Machine learning and high-frequency algorithms during batch auctions

By

Gabriel Yergeau[1]

[1] The author is an IVADO postdoctoral fellow at Université de Montréal, HEC Montreal, and National Bank of Canada. He is financially supported by the Canada Research Chair in Risk Management.

Author e-mail: gabriel.yergeau@umontreal.ca

Machine learning and high-frequency algorithms during batch auctions

Abstract

We present the first direct evidence of algorithmic imprints during batch auctions. Order anticipation is an integral part of high-frequency traders' strategies. Hence, some participants may have economic incentive to encrypt noise in the data. We use machine learning to identify five types of algorithmic imprints that hinder the processing of auction information and have the encrypted noise characteristics. Our approach rests on the shifted wavelet tree (Yunyue and Shasha (2003)), a burst detection indicator, and the dynamic time warping similarity measure (Skutkova, Vitek, et al. (2013)). We show that market participants can adapt their trading to the presence of encrypted noise by filtering data in real time, thus clarifying the price discovery process. This could reveal the presence of informed traders. The methodology deployed is adaptable to different environments, including continuous trading.

Keywords: algorithmic trading; limit order book; call auction; investment decisions

JEL classification: G02, G10, G11, G14

1 Introduction

We present the first direct evidence of algorithmic imprints during batch auction (the auction hereafter). Auction are widely used at the opening and closing of stock markets. Adopted by almost all stock exchanges around the world, they are crucial trading mechanisms. The beneficial effect of auctions on market quality has been well established by event studies associated with the inauguration of auctions on various stock exchanges. Pagano and Schwartz (2003) determine that auctions implemented in 1996 and 1998 on the Paris stock exchange lower execution costs for participants. The Singapore Exchange introduced opening and closing auctions in August 2000. These introductions are investigated by Comerton-Forde, Ting Lau et al. (2007) and Chang, Rhee, et al. (2008).They note an improvement in market quality and a decrease in end-of-day price manipulation in the market. The closing auction started at the London Stock Exchange in May 2000 and Chelley-Steeley (2008) notes an improvement in market quality at that exchange. Pagano, Peng, et al. (2013) analyze the impact on bid-ask spread and price volatility of auctions introduced in 2004 on NASDAQ. Their results suggest positive spillovers on price formation dynamic behavior. In June 2009, Nasdaq-OMX launched index futures auctions. Hagströmer and Nordén (2014) conclude that Nasdaq-OMX auctions improved closing price accuracy and end-of-day liquidity.

Algorithmic trading (AT), of which high-frequency trading is a subset, is defined as a trading system whose decision-making process does not involve human intervention (Bates (2017)). It is the expression of a fundamental trend centered on technological development. The nature of competition evolves as high-frequency traders change speed into information (O'Hara (2015)). J.P. Morgan (2017) notes that humans already play a very small role in short-term trading. During continuous time sessions, Bouveret, Guillaume, et al. (2014) estimate the value traded by high-frequency traders at 24% in Europe and 21% on Xetra.

We identify and classify trading algorithm activities from the Frankfurt stock exchange auction data. Information opacity is more important than that of continuous trading. We present a methodology to infer the order's characteristics. Our algorithmic imprint recognition focuses on Abrol, Chesir, et al.'s (2016) negative price loops (NPLs). NPLs exhibit Stiglitz's (2014) encrypted noise characteristics. Specifically, they blur the price discovery process. Public information flow is monitored to detect high-frequency activity bursts. Yunyue and Shasha's (2003) shifted wavelet tree (SWT) structures the data, and we apply a burst indicator to reveal activity eruptions. Supervised learning is used to compare burst sequences with pre-identified NPL sequences to determine their similarity. We identify five types of NPLs: 456,772 events are uncovered, representing more than 11% of all auction events. NPL users can be either informed traders or proprietary firms and are a by-product of low latency trading. They are conceptualized to generate redundant information and they blur the price and quote discovery processes.

The paper is structured as follows: Section 2 presents a review of the literature. Section 3 characterizes the institutional context of trading on Xetra, the electronic platform of the German stock exchange. Section 4 defines the concept of algorithmic sequences and develops the methodologies that allow their identification. Section 5 presents the data and transaction costs. Section 6 reports the results. Section 7 discusses the results and concludes.

2 Literature review

2.1 Behavior

Numerous studies investigate high-frequency traders' (HFTers) behavior during continuous double auctions. Brogaard, Carrion, et al. (2016) conclude that HFTers supply liquidity during extreme price movements. Subrahmanyam and Zheng (2016) note HFTers' ability to manage limit orders in anticipation of short-term price movements. Goldstein, Kwan, et al. (2016) find that HFTers provide liquidity on the thick side of the order book and demand liquidity on the thin side. Hirschey (2016) states that HFTers can anticipate the order flow from other investors. Menkveld and Yueshen (2016) emphasize the importance of inter-market arbitrage as a behavioral characteristic. These studies shed light on the industry aggregate behavior but do not distinguish between specific traders' activities that may exhibit heterogeneous behaviors (Carrion (2013)). There are exceptions: Menkveld (2013) highlights the positive contribution of a high-frequency market maker's (HFMM) arrival on Chi-X Europe, and Yergeau (2016) analyzes the behavior of an endogenous liquidity provider using the dynamic inventory management model of Ait-Sahalia and Saglam (2014). A review of the high-frequency trading industry is presented in Chung, K. H. and A. J. Lee (2016).

2.2 Machine learning

Yang, Qiao, et al. (2012) utilize machine learning to identify Kirilenko, Kyle, et al.'s (2016) categories of traders. Variables that motivate all decisions are the inventory position (Kyle (1985); Glosten and Milgrom (1985); Huang and Stoll (1997); among others) and imbalances at the first and third levels of the order book (Cont, Kukanov, et al. (2014)). Yang, Qiao, et al. (2012) attempt to obtain trader categories' reward functions from inverse reinforcement learning. Eighteen simulations of approximately

300,000 E-Mini S&P 500 LOB activities serve as learning. Simulations come from Hayes, Paddrik, et al.'s (2012) agent-based model. The authors show a clear connection between the trader classification done by Kirilenko, Kyle, et al. (2016) and their results from machine learning approach.

3 Institutional context: trading on Xetra

Table 1 shows that a hybrid market model with three auctions and two continuous trading periods characterizes the DAX and MDAX segments. During auctions, traders can submit limit and market orders. After the auction ends, matching orders are executed at a single price and unexecuted orders transferred to the next trading stage.

[insert Table 1 here]

[insert Table 2 here]

Table 2 shows the public information available during auctions. Public information consists of eight elements: the stock identifier, the date, the timestamp in microseconds, the status of the auction (opening, intraday or closing), the indicative price, the quantity matched at the indicative price, the surplus side (imbalance) and its quantity. The identity of the trader, the type of event (creation, modification, or cancellation), and the quantity associated with each event are not published. They are deduced from public information, and are interpreted in Table 3.

[insert Table 3 here]

Any matched quantity variation represents the order quantity. If the matched quantity increases and the indicative price increases (decreases), a buy (sell) order occurs. If the matched quantity decreases and the indicative price increases (decreases), a cancellation of a sell (buy) order occurs. In the absence of a matched quantity change, the surplus variation corresponds to the event quantity. An increase in the surplus variation is due to a limit order creation on the marginal variation side, whereas a decrease is due to a cancellation.

Given that our data do not identify orders specifically, our methodology undoubtedly introduces noise in the association of activity sequences with a given source. This bias could originate from the aggregation of two or more orders reported during the same event. Even if possible, microsecond timestamps suggest that our methodology can infer related algorithmic sequences.

