

ON DETECTING SPOOFING STRATEGIES IN HIGH FREQUENCY TRADING

XUAN TAO, ANDREW DAY, LAN LING, AND SAMUEL DRAPEAU

ABSTRACT. Spoofing is an illegal act of artificially modifying the supply to drive temporarily prices in a given direction for profit. In practice, detection of such an act is challenging due to the complexity of modern electronic platforms and the high frequency at which orders are channeled. We present a micro-structural study of spoofing in a simple static setting. A multilevel imbalance which influences the resulting price movement is introduced upon which we describe the optimization strategy of a potential spoofer. We provide conditions under which a market is more likely to admit spoofing behavior as a function of the characteristics of the market. We describe the optimal spoofing strategy after optimization which allows us to quantify the resulting impact on the imbalance after spoofing. Based on these results we calibrate the model to real Level 2 datasets from TMX, and provide some monitoring procedures based on the Wasserstein distance to detect spoofing strategies in real time.

Keywords: Spoofing, High Frequency Trading, Imbalance, Limit Order Book.

1. INTRODUCTION

The act of spoofing is a specific trading activity that aims at artificially modifying the supply on the market without intent to trade to move it away from its equilibrium. One might profit from the resulting short term price movement by canceling the previous supply while the market comes back to its equilibrium. Such a strategy implies that the spoofer should be able to act anonymously, fast and in a market where all the other agents react to offer and demand. In this regard, with the recent rise of centrally cleared venue and high frequency algo-trading, the ground for the existence of spoofing schemes is rising, see Shorter and Miller [21].¹ In a competitive market where many potential spoofers are present, spoofing behavior might cancel out, but most regulations consider it as illegal. For instance, the 2010 Dodd-Frank Act prohibits spoofing – defined as activity of bidding or offering with the intent to cancel before execution – that can be prosecuted as “a felony punishable by up to \$1 million in penalties and up to ten years in prison for each spoofing count”.²

Yet, for several reasons, detecting and prosecuting spoofing behavior is a challenging problem. First is the sheer amount of data produced from high frequency trading across many financial products and venues. Second, it is usually impossible to trace in real time who

Date: October 1, 2020.

We also thank the Fields Institute for the organization of the many “Fields-China Joint Industrial Problem Solving Workshop” from which this problem stems. Finally, we thank TMX and their data analysis team for supporting this project with profound datasets, high performing computing facilities as well as precious market insights in high frequency trading. The financial supports from the National Science Foundation of China, Grants Numbers: 11971310 and 11671257; as well as from Shanghai Jiao Tong University, Grant number AF0710020; are gratefully acknowledged. Andrew Day gratefully acknowledges the funding provided by Mitacs, the Ontario Graduate Scholarship (OGS), and the Natural Sciences and Engineering Research Council of Canada postgraduate scholarship (NSERC PGS) .

¹In 2010, trader Navinder Singh Sarao was accused of exacerbating a flash crash by placing thousands of E-mini S&P 500 stock index futures contract orders in one day and changed or moved those orders more than 20 million times before they were cancelled.

²In 2019, the high frequency company Tower Research Capital agreed to pay a fine of about \$60 million over spoofing allegations. In 2020, JP Morgan settles spoofing lawsuit alleging fraud for about \$920 million.

is behind every trade. From a CCP viewpoint, they mainly have access to the broker ID through which the trade has been channeled resulting only in aggregated informations. Furthermore, a potential spoofer might post those trades through different venues and brokers. Third, aside from a loose definition, it is unclear how a spoofing strategy differs quantitatively from other strategies and what is the resulting impact. Thus the complexity of quantifying and discriminating spoofing strategies from legitimate ones. Based on the above points, it seems difficult to provide an efficient way to monitor the market for spoofing behavior.³ However, a potential spoofer is also confronted to the constraints of modern electronic trading platforms. Indeed, from the basic spoofing description, the spoofer has to act rapidly in a complex and high frequency environment. Therefore, it must rely on fast – henceforth simple – algorithmic strategies which, due to the complexity of the dynamic structure of a limit order book, is based on aggregated signal.

Along these lines, we intentionally address a quantitative analysis of spoofing in a simple setup. As a basis for this study, let us consider a simple example. We suppose that in the next period the limit order book shifts up by one unit with a probability \bar{p} and down by one unit otherwise. From the perspective of an agent whose objective is to purchase two shares, it faces the following three idealized situations.

1 - Immediately post a buy market order for a total cost of

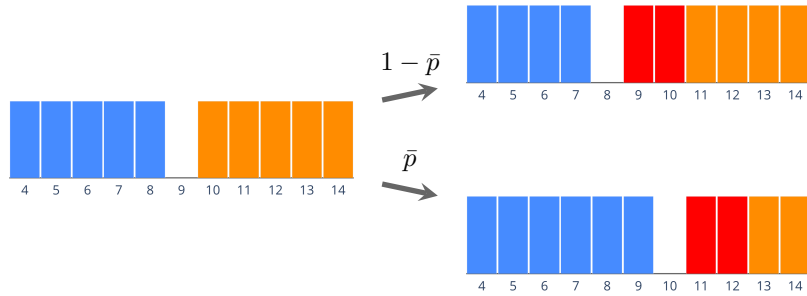
$$\hat{C} = 10 + 11 = 21$$



2 - Delay the buy market order for one period resulting in a total average cost of

$$\tilde{C} = (1 - \bar{p})(9 + 10) + \bar{p}(11 + 12) = (1 - \bar{p})19 + \bar{p}23$$

which is smaller than \hat{C} if and only if $\bar{p} < 1/2$, hence a bearish market.



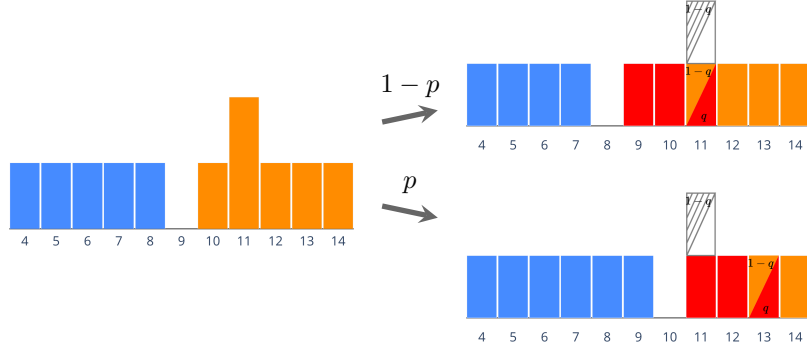
3 - Delay the buy market order and post a sell limit order of one share at a distance of one unit from the best ask price to artificially modify the offer and demand resulting in a temporary more bearish state $p < \bar{p}$.⁴ Doing so, with a probability q , its sell limit order is executed through an incoming market order walking the limit order book beyond one unit. For this executed sell order, the agent receives an average price of $q((1 - p)10 + p12)$ while its inventory increases in average to $2 + q$. The cost of buying back this increased

³Aside from obvious cases or exogenous approaches as for insider prosecution.

⁴We suppose that during this time period, if the market moves through limit order posting/canceling, the agent keeps its sell limit order one unit away from the best ask price by rapidly canceling and posting again.

inventory minus the gain from selling its limit order results in an average net cost of

$$\begin{aligned} C &= (1-p)(9+10+q11) + p(11+12+q13) - q((1-p)10 + p12) \\ &= (1-p)19 + p23 + q \end{aligned}$$



From this simple situation, a profitable spoofing situation depends on the value of \bar{p} , how bearish $p < \bar{p}$ the market reacts to an increase of one share at a distance of one unit from the best ask price, as well as the probability q of being adversarially executed through a market order during this time. The goal of this paper is to study the interplay between these different factors. In particular the impact on the price movement as a function of the spoofing size and depth in the limit order book. Based on this study, we present some approaches to track spoofing behavior and calibrate those to real market data.

We model the impact of the offer and demand on the price through the volume imbalance often taken as the ratio of the volume of the best bid divided by the total volume on the best bid and ask. Since the spoofer never intends to have their orders executed, spoofing is more likely to happen beyond the top of the limit order book since the possibility of getting executed is too high which results in a negative payoff to the spoofer. To take this into account we weight the impact on the imbalance in terms of depth as follows

$$\bar{\iota} = \frac{\sum \bar{v}_k^- w_k}{\sum (\bar{v}_k^- + \bar{v}_k^+) w_k}$$

where \bar{v}_k^\pm represents the volume on the bid/ask k units away from the best bid/ask and w_k is the relative impact on the imbalance at level k . If the agent posts a sell limit order v on the ask side at tick level k_0 , the imbalance moves to

$$\iota(v) = \frac{\sum \bar{v}_k^- w_k}{\sum (\bar{v}_k^- + \bar{v}_k^+) w_k + w_{k_0} v} \leq \bar{\iota}$$

If we denote by dp_n the probability of a price deviation of n units, the dependence on the imbalance ι is given as follows

$$dp_n = \iota dp_n^+ + (1 - \iota) dp_n^-, \quad n = \dots, -2, -1, 0, 1, 2, \dots$$

where dp_n^\pm represents the price deviation when the imbalance is at its extreme. When the imbalance ι is low/high – the offer/demand dominates – the price distribution is biased downwards/upwards through dp^\pm . The agent can influence the price distribution in a non linear way through the volume it posts:

$$v \mapsto dp(v) := \iota(v) dp^+ + (1 - \iota(v)) dp^-$$

Given the probability dq of a sell limit order hitting the limit order book up to a given level the resulting net average cost of spoofing turns out to be

$$C(v) = pH + \underbrace{(1-Q)G(H)}_{\text{Cost for optimal situation}} + \underbrace{H\mu^+(2\iota(v)-1)}_{\text{Spoofing impact}} + \underbrace{QG(H+v)+v\nu}_{\text{Cost for being caught wrong way}}$$

where H is the initial objective of shares to acquire, G is the liquidity costs from walking through the limit order book, $\mu^+ > 0$ is the mean of dp^+ , Q (resp ν) is the probability (resp mean beyond k_0) of being executed beyond k_0 . From this expression, there is a competitive aspect between the risk of being caught on the wrong side and the fact of pushing $\nu(v)$ way beyond $1/2$ to get $H\mu^+(2\nu(v) - 1) \leq 0$.

In a theoretical part we first provide conditions for the limit order book to admit spoofing manipulation⁵ as a function of the initial imbalance \bar{i} , local sensitivity of the imbalance w , overall price impact μ^+ , liquidity cost G as well as the initial objective H . In short, this model confirms several intuitive facts when spoofing is more likely to occur

- if the probability Q of the spoofing order being executed is small;
- if the local sensitivity w_{k_0} or the overall price impact μ^+ is large;
- if the amount of share H to buy is large with respect to the depth of the limit order book;
- if the initial market imbalance is close to $1/2$.
- away from the top of the limit order book;

As for the depth of the limit order book – how liquid the market is – the results are inconclusive. For this to be taken into account, one should account for how the above mentioned parameters depend on the liquidity of the limit order book. The subsequent empirical study shows that it is the case, but we can not derive conclusions from this model as in Shorter and Miller [21] where illiquid markets seems more likely to be prone to spoofing. We then address the impact of spoofing on the resulting imbalance $i_{spoof} = \nu(v_{spoof})$ after spoofing and discuss its dependence as a function of the aforementioned parameters as the imbalance modified by a spoofer can be compared to the steady state one. We characterize and discuss the deviation for the imbalance as a function of the different parameters. In particular as a function of the depth where the spoofing order is posted. We finally address the situation of a market maker using spoofing strategies for a positive round trip payoffs. Based on this study, we can theoretically discriminate spoofed imbalance from steady state one. The main idea for potential spoofing detection is to compare the imbalance before and after each market order. If the market is on steady state, the imbalance after a legitimate market order should come back to its current steady state \bar{i} . On the other hand, if the market order is the result of a spoofing pattern as described in the simplified aforementioned model, the imbalance before the market order should be of the form

$$i^* \approx \frac{b}{b + a + wv} < \frac{b}{a + b} = \bar{i}$$

while returning to its market steady state as soon as the spoofed volumes are canceled. The theoretical part provides a first approximation how i^* deviates from \bar{i} as a function of the market parameters.

