458***	0.0247***	0.0286***	0.0430***	0.0511***	0.0344***	0.0559***	-0.0056	-0.0233***
	(0.0034)	(0.0011)	(0.0021)	(0.0029)	(0.0051)	(0.0010)	(0.0065)	(0.0073)	(0.0044)
DuringSpoof	-0.0005***	0.8258***	0.5875***	0.7365***	-0.3047***	-0.1822***	-0.0076	-0.1221***	-0.0198***
	(7.562e-05)	(0.0856)	(0.0551)	(0.0857)	(0.0600)	(0.0060)	(0.0102)	(0.0174)	(0.0053)
AfterSpoof	0.0002***	0.4379***	0.0399	-0.0625	-0.1127***	-0.0248***	-0.0251***	-0.0187**	0.0174***
	(2.909e-05)	(0.0470)	(0.0279)	(0.0581)	(0.0405)	(0.0047)	(0.0085)	(0.0093)	(0.0037)
dTimestamp	-0.0016***	0.2531***	0.0871***	-0.1168***	0.1090***	0.0286***	0.0275***	-0.2148***	-0.0803***
	(3.117e-05)	(0.0368)	(0.0255)	(0.0266)	(0.0208)	(0.0021)	(0.0025)	(0.0100)	(0.0047)
F-test	1.209e+07***	1.486e+08***	1.415e+08***	1.652e+08***	1.362e+08***	1.399e+08***	1.7e+08***	1.411e+09***	3.55e+08***
R-squared	0.7359	0.9716	0.9703	0.9744	0.9691	0.9699	0.9751	0.9969	0.9879
N	21,706,193	21,706,193	21,706,193	21,706,193	21,706,193	21,706,193	21,706,193	21,706,193	21,706,193
Ask spoof orders									
First lag	0.7140***	0.9615***	0.9606***	0.9482***	0.9370***	0.9523***	0.9366***	0.9921***	1.0312***
	(0.0041)	(0.0008)	(0.0013)	(0.0027)	(0.0046)	(0.0008)	(0.0057)	(0.0078)	(0.0033)
Second lag	0.1352***	0.0256***	0.0271***	0.0373***	0.0487***	0.0345***	0.0482***	0.0059	-0.0492***
	(0.0024)	(0.0008)	(0.0013)	(0.0026)	(0.0045)	(0.0008)	(0.0056)	(0.0078)	(0.0037)
DuringSpoof	2.234e-05	0.6774***	-0.3221***	1.0106***	0.5633***	-0.0214***	-0.2156***	0.0008	-0.1205***
	(8.416e-05)	(0.0781)	(0.0438)	(0.1028)	(0.0562)	(0.0056)	(0.0116)	(0.0180)	(0.0132)
AfterSpoof	0.0003***	-0.0313	-0.0827***	0.3640***	-0.0345	-0.0351***	-0.0363***	0.0392***	-0.0027
	(3.461e-05)	(0.0495)	(0.0315)	(0.0563)	(0.0307)	(0.0044)	(0.0076)	(0.0082)	(0.0078)
dTimestamp	-0.0015***	-0.1655***	0.1570***	0.3165***	0.2000***	0.0347***	0.0183***	-0.1320***	-0.2438***
	(2.502e-05)	(0.0327)	(0.0199)	(0.0364)	(0.0183)	(0.0021)	(0.0029)	(0.0180)	(0.0128)
F-test	9.73e+06***	1.641e+08***	1.74e+08***	1.444e+08***	1.443e+08***	1.606e+08***	1.362e+08***	1.017e+09***	1.259e+08***
R-squared	0.6874	0.9737	0.9752	0.9703	0.9702	0.9732	0.9685	0.9957	0.9661
N	22,140,247	22,140,247	22,140,247	22,140,247	22,140,247	22,140,247	22,140,247	22,140,247	22,140,247

Note: standard errors are reported in parentheses. *, **, and *** indicates statistical significance at 10%, 5% and 1% level, respectively. *First lag* and *Second lag* are the first and second lag of the respective liquidity measure. *DuringSpoof* and *AfterSpoof* are dummy variables that categorize if an observation occurred during spoofing (*DuringSpoof* = 1) or after the spoof order was removed from the limit order book (*AfterSpoof* = 1). This sets the period before the spoof order was added as the reference category.

Table 5.F.2 | Panel data regressions of the impact of spoofing on liquidity in the wheat futures market from CME Group between July 2019 to June 2020.

Variables	Bid-ask spread	Bid depth 1–5	Bid depth 6–10	Ask depth 1–5	Ask depth6–10	Bid orders	Ask orders	APM bid	APM ask
Bid spoof orders									
First lag	0.7508***	0.9389***	0.9330***	0.9447***	0.9409***	0.9340***	0.9312***	1.0421***	1.0497***
	(0.0046)	(0.0021)	(0.0028)	(0.0022)	(0.0029)	(0.0010)	(0.0010)	(0.0027)	(0.0034)
Second lag	0.1504***	0.0519***	0.0513***	0.0455***	0.0463***	0.0585***	0.0611***	-0.0439***	-0.0508***
	(0.0022)	(0.0021)	(0.0027)	(0.0023)	(0.0029)	(0.0010)	(0.0009)	(0.0027)	(0.0034)
DuringSpoof	0.0001	0.2266***	-0.0435**	-0.0456**	0.0073	0.0086***	-0.0373***	-0.0057	0.0575***
	(0.0002)	(0.0221)	(0.0191)	(0.0181)	(0.0219)	(0.0033)	(0.0028)	(0.0062)	(0.0097)
AfterSpoof	-4.005e-05	0.0824***	-0.0147	0.0918***	0.0020	0.0178***	0.0157***	-0.0062***	0.0120***
	(7.153e-05)	(0.0114)	(0.0102)	(0.0118)	(0.0123)	(0.0018)	(0.0015)	(0.0023)	(0.0031)
dTimestamp	-0.0030***	0.1476***	0.0191*	0.0756***	0.0196***	0.0560***	0.0477***	-0.1538***	-0.1539***
	(5.097e-05)	(0.0105)	(0.0099)	(0.0099)	(0.0072)	(0.0015)	(0.0015)	(0.0050)	(0.0050)
F-test	1.092e+07***	1.546+e08***	8.944e+07***	1.462e+08***	1.101e+08***	1.863e+08***	1.818e+08***	1.005e+09***	1.334e+09***
R-squared	0.7861	0.9811	0.9678	0.9801	0.9737	0.9843	0.9839	0.9971	0.9978
N	14,874,143	14,874,143	14,874,143	14,874,143	14,874,143	14,874,143	14,874,143	14,874,143	14,874,143
Ask spoof orders									
First lag	0.7500***	0.9314***	0.9225***	0.9461***	0.9444***	0.9381***	0.9311***	1.0116***	1.0556***
	(0.0047)	(0.0023)	(0.0025)	(0.0025)	(0.0051)	(0.0010)	(0.0010)	(0.0039)	(0.0043)
Second lag	0.1506***	0.0567***	0.0622***	0.0457***	0.0435***	0.0547***	0.0616***	-0.0141***	-0.0569***
	(0.0023)	(0.0022)	(0.0024)	(0.0025)	(0.0051)	(0.0009)	(0.0010)	(0.0037)	(0.0043)
DuringSpoof	0.0012***	0.0315	-0.0539**	0.2854***	0.0078	-0.0343***	0.0421***	0.0538***	0.0134
	(0.0005)	(0.0203)	(0.0220)	(0.0194)	(0.0197)	(0.0029)	(0.0034)	(0.0108)	(0.0099)
AfterSpoof	8.158e-05	0.1237***	-0.0273*	0.1036***	0.0513***	0.0129***	0.0238***	-0.0028	0.0031
	(5.1e-05)	(0.0128)	(0.0141)	(0.0108)	(0.0098)	(0.0016)	(0.0017)	(0.0045)	(0.0040)
dTimestamp	-0.0028***	0.0908***	0.0254**	0.1724***	0.0291***	0.0507***	0.0477***	-0.1289***	-0.1912***
	(5.269e-05)	(0.0083)	(0.0104)	(0.0112)	(0.0063)	(0.0015)	(0.0017)	(0.0057)	(0.0055)
F-test	1e+07***	1.092e+08***	8.349e+07***	1.605e+08***	1.082e+08***	1.817e+08***	1.766e+08***	6.015e+08***	1.182e+09***
R-squared	0.7847	0.9755	0.9682	0.9832	0.9753	0.9851	0.9847	0.9955	0.9977
N	13,733,545	13,733,545	13,733,545	13,733,545	13,733,545	13,733,545	13,733,545	13,733,545	13,733,545