4 Algorithmic sequences

Our goal is to link algorithmic sequences to algorithm types. Hasbrouck and Saar (2013) correlate same quantity limits and/or market orders with short durations to high-frequency algorithms. We apply this concept to auctions. First, we identify algorithmic sequences with the SWT tree structure (Yunyue and Shasha (2003)) and a burst detection indicator. DTW (Skutkova, Vitek et al. (2013)) measures the similarity of these sequences to reference sequences exhibiting NPL characteristics documented in the literature.

Yunyue and Shasha (2003) propose a tree structure, the SWT, to monitor bursts of real-time activities from a data stream. The tree aggregates the time intervals while preserving the original structure of the data. The main contribution of Yunyue and Shasha (2003) is related to the reduction in the number of windows necessary to monitor events. Their structure shrinks from $O(n^2)$ by considering the set of all possible combinations to $O(n)$ where n is the number of windows of the smallest time interval of the sample considered. Figure 1 illustrates the structure of the SWT that exploits the half-overlap of time windows.

[insert Figure 1 here]

Equation (1) uses the duration and number of SWT levels to obtain the basic time interval of the tree, i.e. the level 0 time interval:

$$\mathrm{SWTtime\ interval}_{\mathrm{level0}} = \frac{\mathrm{time}_{\mathrm{end}} - \mathrm{time}_{\mathrm{start}}}{2^{\mathrm{j}}}\,, \qquad (1)$$

where: j= number of SWT levels.

For a closing auction with a total duration of five minutes, we use a 14-level tree with 16,384 windows (2^{14}). The time interval at level 0 is 0.0183 seconds ($(5\ minutes * 60)/16{,}384$). The tree reduces the number of windows to supervise from 268,435,456 ($16{,}384^2$) to 16,384 windows. Burst detection of abnormal activities in a timely manner becomes feasible. Identification of these activities depends on the intraperiod cumulative value of the burst indicator $F(\cdot)$ and a threshold:

$$\mathrm{F}(\mathrm{x}_{\mathrm{i,j}}) > \mathrm{f}(\mathrm{w}_{\mathrm{i}})\,, \qquad (2)$$

where:

$\mathrm{x}_{\mathrm{i,j}}$ = time interval of SWT level *i*, window *j*,
$\mathrm{f}(\mathrm{w}_{\mathrm{i}})$= time interval threshold of SWT level *i*.

The alarm domain, $\mathrm{f}(\mathrm{w}_{\mathrm{i}})$, is equal to $\min(6, 2^{\mathrm{i}})$, i being the SWT level monitored. When an alarm comes from a higher level, efficient streaming algorithms (online and batch) are used (Yunyue and Shasha's (2003) *Lemma 3*). This makes it possible to precisely locate the level 0 sequence involved.

To link the events of activity bursts to a specific type of algorithm, we compare them to reference sequences identified using stylized facts like quote stuffing: (Egginton, Van Ness et al. (2014); Ye, Yao et al. (2013); Crédit Suisse (2012); Brogaard (2010)), and phantom liquidity (Blocher, Cooper, et al. (2016); Korajczyk and Murphy (2016)).

We quantify the similarity between reference sequences and sequences from alarms with the DTW distance. DTW finds an optimal alignment between two data sequences. It minimizes time shift and distortion effects. It can measure the similarity between two series that may differ in length. This similarity measure achieves the best pattern recognition (Petitjean, Forestier, et al. (2014); Ding, Trajcevski, et al. (2008)).

5 Data

Data come from Xetra, the electronic platform of the Frankfurt Stock Exchange. The database contains all events related to auctions sent via the Enhanced Broadcast System, an information flow used by high-frequency traders. Xetra Parser, developed by Bilodeau (2013), is used to reconstruct the event sequence. The timestamps are in microseconds, and trading is anonymous.

Our sample has 60 components: 30 from the DAX index, made up of the stocks with the highest market capitalization; and 30 from the MDAX index, which comprises stocks with average market capitalization and excludes technology. Hereafter, we refer to the 30 DAX (MDAX) components as DAX (MDAX). Auction cover the period of February to July 2013. They account for about 15% of all activities, i.e. 4,094,751 events. The other 85% events occur mainly during continuous trading sessions.
Table 4 shows the statistics by auction.

[insert Table 4 here]

Closing auctions trigger most activities, namely 71% (69%) of the DAX (MDAX) events. The relative importance of the three auctions is qualitatively the same in both indexes. The activities happen predominantly during closing auctions. Differences between average and median matched quantities are due to the presence of frequent extreme values.

Table 5 classifies events according to their impact on matched quantity. 15.03% (35.52%) of DAX (MDAX) events increase the matched quantity. This results from limit orders’ decreasing the existing surplus size, a conservative strategy, or aggressive limit orders’ affecting the indicative price. Quantity additions to an existing surplus represent most events: 52.78% (DAX) and 51.05% (MDAX). This reflects an effort to minimize the impact of orders on indicative price, an institutional trader characteristic. Cancellation of previously matched orders accounts for 1,062,314 (106 576) DAX (MDAX) events.

[insert Table 5 here]

Easley, Lopez de Prado et al. (2012) link temporal cyclicity to institutional traders. Figure 2 shows the behavior of DAX closing auctions event numbers by 5-second time intervals. Six bursts in the number of orders occur simultaneously for all components of the DAX.[2] This is clear evidence of institutional imprints. The very low activity observed during the auctions' last thirty seconds (periods 61 to 66) meaning that random time period addition at the end of the auctions has, at best, a mixed economic contribution. These behaviors are also seen for other auctions and the MDAX.

[insert Figure 2 here]

5.1 Transaction costs

There are many types of traders. Hedgers and institutional investors have heterogeneous investment horizons (Cespay and Vives (2016)). Some may use brokers for their order execution. Accordingly, Battalio, Corwin, et al. (2016) identify US brokers who maximize their revenues by the rebates granted by trading venues, to the detriment of

---

[2] We obtain the same cyclicity pattern when we split the sample in shorter subsamples.

their customers. Deutsche Boerse (2015) states that under the Designated Sponsor Program (Section 2.2.3.2) and the Top Liquidity Provider Program (Section 2.2.3.3), the stock exchange does not charge transaction fees to participants of these programs, and grants them rebates for executed orders (limit and marketable). Top liquidity providers earn a rebate of 0.20 basis points on their traded market value to a maximum of 375k euros per order per day. Other hedgers and institutional investors can benefit from direct market access (DMA); some may use co-location (Malinova and Park (2016); Malinova, Park, et al. (2016)). Direct access allows strategic management of limit and marketable orders using several prices (Upson and Van Ness (2017); Easley, de Prado, et al. (2016)). If these investors are billed directly by the Deutsche Boerse, the Section 2.2.1.1, Table 6 (Deutsche Boerse (2015) op. cit.) establishes a fee model for the DAX based on three activity levels referred to as high, medium, and low volume levels. The cost of the medium volume category is 0.378 basis points based on market values capped at 1.5 million euros per order per day. This model does not qualify for rebates. In the case of billing by the broker, Cappon and Mignot (2014), estimate the cost at 1.5 cents per share, while Menkveld (2016) estimates the cost of executing a marketable order at 7 basis points.

Data during auctions are opaque. Order issuers and their characteristics are not in the public domain. Consequently, we do not identify the exact status of the traders behind the algorithmic activities.

6 Results

NPLs lock the indicative price in a range. They prevent the price discovery process and hamper the disclosure of supply and demand (Abrol, Chesir, et al. (2016)). NPLs may create phantom liquidity, a source of additional costs imposed on investors (BMO Capital Markets (2009)). They are associated with quote stuffing, a cause of latency arbitrage (Brogaard (2010)) and stale quotes (Foucault, Kozhan, et al. (2015); Menkveld and Zoican (2017)).