We calibrate the parameters of this model as well as the weighted imbalance on several Level 2 data sets provided by TMX. Based on the theoretical results we provide possible quantification procedures using real time monitoring of a conditional Wasserstein distance of short term imbalance histories against the long run imbalance history in order to track eventual spoofing behavior.

Before addressing the relevant literature, let us expose the shortcomings and modeling choices of this approach. The micro-structure dynamic of the limit order book at high frequency is highly complex. To excerpt some key impact of spoofing behavior we deliberately concentrate on a static situation where the dynamic of the market is ignored.⁶ For instance, we do not consider situations where compound spoofing behavior happens. We

⁵In other words, better than immediate or delayed market order.

⁶Since we consider the limit order book beyond its top, a dynamic version of the present approach would result into a fairly complex and high dimensional dynamic programming problem.

furthermore assume that there exists a single potential spoofer and that the market is infinitely reactive in the sense that it comes back to its steady state driven by the imbalance. There is no implicit game between the market and the spoofer where the market acknowledge the potential existence of a spoofer. There is also no competitive game between two or more spoofers. Also, even though we shortly address the situation of a round trip for a market maker and the resulting optimal spoofing behavior, we take the viewpoint of a market taker willing to purchase/sell a given amount of shares. The overall goal being a basic understanding of spoofing in its most simple nature, quantify the resulting impact and derive potential detection procedures. Refinement of this approach, other take on, as well as more adequate quantification procedures are topics of further studies.

1.1. Literature review. There exists a solid stream of research showing that even rational speculative activities might destabilize prices, have an adverse effect on market efficiency or eventually leads to different forms of arbitrage; From market speculation based on various form of information asymmetries, for instance Hart and Kreps [14], Allen and Gale [3] or Jarrow [15], to price manipulation in limit order books using different market impact assumptions and trading strategies, as studied in Alfonsi and Schied [1], Alfonsi et al. [2], Gatheral [10], Gatheral and Schied [11]. The specific case of spoofing behavior has not yet been the subject of much theoretical study.

Although many high frequency trading strategies are legitimate, Shorter and Miller [21] point out that high frequency trading firms may engage in potentially manipulative strategies involving the usage of quote cancellations. Lee et al. [16] empirically study the change in spoofing behavior following a change in volume disclosure rules on the Korean Exchange (KRX) at the start of 2002. Up to the end of 2001 the KRX disclosed the total volume of shares on both sides of the book and also the volumes at each tick up to 5 ticks from the best ask/bid. At the start of 2002 the KRX stopped disclosing the total volume on both sides in an effort to stop spoofing, but increased the disclosed volumes at each tick from the first 5 to the first 10 ticks from the best ask/bid. They show that spoofing is profitable and spoofers tend to prefer stocks with higher return volatility, lower market capitalization, lower price and lower managerial transparency. This study suggests the importance of the depth of book on spoofing strategies and potential price manipulation being carried out through a form of "volume imbalance". Wang [24] followed a similar methodology in empirically studying spoofing on Taiwan's index futures market. They found consistent results on the impact of spoofing on the market, but without the novel testing ground on changes in the disclosure of volumes deeper in the limit order book.

Some other studies to detect price manipulations are mainly based on learning algorithms. Among other, Cao et al. [5, 6] based on the definition of spoofing in [17] use K-nearest neighbour, one class support vector machine and adaptive hidden markov models to classify the data. Miranda et al. [18] characterize spoofing and pinging as full and partial observability of Markov decision processes. Under a reinforcement learning framework, they find that in order to maximise the investment growth, a trader will always employ spoofing or pinging orders except when market adds extra transaction costs or fines. In contrast to these empirical studies, our approach focuses on the micro economic features of spoofing behavior, in particular using our main stylized factor $\hat{\Delta}$ imbalance, which measures the difference between bid and ask side.

Concerning the impact of the imbalance on direction of the price movement: Lipton et al. [17] use the definition on the top of the book for the imbalance and study the impact on the trade arrival dynamic and resulting price movement. They fit a stochastic model for this behavior on real market data. Cartea et al. [7] employ volume imbalance as a signal to improve profits on the liquidation of a collection of shares in a dynamic high-frequency trading environment. Gould and Bonart [12] fit logistic regressions between the imbalance and the direction of the subsequent mid-price movement for each of 10 liquid stocks on Nasdaq, and illustrates the existence of a statistically significant relationship.

Xu et al. [26] compute the imbalance at multi position in the limit order book and fit a linear relationship between this imbalance and the mid-price change. They find that the goodness-of-fit is considerably stronger for large-tick stocks than it is for small-tick stocks. The impact of order imbalance on prices has also been studied by Cont et al. [9] and Bechler and Ludkovski [4], for example. Bechler and Ludkovski also found that including characteristics of deeper parts of the book may be necessary for forecasting price impact. However, due to the nature of their dataset, they were only able to look at an aggregated form of the depth of book while we are able to use the exact volumes at all depths in the book. Sirignano [22] also used the book volumes beyond the touch to model price movements in a deep learning setting. Further suggesting the impact of depth of book on predicting future price movements.

Finally, the closest work to the present one in terms of quantitative analysis of spoofing behavior in relationship with imbalance is from Cartea et al. [8]. They adopt a dynamic approach where the trader influences the imbalance to derive the optimal strategy. They calibrate their model to market data and provide trading trajectories for the spoofer showing that spoofing considerably increases the revenues from liquidating a position. While being in a dynamic setting, in contrast to the present study, everything happens at the top of the limit order book for the imbalance to be manipulated. Furthermore, we do not focus here on the resulting gains from the spoofer, but rather on the impact on the imbalance from spoofer as for detection purposes from a regulatory viewpoint.

1.2. Organization of the paper. The first Section introduces the model, the imbalance and the dependence of the price movement on that imbalance. The second Section presents the spoofing strategies, addresses the theoretical conditions for spoofing behavior to happen and provide the resulting imbalance after spoofing together with numerical illustrations. It also addresses the situation of a round trip from a market maker viewpoint in this model. The third Section is dedicated to the calibration procedure of the model on real Level 2 market data from TMX. The last Section discusses and introduces some quantitative approach to track spoofing behavior in real time illustrated on real datasets.

2. LIMIT ORDER BOOK, LIQUIDITY COSTS AND IMBALANCE

The ask price is denoted by p and the limit order book on the ask side by $\bar{v} = (\bar{v}_0, \dots, \bar{v}_N)$, that is, \bar{v}_0 is the volume posted at ask price p , \bar{v}_1 the volume posted at $p + \delta$, etc. where δ is the tick size. We denote by $\bar{v}^- = (\bar{v}_0^-, \dots, \bar{v}_N^-)$ the limit order book on the bid side.⁷ Given a limit order book inventory \bar{v} on the ask side, we define for an amount of share $H \geq 0$ the function

$$F(H) := \inf \left\{ x \in \mathbb{N}_0 : \sum_{k=0}^x \bar{v}_k \geq H \right\}$$

which represents how many positive price tick deviation an order of size H generates. Given an amount of shares H to buy, a bid price p and an ask limit order \bar{v} , the resulting costs of the market order is

$$pH + \sum_{k=0}^{F(H)} k\delta\bar{v}_k - \delta F(H) \left(\sum_{k=0}^{F(H)} \bar{v}_k - H \right) = pH + \delta G(H)$$

The term G on the right hand side represents the liquidity costs depending only on \bar{v} .

Remark 2.1. *Throughout the theoretical part of this work we assume that the limit order book is blocked shaped with an amount $a > 0$ of shares at each price level of the ask side. We then get the continuous approximation*

$$F(H) \approx \frac{H}{a} \quad \text{and} \quad G(H) \approx \frac{H^2}{2a}$$

⁷That is \bar{v}_0^- is the volume posted at bid price $p^- < p$, \bar{v}_1^- the volume posted at $p^- - \delta$, etc.

As for the imbalance of the limit order book, measure of the difference between offer and demand, we proceed as follows. Let w_0, \dots, w_N with $\sum_{k=0}^N w_k = 1$ and $w_k > 0$, a weight for each tick level k , and a limit order book inventory \bar{v}^-, \bar{v} on the bid and ask respectively, we denote by

$$\bar{B} = \sum_{k=0}^N w_k \bar{v}_k^- = \langle w, \bar{v}^- \rangle \quad \bar{A} = \sum_{k=0}^N w_k \bar{v}_k = \langle w, \bar{v} \rangle$$

the weighted average bid and ask volumes. We define the imbalance as

$$\bar{i} := \frac{\bar{B}}{\bar{B} + \bar{A}} \in (0, 1)$$

Remark 2.2. *In a blocked shaped setting, if b denotes the amount of orders on every price level on the bid side, we get*

$$\bar{i} = \frac{\sum_{k=0}^N w_k b}{\sum_{k=0}^N w_k (a + b)} = \frac{b}{a + b}$$

which yields $b = a\bar{i}/(1 - \bar{i})$.

The price deviation in the next period can be triggered by two events. The posting and cancellation of incoming limit orders as well as the posting of market orders. We distinguish between both, since the former does not have an impact on the execution of existing limit orders while the latter has. We generically denote by

$$\begin{aligned} dp &= \{dp_{-N}, \dots, dp_0, \dots, dp_N\} & \mu &= \sum_{x=-N}^N x dp_x \\ dq &= \{dq_{-N-1}, \dots, dq_0, \dots, dq_{N+1}\} & \nu &= \sum_{y=-N-1}^{N+1} y dp_y \end{aligned}$$

the distribution and mean, respectively, of the two possible price movement in the next period. For the sake of simplicity, we assume that the imbalance does not have an impact on incoming market orders and that the price deviation with respect to the market orders is neutral, that is $\nu = 0$. We furthermore assume that they are independent of each others.⁸ To reflect the fact that the imbalance, as an indicator of the offer and demand on the market, has an impact on the market makers, we consider a parametrization $\iota \mapsto dp(\iota)$ of the price movement driven by limit orders as a function of the imbalance ι . Since the imbalance moves between 0 and 1, we assume that the distribution dp moves as a convex combination of ι between the distribution dp^- – distribution when the imbalance is close to 0, that is highly skewed to the left – and the distribution dp^+ – distribution when the imbalance is close to 1, that is highly skewed to the right. Mathematically:

$$dp(\iota) = \iota dp^+ + (1 - \iota) dp^-$$

From the skewness assumptions and symmetry of the imbalance indicator, we assume that

$$dp_x^+ \geq dp_x^- \quad \text{for every } x \geq 0 \quad \text{and} \quad dp_x^+ = dp_{-x}^- \quad \text{for every } x$$

which implies that

$$\mu(\iota) := \sum_{x=-N}^N x dp_x(\iota) = (2\iota - 1) \sum_{x=-N}^N x dp_x^+ = \mu^+(2\iota - 1)$$

⁸The subsequent theoretical study adapt to eventual joint distribution of price movement due to limit and market orders also jointly dependent on the imbalance. The exposition of which is no longer explicit but can be solved numerically.

By assumption, μ^+ is positive, showing that $\mu(i)$ moves between $-\mu^+$ and μ^+ and is equal to 0 for an imbalance of 1/2 when offers equal demand.

