Journal of Financial Econometrics, 2025, **23(2)**, nbae025
https://doi.org/10.1093/jjfinec/nbae025
Article

OXFORD

Statistical Predictions of Trading Strategies in Electronic Markets

Álvaro Cartea[1,2], Samuel N. Cohen[1,2,3], Robert Graumans[1,4], Saad Labyad[1,2,3], Leandro Sánchez-Betancourt[1,2,3] and Leon van Veldhuijzen[4]

[1]Oxford-Man Institute for Quantitative Finance, University of Oxford, Oxford, OX2 6ED, UK
[2]Mathematical Institute, University of Oxford, Oxford, OX2 6GG, UK
[3]British Library, Alan Turing Institute, London, NW1 2DB, UK
[4]Authority for the Financial Markets, Amsterdam, 1017 HS, The Netherlands

Address correspondence to Leandro Sánchez-Betancourt, Mathematical Institute, University of Oxford, Andrew Wiles Building, Woodstock Rd, Oxford, OX2 6GG, UK, or e-mail: sanchezbetan@maths.ox.ac.uk.

Abstract

We build statistical models to describe how market participants choose the direction, price, and volume of orders. Our dataset, which spans 16 weeks for four shares traded in Euronext Amsterdam, contains all messages sent to the exchange and includes algorithm identification and member identification. We obtain reliable out-of-sample predictions and report the top features that predict direction, price, and volume of orders sent to the exchange. The coefficients from the fitted models are used to cluster trading behavior and we find that algorithms registered as Liquidity Providers exhibit the widest range of trading behavior among dealing capacities. In particular, for the most liquid share in our study, we identify three types of behavior that we call (i) directional trading, (ii) opportunistic trading, and (iii) market making, and we find that around one-third of Liquidity Providers behave as market markers.

Keywords: agent-based models, algorithmic trading, limit order book, supervision, statistical prediction

JEL classifications: C51, C52, C53, C55, D53, G10, G17, G18

Modeling economic agent is a fundamental part of understanding financial markets. The increasing use of electronic trading and the vast quantities of data recorded provide opportunities for modeling agents on a fine scale. However, privacy concerns and data access restrictions have led to limited academic research using such data. In this article, we employ a unique dataset to understand trading behavior. The dataset contains member identification and algorithm identification and comprises all activity (orders, transactions,

We thank the editor and two anonymous referees whose comments improved the contents and robustness of the article. We thank Steef Akerboom, Felix Flinterman, Ronald Verhoeven, Blanka Horvath, and Łukasz Szpruch for helping to bring about the collaboration between the Alan Turing Institute and the Autoriteit Financiële Markten. We are grateful to the participants at the Alan Turing Institute Knowledge Share Day, King's College London financial mathematics internal seminar, Oxford-Man institute internal seminar, and the SIAM Financial Mathematics and Engineering 2023 conference. We thank Micha Bender, Paul Besson, Patrick Chang, Charles-Albert Lehalle, and Jose Penalva for comments.

Received: December 21, 2023. **Revised:** June 24, 2024. **Accepted:** September 23, 2024. **Editorial decision:** September 18, 2024.

© The Author(s) 2024. Published by Oxford University Press.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Downloaded from https://academic.oup.com/jfec/article/23/2/nbae025/7826742 by guest on 20 June 2025

cancellations, and amendments) in Euronext Amsterdam. We model the trading behavior of each individual algorithm that sends limit orders (passive and aggressive) to the market. We obtain an accurate description of how algorithms choose direction, price, and volume and identify important market-observable and idiosyncratic features that predict this decision. Our work provides a detailed empirical description of how algorithms respond to market conditions, their own inventory, and their presence in the limit order book (LOB). Lastly, we find clusters of trading behavior and compare them to their dealing capacity in Euronext Amsterdam. A dealing capacity is a contractual arrangement between the trading venue and the trading member, specifying the trading behavior allowed by the member when acting in that capacity. This classification is determined by the trading venue, not the regulator.