SWT and DTW identify five types of NPL algorithms whose characteristics are presented in Table 6. All NPL algorithms share a common structure, namely an uninterrupted creation and cancellation sequence of identical orders (price and quantity). They use limit or market orders, and set the indicative price in a constant range. We categorize NPL algorithms by their impact on information. The impact depends on heterogeneous (LOB depth and granularity) and endogenous (order price and quantity) factors. Types 1 and 2 have the smallest effect. Limit orders are inside the bid-ask spread on the surplus side and do not modify the matched quantity, i.e. they do not match with the LOB's opposite side. Types 3 and 4 have a greater impact because they change the matched quantities. The imbalance (surplus quantity) stays on the same side of the market. This can result from a marketable order with a quantity smaller than the surplus. Type 5 influences all variables.

[insert Table 6 here]

The execution speed of NPLs is too fast for humans to perceive. A graphic presentation whose paradigm is event-driven rather than temporal is revealing. Figure 3 displays Deutsche Bank’s closing auction events on February 12, 2013. The graphs show clockwise: the indicative price, the matched quantity, the surplus quantity, and the duration between events. We identify four NPL sequences. As previously defined, the indicative price is within a constant range. Indicative price volatility changes with sequences. Durations are significantly shorter than their average during these NPLs. It confirms activities related to high-frequency algorithms.

[insert Figure 3 here]

Figure 4 illustrates the analyzed variables for NPL identification. An identical order (price and quantity) is created and canceled six times (twelve events) during a time interval of 0.08 seconds with no intervening activities. We link the sequence to the type 1 algorithm in Table 8: indicative price and sell side surplus vary while the matched quantity remains the same. Durations between events range from 0.4 to 2.0 milliseconds. To act at this speed, the algorithm is probably operated from a co-location site. DTW determines the similarity between reference sequences and the potential NPL sequence.

[insert Figure 4 here]

403,746 DAX events are due to NPLs (Table 7). They mainly occur at the closing auction, where they represent more than 17% of events. These high-frequency events are seen at the first SWT levels: 61.71% of the activities are detected at level 2, where the time interval is 0.08 seconds. Each increase in one SWT level doubles the preceding time interval. The MDAX behavior is qualitatively similar.

[insert Table 7 here]

Table 8 summarizes the NPLs' characteristics. We identify 2,595 DAX NPL sequences. They average 143 repetitions and have a median duration of 0.012 seconds. Sequences mainly arise during closing auctions. MDAX has 134 NPL sequences affecting eight index components. These sequences occur at the end of the day and exhibit 357 repetitions on average.

[insert Table 8 here]

Figure 5 shows the NPL distributions during the DAX and MDAX closing auctions. NPLs are present from the second minute to the penultimate minute of these auctions.

[insert Figure 5 here]

Figure 6 displays the distribution of NPL events for the five algorithm types of Table 8. They differ significantly. For the DAX, all types are used and almost one-third of the NPLs (129,823 events) are generated by algorithm 5. This algorithm exhibits the most aggressive characteristics by far. It implies a change in the surplus side along with variations in the indicative price and the matched quantity. For the MDAX, only types 1 and 2 are used. These types place limit orders inside the bid-ask spread.

[insert Figure 6 here]

7 Discussion and conclusion

Order anticipation is an integral part of the strategies commonly used by high-frequency traders documented in the literature. Baldauf and Mollner (2016) show that high-frequency liquidity providers use order identification to avoid adverse selection by canceling mispriced quotes. Brunnermeier and Pedersen (2005) describe the exploitation of orders from institutional investors that need to liquidate their positions as predatory. Clark-Joseph (2013) documents the behavior of HFTers that probe liquidity in order to obtain information about large incoming orders. Yang and Zhu (2016) introduce the concept of back-running, in which fast traders compete with institutional investors that make large orders, after recognizing their imprints. Thus, each participant has an economic incentive to make it more difficult for other traders to extract information from public data (the Stiglitz (2014) data encryption hypothesis).

SWT and DTW can reveal NPLs. They abound during auctions. NPLs are a by-product of low-latency trading in that they are characterized by short duration. Position management is not their goal; their cancellation rates converge at 100%. They are conceptualized to generate redundant information and they blur the price discovery process. NPLs are consistent with Stiglitz's (2014) data encryption hypothesis. Institutional investors and proprietary firms use AT (Hasbrouck and Saar (2013); Yang and Zhu (2016)). Because we cannot identify the originator of the order, we must consider that the presence of NPLs is due to one of two scenarios. Either informed traders wish to conceal their trading intentions, or proprietary firms induce delays in information processing justified by their desire to take advantage of information extracted from informed traders. Figure 7 shows a histogram of transaction costs of NPL sequences. The transaction cost per NPL sequence is maximum at 56.70 euros. These costs do not deter traders from encrypting the price discovery process. Moreover, although NPL generators are part of the designated sponsor or top liquidity provider programs (Deutsche Boerse (2015), Section2.4.1), they do not incur transaction costs. Rather, they earn rebates.

[insert Figure 7 here]

Regulators face numerous challenges arising from encrypted noise. However, the stock exchanges encourage the use of encrypted noise activities by high-frequency traders through their fee and rebate structure. An encrypted noise ban may hamper low-latency informed traders from executing trades with a minimum price impact. Designing regulations to forbid encrypted noise by constraining order submission can be difficult because it opens up opportunities for regulatory arbitrage (Stiglitz (2014)). Tackling the problem by reducing trading speeds would also affect liquidity takers (Shorter and Miller (2014)). If regulatory bodies target specific behaviors, quants can modify the algorithms.

For both DAX and MDAX, the main concern of the identified sequences is to limit the orders' impact on the indicative price, a characteristic of large investors during continuous time trading (Duffie and Zhu (2017)). Figure 8 illustrates this preoccupation. A time-weighted average price (TWAP) algorithm is active during the May 15, 2015 closing auction on Volkswagen stock (identifier 130). Seven events occur in a 0.073

second timeframe. We now comment on the graphics in clockwise order. None of the seven trades modifies the indicative price. There is no surplus on the offer side. The short position increases steadily by a quantity of 20, to end at -140. This results from seven marketable orders sent against the bid surplus quantity. All durations are in the 0.011 to 0.015 second range. Such regularities require a low latency infrastructure. This behavior is in line with Menkveld's (2016) finding of patterns of liquidity-demanding tradebots from high and low-frequency traders.

[insert Figure 8 here]

NPL detection allows market participants to adapt their trading to the presence of encrypted noise. By defining appropriate burst indicators and targeting specific timeframes, SWT and the supervised learning approach can be used to monitor encrypted noise evolution in streaming and batch environments. Filters can be implemented to mitigate data congestion while clarifying the price discovery process. In this study, we present the first direct evidence of algorithmic imprints during auctions. Our approach distinguishes algorithm types. The information content of NPL algorithms varies, yet all algorithms reduce the information processing speed. Position management differs between the DAX and the MDAX. These differences may stem from a greater trading intensity coupled with hedging operations and arbitrage opportunities between the DAX components and the very liquid DAX futures.

NPL detection can enable market participants to adapt their trading to the presence of encrypted noise by filtering data and clarifying the price discovery process. The methodology deployed makes the approach adaptable to different environments, including continuous time trading.

8 References

Abrol, S., et al. (2016). "High Frequency Trading and US Stock Market Microstructure: A Study of Interactions between Complexities, Risks and Strategies Residing in U.S. Equity Market Microstructure." Fin. Markets, Institutions & Instruments **25**(2): 107-165.