3. SPOOFING STRATEGY

Suppose that at a given time we are given an ask price p and a limit order book inventory (\bar{v}^-, \bar{v}) . A trader willing to buy an amount H of shares faces the following three options.

- **Immediate market order:** for a total costs of

$$pH + \delta G(H)$$

- **Delayed market order:** for an average total cost of

$$\begin{aligned} \sum_{x=-N}^N \sum_{y=-N-1}^{N+1} [(p + \delta(x+y))H + G(H)] dp_x(\bar{v}) dq_y \\ = pH + \delta (G(H) + H\mu^+ (2\bar{v} - 1)) \end{aligned}$$

Clearly if $\bar{v} < 1/2$, then this second option is better than a direct buy.

- **Spoofing and delayed market order:** Book first an ask limit order v at a depth k in $\{0, 1, \dots, N\}$ on top of the ask limit order book \bar{v}_k to increase the liquidity on the ask side and signal a surge in supply to the market. In the next period the price deviates from p to $p + \delta(x+y)$ and two situations may happen:

- $y \leq k$: no market order of sufficient magnitude hits the limit order book and therefore this limit order is not executed against an incoming market order. The previous limit order is canceled and the amount H of shares is acquired for a cost of

$$(p + \delta(x+y))H + \delta G(H)$$

- $y > k$: the limit order is executed against an incoming market order at a price level $p + \delta(x+k)$. The new objective moves to $H+v$ resulting in a net cost of

$$\begin{aligned} (p + \delta(x+y))(H+v) + \delta G(H+v) - (p + \delta(x+k))v \\ = (p + x\delta)H + \delta G(H+v) + \delta(y-k)v \end{aligned}$$

It follows that the spoofing net cost for a price deviation of $p + \delta(x+y)$ is given by

$$(3.1) \quad C_k(v, x, y) = (p + (x+y)\delta)H + \delta G(H + v1_{\{y>k\}}) + \delta(y-k)v1_{\{y>k\}}$$

However, the posting of the selling limit order modifies the imbalance from \bar{v} to

$$v_k(v) := \frac{B}{A + B + w_k v}$$

In other words, the imbalance will move downwards, shifting the distribution dp to more favorable outcomes. Since we assume that $\nu = 0$, it follows that the average net costs are given by

$$\begin{aligned} C_k(v) &:= \sum_{x=-N}^N \sum_{y=-N-1}^{N+1} C_k(v, x, y) dp_x(v_k(v)) dq_y \\ &= \sum_{x=-N}^N (p + \delta x)H dp_x(v_k(v)) + \delta(1 - Q_k)G(H) + \delta Q_k G(H+v) + \delta v \sum_{y=k+1}^{N+1} (y-k) dq_y \\ &= pH + \underbrace{\delta(1 - Q_k)G(H)}_{\text{Cost for optimal situation}} + \underbrace{\delta H \mu^+ (2v_k(v) - 1)}_{\text{Spoofing impact}} + \underbrace{\delta Q_k G(H+v) + \delta v \nu_k}_{\text{Cost for being caught wrong way}} \end{aligned}$$

where

$$Q_k = \sum_{y=k+1}^{N+1} dq_y \quad \text{and} \quad \nu_k = \sum_{y=k+1}^{N+1} (y-k) dq_y$$

Note that this cost functional is convex in v since G and ν_k are convex functions. Note also that in order to take advantage of this spoofing impact, it is necessary to drive the imbalance $\nu_k(v)$ below $1/2$.

Remark 3.1. *Note that we implicitly assume that the spoofing only happens at a given depth k . It is possible to spoof simultaneously at different depths resulting in a slightly more complex cost function that can be solved numerically. The conclusions do not change qualitatively and we use the more general multi-depth spoofing for the analysis of data in the subsequent sections.*

3.1. Existence of Spoofing Manipulation. The question is whether it is possible to push the imbalance as much as possible to 0 in order to offset the costs of posting selling orders, they being executed and paying the liquidity costs of buying them back.

Definition 3.2. *We say that the limit order book (\bar{v}^-, \bar{v}) admits a (market taker) spoofing manipulation if there exists $v > 0$ and $k \in \{0, 1, \dots, N\}$ such that*

$$(3.2) \quad \begin{cases} C_k(v) < pH + \delta(G(H) + \mu^+ H(2\bar{v} - 1)) & \text{if } \bar{v} \leq 1/2 \\ C_k(v) < pH + \delta G(H) & \text{if } \bar{v} > 1/2 \end{cases}$$

According to the average costs of spoofing, these two inequalities turn into

$$(3.3) \quad Q_k [G(H + v) - G(H)] < 2\mu^+ H \left(\bar{v} \wedge \frac{1}{2} - \nu_k(v) \right) - v\nu_k$$

The following results concerns the existence of spoofing manipulation in a blocked shaped setting where the volume on the ask side of the limit order book is a everywhere. For scaling reasons let $H = \rho a$, where ρ represents the ratio of the shares to purchase to the depth of the limit order book. Our first result concerns the existence of spoofing manipulation.

Proposition 3.3. *In a block shaped setting, where the volume on the ask side of the limit order book is a everywhere. The following assertions hold*

- *If $\bar{v} \leq 1/2$, the limit order book admits no spoofing manipulation if and only if (3.4) holds;*
- *If $\bar{v} > 1/2$, the limit order book admits no spoofing manipulation if (3.4) holds.*

Where $H = \rho a$ and

$$(3.4) \quad 2\rho\mu^+ (1 - \bar{v})\bar{v}w_k \leq Q_k\rho + \nu_k \quad \text{for all } k$$

Proof. Let $H = \rho a$, and the imbalance $\nu_k(v) = b/(b + a + w_k v)$ with $b = a\bar{v}/(1 - \bar{v})$. It follows that the gradient of $\nu_k(v)$ is given by

$$\nabla \nu_k(v) = -\frac{b}{(a + b + w_k v)^2} w_k = -\frac{(1 - \bar{v})}{a\bar{v}} \nu_k^2(v) w_k$$

If $\bar{v} \leq 1/2$, from the previous equations, since $g(x) = x^2/(2a)$, it follows that there is no spoofing manipulation if and only if

$$f(v) := Q_k \frac{v^2}{2a} + Q_k \rho v + v\nu_k - 2\rho\mu^+ (\bar{v} - \nu_k(v)) \geq 0$$

for any $v \geq 0$. Taking the gradient for this function yields

$$\nabla f(v) = Q_k \frac{v}{a} + Q_k \rho + \nu_k - 2\rho\mu^+ \frac{1 - \bar{v}}{\bar{v}} \nu_k^2(v) w_k$$

which is a monotone functional in v . Since $f(0) = 0$, it follows that $f(v) \geq 0$ for any v if and only if $\nabla f(0) \geq 0$ which is equivalent to

$$Q_k \rho + \nu_k \geq 2\rho\mu^+ (1 - \bar{v})\bar{v}w_k$$

If $\bar{\nu} > 1/2$, there is no spoofing manipulation if and only if

$$f(v) := q_k \frac{v^2}{2a} + \rho q_k v - q_k k v - 2\rho a \mu^+ (1/2 - \nu_k(v)) \geq 0$$

for any $v \geq 0$ the gradient of which is given by

$$\nabla f(v) = q_k \frac{v}{a} + \rho q_k - q_k k v - 2\rho \mu^+ \frac{1 - \bar{\nu}}{\bar{\nu}} \nu_k^2(v) w_k$$

Since $f(0) > 0$, as previously argued, it follows that $f(v) \geq 0$ as soon as $\nabla f(0) \geq 0$, which yields the same conditions. \square

From this proposition, we deduce that price manipulation is more likely to occur

- if Q_k is small – and as a byproduct ν_k . If the probability to get a spoofing order executed is small, there is relatively no downside at spoofing.
- if $\bar{\nu}$ is close to $1/2$. If the imbalance is close to $1/2$, then $\bar{\nu}(1 - \bar{\nu})$ is maximum. The impact of moving the price in ones favor is maximal there.
- if μ^+ is large: μ^+ represents the mean deviation sensitivity as a function of the imbalance. The more sensitive the price movement is with respect to the imbalance, the more likely spoofing strategies may occur.
- If w_k is large: w_k represents the relative impact at tick level k of a spoofing volume to the imbalance. If one of w_k is large with respect to the corresponding Q_k , then spoofing is more likely to occur there.
- if ρ is relatively large. If the amount of order to buy relative to the overall offer is very large, spoofing is more likely to happen.

Figure 1 represents the spoofing in terms of the initial imbalance $\bar{\nu}$ with varying market parameters.

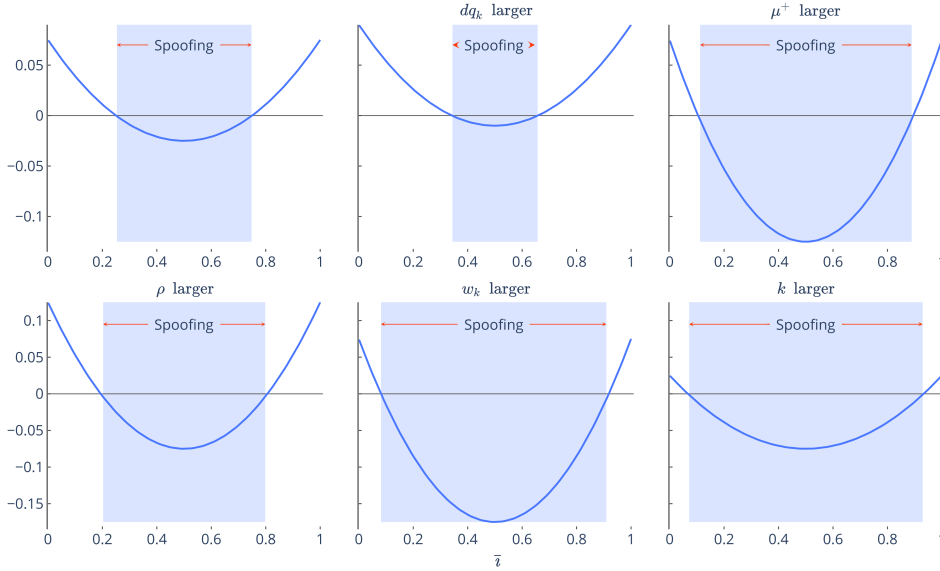


FIGURE 1. Spoofing condition (3.4) as a function of $\bar{\nu}$ and in (a) $\mu^+ = 1$; $\rho = 1$, $k = 3$, $w_k = 0.2$, $dq_i = 0.025$ for all $i \geq k$. One parameter is increased each time with respect to (a) where (b): $dq_i = 0.03$ for all $i \geq k$; (c): $\mu^+ = 2$; (d): $\rho = 2$; (e): $w_k = 0.5$; (f): $k = 4$.

3.2. Optimal Spoofing and Resulting Imbalance Impact. Let us now address the problem of finding the optimal spoofing strategy. In particular as a function of the depth at which the spoofing order is placed.