The models we build have a variety of applications, both in academic and regulatory contexts. From an academic perspective, our findings provide statistical evidence for many microstructural stylized facts discussed in the literature, and our study is the first to confirm these facts using agent-based models trained on a dataset that contains both member and algorithm identification. Furthermore, by understanding the behavior of algorithms, our models can inform mathematical and algorithmic modeling of market participants. In an agent-based modeling context, our results provide guidance for calibration of models of the behavior of agents at a microstructure level, rather than through their aggregate behavior in a market. Our results assess the stability of agents through time, and report the key market and agent variables which need to be modeled.

From a regulatory perspective, our models provide guidance for surveillance practices. In particular, our models show the most important features affecting individual trading behavior. Supervisors could use this information to improve their anomaly detection, focusing on outliers in features that are especially important predictors of individual trading behavior. Also, our work is a starting point to build an agent-based simulator where market dynamics can be replicated at a granular level to understand how each message to the order book may affect individual behavior and therefore affect collective dynamics. Our statistical framework is devised to capture the fine microstructural facts that drive the trading decisions of individual algorithms, which is key to understanding how trading algorithms react to new market information and to messages to the LOB. With such a market simulator, firms will be able to test their trading algorithms and regulators will be able to study the effect of new trading algorithms on the quality of the market. Similarly, financial authorities will have a tool to build counterfactual trading tapes to gain insights into market behavior in the absence of certain trading algorithms or with potential new entrants in the market.

1.1 Insights

We build models to predict the behavior of individual trading algorithms with a dataset of beyond-level-three LOB data that contains the activity of all market participants active in Euronext Amsterdam. For each ticker and for each algorithm in our study, we propose models to predict the three principal characteristics of an incoming order: direction (buy or sell), price, and volume. By implication, the price, the time-in-force, and state of the LOB, determine whether orders are aggressive or passive. We build a comprehensive feature space comprising information available in the market and information available to the algorithm at the moment of making decisions. We do not account for the latency of market participants when constructing the features; instead, we employ information from the time "just before" the order reaches the market.[1] Specifically, we endow each algorithm with

[1] We expect that accounting for latency when constructing the features would improve the out-of-sample accuracies we obtain. See Cartea and Sánchez-Betancourt (2021) for examples of how one can compute implied latency at a market participant level.

Downloaded from https://academic.oup.com/jfec/article/23/2/nbae025/7826742 by guest on 20 June 2025

fifty-three features that are updated in continuous time and that describe the state of the LOB, the algorithm's past actions, and the member's past actions. The main contributions of this article are:

i) To obtain high out-of-sample accuracies when one predicts (i) the direction (buy or sell) of limit orders, (b) the price limit of the limit order (irrespective of its time-in-force), and (c) the volume of the order.
ii) To show the decrease in accuracy of predictions when we build features that do not use the identity of the member and do not use the identity of the algorithm.
iii) To provide a detailed description of the features that are important to predict the direction, price, and volume of limit orders sent by algorithms. For example, (a) the imbalance of the algorithm's limit orders resting in the LOB, the best quotes, and the intraday accumulated inventory variables, are the most important features determining the direction of the order; (b) the volumes at the best bid and ask prices, the imbalance of the algorithm's limit orders resting in the LOB, and the bid–ask spread in the LOB are the most important features determining the price of the order, and (c) the intraday accumulated inventory and cash variables of both algorithms and members are the most important features determining the volume of the order.
iv) To show three clusters of "stylised trading behaviour." We use regression coefficients to build the clusters and to show how these clusters compare with the registered dealing capacity of market participants.[2] For the most liquid share considered, we find that:
    a) Liquidity Providers have the widest range of trading behavior in the dataset.
    b) One-third of the Liquidity Provider algorithms fall within a cluster that we label "market markers." These algorithms tend to maintain a balanced provision of liquidity in the LOB, provide liquidity inside the spread, revert their inventories to zero, and often provide liquidity at the best quotes available in the book.[3] The other two-thirds of Liquidity Providers are split roughly in half according to their trading behavior and fall in clusters that we call "opportunistic traders" and "directional traders."
    c) Across clusters, the wider the bid–ask spread the more likely algorithms are to send eager-to-trade orders. In this article, an "eager-to-trade" order is either a buy limit order with limit price higher than the best bid price, or a sell limit order with limit price lower than the best ask price—essentially (as we see in Section 4.5.8), these are orders that either cross the spread or provide liquidity inside the spread. Similarly, an increase in the number of messages over the last 1–100 milliseconds reduces the probability of posting at-the-touch in either direction. In this article, an "at-the-touch" order is either a buy limit order with a limit price equal to the best bid price, or a sell limit order with limit price equal to the best ask price. Lastly, for most horizons used to compute the quadratic variation of transaction prices, the larger the quadratic variation, the less likely algorithms will show aggressive behavior, and the more likely it is that they will post liquidity inside the spread.
        i) Algorithms in the first cluster (directional traders) have the lowest order to trade ratios, do not maintain a balanced provision of liquidity in the LOB, seem indifferent to their own imbalance in the book when they decide to send an eager-to-trade order, have a small baseline probability of posting at-the-touch orders or improving the bid–ask spread, and often send orders in the direction of their