Ait-Sahalia, Y. and M. Saglam (2014). "High Frequency Traders : Taking Advantage of Speed." SSRN Electronic Journal.

Baldauf, M. and J. Mollner (2016). "Fast Traders Make a Quick Buck: The Role of Speed in Liquidity Provision." SSRN Electronic Journal.

Bates, T. (2017). "Topics in Stochastic Control with Applications to Algorithmic Trading." Ph. D. Thesis , LSE.

Battalio, R. H., et al. (2016). "Can Brokers Have It All? On the Relation between Make-Take Fees and Limit Order Execution Quality." J. Finance, Forthcoming.

Bilodeau, Y. (2013). "Xetraparser [computer software]."

Blocher, J., et al. (2016). "Phantom Liquidity and High Frequency Quoting." J. Trading **11**(3): 6-15.

BMO Capital Markets, Q. E. S. (2009). "The Hidden Costs of Locked Markets." Research Report.

Bouveret, A., et al. (2014). "High-frequency Trading Activity in EU Equity markets." ESMA Economic Report.

Brogaard, J. (2010). "High Frequency Trading and its Impact on Market Quality." SSRN Electronic Journal.

Brogaard, J., et al. (2016). "High-Frequency Trading and Extreme Price Movements." SSRN Electronic Journal.

Brunnermeier, M. and L. H. Pedersen (2005). "Predatory Trading." J. Finance **60**(4): 1825-1863.

Cappon, A. and S. Mignot (2014). "The Brokerage World Is Changing, Who Will Survive?" Forbes **April 16** (https://www.forbes.com/sites/advisor/2014/04/16/the-brokerage-world-is-changing-who-will-survive/#68e93b9068a7).

Carrion, A. (2013). "Very Fast Money: High-frequency Trading on the NASDAQ." J. Fin. Markets **16**(4): 680-711.

Cespay, G. and X. Vives (2016). "The Welfare Impact of High Frequency Trading." American Finance Association - Conference.

Chang, R. P., et al. (2008). "How Does the Call Market Method Affect Price Efficiency? Evidence from the Singapore Stock Market." J. Bank. Financ. **32**(10): 2205-2219.

Chelley-Steeley, P. L. (2008). "Market Quality Changes in the London Stock Market." J. Bank. Financ. **32**(10): 2248-2253.

Chung, K. H. and A. J. Lee (2016). "High-frequency Trading: Review of the Literature and Regulatory Initiatives around the World." Asia-Pacific J. Fin. Stud. **45**(1): 7-33.

Clark-Joseph, A. D. (2013). Exploratory Trading. Working Paper, Harvard.

Comerton-Forde, C., et al. (2007). "Opening and Closing Behavior Following the Introduction of Call Auctions in Singapore." Pacific-Basin Financ. J. **15**(1): 18-35.

Cont, R., et al. (2014). "The Price Impact of Order Book Events." J. Fin. Economics. **12**(1): 47-88.

Crédit Suisse (2012). What Does "Bad" HFT Look Like, How often Does It Happen, and How Do We Detect It? AES Analysis.

Deutsche Boerse (2015). "Price List for the Utilization of the Exchange EDP of FWB Frankfurt Stock Exchange and of the EDP XONTRO." Research Paper.

Ding, H., et al. (2008). "Querying and Mining of Time Series Data: Experimental Comparison of Representations and Distance Measures." PVLDB **1**(2): 1542-1552.

Duffie, D. and H. Zhu (2017). "Size Discovery." R. Financ. Studies. **30**(4): 1095-1150.

Easley, D., et al. (2016). "Discerning Information from Trade Data." J. Fin. Economics. **120**(2): 269-285.

Easley, D., et al. (2012). "The Volume Clock: Insights into the High-frequency Paradigm." J. Portfolio Manage. **39**(1): 19-29.

Egginton, J. F., et al. (2014). "Quote Stuffing." SSRN Electronic Journal.

Foucault, T., et al. (2015). "Toxic Arbitrage." HEC Paris Research Paper No. FIN-2014-1040.

Glosten, L. R. and P. R. Milgrom (1985). "Bid, Ask and Transaction Prices in a Specialist Market with Heterogeneously Informed Traders." J. Fin. Economics. **14**(1): 71-100.

Goldstein, M., et al. (2016). "High-frequency Trading Strategies." SSRN Electronic Journal.

Hagströmer, B. and L. Nordén (2014). "Closing Call Auctions at the Index Futures Market." J. Fut. Markets **34**(4): 299-319.

Hasbrouck, J. and G. Saar (2013). "Low-latency Trading." J. Fin. Markets. **16**(4): 664-678.

Hayes, R., et al. (2012). "Agent Based Model of the E-Mini S&P 500 Future: Application for Policy Making." Computational Intelligence for Financial Engineering & Economics Conference.

Hirschey, N. (2016). "Do High-Frequency Traders Anticipate Buying and Selling Pressure?" SSRN Electronic Journal.

Huang, R. D. and H. R. Stoll (1997). "The Components of the Bid-Ask Spread: A General Approach." R. Financ. Studies **10**(4): 995-1034.

J.P. Morgan (2017). "Big Data and AI Strategies: Machine Learning and Alternative Data Approach to Investing." White Paper: 280 pages.

Kirilenko, A. A., et al. (2016). "The Flash Crash: the Impact of High Frequency Trading on an Electronic Market." J. Finance, Forthcoming.

Korajczyk, R. A. and D. Murphy (2016). "High Frequency Market Making to Large Institutional Trades." SSRN Electronic Journal.

Kyle, A. S. (1985). "Continuous Auctions and Insider Trading." Econometrica **53**(6): 1315-1335.

Malinova, K. and A. Park (2016). "“Modern” Market Makers." Working Paper, University of Toronto.

Malinova, K., et al. (2016). "Taxing High Frequency Market Making: Who Pays the Bill?" SSRN Electronic Journal.

Menkveld, A. J. (2013). "High Frequency Trading and the New Market Makers." J. Fin. Markets. **16**(4): 712-740.

Menkveld, A. J. and B. Z. Yueshen (2016). "The Flash Crash: A Cautionary Tale about Highly Fragmented Markets." SSRN Electronic Journal.

Menkveld, A. J. and M. A. Zoican (2017). "Need for Speed? Exchange Latency and Liquidity." R. Financ. Studies. **30**(4): 1188-1228.

O'Hara, M. (2015). "High Frequency Market Microstructure." J. Fin. Economics. (116): 257-270.

Paddrik, M. E., et al. (2016). "Visual Analysis to Support Regulators in Electronic Order Book Markets." Environment Systems and Decisions **36**(2): 167-182.

Pagano, M. S., et al. (2013). "A Call Auction's Impact on Price Formation and Order Routing: Evidence from the NASDAQ Stock Market." J. Fin. Markets. **16**(2): 331-361.

Pagano, M. S. and R. A. Schwartz (2003). "A Closing Call's Impact on Market Quality at Euronext Paris." J. Fin. Economics. **68**(3): 439-484.

Petitjean, F., et al. (2014). "Dynamic Time Warping Averaging of Time Series Allows Faster and more Accurate Classification." IEEE International Conference on Data Mining, at Shenzhen, China.

Shorter, G. and R. S. Miller (2014). High-frequency Trading: Background, Concerns, and Regulatory Developments. Congressional Research Service, Washington DC.

Skutkova, H., et al. (2013). "Classification of Genomic Signals using Dynamic Time Warping." BMC Bioinformatics **14** (10).

Stiglitz, J. (2014). "Tapping the Brakes: Are Less Active Markets Safer and Better for the Economy?" Speech at the 2014 Financial Markets Conference at the Federal Reserve Bank of Atlanta.