Note: standard errors are reported in parentheses. *, **, and *** indicates statistical significance at 10%, 5% and 1% level, respectively. *First lag* and *Second lag* are the first and second lag of the respective liquidity measure. *DuringSpoof* and *AfterSpoof* are dummy variables that categorize if an observation occurred during spoofing (*DuringSpoof* = 1) or after the spoof order was removed from the limit order book (*AfterSpoof* = 1). This sets the period before the spoof order was added as the reference category.

Table 5.F.3 | Panel data regressions of the impact of spoofing on liquidity in the soybean futures market from CME Group between July 2019 to June 2020.

Variables	Bid-ask spread	Bid depth 1–5	Bid depth6–10	Ask depth 1–5	Ask depth 6–10	Bid orders	Ask orders	APM bid	APM ask
Bid spoof orders									
First lag	0.6456***	0.9148***	0.9031***	0.9083***	0.9047***	0.9026***	0.8980***	1.0120***	0.9845***
	(0.0016)	(0.0010)	(0.0014)	(0.0013)	(0.0017)	(0.0006)	(0.0020)	(0.0191)	(0.0023)
Second lag	0.1391***	0.0623***	0.0774***	0.0684***	0.0704***	0.0736***	0.0743***	-0.0228	0.0105***
	(0.0013)	(0.0009)	(0.0015)	(0.0013)	(0.0018)	(0.0005)	(0.0022)	(0.0182)	(0.0020)
DuringSpoof	0.0007***	0.6763***	-9.04e-05	0.5141***	0.0869***	-0.1005***	0.0456***	-0.1457***	-0.0410***
	(7.553e-05)	(0.0234)	(0.0158)	(0.0176)	(0.0194)	(0.0028)	(0.0045)	(0.0382)	(0.0013)
AfterSpoof	2.095e-05	-0.0108	-0.0065	0.0131	-0.0261**	-0.0184***	-0.0224***	-0.0001	0.0035***
	(2.262e-05)	(0.0105)	(0.0076)	(0.0096)	(0.0102)	(0.0017)	(0.0025)	(0.0057)	(0.0004)
dTimestamp	-0.0027***	0.3943***	-0.0019	0.0305***	-0.0214***	0.0476***	0.0343***	-0.1696***	-0.0529***
	(2.317e-05)	(0.0069)	(0.0042)	(0.0055)	(0.0064)	(0.0008)	(0.0012)	(0.0053)	(0.0011)
F-test	1.855e+07***	2.78e+08***	3.232e+08***	2.706e+08***	2.524e+08***	2.636e+08***	2.247e+08***	6.536e+08***	1.278e+09***
R-squared	0.5709	0.9522	0.9586	0.9510	0.9476	0.9498	0.9416	0.9791	0.9892
N	69,775,976	69,775,976	69,775,976	69,775,976	69,775,976	69,775,976	69,775,976	69,775,976	69,775,976
Ask spoof orders									
First lag	0.6447***	0.9149***	0.9075***	0.9125***	0.9073***	0.9011***	0.8912***	1.0523***	0.9381***
	(0.0014)	(0.0011)	(0.0015)	(0.0011)	(0.0021)	(0.0007)	(0.0021)	(0.0172)	(0.0030)
Second lag	0.1359***	0.0633***	0.0730***	0.0619***	0.0692***	0.0746***	0.0822***	-0.0588***	0.0536***
	(0.0012)	(0.0010)	(0.0015)	(0.0010)	(0.0021)	(0.0006)	(0.0020)	(0.0172)	(0.0029)
DuringSpoof	0.0004***	0.2889***	0.0314*	0.7821***	-0.0644***	0.0233***	-0.1288***	0.0011	-0.0590***
	(9.347e-05)	(0.0201)	(0.0176)	(0.0237)	(0.0182)	(0.0030)	(0.0046)	(0.0121)	(0.0041)
AfterSpoof	0.0001***	0.0187**	0.0052	0.0219**	-0.0421***	-0.0035**	-0.0276***	0.0085*	0.0053***
	(2.332e-05)	(0.0095)	(0.0077)	(0.0105)	(0.0084)	(0.0017)	(0.0022)	(0.0051)	(0.0007)
dTimestamp	-0.0027***	0.1145***	-0.0109***	0.3334***	0.0002	0.0450***	0.0373***	-0.0932***	-0.1149***
	(2.416e-05)	(0.0055)	(0.0041)	(0.0076)	(0.0046)	(0.0008)	(0.0012)	(0.0043)	(0.0022)
F-test	1.738+e07***	2.801e+08***	3.114e+08***	2.374e+08***	2.569e+08***	2.479e+08***	2.237e+08***	1.077e+09***	7.273e+08***
R-squared	0.5649	0.9544	0.9588	0.9466	0.9505	0.9488	0.9435	0.9877	0.9819
N	67,000,137	67,000,137	67,000,137	67,000,137	67,000,137	67,000,137	67,000,137	67,000,137	67,000,137

Note: standard errors are reported in parentheses. *, **, and *** indicates statistical significance at 10%, 5% and 1% level, respectively. *First lag* and *Second lag* are the first and second lag of the respective liquidity measure. *DuringSpoof* and *AfterSpoof* are dummy variables that categorize if an observation occurred during spoofing (*DuringSpoof* = 1) or after the spoof order was removed from the limit order book (*AfterSpoof* = 1). This sets the period before the spoof order was added as the reference category.

Table 5.F.4 | Panel data regressions of the impact of spoofing on liquidity in the soybean meal futures market from CME Group between July 2019 to June 2020.