Proposition 3.4. *The optimal spoofing volume v_{spoof} at a given level k and resulting imbalance ι_{spoof} – adopting the notations $w := w_k$, $Q := Q_k$ and $\nu := \nu_k$ – are given by:*

$$v_{spoof} = \frac{a}{Q} \left[2\rho w \mu^+ \frac{1-\bar{\iota}}{\bar{\iota}} \iota_{spoof}^2 - (Q\rho + \nu) \right]^+$$

where ι_{spoof} is the unique cubic root solution in $(0, \bar{\iota}]$ of

$$\frac{\bar{\iota}}{\iota} = 1 + \frac{(1-\bar{\iota})w}{Q} \left[2\rho w \mu^+ \frac{1-\bar{\iota}}{\bar{\iota}} \iota^2 - (Q\rho + \nu) \right]^+$$

Proof. Adopting the notations $Q := Q_k$, $\nu = \nu_k$, $w = w_k$, $H = \rho a$, the goal is to optimize over $v \geq 0$ the objective function

$$\begin{aligned} f(v) &= (1-Q) \frac{(\rho a)^2}{2a} + Q \frac{(\rho a + v)^2}{2a} + \rho a \mu^+ (2\iota(v) - 1) + v\nu \\ &= \frac{(\rho a)^2}{2a} + Q \frac{v^2}{2a} + (Q\rho + \nu)v + \rho a \mu^+ (2\iota(v) - 1) \end{aligned}$$

First order condition with Lagrangian λ yields

$$Q \frac{v}{a} + (Q\rho + \nu) - 2\rho w \mu^+ \frac{1-\bar{\iota}}{\bar{\iota}} \iota^2 = \lambda$$

where $\iota := \iota(v)$. Solving as a function of ι in $(0, 1)$, we get

$$\begin{aligned} \lambda(\iota) &= \left[(Q\rho + \nu) - 2\rho w \mu^+ \frac{1-\bar{\iota}}{\bar{\iota}} \iota^2 \right]^+ \\ v(\iota) &= \frac{a}{Q} \left[2\rho w \mu^+ \frac{1-\bar{\iota}}{\bar{\iota}} \iota^2 - (Q\rho + \nu) \right]^+ \end{aligned}$$

Given now the optimal $v(\iota)$ as a function of ι , we solve for ι such that

$$\frac{1}{\iota} = \frac{a + b + wv(\iota)}{b} = \frac{1}{\bar{\iota}} + w \frac{1-\bar{\iota}}{a\bar{\iota}} v(\iota) = \frac{1}{\bar{\iota}} + \frac{w}{Q} \frac{1-\bar{\iota}}{\bar{\iota}} \left[2\rho w \mu^+ \frac{1-\bar{\iota}}{\bar{\iota}} \iota^2 - (Q\rho + \nu) \right]^+$$

Since the left hand side is strictly decreasing on from ∞ to $\frac{1}{\bar{\iota}}$ on $(0, \bar{\iota}]$ and the right hand side is increasing from $\frac{1}{\bar{\iota}}$, on $(0, \bar{\iota}]$, there exists a unique solution which is a cubic root. \square

Though the solution is implicit, we can inspect the spoofing behavior as a function of the distance to the top of the limit order book. Note first that $\iota_{spoof} = \bar{\iota}$ if and only if $2\rho w \mu^+ (1-\bar{\iota})\bar{\iota} \leq Q\rho + \nu$ which results into $v_{spoof} = 0$. This coincide with the no spoofing condition of the previous proposition. We are interested on the relative spoofing size as a function of these parameters. From the definition of the imbalance, $\iota(v)$ increases if and only if v decreases, so we get more spoofing volume as ι_{spoof} gets smaller. Now from the implicit function it holds that

$$\frac{1}{\iota} = \frac{1}{\bar{\iota}} + \frac{w}{Q} \frac{1-\bar{\iota}}{\bar{\iota}} \left[2\rho w \mu^+ \frac{1-\bar{\iota}}{\bar{\iota}} \iota^2 - (Q\rho + \nu) \right]^+ =: f(w, \mu^+, \nu, Q, \bar{\iota}, \rho, \iota)$$

where the function f is an increasing function of ι greater than $1/\bar{\iota}$.

- Since f is increasing as a function of w and μ^+ , it follows that ι_{spoof} is decreasing as a function of w and μ^+ . Hence, spoofing behavior increases as a function of the impact w at level k on the imbalance as well as a function of the overall sensitivity μ^+ of the price movement with respect to the imbalance.
- Since f is decreasing as a function of Q and ν , it follows that ι_{spoof} is increasing as a function of Q and ν . From an empirical viewpoint, $Q = Q_k$ as well as $\nu = \nu_k$ decreases as a function of the depth k . It follows that spoofing behavior is more likely to happen and increase deeper in the limit order book. However, this conclusion is short of the fact that local sensitivity of the imbalance $w = w_k$ also depends on the depth with an inverse impact. According to empirical analysis, it

turns out that w does not exert this decreasing behavior as a function of k at least within a reasonable depth in the limit order book. It seems that spoofing behavior is more likely to happen at a reasonable distance from the top of the limit order book.

The behavior of the resulting imbalance ι_{spoof} as a function of the initial imbalance $\bar{\iota}$ is more difficult to stress out. We know that $\iota_{spoof} \leq \bar{\iota}$ and for the same reasons as before it is increasing as a function of $\bar{\iota}$. The same holds for the dependence on the relative number of shares to purchase ρ . Figure 2 represents the curves of the spoofed imbalance $\iota_{spoof}(\bar{\iota})$ as a function of the initial imbalance $\bar{\iota}$ for different depths with varying market parameters.

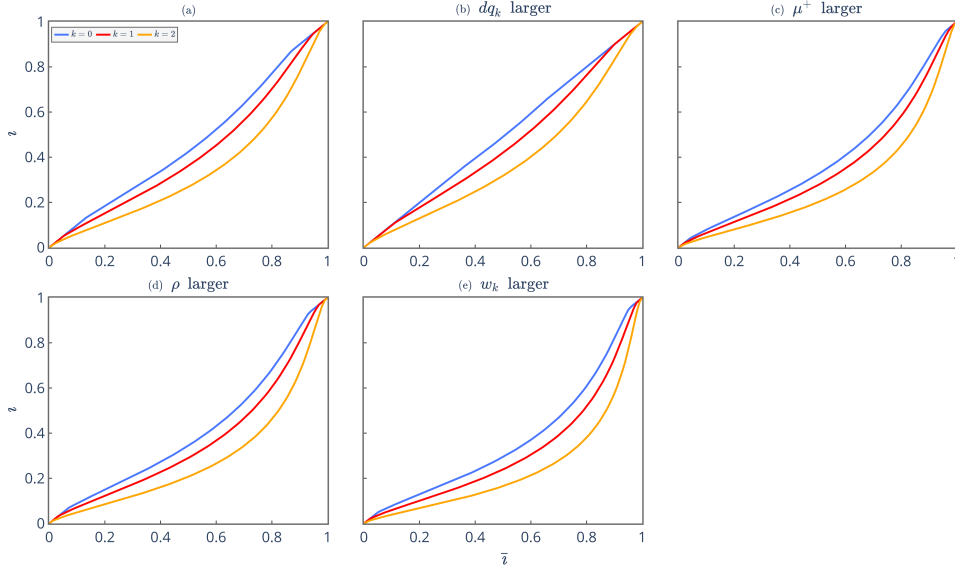


FIGURE 2. ι as a function of $\bar{\iota}$ and in (a) $\mu^+ = 1, \rho = 1, w_k = 0.2, dq_i = 0.003$ for all $i \geq k$. One parameter is increased each time with respect to (a) where (b): $dq_i = 0.006$ for all $i \geq k$; (c): $\mu^+ = 3$; (d): $\rho = 3$; (e): $w_k = 0.5$; Blue line : $k = 0$; red line: $k = 2$; orange line: $k = 4$.

3.3. Round Trip Situation. In this paper we mainly focus on the spoofing behavior from a market taker's viewpoint. As for a market maker, spoofing behavior might be rewarding as well. However, as seen in the following subsection, the rewards from spoofing are intertwined with the ones from pure market making.

We present a simple situation together with the numerical analysis in a blocked shape setting with the same model assumptions as before. We assume that the potential market maker spoofer acts as follows: At the first stage it decides to spoof with a volume v at depth k on the ask side to drive the price down and acquire an amount H of shares after this price movement. When the market comes back to its steady state, it liquidates H and eventually buys back v if it has been executed. We assume that v and H are decided at the very beginning.⁹

⁹This stylised situation makes strong assumptions and simplifications. First H is decided at time 0 even if it is executed after the price movement. This is to prevent conditional optimization. Second, the liquidation of the inventory H and v occurs separately. Once again, to provide simplified optimization problem, while we could numerically consider a liquidation of the net inventory $H - v$. Finally, a second spoofing could happen at the second stage as in the previous section to liquidate the inventory.

After spoofing a volume v at level k , as soon as the price moves the spoofer executes its market order H for a revenue of

$$-(H - v1_{\{y>k\}})(p + \delta\Delta + \delta(x + y)) - \delta G_a(H) - \delta(y - k)v1_{\{y>k\}}$$

where $G_a(H) = H^2/(2a)$, $p = (p^+ + p^-)/2$, and $\Delta = (p^+ - p^-)/(2\delta)$ is the effective spread in ticks. The spoofer then waits for the market to return to its steady state and liquidate the resulting inventory with market orders. For ease of computation, we assume that it executes two market orders: One for H and one for v if it has been executed¹⁰ for a revenue of

$$Hp - \delta\Delta H - G_b(H) - 1_{\{y>k\}}(v(p + \delta\Delta) + \delta G_a(v))$$

Adding both and integrating yields an average net revenue of

$$\begin{aligned} \frac{R(H, v)}{\delta} &= -H(2\Delta + \mu^+(2\nu_k(v) - 1)) - G_a(H) - G_b(H) \\ &\quad + Q_k[(k + \mu^+(2\nu_k(v) - 1))v - G_a(v)] \\ &= -H(2\Delta + \mu^+(2\nu_k(v) - 1)) - \frac{1}{a\bar{v}}\frac{H^2}{2} + Q_k v[k + \mu^+(2\nu_k(v) - 1)] - \frac{Q_k}{a}\frac{v^2}{2} \end{aligned}$$

From this equation, we can derive the following remarks concerning the decision of the spoofer:

- If $v = 0$: This corresponds to the classical situation where a market maker takes advantage of the temporary market movement to execute a market order and cash out at a later time when the market comes back to its steady state. Clearly, it gets a positive gain if and only if

$$\bar{v} \leq \frac{1}{2} - \frac{\Delta}{\mu^+}$$

In particular, if the effective spread Δ is large, or if μ^+ is small, then it is impossible or the initial imbalance should be very small. In the case where this happens, then \bar{H}^* is given by

$$\bar{H}^* = a\bar{v}(2\Delta + \mu^+(2\bar{v} - 1))^-$$

with corresponding revenue of

$$\bar{R}^* = \frac{1}{2} \left[a\bar{v}(2\Delta + \mu^+(2\bar{v} - 1))^- \right]^2$$

- If $H = 0$: This corresponds to the classical situation where a market maker posts limit orders at a given depth to gain from possible fluctuations. This results in corresponding average revenue given v of

$$\hat{R}(v) = Q_k v(k + \mu^+(2\nu_k(v) - 1)) - \frac{Q_k}{a}\frac{v^2}{2}$$

From this equation, even if the spoofer gets a positive gain of k ticks buy executing its order, it will drive the imbalance $\nu_k(v)$ below $1/2$ and face adverse price movement that will offset its gains. The optimal $\hat{v}^* = \hat{v}^*(\bar{v})$ in that situation is not explicit, but can be easily numerically implemented and corresponds to an optimal revenue of

$$\hat{R}^* = Q_k \hat{v}^*(k + \mu^+(2\nu_k(\hat{v}^*) - 1)) - \frac{Q_k}{a}\frac{\hat{v}^2}{2}$$

¹⁰Combining both in terms of $H - v1_{\{y>k\}}$ is cost effective but complicates the exposition of the result.