Subrahmanyam, A. and H. Zheng (2016). "Limit Order Placement by High-Frequency Traders." SSRN Electronic Journal.

Upson, J. and R. A. Van Ness (2017). "Multiple Markets, Algorithmic Trading, and Market Liquidity." J. Fin. Markets. **32**: 49-68.

Yang, S. Y., et al. (2012). "Gaussian Process Based Trading Strategy Identification." SSRN Electronic Journal.

Yang, L. and H. Zhu (2016). "Back-Running: Seeking and Hiding Fundamental Information in Order Flows." Working Paper , Rothman School of Management No. 2583915.

Ye, M., et al. (2013). "The Externalities of High Frequency Trading." SSRN Electronic Journal.

Yergeau, G. (2016). "Profitability and Market Quality of High Frequency Market-Makers: An Empirical Investigation." SSRN Electronic Journal.

Yunyue, Z. and D. Shasha (2003). "Efficient Elastic Burst Detection in Data Streams." Proceedings of the ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining: 336-345.

Table 1. Classification of traders

	Position	**Inventory**	**Types**
Intraday intermediairies	Small relative to volume	Mean-reverting	HFTers' inventories are positively correlated to stock prices HFMMs' inventories are negatively correlated to stock prices
Fundamental traders	More than 15% of total trading volume	Long Short	Buyer Seller
Small traders	Small	Small if any	Less than 10 contracts per day

This table shows the classification of traders of Kirilenko, Kyle et al. (2016) inferred from the CFTC audit data of the Chicago Mercantile Exchange E-mini S & P 500 futures contract for the period May 3 to 6, 2010. Abbreviations: HFTers: high-frequency traders; HFMMs: high-frequency market makers.

Table 2. Xetra market model

	Opening auction		Intraday auction		Closing auction
DAX	8:50-9:00	Continuous trading	13:00-13:02	Continuous trading	17:30-17:35
MDAX	8:50-9:02		13:05-13:07		17:30-17:35

Regular auction periods are followed by a random end of 30 seconds maximum.

Table 3. Auctions: public information

Stock ID	Date	Time stamp	Status	Matched Q	Indicative Price	Surplus Ask	Surplus Bid
2841	20130201	31800142084	5	20,286	€ 37.925	-	5100
2841	20130201	31800827109	5	25,286	€ 37.925	-	100
2841	20130201	31802184578	5	25,973	€ 37.910	-	413
2841	20130201	31805024087	5	32,291	€ 38.000	-	427
2841	20130201	31805026280	5	32,321	€ 38.100	-	2274
2841	20130201	31810379770	5	33,266	€ 38.200	1,365	-

Stock ID: Stock identifier; Status: 5: opening auction.

Table 4. Interpretation of public information on auctions

Δ I. Q.	**Δ I.P.**	**Δ Ask Surp.**	**Δ Bid Surp.**	**Code**	**Interpretation**
+	+			1	Marketable order buy
+	-			1	Marketable order sell
-	-			2	Cancellation buy
-	+			2	Cancellation sell
0	-	≠ 0		1	Sell limit order < indicative price and > best bid
0	+		≠ 0	1	Buy limit order > indicative price and < best ask
0	+	≠ 0		2	Sell limit order cancellation < indicative price
0	-		≠ 0	2	Buy limit order cancellation > indicative price
0	0	(Δask-Δbid)>0		1	Buy (Sell) limit order if Δbid>Δask (Δask>Δbid)
0	0	(Δask -Δbid)<0		2	Cancellation buy (sell) limit order if Δbid<Δask (Δask<Δbid)

Code 1: creation; Code 2: cancellation; Δ I.Q.: variation in the indicative quantity; Δ I.P.: variation in the indicative price; Δ Ask Surp.: variation in the ask surplus; Δ Bid Surp.: variation in the bid surplus.

Table 5. Statistics - all auctions

		# Stocks	# Events	% of total events	Mean E.O.A. Matched Q	Median E.O.A. Matched Q
DAX	**Open**	30	582,330	17.64%	50,731	18,636
	Midday	30	370,963	11.24%	63,488	8,609
	Close	30	2,347,037	71.12%	264,939	32,484
	# Events	30	**3,300,330**	**100.00%**		
MDAX	**Open**	30	185,020	23.29%	3,235	1,102
	Midday	30	58,384	7.35%	2,273	523
	Close	30	551,017	69.36%	16,935	1,315
	# Events		**794,421**	**100.00%**		

Stocks: total number of stocks in the sample; # events: total number of events; Mean E.O.A. Matched Q: Mean end of auction matched quantity; Median E.O.A. Matched Q: Median end of auction matched quantity.

Table 6. Indicative quantity - all auctions

	Variation in Matched Q	# events I.Q.	% events I.Q.	Mean I.Q.	Median I.Q.
DAX	**= 0**	1,741,847	52.78%		
	> 0	496,199	15.03%	3,815	430
	< 0	1,062,314	32.19%	- 3,061	- 302
	Total	**3,300,360**	**100.00%**		
MDAX	**= 0**	405,631	51.06%		
	> 0	282,214	35.52%	764	152
	< 0	106,576	13.42%	-524	-136
	Total	**794,421**	**100.00%**		

events I.Q.: number of events affecting the indicative quantity; % events I.A.: percentage of total number of events for the category; Mean I.Q.: average indicative quantity; Median I.Q.: median indicative quantity.

Table 7. NPL algorithm types

Type of algorithm	**Matched quantity**	**Surplus bid**	**Surplus ask**
1	0	0	< >
2	0	< >	0
3	< >	< >	0
4	< >	0	< >
5	< >	< >	< >

< > : a positive or negative variation in the variable; 0: no impact.

Table 8. Negative price loops and SWT levels

SWT levels	2	3	4	5	NPL	All	% all
Open	170	268	788	-	1,226	582,330	**0.21%**
Midday	-	-	9	11	20	370,963	**0.01%**
Close	248,974	56,826	96,700	-	402,500	2,347,037	**17.15%**
NPL	**249,144**	**57,094**	**97,497**	**11**	**403,746**	**3,300,330**	**12.23%**
% by level	**61.71%**	**14.14%**	**24.15%**	**0.00%**			

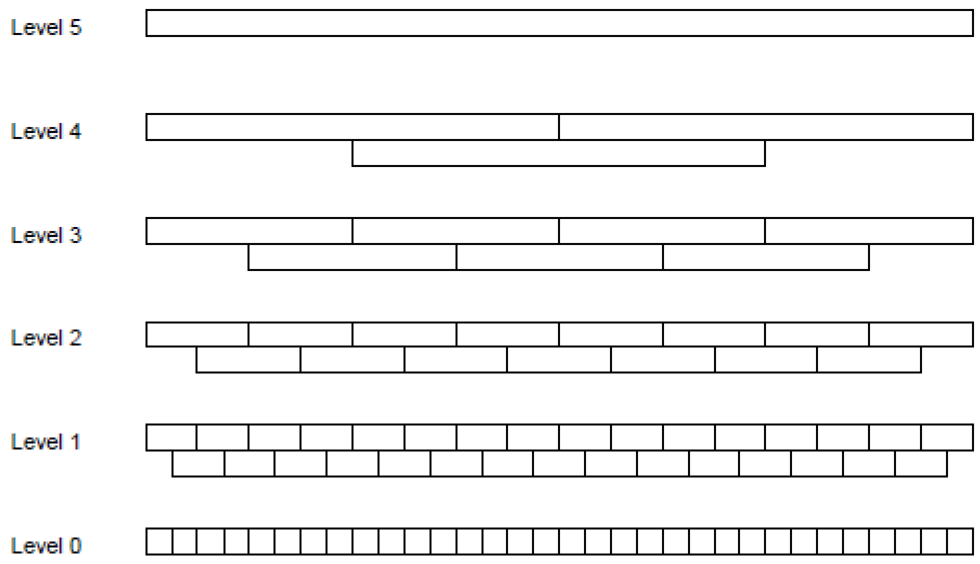
SWT levels: shifted wavelet tree level at which algorithmic sequences have been identified. A measure of latency.