Variables	Bid-ask spread	Bid depth 1–5	Bid depth6–10	Ask depth 1–5	Ask depth 6–10	Bid orders	Ask orders	APM bid	APM ask
Bid spoof orders									
First lag	0.7088***	0.9070***	0.9088***	0.9069***	0.9124***	0.9105***	0.9099***	0.9550***	1.0267***
	(0.0012)	(0.0030)	(0.0030)	(0.0021)	(0.0034)	(0.0013)	(0.0010)	(0.0022)	(0.0020)
Second lag	0.1305***	0.0704***	0.0698***	0.0772***	0.0759***	0.0766***	0.0802***	0.0391***	-0.0280***
	(0.0007)	(0.0028)	(0.0030)	(0.0019)	(0.0033)	(0.0012)	(0.0009)	(0.0022)	(0.0020)
DuringSpoof	0.0003***	0.3035***	-0.1078***	0.1060***	1.266e-05	-0.0533***	-0.0031	-0.1179***	0.0362***
	(3.98e-05)	(0.0327)	(0.0273)	(0.0146)	(0.0119)	(0.0030)	(0.0024)	(0.0065)	(0.0080)
AfterSpoof	0.0002***	-0.0070	-0.0730***	0.0415***	-0.0079	-0.0308***	-0.0137***	0.0111***	0.0055**
	(1.282e-05)	(0.0151)	(0.0117)	(0.0081)	(0.0054)	(0.0015)	(0.0011)	(0.0030)	(0.0026)
dTimestamp	-0.0012***	-0.0090	0.0829***	-0.0107*	0.0538***	0.0067***	0.0140***	-0.1878***	-0.1786***
	(1.547e-05)	(0.0100)	(0.0081)	(0.0058)	(0.0044)	(0.0012)	(0.0011)	(0.0057)	(0.0054)
F-test	1.111e+07***	1.093e+08***	1.157e+08***	1.569e+08***	2.142e+08***	1.964e+08***	2.514e+08***	4.542e+08***	2.165e+09
R-squared	0.6704	0.9524	0.9549	0.9664	0.9751	0.9729	0.9787	0.9881	0.9975
N	27,336,963	27,336,963	27,336,963	27,336,963	27,336,963	27,336,963	27,336,963	27,336,963	27,336,963
Ask spoof orders									
First lag	0.7019***	0.9106***	0.9204***	0.9109***	0.9031***	0.9047***	0.9091***	0.9742***	0.9736***
	(0.0011)	(0.0040)	(0.0031)	(0.0015)	(0.0024)	(0.0013)	(0.0009)	(0.0054)	(0.0025)
Second lag	0.1334***	0.0676***	0.0575***	0.0699***	0.0832***	0.0814***	0.0808***	0.0231***	0.0238***
	(0.0006)	(0.0041)	(0.0033)	(0.0013)	(0.0023)	(0.0013)	(0.0008)	(0.0054)	(0.0024)
DuringSpoof	0.0004***	0.3203***	-0.0550**	0.2973***	-0.0098	0.0178***	-0.0354***	0.0373***	-0.1136***
	(3.861e-05)	(0.0275)	(0.0266)	(0.0153)	(0.0103)	(0.0030)	(0.0022)	(0.0033)	(0.0076)
AfterSpoof	0.0002***	-0.0403***	-0.0602***	0.0266***	-0.0455***	-0.0266***	-0.0196***	0.0042***	0.0087**
	(1.238e-05)	(0.0156)	(0.0133)	(0.0083)	(0.0046)	(0.0015)	(0.0010)	(0.0011)	(0.0035)
dTimestamp	-0.0011***	-0.1241***	0.0683***	0.0702***	0.0367***	0.0020*	0.0138***	-0.0742***	-0.2526***
	(1.369e-05)	(0.0087)	(0.0078)	(0.0055)	(0.0044)	(0.0011)	(0.0010)	(0.0021)	(0.0089)
F-test	1.114e+07***	1.186e+08***	1.184e+08***	1.345e+08***	1.893e+08***	1.854e+08***	2.601e+08***	1.059e+09***	1.059e+09***
R-squared	0.6622	0.9543	0.9542	0.9594	0.9708	0.9702	0.9786	0.9947	0.9947
N	28,446,147	28,446,147	28,446,147	28,446,147	28,446,147	28,446,147	28,446,147	28,446,147	28,446,147

Note: standard errors are reported in parentheses. *, **, and *** indicates statistical significance at 10%, 5% and 1% level, respectively. *First lag* and *Second lag* are the first and second lag of the respective liquidity measure. *DuringSpoof* and *AfterSpoof* are dummy variables that categorize if an observation occurred during spoofing (*DuringSpoof* = 1) or after the spoof order was removed from the limit order book (*AfterSpoof* = 1). This sets the period before the spoof order was added as the reference category.

Table 5.F.5 | Panel data regressions of the impact of spoofing on liquidity in the soybean oil futures market from CME Group between July 2019 to June 2020.

Variables	Bid-ask spread	Bid depth 1–5	Bid depth6–10	Ask depth 1–5	Ask depth 6–10	Bid orders	Ask orders	APM bid	APM ask
Bid spoof orders									
First lag	0.7482***	0.9068***	0.9132***	0.9210***	0.9226***	0.9122***	0.9138***	0.9881***	1.0218***
	(0.0022)	(0.0027)	(0.0038)	(0.0016)	(0.0033)	(0.0008)	(0.0011)	(0.0029)	(0.0020)
Second lag	0.1169***	0.0576***	0.0533***	0.0636***	0.0667***	0.0700***	0.0708***	0.0072**	-0.0234***
	(0.0018)	(0.0026)	(0.0035)	(0.0015)	(0.0032)	(0.0007)	(0.0010)	(0.0029)	(0.0019)
DuringSpoof	-4.806e-05***	0.3320***	0.0315	-0.0463***	-0.0377***	-0.0778***	-0.0613***	-0.0308***	0.0546***
	(5.354e-06)	(0.0346)	(0.0308)	(0.0100)	(0.0067)	(0.0031)	(0.0024)	(0.0036)	(0.0044)
AfterSpoof	9.082e-06***	-0.0550***	-0.1676***	-0.0563***	-0.0564***	-0.0529***	-0.0476***	0.0045**	0.0071***
	(2.197e-06)	(0.0179)	(0.0176)	(0.0068)	(0.0043)	(0.0015)	(0.0014)	(0.0022)	(0.0018)
dTimestamp	-0.0001***	0.0273***	0.0515***	0.0165***	0.0064*	0.0111***	0.0185***	-0.0787***	-0.1228***
	(1.585e-06)	(0.0091)	(0.0076)	(0.0041)	(0.0034)	(0.0012)	(0.0011)	(0.0020)	(0.0024)
F-test	9.533e+06***	4.616e+07***	4.962e+07***	1.097e+08***	1.585e+08***	9.606e+07***	1.124e+08***	3.96e+08***	1.164e+09***
R-squared	0.7218	0.9263	0.9311	0.9676	0.9773	0.9632	0.9683	0.9908	0.9969
N	18,388,956	18,388,956	18,388,956	18,388,956	18,388,956	18,388,956	18,388,956	18,388,956	18,388,956
Ask spoof orders									
First lag	0.7477***	0.8796***	0.8948***	0.9225***	0.9365***	0.9144***	0.9156***	1.0079***	0.9990***
	(0.0017)	(0.0088)	(0.0066)	(0.0013)	(0.0020)	(0.0008)	(0.0009)	(0.0084)	(0.0015)
Second lag	0.1139***	0.0939***	0.0789***	0.0611***	0.0535***	0.0697***	0.0698***	-0.0123*	-0.0013
	(0.0013)	(0.0082)	(0.0063)	(0.0010)	(00019)	(0.0008)	(0.0008)	(0.0069)	(0.0015)
DuringSpoof	-2.59e-05***	0.0108	-0.0642***	0.1687***	-0.0271***	-0.0597***	-0.0588***	0.0363***	-0.0424***
	(5.664e-06)	(0.0205)	(0.0208)	(0.0144)	(0.0075)	(0.0028)	(0.0029)	(0.0059)	(0.0046)
AfterSpoof	2.666e-06	-0.0054	-0.1111***	-0.0057	-0.0737***	-0.0495***	-0.0536***	0.0123**	0.0175***
	(1.78e-06)	(0.0136)	(0.0121)	(0.0064)	(0.0039)	(0.0014)	(0.0014)	(0.0049)	(0.0019)
dTimestamp	-0.0001***	-0.0495***	0.0022	0.0421***	0.0138***	0.0110***	0.0110***	-0.0503***	-0.1459***
	(1.685e-06)	(0.0075)	(0.0085)	(0.0044)	(0.0029)	(0.0012)	(0.0012)	(0.0113)	(0.0030)
F-test	9.876e+06***	6.492e+07***	6.671e+07***	1.095e+08***	1.831e+08***	1.157e+08***	1.263e+08***	5.856e+08***	7.091e+08***
R-squared	0.7156	0.9430	0.9444	0.9654	0.9790	0.9672	0.9699	0.9933	0.9945
N	19,640,536	19,640,536	19,640,536	19,640,536	19,640,536	19,640,536	19,640,536	19,640,536	19,640,536

Note: standard errors are reported in parentheses. *, **, and *** indicates statistical significance at 10%, 5% and 1% level, respectively. *First lag* and *Second lag* are the first and second lag of the respective liquidity measure. *DuringSpoof* and *AfterSpoof* are dummy variables that categorize if an observation occurred during spoofing (*DuringSpoof* = 1) or after the spoof order was removed from the limit order book (*AfterSpoof* = 1). This sets the period before the spoof order was added as the reference category.

