TAMPEREEN TEKNILLINEN YLIOPISTO
TAMPERE UNIVERSITY OF TECHNOLOGY

JANNE RAJAMÄKI

DETECTING POTENTIAL SPOOFING BEHAVIOUR

Master of Science Thesis

Examiner: Professor Kari Systä
Examiner and topic approved by the
Faculty Council of the Faculty of
Computing and Electrical Engineering
on Day Month Year

ABSTRACT

TAMPERE UNIVERSITY OF TECHNOLOGY
Master's Degree Programme in Information Technology
JANNE RAJAMÄKI: Detecting potential spoofing behaviour
Master of Science Thesis, 16 pages
April 2018
Major: Software engineering
Examiner: Professor Kari Systä
Keywords: keywords

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. $\sin^2(\alpha) + \cos^2(\beta) = 1$. If you read this text, you will get no information $E = mc^2$. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. $\sqrt[n]{a} \cdot \sqrt[n]{b} = \sqrt[n]{ab}$. This text should contain all letters of the alphabet and it should be written in of the original language. $\frac{\sqrt[n]{a}}{\sqrt[n]{b}} = \sqrt[n]{\frac{a}{b}}$. There is no need for special content, but the length of words should match the language. $a\sqrt[n]{b} = \sqrt[n]{a^n b}$.

TIIVISTELMÄ

TAMPEREEN TEKNILLINEN YLIOPISTO
Tietotekniikan DI-tutkinto-ohjelma
JANNE RAJAMÄKI: Potentiaalisten valekauppojen tunnistaminen
Diplomityö, 16 sivua
Huhtikuu 2018
Pääaine: Ohjelmistotuotanto
Tarkastaja: Professori Kari Systä
Avainsanat: avainsanat

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. $\sin^2(\alpha) + \cos^2(\beta) = 1$. If you read this text, you will get no information $E = mc^2$. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. $\sqrt[n]{a} \cdot \sqrt[n]{b} = \sqrt[n]{ab}$. This text should contain all letters of the alphabet and it should be written in of the original language. $\frac{\sqrt[n]{a}}{\sqrt[n]{b}} = \sqrt[n]{\frac{a}{b}}$. There is no need for special content, but the length of words should match the language. $a\sqrt[n]{b} = \sqrt[n]{a^n b}$. finnish

CONTENTS

1. INTRODUCTION

TODO: Will be written later, about one to two page

1.1 Thesis statement

TODO: More details about research goals and needs proper conclusion

The main objective of this thesis is to study spoof trading in Northern European stock exchanges. This research attempts to explain phenomena of spoof trading itself, market conditions in which manipulator is likely to commence spoof trading and examine differences of the orders that are utilised in spoof trading compared to regular stock market trading.

In spoof trading trader intentionally submits deceptive trade order into exchange in order to falsify true demand or supply of a security. The categorization of an order as genuine or artificial beyond no doubt via data analytical means is difficult, since in isolation any kind of trade is permitted, even if the trade does not conform to common trading behaviour or is even harmful to the trader. As a result the assessment of order's genuineness through quantitative analysis is always at the mercy of an inspector.

For above reason, estimate about existence of spoof trading is at best vague and it is not possible with data analytical means to truly confirm spoof trading phenomena. Hence, it is necessary to study logical connections that are interconnected with spoof trading. In spoof trading the most defining aspect is the link between trade imbalance and price changes, as manipulator attempts to influence price of a security through deceptive orders.

This thesis will firstly explore how trade imbalances do affect security prices. More precisely whether increased demand of an asset will lead to rise in the price, and conversely increase in supply of an asset will cause price to decrease. This behaviour would be in accordance with the laws of supply and demand, but the difference is in the essence of buy-sell-imbalance of an asset.

In spoof trading this imbalance is artificial and won't be corrected until cancellation of the spoof order, while in normal trading imbalances should fluctuate more smoothly. Then by studying the nature of these trade imbalances, it is potentially possible to determine whether spoof trading occurs in Northern European stock exchanges.