[2] The three dealing capacities of market participants in our study are (i) Liquidity Provider, (ii) House, and (iii) Client; we return to this in Section 2.2.

[3] An agent has a balanced provision of liquidity in the LOB if the volume that the agent has posted on the ask side of the LOB is similar to the volume that the agent has posted on the bid side of the LOB.

Downloaded from https://academic.oup.com/jfec/article/23/2/nbae025/7826742 by guest on 20 June 2025

intraday accumulated inventories (i.e., buy orders when they have a positive intraday accumulated inventory; sell orders when they have a negative intraday accumulated inventory). More than 80% of their eager-to-trade orders trade aggressively against the other side of the LOB. This behavior is consistent with execution or position-building algorithms; the split of aggressive and passive trading shows how algorithms choose to deploy their execution schedule.[4]

ii) Algorithms in the second cluster (opportunistic traders) trade aggressively (73% of their transactions are liquidity taking, compared to 51% for cluster 1 and 26% for cluster 3), and similar to algorithms in cluster 1, more than 80% of their eager-to-trade orders trade aggressively against the other side of the LOB. These algorithms do not mean-revert their intraday accumulated inventory and their own imbalance in the book is an indicator of their trading direction. These algorithms have the lowest (across clusters) percentage of eager-to-trade orders that improve the spread, the lowest baseline probability of posting orders at-the-touch, and the highest baseline probability of posting limit orders deeper in the LOB. These algorithms post most of their orders away from the best quotes in an attempt for higher but less frequent profits.

iii) Algorithms in the third cluster (market makers) have the highest order to trade ratio across clusters, are keen to maintain a balanced provision of liquidity in the LOB, have the highest baseline probability of posting at-the-touch (when compared with the algorithms in the other two clusters), and show eagerness to revert their intraday accumulated inventory to zero. Algorithms in this cluster have the highest baseline probability of sending an order that provides liquidity inside the spread. More precisely, 81% of the eager-to-trade orders sent by algorithms in this cluster were limit orders that provided liquidity inside the spread. This behavior is consistent with that of market makers who are averse to holding inventories (e.g., due to asymmetry of information).[5]

Using the above we discuss how supervisors can use the results of our study for the benefit of orderly trading and to have oversight on the integrity of the market.

1.2 Connection with Survey Data

The recent article AFM (2023) reports the results of a survey conducted with a sample of Dutch proprietary trading firms using algorithmic trading.[6] The article reports that (i) more than 80% of their trading algorithms use machine learning, (ii) their trading algorithms use features such as quantities in the LOB, price trends, volatility, and volume imbalances to make short-term predictions, (iii) model parameters or model weights are updated frequently, (iv) information of the recent past has predictive power in the short future, (v) the list of possible actions that an algorithm can take after a prediction is limited and hard coded, and (vi) firms tend to use simple models, such as linear and logistic regressions, to make predictions.