Table 5.F.6 | Panel data regressions of the impact of spoofing on liquidity in the live cattle futures market from CME Group between July 2019 to June 2020.

Variables	Bid-ask spread	Bid depth 1–5	Bid depth6–10	Ask depth 1–5	Ask depth 6–10	Bid orders	Ask orders	APM bid	APM ask
Bid spoof orders									
First lag	0.8303***	0.9724***	0.9474***	0.9605***	0.9633***	0.9458***	0.9278***	1.0979***	1.0089***
	(0.0131)	(0.0080)	(0.0043)	(0.0059)	(0.0048)	(0.0018)	(0.0019)	(0.0184)	(0.0071)
Second lag	0.0692***	0.0226***	0.0404***	0.0287***	0.0165***	0.0505***	0.0682***	-0.1000***	-0.0107
	(0.0072)	(0.0079)	(0.0042)	(0.0063)	(0.0055)	(0.0017)	(0.0019)	(0.0184)	(0.0072)
DuringSpoof	-0.0003***	0.1004***	0.0891***	0.2200***	0.1252***	0.0118**	0.0107*	-0.0507	0.0439
	(8.453e-05)	(0.0273)	(0.0143)	(0.0469)	(0.0425)	(0.0055)	(0.0057)	(0.0429)	(0.0387)
AfterSpoof	-0.0001***	0.0642***	0.0758***	0.1333***	0.1837***	0.0253***	0.0382***	-0.0576***	-0.0272**
	(4.216e-05)	(0.0083)	(0.0085)	(0.0198)	(0.0232)	(0.0026)	(0.0027)	(0.0130)	(0.0115)
dTimestamp	-0.0008***	0.0114	0.0302***	0.0282**	-0.0170	0.0158***	0.0302***	-0.1540***	-0.2086***
	(2.948e-05)	(0.0132)	(0.0059)	(0.0132)	(0.0119)	(0.0035)	(0.0031)	(0.0241)	(0.0115)
F-test	1.961e+06***	5.412e+07***	2.01e+07***	2.393e+07***	1.192e+07***	7.28e+07***	6.98e+07***	1.67e+08***	2.039e+08***
R-squared	0.7970	0.9909	0.9758	0.9796	0.9598	0.9932	0.9929	0.9970	0.9976
N	2,499,065	2,499,065	2,499,065	2,499,065	2,499,065	2,499,065	2,499,065	2,499,065	2,499,065
Ask spoof orders									
First lag	0.8316***	0.9529***	0.9281***	0.9508***	0.9584***	0.9405***	0.9223***	1.1170***	0.9670***
	(0.0087)	(0.0055)	(0.0045)	(0.0064)	(0.0056)	(0.0012)	(0.0015)	(0.0123)	(0.0034)
Second lag	0.0700***	0.0424***	0.0591***	0.0385***	0.0231***	0.0559***	0.0736***	-0.1185***	0.0301***
	(0.0048)	(0.0054)	(0.0044)	(0.0068)	(0.0064)	(0.0012)	(0.0015)	(0.0124)	(0.0034)
DuringSpoof	-0.0002***	-0.0308	0.0994***	0.2243***	0.1792***	0.0037	0.0134***	0.2026***	-0.0900***
	(4.731e-05)	(0.0261)	(0.0158)	(0.0284)	(0.0212)	(0.0045)	(0.0040)	(0.0472)	(0.0324)
AfterSpoof	-0.0002***	0.0351***	0.0997***	0.1428***	0.1464***	0.0244***	0.0268***	-0.0372***	-0.0195*
	(4.056e-05)	(0.0069)	(0.0076)	(0.0145)	(0.0145)	(0.0017)	(0.0016)	(0.0094)	(0.0118)
dTimestamp	-0.0010***	-0.0156*	0.0480***	0.1219***	0.0243***	0.0300***	0.0607***	-0.2053***	-0.3248***
	(2.404e-05)	(0.0092)	(0.0080)	(0.0093)	(0.0050)	(0.0028)	(0.0027)	(0.0162)	(0.0130)
F-test	3.497e+06***	1.067e+08***	3.319e+07***	4.259e+07***	2.261e+07***	1.351e+08***	1.132e+08***	5.025e+08***	2.05e+08***
R-squared	0.8010	0.9919	0.9745	0.9800	0.9630	0.9936	0.9924	0.9983	0.9958
N	4,344,793	4,344,793	4,344,793	4,344,793	4,344,793	4,344,793	4,344,793	4,344,793	4,344,793

Note: standard errors are reported in parentheses. *, **, and *** indicates statistical significance at 10%, 5% and 1% level, respectively. *First lag* and *Second lag* are the first and second lag of the respective liquidity measure. *DuringSpoof* and *AfterSpoof* are dummy variables that categorize if an observation occurred during spoofing (*DuringSpoof* = 1) or after the spoof order was removed from the limit order book (*AfterSpoof* = 1). This sets the period before the spoof order was added as the reference category.

Appendix 5.G

Table 5.G.1 | Coefficients for *dTimestamp* variable for each panel data regression per agricultural futures market of CME Group between July 2019 and June 2020.

	Corn		Wheat		Soybean		Soybean meal		Soybean oil		Live cattle	
Spoof orders	**Bid**	**Ask**	**Bid**	**Ask**	**Bid**	**Ask**	**Bid**	**Ask**	**Bid**	**Ask**	**Bid**	**Ask**
Bid-ask spread	−0.002***	−0.002***	−0.003***	−0.003***	−0.003***	−0.003***	−0.001***	−0.001***	−0.0001***	−0.0001***	−0.001***	−0.001***
Bid depth $_{1\text{-}5}$	0.253***	−0.166***	0.148***	0.091***	0.394***	0.115***	−0.009	−0.124***	0.027***	−0.050***	0.011	−0.016*
Bid depth $_{6\text{-}10}$	0.087***	0.157***	0.019*	0.025**	−0.002	−0.011***	0.083***	0.068***	0.052***	0.002	0.030***	0.048***
Ask depth $_{1\text{-}5}$	−0.117***	0.317***	0.076***	0.172***	0.031***	0.333***	−0.011*	0.070***	0.017***	0.042***	0.028**	0.122***
Ask depth $_{6\text{-}10}$	0.109***	0.200***	0.020***	0.029***	−0.021***	0.0002	0.054***	0.037***	0.006*	0.014***	−0.017	0.024***
Bid orders	0.029***	0.035***	0.056***	0.051***	0.048***	0.045***	0.007***	0.002*	0.011***	0.011***	0.016***	0.030***
Ask orders	0.028***	0.018***	0.048***	0.048***	0.034***	0.037***	0.014***	0.014***	0.019***	0.011***	0.030***	0.061***
APM bid	−0.215***	−0.132***	−0.154***	−0.129***	−0.170***	−0.093***	−0.188***	−0.074***	−0.079***	−0.050***	−0.154***	−0.205***
APM ask	−0.080***	−0.244***	−0.154***	−0.191***	−0.053***	−0.115***	−0.179***	−0.253***	−0.123***	−0.146***	−0.209***	−0.325***

Note: This table shows the coefficients for the *dTimestamp* variable for each panel data regression per futures market from CME Group between July 2019 and June 2020, split by bid and ask spoof orders. *dTimestamp* is the time difference (in seconds) between two consecutive messages. Numbers are rounded to the third decimal, unless the fourth decimal is necessary for interpretation purposes.