In case spoof trading was detected, the last step is to study other characteristics of the spoof orders and the market circumstances in which spoof trading is likely to occur.

In conclusion this thesis pursues following research goals:

1. Study whether trade imbalances lead to price changes.
2. Attempt to identify spoof trading and confirm its' existence in Northern European stock market.
3. Study how spoof trading differs from regular trading.
4. Describe the market situations in which spoof trading is likely to happen.

1.2 Literature review

TODO: Needs new ending, and more (technical) sources in comments

Allen and Gale (1992) initially argued that under asymmetrical buy and sell liquidity of a stock, buy and sale orders lead to different price responses, which can create opportunities for profitable price manipulation. They developed theoretical framework to prove that an uninformed trader is able profit by commencing in transaction-based stock price manipulation. Uninformed trader does not possess any insider information about the asset. Transaction-based stock manipulation is based on influencing stock prices only through orders without doing any external actions such as misinforming other traders.

Jarrow (1992) studied potential market manipulation strategies by large traders, showing large traders, even if they do not possess insider information, have enough market power to be able to manipulate market prices through their sheer size of orders. Allen and Gale in addition to Jarrow's research proved theoretical existence of spoof trading, enabling researchers to study the phenomena in real market.

Numerous empirical studies have shown market manipulation is worldwide phenomena: Aggarwal and Wu (2006), Aktaş and Doğanay (2006, see Ögüt et al. 2009), Lee at al. (2009), and Kong and Wang (2014) showed evidence of spoof order manipulation in US, Turkey, Korea, and China respectively.

Aggarwal and Wu (2006) extended framework developed by Allen and Gale and studied how information seekers affect trading. They found that informed parties, corporate insiders, brokers, underwriters, large shareholders and market makers, may worse market efficiency and are more likely to be manipulators. When studying U.S equity market, they found that manipulating mostly happens in small, illiquid and inefficient markets.

Lee at al (2009) studied buy-orders in KRX Korea Exchange. They found the susceptible stocks to spoofing behaviour tend to have higher return volatility, lower market capitalization, lower price levels and lower managerial transparency. Also when exchange stopped disclosing the total quantity of buy and sell sides, so that spoofing orders could no longer be used to mislead investors, spoofing dramatically declined. In general spoofing orders were much bigger than typical orders and usually at least 10 ticks away from the current price, thus executability of spoofing orders were extremely low, consequently confirming the intent of manipulation. (Lee et al. 2009)

However in the study of Lee et al. spoofing order was defined as "*a bid/ask with a size at least twice the previous day's average order size and with an order price at least 6 ticks away from the market price, followed by an order on the opposite side of the market, and subsequently followed by the withdrawal of the first order*" (Lee et al. 2009, p. 11). This kind of precise definition of spoofing order surely affects the results they will gather from their analysis.

Kong and Wang (2014) found that market statistics of manipulation cases, like price, volume, turnover ratio volatility and liquidity, were in accordance with previous studies. Also market manipulators tend to target specific time periods of greater information asymmetry, then manipulators are able to present themselves as informed traders and affect behaviour of other investors in the short-term, while in the long-term the effects of manipulation disappear. (Kong, Wang 2014)

In the previously mentioned studies data was labelled in advance or labelled using fixed rules as manipulative or non-manipulative. Lee et al. used rule-based approach, other studies used court rulings to define manipulative trades. This inherently causes the subjects of studies were only incompetent manipulators who were caught or classified using arbitrary manual rules. None of the studies adopted pure data analytical viewpoint of detecting spoofing behaviour in stock-wide context, which I will study in my thesis.

To date most of the spoofing behaviour research only concerns study of spoofing orders. The orders are deterministically identified either as normal stock market behaviour or spoofing behaviour, then spoofing orders' characteristics and impact is further studied. Mostly used identification methods are technical rule-based methods or based on court rulings orders are identified as illegal spoofing behaviour. This however limits the study of

1.3 Data

TODO: Provide more details about data

This thesis utilises stock market data in Northern European exchanges to study spoofing behaviour. The data was provided by thesis director Professor Juho Kanninen from department of industrial management in Tampere University of Technology. The data of each stock was collected from 1.6.2010 to 31.5.2013, which results to the total of 765 trading days worth of data per stock. **QUESTION:** Is this date format correct?