In this article, we find evidence for a number of these items. For example, our analysis demonstrates that one can predict their trading actions based on features containing quantities in the LOB, price trends, volatility, and their accumulated intraday inventory. We validate that more recent information has higher predictive power than older information, and that simple models, such as the one predicting the most used price range in the past or

[4] See Cartea and Jaimungal (2015) for a model of how traders optimally deploy such an execution schedule using aggressive and passive orders.

[5] See the early work Amihud and Mendelson (1980), Ho and Stoll (1981), Kyle (1985), Glosten and Milgrom (1985), and the more recent work Avellaneda and Stoikov (2008), Guéant, Lehalle, and Fernandez-Tapia (2013), and Cartea, Jaimungal, and Penalva (2015) and Guéant (2016).

[6] Trading firms dealing exclusively on own account.

Downloaded from https://academic.oup.com/jfec/article/23/2/nbae025/7826742 by guest on 20 June 2025

the most used volume range in the past, have high accuracies. This lends strong support to the assertion that the actions available to algorithms are hard coded and the number of actions limited. Lastly, we show that (multinomial) logistic regression models have high out-of-sample accuracies, which lends support to algorithms using simpler, faster models.

1.3 Existing Literature

The LOB plays a crucial role in the microstructure of many modern financial markets. A LOB contains information available on a specific market and reflects the past trading decisions of its participants. Abergel et al. (2016) discuss several models for the LOB covering important techniques such as agent-based modeling. Bouchaud, Mézard, and Potters (2002) investigate empirical statistical properties of equity LOBs. Cont, Stoikov, and Talreja (2010) propose a model for LOB dynamics that recovers aspects of the empirical properties satisfied by LOBs; further work in this direction is in Hambly, Kalsi, and Newbury (2020) where the authors provide a probabilistic description of LOB dynamics. Gould et al. (2013) present a detailed account of LOBs, together with statistical analyses of historical LOB data. They discuss how LOB models provide insights into a number of economic questions (e.g., questions regarding market efficiency, price formation, or the rationality of traders), but poorly resemble real LOBs and that several empirical facts have yet to be reproduced satisfactorily.

Agent-based models comprise a number of decision-makers (agents) interacting through prescribed rules. When applied to the economy as a whole, agents can be as diverse as needed—from consumers to policy-makers to investment banks. These models do not assume that the economy will move towards a predetermined equilibrium state, as other models do; see for example, Farmer and Foley (2009). Instead, at any given time, each agent acts according to its current situation, the state of the environment, and the rules governing its behavior.

Farmer, Patelli, and Zovko (2005) apply agent-based modeling to financial markets. They use data from the London Stock Exchange to test a simple model in which zero intelligence agents place orders to trade at random. The model recovers simple stylized facts about the arrival rates of orders; see Lehalle, Guéant, and Razafinimanana (2011) for an overview of the distinction between a zero-intelligence approach to LOB modeling and agent-based models. Byrd, Hybinette, and Balch (2020) introduce the design and implementation of ABIDES, a multi-agent equity market simulator. Assefa et al. (2020) discuss the importance of having reliable synthetic data generators for financial markets. Cohen, Snow, and Szpruch (2023) discuss the role of synthetic data in a modern machine-learning finance context. Vyetrenko et al. (2020) survey the literature and collect statistics and stylized facts seen in real markets that a market simulator should be able to recreate. They show that these models explain a large part of the variance in the bid–ask spread and price diffusion rate. In our work, the coefficients of the regression models open the door to taking agent-based modeling one step further; we enable building a market model with agents whose actions (direction, price, and volume) are learned from data.