Summary

The digitalization of financial markets has transformed financial trading, as it transitioned to electronic trading platforms such as the electronic limit order book (LOB). The LOB is a central marketplace that records all prices and quantities traders are willing to buy and sell for. The shift resulted, among others, in the possibility to automate trading actions, and it introduced new market participants such as algorithmic and high-frequency traders, for whom speed is key. This also means that new types of market manipulation are now possible, occurring faster than visible to the naked eye.

These high-speed markets generate large amounts of data ('big data'), making the identification and study of specific events challenging, particularly for regulators aiming to detect market manipulation. Market manipulation can severely harm the functioning of financial markets – for example, the price discovery and liquidity – an effect that can trickle down into society, affecting agents such as farmers, manufacturers, processors and eventually consumers. Moreover, identifying market manipulation is challenging due to a lack of guidance in the legal framework on differentiating between legitimate and illegitimate behavior. Together, these challenges of analyzing high-frequency data and detecting market manipulation present a distinct issue: the need to effectively identify and analyze market manipulation in a high-frequency context. The aim of this dissertation is to address this issue by answering the research question '*How can market manipulation in a high-frequency context be identified and analyzed?*' To answer this question, we reap the benefits from the decades of big data experience in particle physics, by applying particle physics methodologies and tools to financial market big data. The focus is on the market manipulation practice of 'spoofing' in U.S. futures markets, but methodologies in this dissertation are applicable to other market manipulation types and any market that uses a LOB – for example, stock, crypto, options and spot markets.

To improve the understanding of high-frequency markets, Chapter 2 uses the particle physics data-analysis framework ROOT to introduce a novel visualization methodology capable of visualizing high-frequency LOB data. It uses all information embedded in the irregularly spaced message data and is not dependent on fixed time-intervals. This makes high-frequency trading and information-dense events more apparent and puts individual trading actions into the perspective of the full market. Multiple datasets are linked, and it offers a high degree of customizability with variables and the ability to visualize multiple markets simultaneously. Its wide applicability and scope beyond traditional methods make this methodology useful for many financial-market stakeholders with an interested in understanding high-frequency market data and behavior.

Chapter 3 applies the new visualization methodology from Chapter 2 to a well-known public spoofing case, demonstrating the behavior of high-frequency markets around spoofing.

Spoofing appears as a visible ‘hotspot’ in the visualizations, and the results 1) demonstrate how various markets respond to spoofing; 2) offer possible characteristics to identify spoofing; 3) show how well-hidden spoofing can be; 4) provide insights into the complexities of the techniques required to recognize spoofing; 5) put a value on the miniscule price changes that makes spoofing economically viable; and 6) offer an alternative motivation for spoofing other than moving the price: attracting liquidity.

Chapter 4 delineates spoofing from both an economic and legal perspective by providing a comprehensive overview of spoofing types, legislation, literature and rulings. These aspects are then combined with expert knowledge from the International Expert Group on Market Surveillance to develop a conceptual framework with spoofing dimensions and attributes, which can be used to characterize spoofing or delineate existing documents related to spoofing. The conceptual framework is subsequently used to analyze 204 U.S. spoofing cases in futures markets. The results highlight various characteristics of spoofing – such as the cancellation time, placement in the LOB and hit rate – and its impact.

Using a selection of spoofing characteristics from Chapter 4, Chapter 5 studies the frequency of spoofing in agricultural futures markets and its impact on liquidity. Focusing on the single-spoofing strategy in CME Group data from July 2019 to June 2020, spoof orders comprised approximately between 0.02% and 0.16% of all added orders in agricultural futures markets. These orders differ in impact on liquidity: liquidity generally improves after spoofing in the wheat and live cattle markets and worsens in the corn, soybean, soybean oil and soybean meal markets. Moreover, results suggest an inverse relationship between spoofing frequency and impact on liquidity costs.

Together, the chapters integrate the areas of managing financial big data and studying market manipulation, demonstrating that market manipulation in a high-frequency context can be better identified and analyzed using a holistic approach. Specifically, the chapters together cover 1) effective methods for storing, processing, visualizing and analyzing high-frequency data using the ROOT data-analysis framework from particle physics; 2) strategies to delineate, characterize, identify, visualize and analyze market manipulation; and 3) the synthesis of these approaches. In doing so, this dissertation provides a comprehensive framework for identifying and analyzing market manipulation in a high-frequency context.

Samenvatting

De digitalisering van financiële markten heeft de wijze van handelen getransformeerd, waarbij ze zijn overgestapt naar elektronische handelsplatformen zoals het elektronische orderboek (LOB). Het LOB is een centrale marktplaats waar alle prijzen en volumes die handelaren willen (ver)kopen worden vastgelegd. Deze overgang heeft onder andere geleid tot de mogelijkheid om handelsacties te automatiseren en het heeft nieuw deelnemers aangetrokken, zoals algoritmische en high-frequency handelaren, voor wie snelheid cruciaal is. Ook zijn nieuwe vormen van marktmanipulatie mogelijk geworden, die sneller plaatsvinden dan het menselijk oog kan waarnemen.

Deze "snelle" markten genereren enorme hoeveelheden data ('big data'), wat het identificeren en bestuderen van specifieke gebeurtenissen bemoeilijkt, vooral voor toezichthouders die marktmanipulatie proberen op te sporen. Marktmanipulatie kan de werking van financiële markten ernstig schaden – zoals de prijsvorming en liquiditeit – en dit effect kan doorsijpelen naar de samenleving met gevolgen voor bijvoorbeeld landbouwers, fabrikanten, verwerkers en uiteindelijk consumenten. Daarnaast is het lastig om marktmanipulatie te identificeren door een gebrek aan richtlijnen in het juridische kader die helpen om legitiem van onwettig gedrag te onderscheiden. Samen vormen de uitdagingen van het analyseren van high-frequency data en het opsporen van marktmanipulatie een specifiek probleem: de noodzaak om marktmanipulatie effectief te identificeren en te analyseren in een high-frequency context. Het doel van dit proefschrift is om deze kwestie aan te pakken door de volgende onderzoeksvraag te beantwoorden: *'Hoe kan marktmanipulatie in een high-frequency context worden geïdentificeerd en geanalyseerd?'* Om deze vraag te beantwoorden plukken we de vruchten van de decennia aan big data-ervaring uit de deeltjesfysica, door methodologieën en tools uit deze discipline toe te passen op big data van financiële markten. De focus ligt op de marktmanipulatie genaamd 'spoofing' in Amerikaanse termijnmarkten, maar de methodologieën in dit proefschrift zijn toepasbaar op andere vormen van marktmanipulatie en elke markt die een LOB gebruikt – bijvoorbeeld aandelen-, crypto-, opties- en spotmarkten.