For each stock data consists of daily message book and order book. Message books includes all the orders of a stock submitted to the exchange and order book data describes changes in order book during the day. In addition daily summaries of successful trades, submitted limit orders, and cancellations made by traders are provided.

Table 1.1 *Stock information (Nasdaq Nordic 2017).*

Name	ISIN	Ticker	Market exchange
Atlas Copco A	SE0006886750[1]	ATCO A	OMX Stockholm
Metso Oyj	FI0009007835	METSO	OMX Helsinki
Nokian Renkaat Oyj	FI0009005318	NRE1V	OMX Helsinki
Vestas Wind Systems	DK0010268606	VWS	OMX Copenhagen
Volvo B	SE0000115446	VOLV B	OMX Stockholm

.

Table 1.1 introduces stock market information about companies researched in this thesis. Two of the companies are listed in OMX Nordic Helsinki, two in OMX Nordic Stockholm and one is listed in OMX Nordic Copenhagen exchanges, hence all the biggest northern European stock exchanges are covered.

However, within the exchanges sample size is very small. By November 2017 in total 320 companies were listed in OMX Nordic Stockholm, 53 companies in OMX Nordic Helsinki, and 137 companies in OMX Nordic Copenhagen (Nasdaq GlobeNewswire 2018a). For example in OMX Nordic Helsinki, the exchange with fewest listed stocks

[1] At the time of data gathering ISIN was SE0000101032, but due to stock split Atlas Copco A was assigned new ISIN (Nasdaq GlobeNewswire 2018b)

of mentioned three exchanges, two companies out of 53 is only coverage of few percentage points. Due to small sample size the predictive power of this study is low. Inference of spoofing behaviour in country-wise exchanges is limited, thus it is not possible to generalize the results and estimate the scale of spoofing behaviour in the exchanges.

Table 1.2 *Stock information (Nasdaq Nordic 2017).*

			Daily mean	
Name	Market cap	Currency	Avg price	Volume
Atlas Copco A	Large	SEK	153.59	$3.9 \cdot 10^6$
Metso Oyj	Mid	EUR	31.96	$0.8 \cdot 10^6$
Nokian Renkaat Oyj	Large	EUR	29.36	$0.7 \cdot 10^6$
Vestas Wind Systems	Large	DKK	110.07	$2.3 \cdot 10^6$
Volvo B	Large	SEK	92.66	$9.4 \cdot 10^6$

Stock-wise information of companies is presented in table 1.2. The table clearly shows similarity of data: market capitalizations, daily average prices (per currency), and daily trade volumes all are at the same magnitude. Similarity of data further hinders predictive power of this study. The sample is not diverse, and thus does not provide good sample of all the stocks, which are much more diverse.

In addition to data similarity, all of the stocks' market capitalisations are relatively big, daily average prices rather expensive and the stocks are traded very actively. According to research, susceptible stocks to spoofing behaviour on the contrary tend to have lower market capitalisation and lower average price (Aggarwal, Wu 2006; Lee et al. 2009; Mei et al. 2004). As a consequence, the stocks in this research are not very probable targets of spoof order manipulators.

Since the occurrences of spoofing behaviour is low, especially in this data set, false positives might over-populate findings effectively reducing the probability of detecting true spoof trading.

The most limiting constraint of the data is the lack of user identity. Trades in the daily order books have unique id, but the traders of the orders are not identified, thus destroying ability to track series of orders. As a result, it is not possible to establish logically valid patterns to study spoofing behaviour, instead each order has to be considered in isolation, only examining single order at time and determining whether it is spoofing order or not.