In general, solving for the optimal H is straightforward with

$$H^* = a\bar{v} (2\Delta + \mu^+ (2v_k(v) - 1))^-$$

and corresponding average revenue:

$$\frac{R(v)}{\delta} = \frac{1}{2} \left[a\bar{v} (2\Delta + \mu^+ (2v_k(v) - 1))^- \right]^2 + Q_k v (k + \mu^+ (2v_k(v) - 1)) - \frac{Q_k}{a} \frac{v^2}{2}$$

These two effects are difficult to disentangle from a truly spoofing gain when H as well as v are strictly positive. However, this can be done numerically and the results are presented in Figure 3, where the spoofing region $-H > 0$ as well as $v > 0$ is indicated. We also provide the plots of \bar{H}^* as well as \hat{v}^* against H^* and v^* .

We can however draw some stylised facts about the spoofing behavior from this market maker viewpoint. The impact of the different parameters – initial imbalance \bar{v} , probability of getting executed Q_k , local sensitivity of imbalance on the price impact $w = w_k$ as well as overall price deviation μ^+ are similar to the previous case. However, in addition to the previous part, the effective spread Δ acts negatively on the spoofing opportunity in that context. Indeed, a positive market order H is only triggered if $\mu^+ (2v^* - 1) \leq -2\Delta$, which requires a spoofed imbalance satisfying

$$v_{spoof} \leq -\frac{\Delta}{\mu^+} + \frac{1}{2}$$

If Δ is too large or μ^+ too low, a spoofing strategy is no longer rewarding.

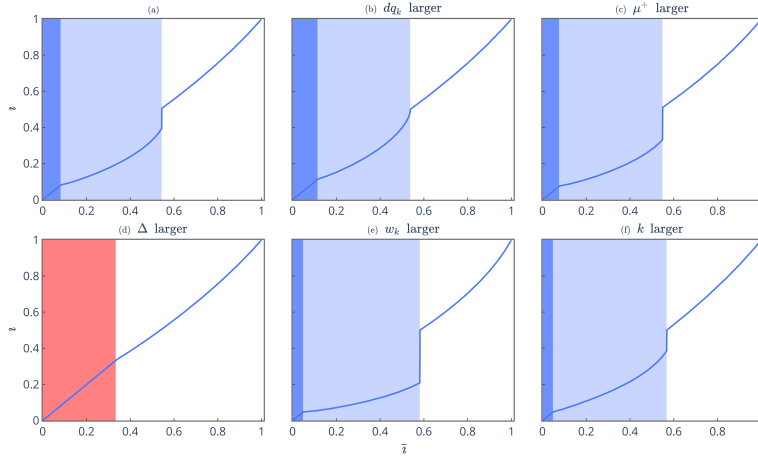


FIGURE 3. v_{spoof} as a function of \bar{v} and in (a) $\mu^+ = 3, \Delta = 0, k = 1, w_k = 0.1, dq_i = 0.001$ for all $i \geq k$. One parameter is increased each time with respect to (a) where (b): $dq_i = 0.002$ for all $i \geq k$; (c): $\mu^+ = 4$; (d): $\Delta = 2$; (e): $w_k = 0.4$; (f): $k = 4$. Red area: $v^* = 0, H^* = 0$; Dark blue area: $v^* = 0, H^* > 0$; Light blue area: $v^* > 0, H^* > 0$; White area: $v^* > 0, H^* = 0$.

4. CALIBRATION

According to this model, we calibrate the imbalance as well as dp and dq on real data provided by TMX. These datasets consists of level 2 data from June to September 2017. Among the 1500 equities we selected 9 varying in company background, market capitalization as well as trading frequency. The level 2 datasets include time, order price, volume, type — buy/sell; booked/cancelled/traded —, order ID and counterpart order ID in case of a trade, see Figure 4

time	side	price	order_id	reason	change
09:30:00.003886	Sell	62	S20170417000000020	Partial Fill	-100
09:30:00.042634	Sell	62.33	S20170417000000031	New Order	200
09:30:00.081079	Buy	61.66	B20170417000000017	Cancelled	-100

FIGURE 4. Original Level 2 dataset of stock AEM provided by TMX.

Since these are provided in diff form, therefore a cumulative aggregation allows to construct the full limit order book at any time as in Figure 5

time	side	price	change	order_id	reason	p_b	p_a	vb5	vb4	vb3	vb2	vb1	va1	va2	va3	va4	va5
09:31:01.279206	Sell	62.1	-100	S201704170000000117	Cancelled	61.95	62.1	100	200	100	200	100	600	500	100	100	200
09:31:01.475385	Sell	62.08	100	S201704170000000024	New Order	61.95	62.1	100	200	100	200	100	500	500	100	100	200
09:31:01.475612	Sell	62.08	100	S201704170000000021	New Order	61.95	62.08	100	200	100	200	100	100	0	500	500	100

FIGURE 5. Generation of the full limit order book out of the Level 2 data.

This operation is computationally heavy and therefore has been realized on a distributed data cluster of TMX with spark.

With the full limit order books at hand we divide the calibration into two steps:

- Find a normalized sample frequency and depth N ;
- Calibrate the imbalance generically, that is, as a function of the weights $w = \{w_k\}_{k=0,\dots,N}$;
- Estimate dq , dp^\pm and weights $w = \{w_k\}_{k=0,\dots,N}$.

The sample frequency should be large enough such that there exists enough variance in price change, see Figure 6

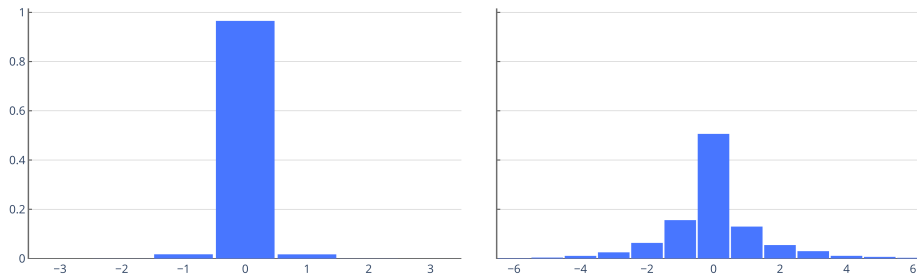


FIGURE 6. Left panel: Histogram of original AEM price change. Right panel: Histogram of AEM price change after sampling.

To compare across markets with different trading activity—and eventually time during the day—we fix a target variance of σ^2 for the price movement and select the optimal frequency f for each stock as to minimize the square distance between σ_f and σ . For a target variance $\sigma^2 = 2$, Table 1 is the sample frequency for different stocks with a summary

statistics of the average volume and arrival rate for Market/Limit Orders on the bid and ask side. As for the depth N , we take the 99% quantile of empirical sampled price change distribution.

Stock	f	N	Market Orders				Limit Orders			
			Buy		Sell		Buy		Sell	
			Vol	Rate	Vol	Rate	Vol	Rate	Vol	Rate
AEM	2	5	119	0.076	120	0.070	148	6.655	148	6.856
BMO	11	4	172	0.110	168	0.130	150	2.480	152	2.560
CNR	6	4	141	0.104	134	0.100	139	2.400	142	2.401
CPG	53	5	311	0.131	95	0.116	873	1.766	885	1.746
HVU	16	4	518	0.079	464	0.078	579	1.823	577	1.810
HXU	26	3	845	0.004	902	0.004	2498	0.905	2578	1.195
PPL	26	4	142	0.101	152	0.110	167	2.312	184	2.470
SSO	28	3	133	0.031	38	0.033	353	1.992	360	1.916
VUN	50	2	305	0.001	473	0.001	2561	0.809	2690	0.738
XEG	213	3	1422	0.020	1587	0.018	3573	1.854	3659	1.926

TABLE 1. Stock data from June 5, 2017 to June 9, 2017. The Vol columns corresponds to the average volume per seconds and the Rate columns corresponds to the number of orders per second.

With the sampling frequency f and depth N , we define the average imbalance at time t as

$$\hat{i}(w, t) = \frac{\sum_{k \leq N} \sum_{t-f \leq s < t} \bar{v}_k^-(s) w_k}{\sum_{k \leq N} \sum_{t-f \leq s < t} (\bar{v}_k^-(s) + \bar{v}_k^+(s)) w_k}$$

which sums up order book volumes within a certain time interval weighted by time difference $\Delta s_i = s_{i+1} - s_i$. The weighting parameter w impacts the average imbalance distribution, see Figure 7.

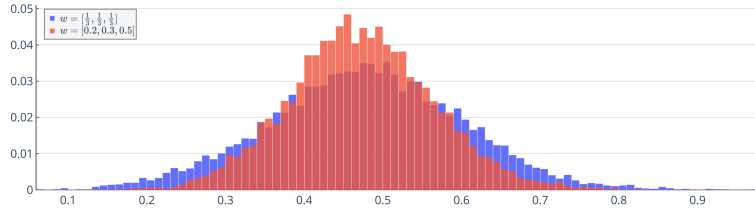


FIGURE 7. Distribution of the average imbalance for different weights for the BMO stock.

Nevertheless for each weight vector, the resulting distribution is close to a skewed normal distribution. For a given weight w , using maximum likelihood, we fit the empirical distribution to the corresponding skew normal distribution $\mathcal{SN}(\alpha(w), \xi(w), \omega(w))$, the fit of which is particularly good, see figure 8 for an example.

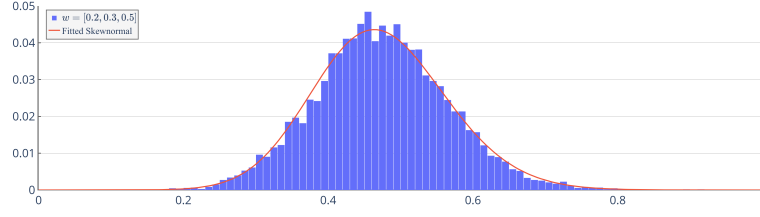


FIGURE 8. Histogram of BMO average imbalance and fitted skewnormal distribution: $w = [0.2, 0.2, 0.2, 0.2, 0.2]$.

The third step is to determine dq , dp^\pm with the optimal weights w . As for dq , it is the probability that the price moves by k ticks triggered by market orders. Thus for each market order, compute

$$F(H) = \inf\{x \in \mathbb{N}_0 : \sum_{k=0}^x v_k \geq H\}$$

where H is the volume of the market order. This represents exactly how many positive tick price deviation an order of size H will produce. We derive dq from the empirical distribution.

As for dp^\pm and w , a maximum likelihood estimation is implemented to solve

$$(4.1) \quad dp^*, w^* = \arg \min_{dp^+, w} \left[-\frac{1}{M} \sum_{m=1}^M \log p(x_m, \hat{i}_m, w) \right]$$

where x_m is the empirical price change, \hat{i}_m the average imbalance for a given weight w ,

$$p(x_m, \hat{i}_m) = dp_{x_m}(\hat{i}_m) p(\hat{i}_m)$$

where $dp_{x_m}(\hat{i}_m) = \hat{i}_m dp_{x_m}^+ + (1 - \hat{i}_m) dp_{x_m}^-$ represents the conditional probability of price change equal to x_m given \hat{i}_m , and $p(\hat{i}_m)$ is the density of the fitted skewnormal distribution for a weight w evaluated at \hat{i}_m .

Figure 9, illustrating the value of the optimal weight w for selected stocks, shows different patterns. Overall, it turns out that the relative impact of the imbalance to the price distribution is more important away from the top of the limit order book.

We also performed this calibration procedure on stock BMO weekly from June 5th to June 30th, as well as for the first hour of trading monthly from June to September. Figure 10 provide the optimal weights in each case for BMO.