Om het begrip van high-frequency markten te verbeteren, introduceert Hoofdstuk 2 een nieuwe visualisatiemethodologie die in staat is om high-frequency LOB-data te visualiseren, met behulp van het data-analyse framework ROOT uit de deeltjesfysica. Het maakt gebruik van alle informatie die is ingebed in de onregelmatig verdeelde message data en is niet afhankelijk van vaste tijdsintervallen. Hierdoor worden high-frequency handel en informatierijke gebeurtenissen zichtbaarder, en individuele acties van handelaren worden in de context van de volledige markt geplaatst. De methodologie koppelt meerdere datasets, biedt een hoge mate van aanpasbaarheid in het toevoegen van variabelen en de mogelijkheid om meerdere markten tegelijkertijd te visualiseren. De brede toepasbaarheid en mogelijkheden die verder reiken dan traditionele methoden, maken het nuttig voor belanghebbenden die geïnteresseerd zijn in het begrijpen van high-frequency marktdata en -gedrag.

Hoofdstuk 3 past de nieuwe visualisatiemethodologie uit Hoofdstuk 2 toe op een bekende publieke spoofing zaak, waarbij het gedrag van high-frequency markten rond spoofing wordt uitgelicht. Spoofing verschijnt als een zichtbare hotspot in de visualisaties, en de resultaten 1) laten zien hoe verschillende markten reageren op spoofing; 2) bieden mogelijke kenmerken om spoofing te identificeren; 3) tonen hoe goed verborgen spoofing kan zijn; 4) geven inzicht in de complexiteit van de technieken die nodig zijn om spoofing te herkennen; 5) geven een waarde aan de minuscule prijsveranderingen die spoofing economisch rendabel maken; en 6) bieden een alternatieve motivatie voor spoofing naast het beïnvloeden van de prijs: het aantrekken van liquiditeit.

Hoofdstuk 4 belicht spoofing vanuit zowel een economisch als juridisch perspectief door een uitgebreid overzicht te geven van vormen van spoofing, wetgeving, literatuur en uitspraken. Deze aspecten worden vervolgens gecombineerd met expert kennis van de International Expert Group on Market Surveillance, om een conceptueel raamwerk te ontwikkelen met spoofing-dimensies en kenmerken. Dit raamwerk kan gebruikt worden om spoofing te karakteriseren of bestaande spoofing documenten te analyseren, en wordt in het hoofdstuk toegepast op 204 Amerikaanse spoofing zaken in termijnmarkten. De resultaten benadrukken verschillende kenmerken van spoofing – zoals hoe snel een spoof order wordt verwijderd, waar deze in het LOB geplaatst wordt en hoe vaak een spoof order tot een transactie leidt – en de impact ervan.

Een selectie van spoofing-kenmerken uit Hoofdstuk 4 wordt gebruikt om in Hoofdstuk 5 te onderzoeken hoe vaak spoofing in agrarische termijnmarkten voorkomt en wat de impact ervan is op de liquiditeit. Er wordt gericht op de single-spoofing strategie in de data van CME Group van juli 2019 tot juni 2020. De resultaten laten zien dat ongeveer tussen 0,02% en 0,16% van alle toegevoegde orders in agrarische termijnmarkten bestonden uit spoof orders. Deze orders verschillen in impact op liquiditeit: over het algemeen verbetert liquiditeit in de tarwe- en levend rundvee markten en verslechtert in de maïs-, soja-, sojaolie- en sojameelmarkten. Bovendien suggereren de resultaten een omgekeerde relatie tussen de frequentie van spoofing en de impact op liquiditeitskosten.

Samen integreren de hoofdstukken de gebieden van het managen van financiële big data en het bestuderen van marktmanipulatie, waarbij wordt aangetoond dat marktmanipulatie in een high-frequency context beter kan worden geïdentificeerd en geanalyseerd met een holistische aanpak. De hoofdstukken behandelen samen 1) effectieve methoden voor het opslaan, verwerken, visualiseren en analyseren van high-frequency data met behulp van het ROOT data-analyse framework uit de deeltjesfysica; 2) strategieën om marktmanipulatie te omschrijven, karakteriseren, identificeren, visualiseren en analyseren; en 3) de synthese van deze benaderingen. Hiermee biedt dit proefschrift een uitgebreid kader voor het identificeren en analyseren van marktmanipulatie in een high-frequency context.

Acknowledgements

At a job fair in 2016, a news reporter asked me on camera – with slight disbelief – why I was seeking a job relating to derivative markets. I replied with a big smile “because it makes me happy”. Now eight years later, I feel so lucky for the opportunity to have worked on something that brought, and still brings, me so much joy.

My PhD journey started with an enthusiastic phone call from Joost in which he mentioned something about “CERN”, “similar to financial markets” and “market manipulation”. After five years I can finally come clean and tell you, Joost, that I had no idea what you were talking about and I did not even know what this “CERN” was. When I later told my parents about this confusing – but enthusiastic – conversation, I could feel their slight disappointment through the phone. How could I not know CERN? Have I never read “Het Bernini Mysterie” by Dan Brown?! (English title: “Angels & Demons”) If you happen to be a CERN-employee reading this, please consider this as my formal apology. That phone call resulted in my first trip to CERN and, six months later, the start of my PhD in 2019. As with any PhD, mine has also come with ups and downs, laughter and tears, breakthroughs and setbacks. Let me give you a low and a high for the personal touch. One of my lows was when life, or the world, became too overwhelming and I had to recharge my battery. Not fun, but important lessons were learned. One of my highs was the kick-off event of the International Expert Group on Market Surveillance (IMS Group). It was amazing to experience the enthusiasm and openness of people wanting to achieve the same goal, and the friendly and warm atmosphere they created.

Overall, if you have asked me in the past years how my PhD-experience was so far, nine out of ten times you would have heard the answer “with these people around me, I would do another PhD in a heartbeat”. I have been so fortunate with the people I am working and collaborating with, the people who are guiding and supporting me, both at work and in my personal life. Enthusiastic people who encouraged me and pushed me to bring out my best. You all made me into the person I am today, and I am proud of that! I am extremely grateful for all of you and want to thank you this way.

I will start with **Joost**, who kickstarted my interest in futures markets ten years ago, bringing me to where I am today. I am so thankful for all the opportunities you have given me and for everything you did throughout my whole PhD journey. You are not only a mentor to me, but also a role model and friend. I have so much respect for how you balance your busy professional responsibilities, but never losing sight of what truly matters in life: family, friends, the people you love. You made my PhD enjoyable, provided an environment in which I could flourish and pushed me to challenge myself and continue to grow. Thank you for your contagious enthusiasm in your profession, and encouragements in thinking outside of the box. For the faith and trust you put in me, which in turn made me have more faith and trust in my own abilities. You pushed me out of my comfort zone in a gentle manner, allowing me to expand on my strengths. Thank you for all the fun, productive, creative and inspiring trips to CERN. I especially enjoyed the "what to do next" brainstorm sessions with you and Philippe, where we could come up with any idea, no matter how wild or out-of-the box, as nothing was off-limits. And then after a long day, the casual and relaxed talks with both you, discussing the world, personal matters, your exciting adventures, Philippe's food reviews and my cat café ambitions. No matter the time of the day, I could always call you regardless if it was work related or something personal. You always had my best interest in mind and prioritized my well-being in my PhD. I am forever grateful that you introduced me to these fascinating but complex markets, and for your dedication to my PhD, me as a person and to your profession.