This thesis only focuses on stock market, since only stock market data was provided for this research. As a consequence spoofing behaviour detection of intermarket manipulation is not possible. Manipulator might attempt to manipulate underlying of a derivative, e.g. submit spoofing orders at the stock market while gaining profits in derivatives of the manipulated assets.

In conclusion because of these limitations it is very unlikely to find spoofing behaviour in this data set, and further draw any interference how common spoofing behaviour is in Northern European stock market. However, if this thesis is able to capture spoofing trading, then it is rather probable that the spoof trading phenomena exists in Northern European stock markets.

2. FINANCIAL MARKETS

PLAN: Overall introduction to financial markets, introduction of order types and other essential financial terms

Financial market is a market in which its' participants buy and sell financial assets. The mission of a financial market is to bring together buyers and the sellers in the market. The trades in financial markets are bilateral, two sides of an trade are connected through the market but the underlying asset and payment are exchanged only between the trading sides, markets do not engage in the trades.

Financial markets are divided into exchange-traded markets and over-the-counter (OTC) markets. In exchange markets market participants trade exchange standardized contracts while in OTC markets two parties are free to define contracts by themselves. Exchange markets are mainly divided by their geographical location, but also further categorized by the instruments they trade. Most common types of exchange markets include stock markets, bond markets, and various derivatives markets such as futures markets and options markets. (Hull 2017)

In exchange-traded markets traders mainly submit limit orders or market orders to buy and sell financial assets. In limit orders trader makes a bid to buy or sell a decided amount of financial instruments at a maximum purchase price or minimum selling price. Market order buys a financial instrument at lowest selling prices or sells one at the highest buying prices, all until whole volume of the market order is consumed.

2.1 Market micro-structure

PLAN: Explain how trading works, how the price of an asset is formed

Each asset in the market place generally has many active limit orders, both buy and sell side, waiting to be consumed. These set of limit orders then creates an order book for the underlying asset. Each asset in the market place generally has many active limit orders, both buy and sell side, waiting to be consumed. These set of limit orders then creates an order book for the underlying asset. Each asset in the market place generally has many active limit orders, both buy and sell side, waiting to be consumed. These set of limit orders then creates an order book for the underlying asset.

Table 2.1 *An order book example.*

Price	Quantity	Order type
54.50	5000	SELL
54.40	**1000**	**SELL**
54.30	2000	BUY
54.10	10000	BUY

Each asset in the market place generally has many active limit orders, both buy and sell side, waiting to be consumed. These set of limit orders then creates an order book for the underlying asset. Each asset in the market place generally has many active limit orders, both buy and sell side, waiting to be consumed. These set of limit orders then creates an order book for the underlying asset. Each asset in the market place generally has many active limit orders, both buy and sell side, waiting to be consumed. These set of limit orders then creates an order book for the underlying asset.

Table 2.2 *Example order book after new buy-side limit order of 2500 units at 54.20.*

Price	Quantity	Order type
54.50	5000	SELL
54.40	**1000**	**SELL**
54.30	2000	BUY
54.20	2500	BUY
54.10	10000	BUY

Table 2.3 *Example order book after new sell-side market order of 1500 units.*

Price	Quantity	Order type
54.50	5000	SELL
54.40	1000	SELL
54.30	**500**	**BUY**
54.20	2500	BUY
54.10	10000	BUY

- four? pictures of stock trading: vanha tilanne, uusi limit order, uusi market order, uusi tilanne

One of the most important functions of market exchange is price discovery process. Nasdaq glossary () defines price discovery process as a process that produces the prices of assets in the marketplace through buyers' and sellers' interactions.

2.2 Spoof trading

PLAN: Explain spoof trading, confirm its' illegality, define spoofing order

Market manipulation is considered as a method of artificially affecting price of a security. FIX: SOURCE???