Agent-based modeling can also provide relevant insights for supervisors looking at market manipulation. Wang et al. (2021) present an agent-based model where agents spoof the LOB to manipulate prices: submit spurious orders to mislead market participants. They demonstrate that traders who use historical data to predict prices improve price discovery and social welfare, but their existence in equilibrium renders a market vulnerable to manipulation. That is, spoofing strategies can effectively mislead traders, distort prices, and reduce total surplus. Our study provides the foundations to create more realistic agent-based models, because we show that the trading decisions of agents can be modeled and predicted well. Furthermore, we show which features are most important in predicting the direction,

Downloaded from https://academic.oup.com/jfec/article/23/2/nbae025/7826742 by guest on 20 June 2025

price, and volume of an order; see also Cartea, Jaimungal, and Wang (2020), Tao et al. (2022), and Williams and Skrzypacz (2020).

Roşu (2009) presents a model of an order-driven market in which fully strategic, symmetrically informed liquidity traders dynamically choose between limit and market orders. The model makes a number of empirical predictions, such as (i) higher trading activity and higher trading competition cause smaller spreads and lower price impact, and (ii) market orders lead to a temporary price impact larger than the permanent price impact. Sirignano and Cont (2019) find evidence of a universal relationship between order flow and the direction of price moves—they find stable out-of-sample accuracies across shares and periods. Our results also show stable out-of-sample accuracies when predicting direction, price, and volume of orders sent to the exchange by market participants. The models we build in this article can be used to answer questions in the two aforementioned studies. In particular, we report how the probability of posting liquidity inside the spread and the probability of crossing the spread change according to key features (e.g., volume imbalances). We find that the higher the recent activity the more likely algorithms are to submit orders that are eager to trade.

Bouchaud et al. (2003) use trades and quotes data from the Paris stock market to show that the random walk nature of traded prices results from an interplay between two opposite tendencies: long-range correlated market orders and mean reverting limit orders. They define and study a model where the price is the result of the impact of all past trades. Aït-Sahalia et al. (2022) use machine learning to study the predictability of ultra-high-frequency share returns. They find that, contrary to low-frequency and long horizon returns, predictability of high-frequency returns is large, systematic, and pervasive over short horizons. Our study employs a number of the features they use, specifically by taking features over various short time horizons. We show how these features remain predictive, even when one takes into account algorithm-specific information that is not visible to the market.

Hendershott, Jones, and Menkveld (2011) focus on the impact of algorithmic trading on market quality, and find that algorithmic trading narrows spreads, reduces adverse selection, and reduces trade-related price discovery. Brogaard, Hendershott, and Riordan (2014) find that high-frequency traders help to stabilize prices during transitory shocks. Menkveld (2016) argues that increased speed in processing information reduces information asymmetry, at least for informed investors who obtain their information advantage from public news. Aït-Sahalia and Saglam (2013) find that if a market maker can predict the direction of future market orders, then the liquidity provided by the market maker improves. Cartea and Penalva (2012) propose that greater speed could allow fast traders to profitably intermediate between liquidity demanders and liquidity suppliers; see also Cohen and Szpruch (2012). This additional intermediation layer would increase execution costs and microstructure volatility. Furthermore, Cartea et al. (2019) find that an increase in ultra-fast machine-driven activity is associated with lower intraday market quality (greater quoted and effective spreads and lower depth). Closest to our work is that by Goldstein, Kwan, and Philip (2023), who employ five LOB features to predict whether a submitted order would be a passive execution, a limit order submission, an amendment, or a cancellation. They also conduct a number of additional regressions to describe variables such as order imbalance and probability of fills. In our article, we build regression models to predict the direction, price, and volume attached to an incoming order based on market features and private information at the time just before the order is processed by Euronext. Unlike in Goldstein, Kwan, and Philip (2023), we model passive execution and limit order submission together, and we exclude amendments and cancellations. We show that our

Downloaded from https://academic.oup.com/jfec/article/23/2/nbae025/7826742 by guest on 20 June 2025

models obtain reliable out-of-sample accuracies and we discuss how our models can be used to build agent-based market simulators.