Koos, I am so grateful you stepped in as my second promotor. Your expert knowledge, combined with your warm and friendly personality, made it so enjoyable to work with you. I am sure I already told you this, but econometrics scared the hell out of me. But with your patient, friendly and clear explanations you helped me conquer my fears. Or at least partially, as I still get sweaty palms thinking about time-series econometrics. I had fun in our meetings. Not only when discussing analysis, strategies, python code or my lengthy writing style, but also when discussing life or your dislike of horses. You always thought along with me, came up with creative ideas and pushed me to challenge myself. Especially during the last year, when we worked more intensively together. Looking back, you gave me exactly what I needed to finish my last paper and complete the final stretch of my PhD. I also appreciate that you tried to keep my last article short and sweet – something I desperately needed – despite it getting longer and longer as Joost got more involved. Thank you for your support and guidance, your patience in teaching, and your warm and kind conversations.

Joost and Koos, your guidance and support are a blueprint for how all PhD supervision should be. I would wish anyone pursuing a PhD a Joost and Koos as their promotors. You pushed me to give my absolute best, but in a gentle way and never losing sight of my personal development and well-being. Thank you.

I would also like to thank my reading committee – **Bedir Tekinerdogan**, **Guillaume Bagnarosa**, **Hélène Vletter-van Dort** and **Steven Duivenvoorden** – for reading and evaluating my dissertation, as well as for your attendance and questions during my PhD defense.

A massive thank you to everyone in the **HighLO** team, it has been a pleasure to work with you, learn together and combine our expertise to create something innovative. **Philippe**, you are my partner-in-crime (in the figurative sense, of course!) and indispensable colleague and friend. I could not have completed my PhD without you, since your expertise in data science has been the backbone of our research. You have been my go-to person for anything related to my PhD, always ready to collaborate, share insights and push our research forward. But also on a personal level, our conversations about stock markets, good food, plants, cats and life in general made my PhD much more enjoyable. Thank you for being a teammate and friend throughout this process. **Tarek**, you introduced me to the wondrous world of energy markets. You are such a warm and kind person, and I really appreciate how you always involve me. I am excited about the future of your energy market research, and I look forward to continue working with you. **Serdar**, working with you was fun and you always made me laugh. **Axel**, you are so inspiring to work with. Not only are you highly skilled and knowledgeable in your work, but you are also incredibly supportive and encouraging, both professionally and personally. Thank you for always thinking along, your encouragement, your feedback and your enthusiasm. **Stephan**, while you were part of HighLO for a relatively short time, your insights were invaluable. It was fun to work with you, and I am glad you still join the social activities and entertain us with your crazy stories. **Lorenzo** and **Jonas**, thank you for all your sharp insights and support during my PhD. You guided me through the particle physics methodologies and tools and I could always count on your help. It has been a pleasure working with you, and I hope we continue to do so for many more years. **Han**, thank you for all the efforts you put into HighLO. Witnessing how CERN innovations can be applied to other industries has been inspiring.

Olaf, **Arjen, Titia** and **Mieke**, thank you for giving me the opportunity to work on my PhD at WEcR and taking me under your wing. **Arjen**, your enthusiasm and willingness to entertain any idea, no matter how crazy, were exciting. Thank you for all the fun brainstorms in how to further develop Project HighLO, market intelligence, T-25, tipping points and many more initiatives. I appreciate the effort you put into embedding me in the organization. My personal favorite was your pizza-party-with-grape-juice approach. It was great going into the details of my research, looking at things together, and also sharing personal stories. **Kristel**, thank you for all the fun catch-up calls and your hard work in promoting my research and Project HighLO. Not only could I talk to you about the details of my research, which I really appreciated, but you also made time for me to discuss life. (And of course, to discuss nails and personal stylists, which I desperately needed with all these men around me. May I ever be as stylish as you!). I would also like to thank all the **WEcR colleagues** who involved me in the company and encouraged me along the process.

All the wonderful people from **MCB**, thank you for your kind, warm, inclusive and welcoming spirit. Although my research was different from the majority of the group, I could always count on you if I ever needed help. The MCB PhD days hold a special place in my heart, as I learned so much from your world, the critical thinkers and your encouraging feedback. I want to extend a special thank you to **Ellen V.**, who was always there to help me navigate the organization. I could always count on you, and you saved me so much stress by being there.

Vic, you are one of the few people who have read my dissertation from front to back. That is impressive by itself, but even more so considering it still had all my tense mix-ups and grammar mistakes. You improved the quality of my dissertation and, by learning from my mistakes, the quality of my writing. I hope this uncorrected chapter proves it right by not containing too many mistakes! Thanks to your guidance in improving my presentation skills, I now feel more comfortable letting my enthusiasm – complete with all the accompanying hand gestures – flow freely while presenting. **Mathijs**, I was so happy to finally meet someone without a data scientist background with whom I could have in-depth discussions about financial markets. Thank you for the discussions, feedback on my research, involving me in your PhD and your enthusiasm. You know where to find me if you ever want to write something together! I would also like to thank members of the **IMS Group**. I have learned, and I am still learning, so much from the discussions and I appreciate your help and guidance in my research. I hope we can continue to collaborate and make some positive changes together.

Thank you to **Merije** and **Else**, who are my paranymphs together with Philippe. **Merije**, can you believe it has already been over four years since we reunited at WUR? I am so happy I send you a message when I recognized you, which resulted in our biweekly thee-koffie-of-water-maar-geen-cola-call. I am always looking forward to these calls as I enjoy catching up and laughing with you, talking about work and life, and of course hearing your opinion on the Eurovisie Songfestival. You may not have realized it, but you have been a constant source of joy throughout my PhD. **Else**, you joined WEcR shortly after I started my PhD, and although I cannot pinpoint the exact moment we met, we clicked instantly. Our academic topics are completely different – sorry, I still only understand half of what you do! – but we connected on a more personal level. Our PhD’s, work projects, cats, creative hobbies and more serious topics: I feel like I can talk to you about anything. Your considerate, thoughtful and empathetic qualities are something I really appreciate.

Thank you **Karlijn**, for being one of my biggest cheerleaders throughout my PhD. I can always count on you for a good time, and we share so many memorable moments together. You made sure we always celebrated milestones, even if they were small. You were there for me when things were great, but also when they were not. I know I can always count on you. Your support and enthusiasm mean so much to me, and have given me a lot of energy in my PhD. Thank you for being such a wonderful friend. **Eline**, you have become a very close friend to me ever since we met in the second year of my PhD. We can talk nonstop for hours and hours, and still when we say goodbye, tell each other that our time together was

too short. We can dive deep into personal matters, but also be silly and have fun together. I admire how you dance your way through life. This year you started a PhD yourself, and I hope I can provide the same level of support and friendship that you have given me.

A special thanks to **my family**, who have always been there for me. I am grateful that we are so close and that I can always count on you. You have provided a rock-solid foundation for me to rely on. I love you all so much. Thank you to my kind-hearted and perceptive mama, **Ingrid**, and my sweet and knowledgeable papa, **John**. You are both so caring, smart, kind and funny. You have offered me so many opportunities and always encouraged and motivated me to do the best I can. You also gave me the well-needed relaxing moments to stay balanced. I could not have finished my dissertation without you. Thank you to my big brother **Paul**, your free-spirited, compassionate and funny personality has been a great example for me and is something I truly admire. Thank you to my little (yes, little!) brother **Mark**, for your caring nature, your sincerity and genuine interest. Also for your dry humor that makes my eyes roll so much I am amazed they are not stuck. Thank you to my strong and fun little sister **Nicolette**, you are a true powervrouw. I can always count on your wise, down-to-earth advice, and you never fail to make me laugh until tears stream down our faces. From the bottom of my heart, thank you all. I love how silly we can be and I am so proud of all of you.