Table 9. NPL algorithmic sequences

DAX		Total	Open	Midday	Close
Sequences		2,595	79	2	2,514
Stocks		30/30	17/30	2/30	30/30
Repetition	**Mean**	143	9	8	147
	Median	61	8	8	65
Duration (sec.)	**Mean**	21.150	0.479	0.169	21.818
	Median	2.230	0.097	0.169	2.422
Duration (sec.)	**Mean**	0.460	0.040	0.022	0.473
Between events	**Median**	0.020	0.012	0.022	0.020
MDAX					
Sequences		134	1	-	133
Stocks		8/30	1/30	-	7/30
Repetition	**Mean**	355	52	-	357
	Median	144	52	-	144
Duration (sec.)	**Mean**	22.656	2.108	-	22.719
	Median	3.535	2.108	-	3.616
Duration (sec.)	**Mean**	0.186	0.041	-	0.187
Between events	**Median**	0.018	0.041	-	0.018

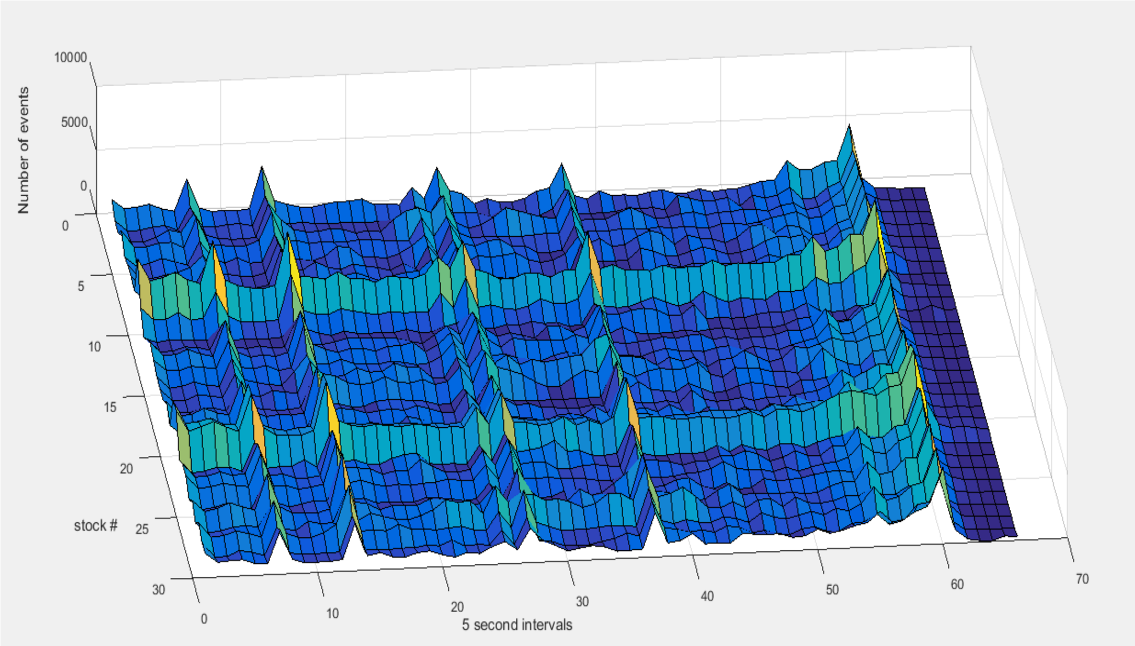
Sequences: algorithmic sequences from NPL; Repetition: Number of repetitions during the sequence; Duration (sec.): duration in seconds of a sequence; Duration (sec.) mean, median: average (median) duration between events in a sequence.

Figure 1. Shifted wavelet tree - structure



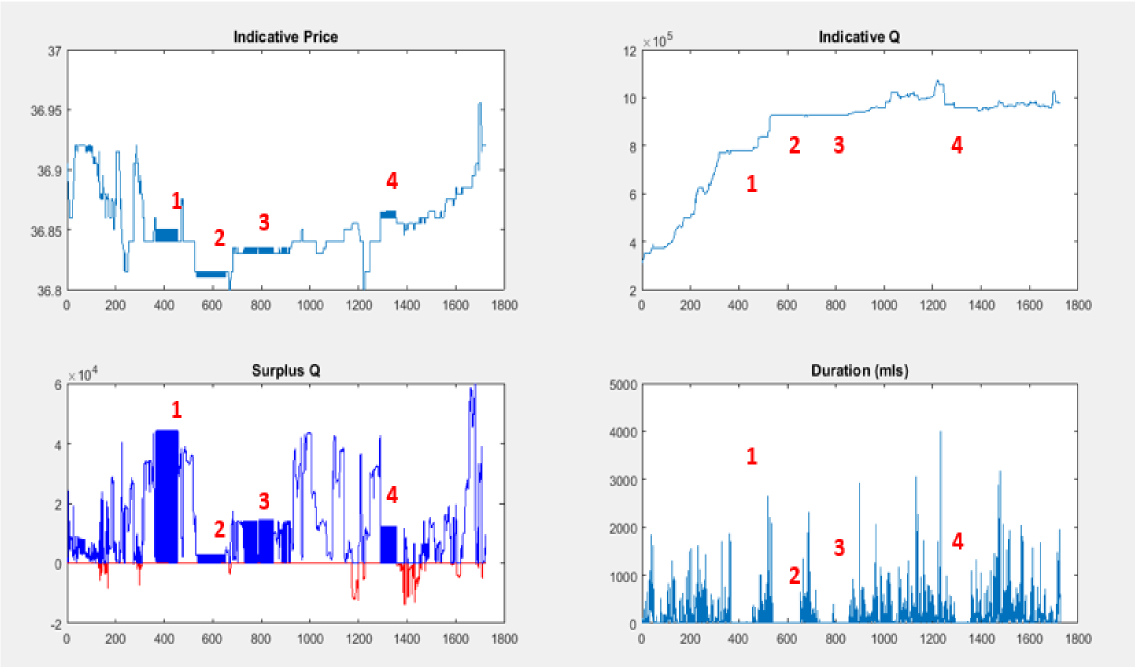
A graphical representation of SWT data structure. Level 0 contains aggregates from the smallest discrete time interval. Level 1 highest row is the pairwise sum of each pair of two consecutive data at level 0, starting with the first time interval of level 0. The lowest row of level 1 is the pairwise sum of pair of two consecutive data at level 0, starting with the second time interval of level 0. This creates the observed overlapping. The process is repeated for higher levels.