(Kyle, Viswanathan 2008)

In spoof trading manipulator attempts to simulate fake demand or supply for an asset and after other investors utilize this fake information in their trading, consequently causing price of the asset change,

A successful case of spoof trading divides into following separate stages: manipulator feigns demand or supply of an asset, price of the assets distorts favourably for manipulator, manipulator makes her genuine trade at the distorted price, manipulator cancels her original fake order, markets adapt and price returns to normal, then manipulator is able to gain profit since she was able to dictate the price process of manipulated asset.

Lee at al () defines spoofing order as "*a bid/ask with a size at least twice the previous day's average order size and with an order price at least 6 ticks*[1] *away from the market price, followed by an order on the opposite side of the market, and subsequently followed by the withdrawal of the first order.*

spoofing order as an order that a trader submits to exchange without intention of execution, instead aiming to distort other market participants' perception of asset's price by providing misleading information about true demand or supply of a asset.

The market abuse regulation set by European Council states "market manipulation shall comprise ... a transaction, placing an order to trade or any other behaviour which: (i) gives, or is likely to give, false or misleading signals as to the supply of, demand for, or price of, a financial instrument ..." (Council of European Union 2014) In accordance with the market abuse regulation spoofing behaviour is indisputably prohibited method of market manipulation.

Allen and Gale categorized market manipulation into action-based, information-based and transaction-based manipulation as follows:

1. *action-based manipulation* is based on external actions (other than trading) that can influence price of an asset.
2. *information-based manipulation* is based on releasing false information to affect price of an asset.
3. *transaction-based manipulation* is based on manipulator affecting price of an asset through trading.(1992)

[1]Tick size determines smallest allowed change in security price.

Kong and Wang **KongWang2013** introduced another category of manipulation, namely order-based manipulation which involves order actions but not trade actions. They further asserted

However I will argue this further classification is redundant. In Allen and Gale's categorization transaction-based manipulation can be interpret as manipulation that is not neither action-based nor information-based, that is manipulation without any external actions or information. Consequently making spoofing as a form of transaction-based manipulation, rather than belonging to a new category of manipulation on its own.

In conclusion, regardless of whether applying further categorization by Kong and Wang or not, it is more significant to remark what spoofing behaviour is not. Spoofing behaviour does not belong neither action-based nor information-based manipulation, for this reason it is manipulation that does not contain any external actions taken nor exploitation of information outside of trading. As a consequence, the analysis of spoofing behaviour solely on trading data is justified.

3. UNSUPERVISED ANOMALY DETECTION

PLAN: Overall introduction to anomaly detection Intro to: - what is anomaly detection - detection methods: supervised, semi-supervised and unsupervised anomaly detection - anomaly types: global and local

3.1 Application constraints

PLAN: Note characteristics of anomaly detection in stock exchanges in order to be applicable, (time series, big data)

3.1.1 Detection in time series

PLAN: time-dependency and seasonality (non-stationary), transformation of data (paper from English Uni by Chinese) - point anomalies, collective anomalies, contextual anomalies

3.1.2 Detection in big data

PLAN: Issues with big data Stock exchange data is prime example of XXX with big data issues.

Laney (2001) lists volume, velocity and variety as defining characters of big data. *Volume* refers to the big amount of data. *Velocity* refers to the great speed at which data is created. *Variety* of big data refers to inconsistency of data from different sources.

Application field of stock exchanges ticks off two of the three defining characteristics of big data, both great volume and velocity of big data bring challenges to developers. Variety is not so big factor

- can't handle at once

- constant stream of data - online detection system

3.2 Methods

PLAN: Explanation of available methods, estimate how well they work in big data and time-series, length depends on the level of details. However these kinds of classifications are vague and generalisations are hard to make, since the algorithms, even in the same subcategory, are very different. **QUESTION:** Should I introduce many different kinds of methods or just the I will apply in this thesis? (see below for categorization)

- different types of methods (that are only used in this thesis or all kinds of?)