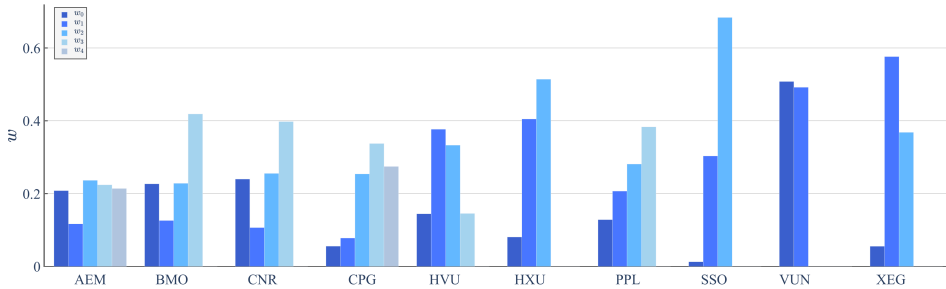


FIGURE 9. w for stock AEM, BMO, CNR, CPG, HVU, HXU, PPL, SSO, XEG from June 5th, 2017 to June 9th, 2017.

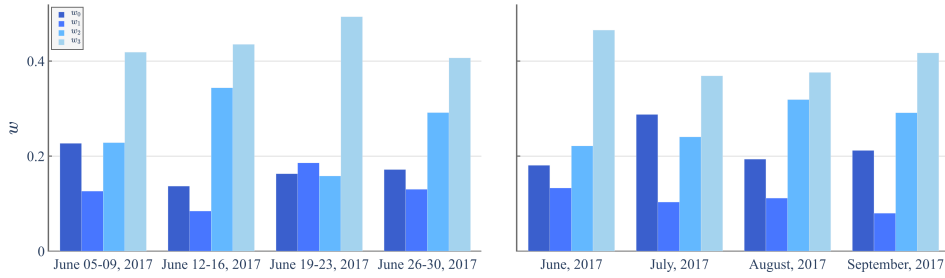


FIGURE 10. Left panel: w for stock BMO each week in June, 2017. Right panel: w for stock BMO each month in June, 2017 only using first hour trading data.

Notice that for the first hour of trading the optimal weights are more consistent across time, but all show that the weight impact on the price movement happens deeper in the limit order book.

As for the corresponding dp^+ and dq , they are represented in Figure 11 for stock BMO from June 5th, 2017 to June 9th, 2017. As expected, dp^+ , representing the price movement as the imbalance is large, is skewed to the right.

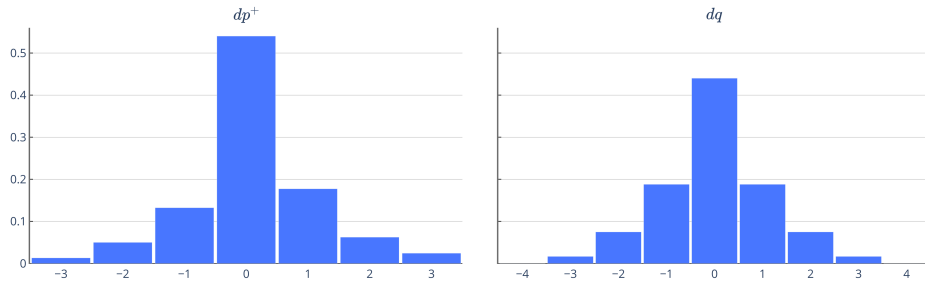


FIGURE 11. dp^+ and dq for BMO from June 5th to June 9th.

5. APPROACHES TO SPOOFING DETECTION

For reasons mentioned in the introduction, it is difficult from a regulatory viewpoint to figure out whether or not spoofing happened a-posteriori. According to the theoretical part, the act of spoofing will influence the resulting imbalance. However, to monitor the imbalance is akin to contemplate pure noise as shown in Figure 12.

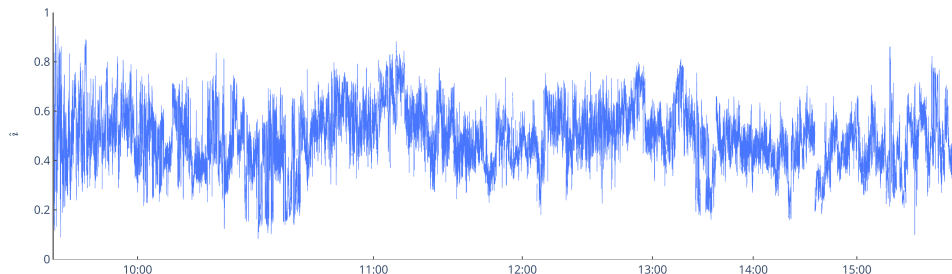


FIGURE 12. Imbalance of stock BMO from 09:30 to 16:00 on June, 7, 2017.

In the following, we propose some possible ways to perform such a monitoring based on the theoretical results. The strategy comes from the following observation: For a spoofing

strategy to be successfully fulfilled, a market order has to be executed.¹¹ Hence when observing an executed market order two situations may happen:

- 1- The market order is a legitimate one. In that case, the imbalance before this market order ι_- and after ι_+ should follow statistically the classical long run behavior. In other words, in a legitimate situation, we should observe statistically the pair

$$(\iota_-, \iota_+)$$

for each market order.

- 2- The market order is the result of a spoofing behavior. The implicit equilibrium without spoofing would be the pair (ι_-, ι_+) . After the market order is executed, the market imbalance should be back to its equilibrium ι_+ . However before the market order, the spoofer observes the implicit imbalance ι_- and decides to spoof according to this information, sending to the market $\iota_{spoof}(\iota_-)$ instead of ι_- . The resulting observation for those spoofed market orders is therefore the pair

$$(\iota_{spoof}(\iota_-), \iota_+)$$

Furthermore, spoofing strategies are supposed to happen sporadically but intensively within a short time horizon. Before presenting some strategies, let us fix some notations:

- $\Pi = \{t_1 < t_2 < \dots < t_M\}$ represents the time stamps of each (buy) market orders in a long sample.¹² For ease of notations, we relabel it by $\{1, \dots, M\}$.
- $\hat{\iota}_-(t)$ and $\hat{\iota}_+(t)$ represents the imbalance before and after the market order happening at time t in Π .
- (ι_-, ι_+) represents the joint distribution of the imbalance right before and after each market orders fitted to the overall data. We assume that these represents the stable behavior of the market without spoofing.
- $(\hat{\iota}_-^N(t), \hat{\iota}_+^N(t))$ represents the empirical distribution generated by the last $s = t, \dots, t - N + 1$ observed imbalances $(\hat{\iota}_-(s), \hat{\iota}_+(s))$, where $N \ll M$ is a short horizon sample size.
- The previous theoretical part, even if not explicit in terms of solution allows us to compute numerically $\iota_{spoof}(\iota_-)$ for a given implicit imbalance ι_- .

5.1. Monitoring $\hat{\iota}_-^N$. A first idea is to monitor the short behavior $\hat{\iota}_-^N$ as times passes to test if it is statistically different from the equilibrium ι_- . This is however not adequate for the following reasons. First, this is not related to spoofing behavior and might reflects some other market patterns. It is also not clear how to derive the magnitude of a potential spoofing behavior. Second, and more importantly, the sequence of $\hat{\iota}_-(t)$ for each market order is highly dependent. Indeed, there might exists market conditions – bullish/bearish, etc – such that a short horizon sample $\hat{\iota}_-^N$ differs strongly from the long term behavior. Figure 13 provides empirical evidence about the sequential dependence of the imbalance $\hat{\iota}_-$ as well as $\hat{\iota}_+$ over time.

¹¹In this paper, we do not consider spoofing strategies involving only limit orders.

¹²Several weeks, either for the full day of selected time of the day.

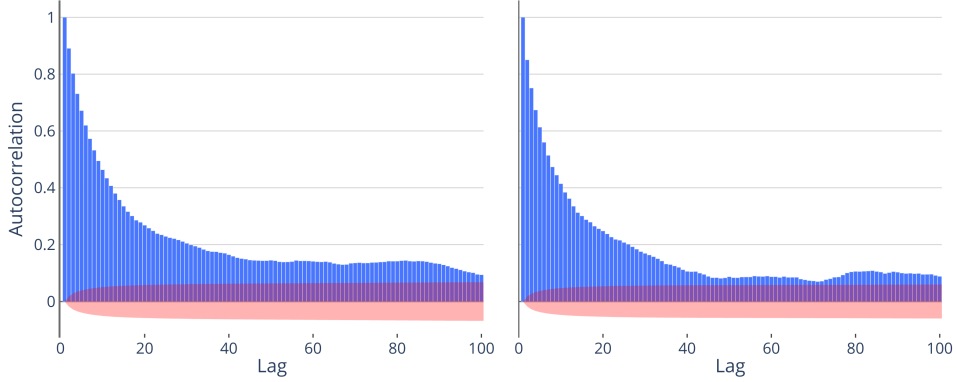


FIGURE 13. Left panel: \hat{i}_- autocorrelation of stock BMO on June 7, 2017. Right panel: \hat{i}_+ autocorrelation of stock BMO on June 7, 2017. Red area is the 95% confidence interval of the autocorrelation.

5.2. **Monitoring** (\hat{i}_-^N, \hat{i}_+^N). The statistical link towards spoofing is the additional observation of the imbalance after the spoofing happen. This provides statistical a-posteriori information about the implicit market equilibrium before spoofing which in case of spoofing can not be observed. Figure 14 shows on the left panel the joint distribution (i_-, i_+) while the right panel represents, based on the model of the theoretical part and calibration, the joint distribution $(i_{spoof}(i_-), i_+)$ would be in the case of spoofing. The spoofed joint distribution is skewed to the left in comparison to the non-spoofed one, in accordance to the theoretical analysis that spoofing decreases the imbalance – in the buy order case – before a market order.

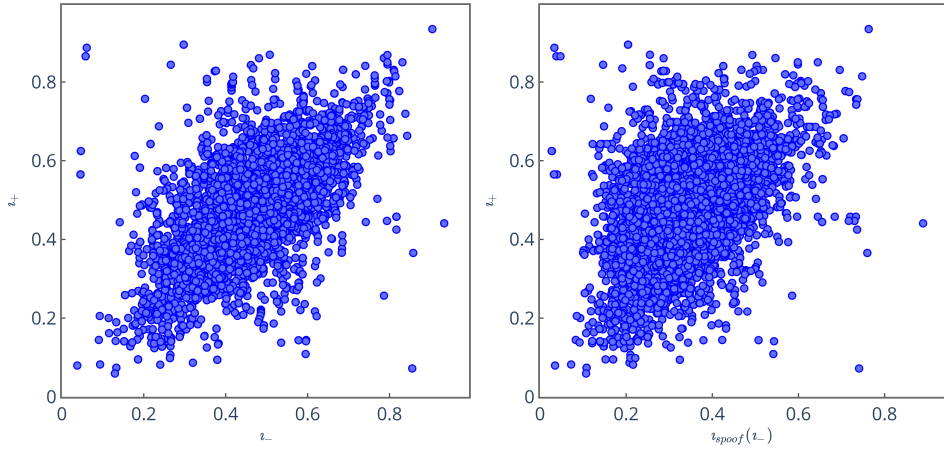


FIGURE 14. Left panel: Empirical joint distribution of (i_-, i_+) . Right panel: Joint distribution of $(i_{spoof}(i_-), i_+)$

A possible way to detect spoofing is therefore comparing the long run distribution (i_-, i_+) with the short term empirical distribution $(\hat{i}_-^N, \hat{i}_+^N)$. These two distributions encode the possibility to disentangle legitimate market behavior from spoofed ones. However, as in the previous approach, the sequence of joint observation is once again not iid as previously mentioned and pictured in Figure 13. For short time horizon, the market may be legitimate, though far away from the long run distribution.