Figure 2. DAX components - closing auctions



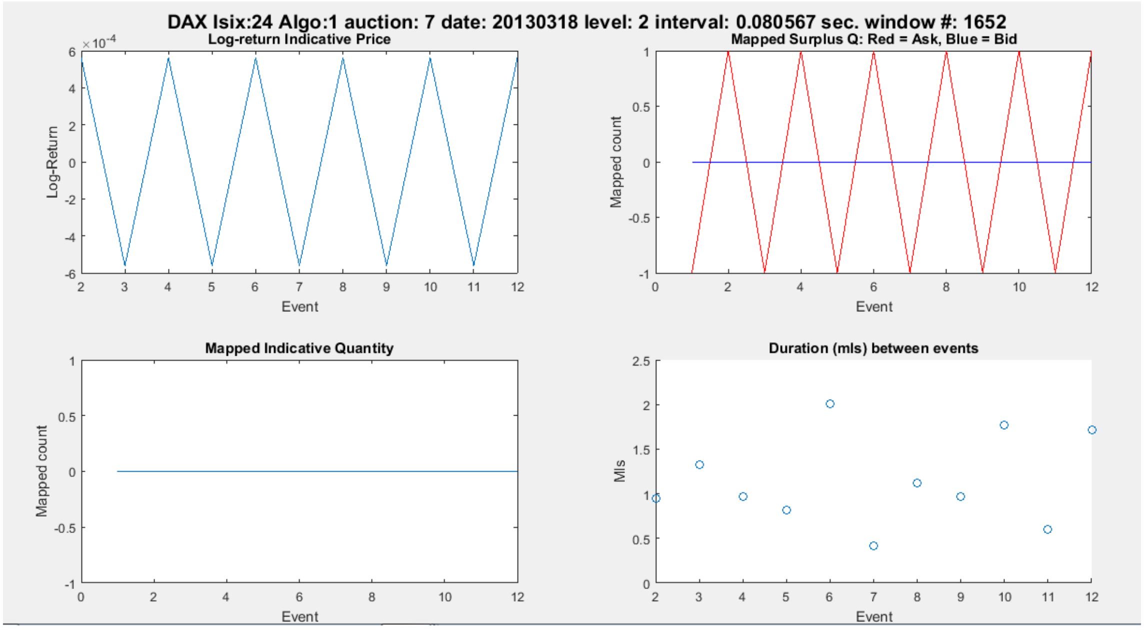
Cyclicity in the number of events affecting all stocks: institutional investors' imprints (Easley, Lopez de Prado et al. (2012)).

Figure 3. Deutsche Bank 20130212 closing auction



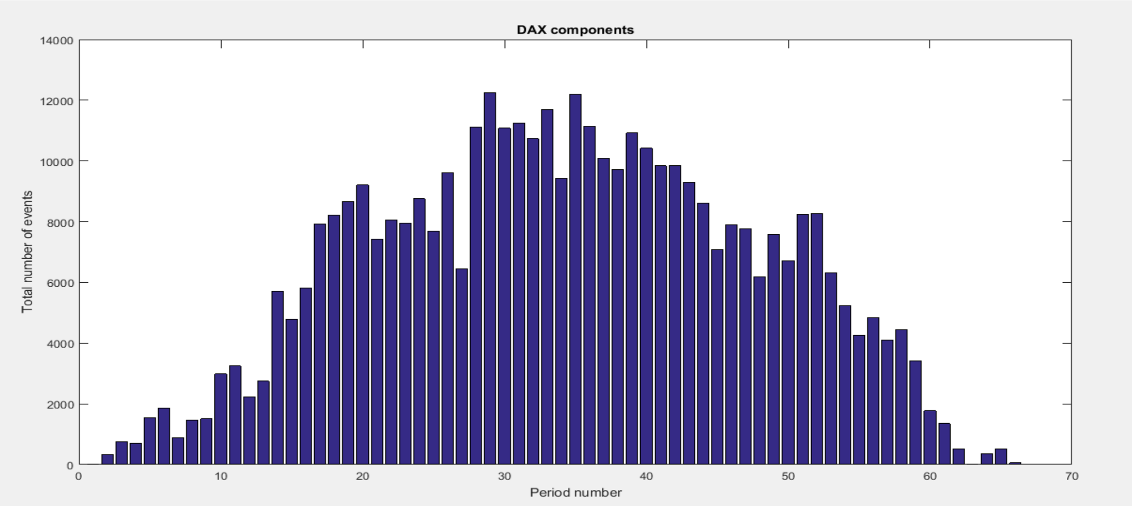
Indicative Q: matching quantity at the indicative price; Indicative price: price maximizing the matching quantity; Surplus Q: imbalance; Duration: time lapse between two events; A graphical representation of four NPLs labeled from 1 to 4. We comment on the graphics clockwise. Indicative price volatility changes with sequences. This volatility influences the indicative quantities marginally. NPL aggressiveness and/or LOB depth have significantly different impacts on imbalances (surplus quantities). All NPLs are executed via ultra high-frequency algorithms.

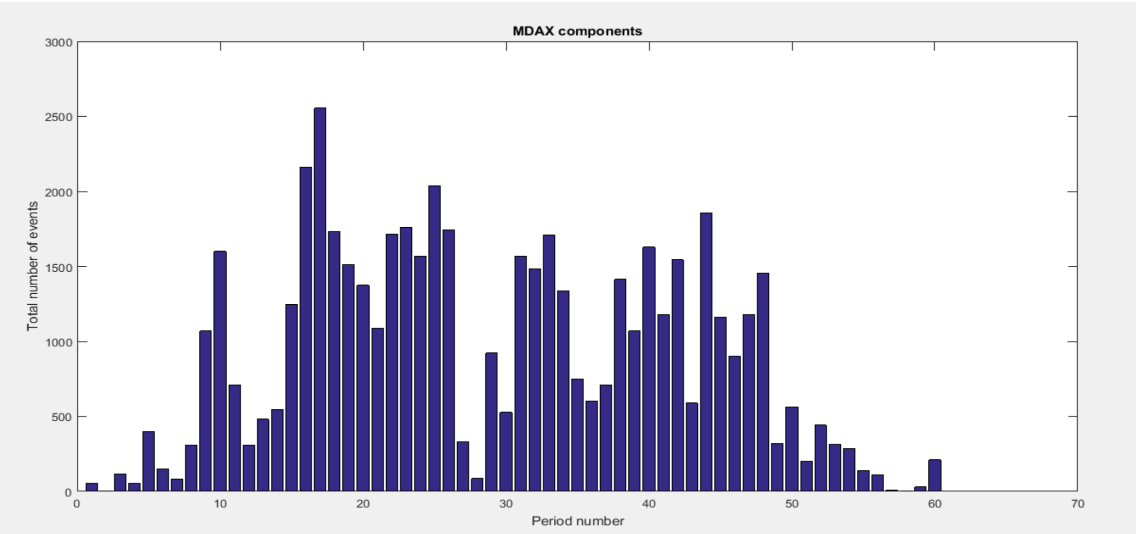
Figure 4. Identification parameters for negative price loops



Mapped surplus q ∈ (−1,1); mapped indicative q ∈ (−1,1); 0: no variation in mapped variable; mls: millisecond; Following a burst indicator alarm from a SWT tree, characteristics from a potential NPL are graphed. Events come from a time interval of 0.08 seconds (SWT level 2). We interpret the charts clockwise. The indicative price logarithmic returns and the surplus quantities vary symmetrically, confirming the NPLs' creation-cancellation sequence characteristics. Limit orders involved are inside the bid-ask spread on the offer side because matched (indicative) quantities do not fluctuate. Duration requires high-frequency technologies. DTW compares the sequence to pre-identified ones and diagnoses NPL existence.