Margot, **Jan**, **Jorg**, **Janjira**, **Valeri** and **Dave**, thank you for your constant interest and curiosity in my research. It means so much to me that you genuinely want to understand my work. Some of you even read my first published article! You have always made me feel welcome and you supported me from the moment I met Jesper. Thank you for being the kind, warm and curious people you are.

A special thanks to my sweetest **Does**. You were a constant source of comfort and joy throughout my PhD. Your unwavering positivity and ability to find happiness in every situation taught me invaluable lessons. Thank you for your unconditional love and the trust you put in me. I love you so much and miss you every day. **Louka**, your persistent nudges for attention and demands for outdoor adventures kept my spirits high and stress levels low. Thank you for making me laugh everyday and I am honored – yet grossed out please stop – for being your go-to person for your after-dinner-burp. **Fleur**, **Charlie**, **Joris**, **Teddy** and **Misty**, thank you for your silent encouragements and reminders to take things easy.

The love of my life, allerliefste **Jesper**, thank you for being the wonderful person you are. We met shortly after I started my PhD, and I probably have to thank CERN for this – maybe also apologize again – since I swipe-baited you by mentioning them in my profile. From our very first date, I knew you were one of a kind; we share so many interests, personality traits

and this stuiterbal-like-energy when we get excited. Who knew that my "I am staying at your place for one week" – due to my job banning me from the office because I lived in Brabant and had a stuffy nose, remember COVID-19? – would turn into us moving in together, buying a house, renovating that house, and now buying a "piece of grass" to realize our dream. During all of this and my PhD, you were a constant powerful source of support who I could always rely on. You always believed in me and helped me stay sane in my PhD. You never fail to make me laugh, whether it is with your (dumb) jokes, our dancing together, your silly faces and impersonations, or simply goofing around. Thank you for all the happiness and love, and for making me a better person. I love you and I cannot wait to see what the future holds for us.

Last, I would like to thank **myself**. My enthusiasm, my creativity, my curiosity and my determination have brought me this far. I am proud of the person I became over the last years.

liefs,
Marjolein

About the author

Marjolein Eveline Verhulst was born on March 26, 1992 in Amsterdam, the Netherlands. She completed the Bachelor's program Business Administration at Radboud University in Nijmegen in 2014, and also took part in the interdisciplinary Honours Programme at the Radboud Honours Academy. Dedicating the final year of her bachelor to internships, Marjolein developed a strong interest in futures markets after a tour at the trading pits at the Chicago Mercantile Exchange Group. In 2015, she finished the Master International Business: Marketing-Finance at Maastricht University, and wrote her thesis on what the best futures markets are for European agricultural hedgers to manage their price risk exposure.

Marjolein started working for the cooperative ZON fruit & vegetables in 2016 as business development analyst. Here, she was responsible for researching, designing and developing European vegetable futures contracts and negotiated with the leading stakeholders in Europe to implement these contracts. In 2019, she started working as a financial market researcher at Wageningen Economic Research and simultaneously began her PhD trajectory as an external PhD student with the Marketing and Consumer Behaviour Group at Wageningen University.

Marjolein's research focused on how market manipulation in high-frequency markets can be identified and analyzed, by applying particle physics tools and methodologies to futures market data. In doing so, she closely collaborated with the European Organization for Nuclear Research (CERN). She made additional efforts to ensure her research had practical applications, to make her research and findings accessible to the general public, and initiated connections with key industry stakeholders. The latter resulted in Marjolein co-founding the International Expert Group on Market Surveillance (IMS Group). IMS Group bridges science and industry to tackle financial market surveillance challenges, underlining the need for cooperation in the supervision of these markets. At the time of writing, IMS Group consists of eighteen regulators and exchanges across Europe and the United States. During her PhD trajectory, Marjolein followed numerous courses, was involved in teaching, supervised BSc and MSc thesis students, organized a conference and a scientific panel session, and presented her work at national and international conferences.

After her PhD defense, Marjolein aspires to keep one foot in academia and the other in industry, blending scientific research with practical implementation. She is passionate about financial risk management instruments such as futures contracts, enjoys bridging multiple disciplines, and is eager to work in (futures) market surveillance to share her enthusiasm with others. She also has a great fondness for animals, so do not be too surprised if one day you find Marjolein pouring her enthusiasm into running her own cat café with adoptable cats.

TRAINING AND SUPERVISION PLAN

Marjolein E. Verhulst
Wageningen School of Social Sciences (WASS)
Completed Training and Supervision Plan

Name of the learning activity	Department/Institute	Year	ECTS*
Project related competences **A1 Managing a research project**			
WASS Introduction	WASS/WUR	2019	1
Writing research proposal	MCB/WUR	2019	6
'The Effect of Liquidity Shocks on the Resiliency of Agricultural Futures Markets'	Agricultural & Applied Economics Association, Atlanta, USA	2019	1
'The Impact of Roll Strategies and Commodity Futures Characteristics on Long-Only Strategies'	Agricultural & Applied Economics Association, Atlanta, USA	2019	1
'The Effect of Price Limits on Liquidity: A Natural Experiment'	European Association of Agricultural Economists	2021	1
'Applying Particle Physics Data Analysis to Limit Order Book Data to Detect Malicious Anomalies in Futures Markets'	PhD Symposium Day 2019: Science with Impact, Wageningen, the Netherlands	2019	1
A2 Integrating research in the corresponding discipline			
Capita Selecta AEP – *Econometrics and Machine Learning for Time Series*	WUR, Wageningen, the Netherlands	2023	6
Advanced ROOT Training	CERN, Meyrin, Switzerland	2021	6
General research related competences **B1 Placing research in a broader scientific context**			
Biases and Portfolio Selection	Coursera, online	2022	0.6
Organise and chair organised session at European Association of Agricultural Economists (EAAE)	EAAE, online	2021	3
Lovelace Machine Learning Bootcamp	LovelaceCamp by Rotterdam.AI, Rotterdam, the Netherlands	2019	2
B2 Placing research in a societal context			
'Wat deeltjesfysica ons kan leren bij het opsporen van afwijkend gedrag op financiële markten'	Institute for Financial Crime: FinCrimeWeek 2022, online	2022	1
Popular Science Writing	WASS/WUR, Wageningen, the Netherlands	2021	1.5
Foundation of the 'International Expert Group on Market Surveillance', the organisation of a two-day kick-off event thereof and a research presentation at this event.	IMS Group, CERN	2023	1.5
Career related competences/personal development **C1 Employing transferable skills in different domains/careers**			
Introduction to LaTeX	PE&RC, Wageningen, the Netherlands	2021	0.1
PhD Competence Assessment	WGS/WUR, Wageningen, the Netherlands	2020	0.3
Teaching and assisting MCB-51403 Capita Selecta Commodity Futures and Options Markets	MCB/WUR, Wageningen, the Netherlands	2020-2022	2.0
Supervision MSc Students	WUR, Wageningen, the Netherlands	2021-2023	2.0
Total			**37.0**

*One credit according to ECTS is on average equivalent to 28 hours of study load

Colophon

The research described in this thesis was financially supported by Wageningen Economic Research (WEcR) and the Commodity Risk Management Expertise Centre (CORMEC).

Financial support from Wageningen University for printing this thesis is gratefully acknowledged.

Copy editing by Vic Diederen, STORM Tekst & Training
Cover design by Simone Golob, Simone Golob Illustratie & Vormgeving
Lay-out by Ron Zijlmans, RON Graphic Power
Printed by ProefschriftMaken on FSC-certified paper