5.3. **Monitoring \hat{i}_-^N conditioned on \hat{i}_+^N .** To overcome the previous problem the next approach is to monitor \hat{i}_-^N conditioned on the current \hat{i}_+^N . From our hypothesis, v_+ represents the current steady state of the market at equilibrium. It turns out that conditioned on $\hat{i}_+(t)$ the sequence of $\hat{i}_-(t)$ is closer to iid.

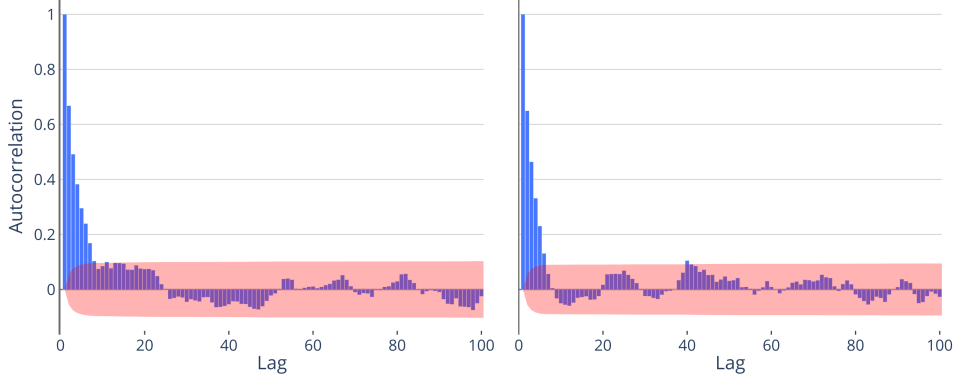


FIGURE 15. Left panel: Autocorrelation of $\hat{i}_-^N | \hat{i}_+^N \in [0.329, 0.378]$ of stock BMO on June 7, 2017. Right panel: Autocorrelation of $\hat{i}_-^N | \hat{i}_+^N \in [0.561, 0.61]$ of stock BMO on June 7, 2017. Red area is the 95% confidence interval of the autocorrelation.

In order to detect spoofing behavior, instead of adopting a statistical test for which some parametric assumptions on the distribution has to be made, we measure the distance between v_- and \hat{i}_-^N conditioned on the current observed imbalance \hat{i}_+^N using a conditional 2-Wasserstein distance.

For two distributions μ and ν the 2-Wasserstein distance is defined as

$$\begin{aligned} W_2(\mu, \nu) &= \left(\inf \left\{ \int (x - y)^2 \pi(dx, dy) : \pi_1 \sim \mu, \pi_2 \sim \nu \right\} \right)^{1/2} \\ &= \left(\int_0^1 (q_\mu(\alpha) - q_\nu(\alpha))^2 d\alpha \right)^{1/2} \end{aligned}$$

From a generic perspective, the conditional distance we consider is as follows: If we assume that

$$(v_-, v_+) \sim K(y, dx) \otimes \mu(dy)$$

where $\mu \sim v_+$ and $K(y, \cdot) \sim v_- | v_+ = y$, it follows that

$$(v_{spoof}(v_-), v_+) \sim K_{spoof}(y, dx) \otimes \mu(dy) \quad \text{where} \quad K_{spoof}(y, \cdot) = K(y, \cdot) \circ v_{spoof}^{-1}$$

Hence given $v_+ = y$, we have

$$W_2(K(y, \cdot), K_{spoof}(y, \cdot)) = \left(\int_0^1 (q_{K(y, \cdot)}(\alpha) - v_{spoof}(q_{K(y, \cdot)}(\alpha)))^2 d\alpha \right)^{1/2}$$

Heuristically we wish to monitor the following two quantities

$$\underbrace{W_2(v_-, \hat{i}_-^N | \hat{i}_+^N)}_{\substack{\text{Distance from the short term imbalance } \hat{i}_-^N \\ \text{to the } \mathbf{equilibrium} \text{ imbalance } v_- \\ \text{given that } v_+ \sim v_+^N}} \quad \text{and} \quad \underbrace{W_2(v_{spoof}(v_-), \hat{i}_-^N | \hat{i}_+^N)}_{\substack{\text{Distance from the short term imbalance } \hat{i}_-^N \\ \text{to the } \mathbf{spoofed} \text{ imbalance } v_{spoof} \\ \text{given that } v_+ \sim v_+^N}}$$

From the data, we can calibrate the joint distribution (ι_-, ι_+) as well as $(\iota_{spooof}(\iota_-), \iota_+)$.¹³ Hence, we have a parametrization of $K(y, dx)$ and $K_{spooof}(y, dx)$ for every y . However for each value $\hat{\iota}_+(l)$ from the discrete distribution $\hat{\iota}_+^N$ we only have a single sample point $\hat{\iota}_-(l)$ at hand. In order to overcome this problem we will bucket the values of $\hat{\iota}_+(l)$ in the sample of $\hat{\iota}_+^N$ to get a Kernel approximation of $\hat{\iota}_-^N$.

The monitoring strategy at a given time t in Π is given as follows

1 - Consider the discrete short term joint distribution $(\hat{\iota}_-^N(t), \hat{\iota}_+^N(t))$ given by the sample

$$(\iota_-(s), \iota_+(s)), \quad s = t, \dots, k - N + 1$$

of the last N pairs of imbalances before time t .

2 - We define the following L buckets of equal cardinality N/L

$$J_l = \left\{ s : s = t, \dots, t - N + 1, q_{\hat{\iota}_+^N} \left(\frac{l-1}{L} \right) \leq \iota_+(s) < q_{\hat{\iota}_+^N} \left(\frac{l}{L} \right) \right\}, \quad l = 1, \dots, L$$

as well as the mid point of each

$$\iota_l = \frac{L}{N} \sum_{s \in J_l} \iota_+(s)$$

3 - For each l , we generate a random sample $\iota_-^{N,L}$ and $\iota_{spooof}^{N,L}$ of N/L points each drawn from $K(\iota_l, \cdot)$ and $K_{spooof}(\iota_l, \cdot)$, respectively.

4 - For each l we compute the Wasserstein distances

$$W_2 \left(\iota_-^{N,L}, \hat{\iota}_-^{N,L} \right) \quad \text{and} \quad W_2 \left(\iota_{spooof}^{N,L}, \hat{\iota}_-^{N,L} \right)$$

where $\hat{\iota}_-^{N,L}$ is the discrete distribution out of the sample $\hat{\iota}_-(s)$ for s in J_l . This is an approximation for the Wasserstein distance

$$W_2 \left(\iota_-, \hat{\iota}_-^N \mid \hat{\iota}_+^N \right) \approx \iota_l \quad \text{and} \quad W_2 \left(\iota_{spooof}, \hat{\iota}_-^N \mid \hat{\iota}_+^N \right) \approx \iota_l$$

5 - we aggregate all together and define the indicators

$$S \left(\iota_-, \hat{\iota}_-^N \mid \hat{\iota}_+^N \right) := \frac{1}{L} \sum_{l=1}^L W_2 \left(\iota_-^{N,L}, \hat{\iota}_-^{N,L} \right)$$

$$S \left(\iota_{spooof}, \hat{\iota}_-^N \mid \hat{\iota}_+^N \right) := \frac{1}{L} \sum_{l=1}^L W_2 \left(\iota_{spooof}^{N,L}, \hat{\iota}_-^{N,L} \right)$$

Remark 5.1. *To enhance the accuracy of this indicator, we run step 3 to 5 a couple of times with different samples and average again.*

¹³The former fits well with a joint normal distribution, while the second one with a skewed normal distribution, see Figure 14. Other parametrization could eventually be used too.

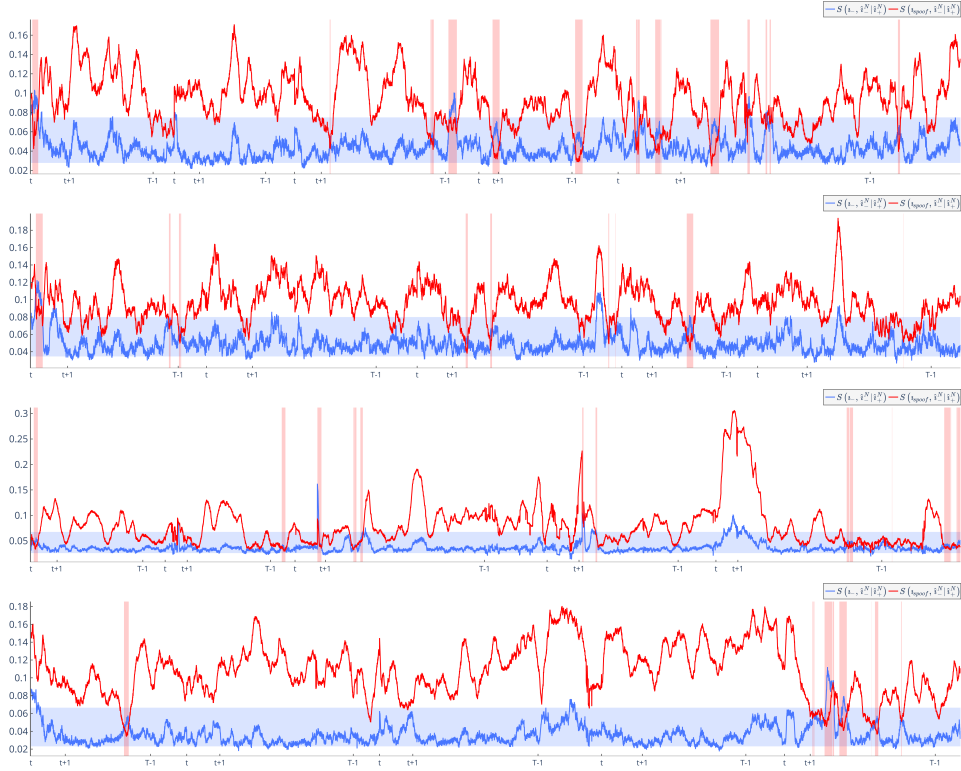


FIGURE 16. $S(\iota_{spoofer}, \hat{i}_+^N | \hat{i}_+^N)$ from June 5, 2017 to June 9, 2017. From top to bottom are stock BMO, CNR, CPG, PPL respectively. Blue area is the 95% confidence interval of $S(\iota_-, \hat{i}_+^N | \hat{i}_+^N)$. Red area is where $S(\iota_{spoofer}, \hat{i}_+^N | \hat{i}_+^N) \leq S(\iota_-, \hat{i}_+^N | \hat{i}_+^N)$.

6. CONCLUSION

In this paper we address the question of assessing quantitatively eventual spoofing behavior in high frequency trading. In a one period setting we present how a spoofing strategy from a market taker or maker is designed by manipulating the imbalance at different depth level to impact the subsequent price movement. We provide and discuss the conditions for the market to allow for spoofing manipulations. We subsequently solve the optimization problem from a spoofer perspective and derive/discuss the resulting imbalance after spoofing as a function of the market parameters. We calibrate the weighted imbalance and price movement impact to Level 2 data provided by TMX. Using these results we propose a quantification instrument to monitor in real time eventual spoofing behavior on the market using a conditional Wasserstein distance. We illustrate these results on the data provided by TMX.

This approach is by no means a definitive answer to spoofing detection but rather a first take on. The dynamic structure of the limit order book and strategy, the memory dependence of the parameters over time, as well as the specificities of one market with respect to another one are left to further study. Furthermore, there might be alternative approaches subject to new research directions – monitoring arrival rates of orders, frequency of book/cancelling, etc. – that could complement such a monitoring approach.