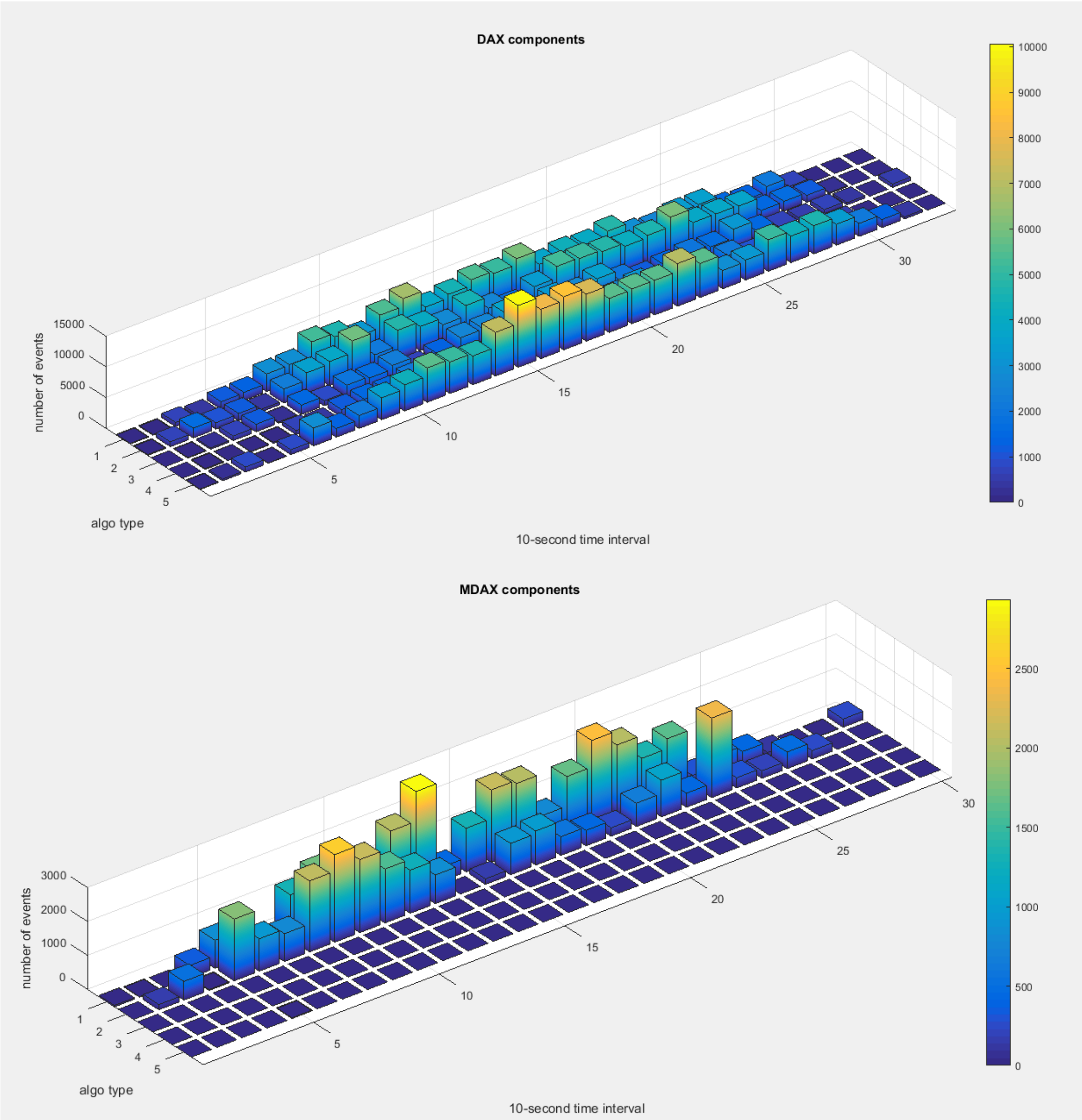
Figure 5. NPL events by 5-second intervals - closing auctions: aggregated statistics





Period number: number of 5-second periods elapsed since the beginning of the auction; total number of events: total number of events in the index for a given 5-second interval. February to July 2013.

Figure 6. NPL events by 10-second interval – closing auctions: type of algorithms



x axis: number of 10-second time intervals elapsed since the beginning of the auction; y axis: algorithmic sequence type; z axis: total number of algorithmic imprint occurrences. February to July 2013.

Figure 7. Transaction cost of NPL sequences

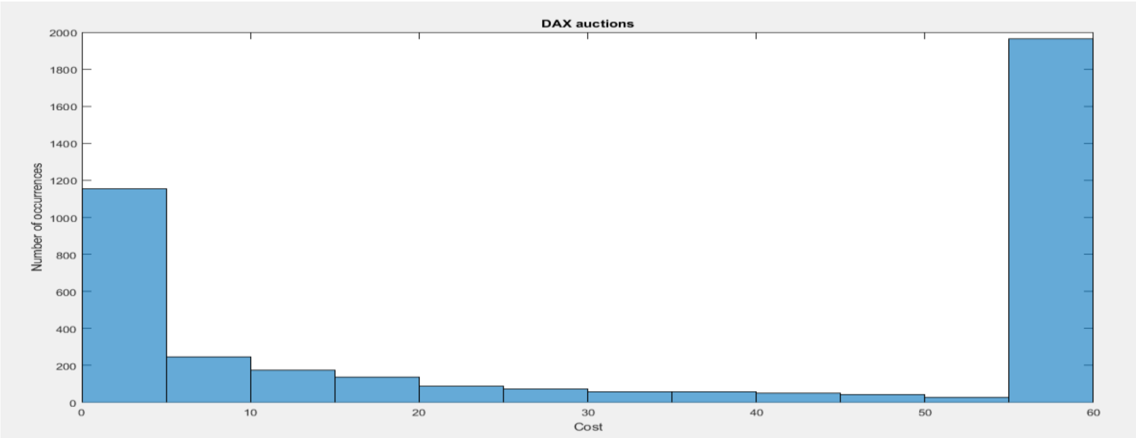
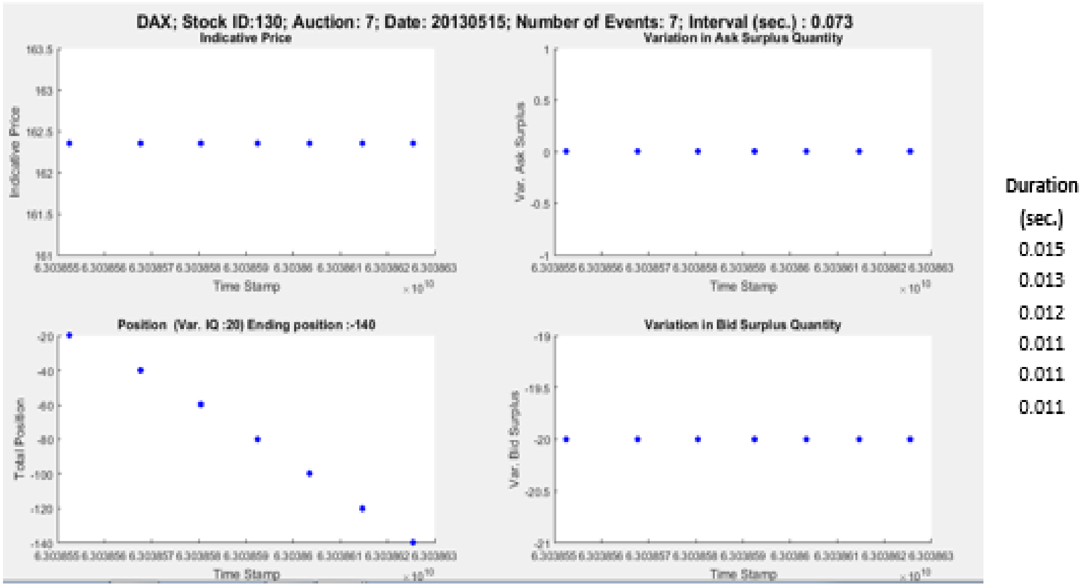


Figure 8. TWAP example - Volkswagen



Stock id: unique stock identifier; auction=7: closing auction; TWAP: time-weighted average price algorithm.