1. Classification based: classifier is trained to detect normal data, if new data point has extreme value it is likely to be anomaly
2. Clustering based: divide data into clusters, check if new data point belongs to any major clusters, anomalies should be separate
3. Nearest neighbour based: data points are given anomaly score, new data has the same score as its' neighbours
4. Statistical based: make statistical analysis of data, i.e. according to some distribution, outliers are likely to be anomalies
5. Subspace based: create new features of data sets, if new features cannot predict new data point it is likely to be anomaly

4. RESULTS

PLAN: Explain practical research process - imbalance in demand and supply lead to favourable price change which can be further exploited

In the context of this research, a case of spoofing trading is considered as follows: Firstly manipulator in order to distort asset prices submits a limit order to supply fake information to the market about the asset. Then after favourable change in price of the asset, manipulator makes her real trade. Subsequently she cancels original limit order hoping that the price will recover, and she is able to ???

In order to empirically study spoof trading, we need to address a few nuances that arise from this scenario. What are defining characteristics of these two intervening order, how they are interconnected; what are the favourable changes in prices; ???.

Manipulator has two two courses of actions, she either attempts to deflate or to inflate the price of an asset. In the case of price of asset falls as a consequence of market manipulation, market had excess supply of the assets, so manipulator submitted an sell side limit order.

In this study it is not possible to identify these two separate orders by trader since order information is anonymous. As a result,

Since it is not possible to determine whether manipulator was or was not able to gain profit through her manipulation, the main phase of this thesis is study are there cases in the data where trade imbalances lead to predictable price behaviour.

PLAN: Results case by case

4.1 Atlas Copco A

4.2 Metso Oyj

4.3 Nokian Renkaat Oyj

4.4 Vestas Wind Systems

4.5 Volvo B

4.6 Summary

PLAN: Write out summary of cases

5. CONCLUSION

PLAN: Final conclusion: results, generalization, further research

REFERENCES

Aggarwal, R., Wu, G. (2006). Stock Market Manipulations. The Journal of Business Vol.79(4), pp. 1915–1954.

Allen, F., Gale, D. (1992). Stock-Price Manipulation. Review of Financial Studies Vol.5(3), pp. 503–529.

Council of European Union (2014). Council regulation (EU) no 596/2014. http://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32014R0596.

Hull, J. (2017). Options, Futures and Other Derivatives. 10th edition. Pearson/Prentice Hall, p. 896.

Jarrow, R. (1992). Market Manipulation, Bubbles, Corners, and Short Squeezes. Journal of Financial and Quantitative Analysis Vol.27(3), pp. 311–336.

Kong, D., Wang, M. (2014). The Manipulator's Poker: Order-Based Manipulation in the Chinese Stock Market. Emerging Markets Finance and Trade Vol.50(2), pp. 73–98.

Kyle, A. S., Viswanathan, S. (2008). How to define illegal price manipulation. American Economic Review, pp. 274–279.

Laney, D. (2001). 3-D Data Management: Controlling Data Volume, Velocity, and Variety. Vol.6.

Lee, E. J., Eom, K. S., Park, K. S. (2009). Microstructure-Based Manipulation: Strategic Behavior and Performance of Spoofing Traders. Journal of Financial Markets Vol.16, pp. 227–252.

Mei, J., Wu, G., Zhou, C. (2004). Behavior Based Manipulation: Theory and Prosecution Evidence.

Nasdaq GlobeNewswire (2018a). Monthly Report - Total Equity Trading November 2017. Available (accessed on 1.1.2018): https://newsclient.omxgroup.com/cdsPublic/viewDisclosure.action?disclosureId=814252.

Nasdaq GlobeNewswire (2018b). Split with redemption and change of ISIN for Atlas Copco AB. Available (accessed on 1.1.2018): https://globenewswire.com/news-release/2015/05/07/733575/0/en/Split-with-redemption-and-change-of-ISIN-for-Atlas-Copco-AB-65-15.html.

Nasdaq Nordic (2017). Stock Market Data. Available (accessed on 31.10.2017): http://www.nasdaqomxnordic.com/.

Ögüt, H., Mete Doganay, M., Aktas, R. (2009). Detecting stock-price manipulation in an emerging market: The case of Turkey. Vol.36, pp. 11944–11949.