7. CREDITS

We thank Fields Institute for the organization of the series ‘‘Fields-China Joint Industrial Problem Solving Workshop’’ through which we got involved into this problem. Thanks to TMX and their data analysis team for providing unique access to large datasets, high performance computing facilities as well as precious insights concerning high frequency trading. Finally, the data analysis could not have been performed without the outstanding work and dedication of the open source community for the development of scientific computing/visualisation libraries/platforms such as NumPy [13], SciPy [23], Mystic [19], Pandas [20, 25], Apache Spark, Plotly to name the most relevant for this work.

APPENDIX A. APPENDIX 1

This section reveals some details in the procedure of calibration and spoofing detection.

A.1. Sampling Frequency f . In real market there exists a large amount of orders which are posted ‘‘far away’’ from the best bid/ask price or cancelled right after booked. These orders have little impact on the price so if the best bid/ask price change is computed as the best bid/ask price difference between 2 consecutive orders, more than 96% of time the price does not move, shown in the left panel of Figure 6. Instead, we extract orders every certain seconds and compute price changes in this ‘‘new’’ dataframe. The sample frequency f is the smallest number of seconds where the variance of sampled price changes σ_f^2 is the closest to a given benchmark σ^2 ,

$$\arg \min_{f>0} |\sigma^2 - \sigma_f^2|$$

When $f = 2s$, stock AEM which is one of the most active stocks on the TSX has a wider price change distribution and its variance is approximately equal to 2. So taking $\sigma^2 = 2$, we conduct tests for different stocks and do see improvements of price change distribution shown in the right panel of Figure 6.

A.2. Computation of $\iota_{spoo f}$. For $t \in \Pi$, $\hat{i}_-(t)$ and $\hat{i}_+(t)$ are computed as

$$\hat{i}_-(t) = \frac{\sum_{k \leq N} \sum_{t-f \leq s < t-0.01} w_k v_k^-(s)}{\sum_{k \leq N} \sum_{t-f \leq s < t-0.01} w_k (v_k^-(s) + v_k^+(s))}$$

$$\hat{i}_+(t) = \frac{\sum_{k \leq N} \sum_{t-0.01 \leq s < t+1} w_k v_k^-(s)}{\sum_{k \leq N} \sum_{t-0.01 \leq s < t+1} w_k (v_k^-(s) + v_k^+(s))}$$

For each $\hat{i}_-(t)$, the optimal spoofing strategy v^* can be solved explicitly from

$$v_k^* = 1 + \frac{(1 - \hat{i}_-(t)) w_k}{Q} \left[2\rho_t w_k \mu^+ + \frac{1 - \hat{i}_-(t)}{\hat{i}_-(t)} \iota^2 - (Q_k \rho_t + \nu) \right]^+$$

where $\iota = \frac{b_t}{b_t + a_t + w_k v_k^*}$, $\rho_t = \frac{H_t}{a_t}$ and H_t is taken as the size of the market order,

$$a_t = \frac{\sum_{k=1}^N \sum_{t-f \leq s < t-0.01} \hat{v}_k^+(s) \Delta s}{Nf}$$

which is the average size of the limit order book between f seconds and 10 milliseconds before a market order. In the same way, we define b_t and

$$\iota_{spoo f}(\hat{i}_-(t)) = \frac{b_t}{b_t + a_t + \sum_k w_k v_k^*}$$

A.3. $q_{\hat{i}_+^N}$. The short horizon sample size N is equal to 100. $(\hat{i}_-^N(t), \hat{i}_+^N(t))$ represents the empirical distribution generated by the last $s = t, \dots, t - N + 1$ observed imbalances $(\hat{i}_-(s), \hat{i}_+(s))$, where $N \ll M$ is a short horizon sample size. For $\hat{i}_-^N(t)$ we bucket $\hat{i}_+^N(t)$ to compute the wassterstein distance where $q_{\hat{i}_+^N}$ is listed in the table below.

l	0	1	2	3	4	5
$q_{\hat{i}_+^N}(\frac{l}{L})$	0.082	0.403	0.512	0.637	0.761	0.934
v_l		0.301	0.473	0.584	0.688	0.822

TABLE 2. $q_{\hat{i}_+^N}(l/L)$ and v_l for $l = 0, \dots, 5$ for the stock BMO at 09:30:07.375 on June 7th, 2017.

REFERENCES

- [1] A. Alfonsi and A. Schied. Optimal trade execution and absence of price manipulations in limit order book models. *SSRN Electronic Journal*, 02 2010.
- [2] A. Alfonsi, A. Fruth, and A. Schied. Optimal execution strategies in limit order books with general shape functions. *Quantitative Finance*, 10, 08 2007.
- [3] F. Allen and D. Gale. Stock-price manipulation. *Review of Financial Studies*, 5: 503–29, 02 1992.
- [4] K. Bechler and M. Ludkovski. Order flows and limit order book resiliency on the meso-scale. *Market Microstructure and Liquidity*, 3(03n04):1850006, 2017.
- [5] Y. Cao, Y. Li, S. Coleman, A. Belatreche, and T. M. McGinnity. Detecting price manipulation in the financial market. In *2014 IEEE Conference on Computational Intelligence for Financial Engineering Economics (CIFER)*, pages 77–84, 2014.
- [6] Y. Cao, Y. Li, S. Coleman, A. Belatreche, and T. M. McGinnity. Adaptive hidden markov model with anomaly states for price manipulation detection. *IEEE Transactions on Neural Networks and Learning Systems*, 26(2):318–330, 2015.
- [7] Á. Cartea, R. Donnelly, and S. Jaimungal. Enhancing trading strategies with order book signals. *Applied Mathematical Finance*, 25(1):1–35, 2018.
- [8] A. Cartea, S. Jaimungal, and Y. Wang. Spoofing and price manipulation in order-driven markets. *Applied Mathematical Finance*, pages 1–32, 02 2020. doi: 10.1080/1350486X.2020.1726783.
- [9] R. Cont, A. Kukanov, and S. Stoikov. The price impact of order book events. *Journal of Financial Econometrics*, 12(1):47–88, 2014.
- [10] J. Gatheral. No-dynamic-arbitrage and market impact. *Quantitative Finance*, 10(7): 749–759, 2010.
- [11] J. Gatheral and A. Schied. Dynamical models of market impact and algorithms for order execution. *SSRN Electronic Journal*, 01 2013.
- [12] M. Gould and J. Bonart. Queue imbalance as a one-tick-ahead price predictor in a limit order book. *Market Microstructure and Liquidity*, 12 2015.
- [13] C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. F. del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, and T. E. Oliphant. Array programming with NumPy. *Nature*, 585(7825):357–362, September 2020. ISSN 1476-4687.
- [14] O. Hart and D. Kreps. Price destabilizing speculation. *Journal of Political Economy*, 94(5):927–52, 1986.

- [15] R. Jarrow. Market manipulation, bubbles, corners, and short squeezes. *Journal of Financial and Quantitative Analysis*, 27:311–336, 09 1992.
- [16] E. J. Lee, K. S. Eom, and K. S. Park. Microstructure-based manipulation: Strategic behavior and performance of spoofing traders. *Journal of Financial Markets*, 16(2): 227–252, 2013.
- [17] A. Lipton, U. Pesavento, and M. Sotiropoulos. Trade arrival dynamics and quote imbalance in a limit order book. *Preprint*, 12 2013.
- [18] E. M. Miranda, P. McBurney, and M. J. Howard. Learning unfair trading: A market manipulation analysis from the reinforcement learning perspective. *2016 IEEE Conference on Evolving and Adaptive Intelligent Systems (EAIS)*, pages 103–109, 2016.
- [19] T. S. A. F. M. A. M.M. McKerns, L. Strand. Building a framework for predictive science. *Proceedings of the 10th Python in Science Conference, 2011*, 2011.
- [20] J. Reback, W. McKinney, jbrockmendel, J. V. den Bossche, T. Augspurger, P. Cloud, gfyong, Sinhrks, A. Klein, M. Roeschke, S. Hawkins, J. Tratner, C. She, W. Ayd, T. Petersen, M. Garcia, J. Schendel, A. Hayden, MomIsBestFriend, V. J. L. Rechenzentrum, P. Battiston, S. Seabold, chris b1, h vetinari, S. Hoyer, W. Overmeire, alimcmaster1, K. Dong, C. Whelan, and M. Mehyar. *pandas-dev/pandas: Pandas*. Software, Feb. 2020.
- [21] G. Shorter and R. Miller. High-frequency trading: Background, concerns, and regulatory developments. Technical report, Congressional Research Service, 04 2015.
- [22] J. A. Sirignano. Deep learning for limit order books. *Quantitative Finance*, 19(4): 549–570, 2019.
- [23] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, Í. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, A. Vijaykumar, A. P. Bardelli, A. Rothberg, A. Hilboll, A. Kloeckner, A. Scopatz, A. Lee, A. Rokem, C. N. Woods, C. Fulton, C. Masson, C. Häggström, C. Fitzgerald, D. A. Nicholson, D. R. Hagen, D. V. Pasechnik, E. Olivetti, E. Martin, E. Wieser, F. Silva, F. Lenders, F. Wilhelm, G. Young, G. A. Price, G.-L. Ingold, G. E. Allen, G. R. Lee, H. Audren, I. Probst, J. P. Dietrich, J. Silterra, J. T. Webber, J. Slavič, J. Nothman, J. Buchner, J. Kulick, J. L. Schönberger, J. V. de Miranda Cardoso, J. Reimer, J. Harrington, J. L. C. Rodríguez, J. Nunez-Iglesias, J. Kuczynski, K. Tritz, M. Thoma, M. Newville, M. Kümmerer, M. Bolingbroke, M. Tartre, M. Pak, N. J. Smith, N. Nowaczyk, N. Shebanov, O. Pavlyk, P. A. Brodtkorb, P. Lee, R. T. McGibbon, R. Feldbauer, S. Lewis, S. Tygier, S. Sievert, S. Vigna, S. Peterson, S. More, T. Pudlik, T. Oshima, T. J. Pingel, T. P. Robitaille, T. Spura, T. R. Jones, T. Cera, T. Leslie, T. Zito, T. Krauss, U. Upadhyay, Y. O. Halchenko, Y. Vázquez-Baeza, and S. . Contributors. *SciPy 1.0: fundamental algorithms for scientific computing in Python*. *Nature Methods*, 17(3):261–272, March 2020. ISSN 1548-7105.
- [24] Y. Wang. Strategic spoofing order trading by different types of investors in the futures markets. *Wall Street Journal*, 2015.
- [25] Wes McKinney. Data Structures for Statistical Computing in Python. In Stéfan van der Walt and Jarrod Millman, editors, *Proceedings of the 9th Python in Science Conference*, pages 56 – 61, 2010. doi: 10.25080/Majora-92bf1922-00a.
- [26] K. Xu, M. Gould, and S. Howison. Multi-level order-flow imbalance in a limit order book. *Preprint*, 07 2019.

SHANGHAI JIAO TONG UNIVERSITY, SHANGHAI, CHINA
E-mail address: taoxuan@sjtu.edu.cn

WESTERN UNIVERSITY, CANADA
E-mail address: aday46@uwo.ca

SHANGHAI JIAO TONG UNIVERSITY (NOW AT PING AN TECHNOLOGY), SHANGHAI, CHINA
E-mail address: lan.ling@sjtu.edu.cn
E-mail address: linglan987@pingan.com.cn

SHANGHAI JIAO TONG UNIVERSITY, SHANGHAI, CHINA
E-mail address: sdrapeau@saif.sjtu.edu.cn
URL: <http://www.samuel-drapeau.info>