Our article focuses on the quoting behavior of trading algorithms, so it complements research into high-frequency trading (HFT).[7] Menkveld (2016) groups traders into high-frequency traders (HFTs) and other traders. HFTs are further grouped into high-frequency market makers, who trade passively by submitting bid and ask quotes, and high-frequency bandits, who trade aggressively by taking out stale quotes. Here, we identify a cluster of algorithms that shows market making behavior; the median resting time of orders sent by algorithms in this cluster is the lowest of all clusters. Also relevant to our article is that by Brogaard, Hendershott, and Riordan (2019), who document that the majority of price discovery occurs with quote updates as opposed to trades. Hasbrouck (2018) argues that regular quote changes are likely the result of HFTs undercutting each other after market orders remove price quotes from the book. Megarbane et al. (2017) study the behavior of HFTs and their role in liquidity provision under market stress scenarios such as high-volatility periods surrounding news announcements. Hoffmann (2014) analyses a model where HFTs quickly cancel their outstanding limit orders after news. Hagströmer and Nordén (2013) use data from Nasdaq-OMX Stockholm to distinguish between two types of HFTs: market makers and opportunistic traders. In our article, we find evidence that the range of trading behavior of Liquidity Providers is wide. In particular, we find that one-third of Liquidity Providers behave as market makers and the other two-thirds are split into opportunistic traders and directional traders.

For a review on statistical approaches to modeling HFT data see Dutta et al. (2022); they focus on modeling the aggregated publicly available data as opposed to individual agents. There are strands of the literature that find that deeper layers of the order book contain useful information; see, for example, Libman, Haber, and Schaps (2021). In our models, the volume posted deeper in the LOB has predicting power within certain clusters; this confirms that the results of Libman, Haber, and Schaps (2021) are applicable to particular types of market participants. We also find that around one-third of Liquidity Provider algorithms behave as market makers and improve the spread around 86% of the times they send an eager-to-trade order. In comparison, directional traders and opportunistic traders improve the spread only 17% and 10% (respectively) of the times they send an eager-to-trade order.

This article also builds upon previous research on the clustering of agents. Mankad, Michailidis, and Kirilenko (2013) develop a dynamic machine learning method that buckets 15,686 traders in the E-mini S&P 500 futures into five persistent categories: high-frequency traders (14 traders), market makers (271 traders), opportunistic traders (7126), fundamental traders (254), and small traders (8021). Wright, Reimherr, and Liechty (2022) introduce machine learning methods that cluster active traders in the market, which can be used for intraday classification. With similar data to ours, Ruan, Bacry, and Muzy (2023) classify short-term strategies and identify market marker and directional trading clusters. Cont et al. (2023) present a granular representation of the LOB that accounts for the origins of different orders. They describe client order flow from a large broker, in particular, they segment clients into different clusters, for which they identify representative prototypes. Our article provides a complementary analysis; we cluster the trading behavior of algorithms, and compare these clusters with the dealing capacity in Euronext Amsterdam.

1.4 Key Contributions

This article constructs models that describe, with high out-of-sample accuracies, how individual agents choose the (i) direction, (ii) price, and (iii) volume, of limit orders sent to

[7] See Menkveld (2016) for a review.

Downloaded from https://academic.oup.com/jfec/article/23/2/nbae025/7826742 by guest on 20 June 2025

Euronext. We compute the decrease in prediction accuracy when we build features that do not use member identification and do not use algorithm identification. For example, without algorithm and member identifications, the out-of-sample accuracy of predicting direction is 4% higher than random guessing (which is 50%), whereas we obtain 70% order-weighted accuracy when we build features that use both member and algorithm identification.

We use the coefficients of our models to cluster the behavior of algorithms. We identify three behavioral clusters (directional, opportunistic, and market making) and provide a number of insights that, to the best of our knowledge, have few analogs in the literature. For example, we find that the range of behavior of Liquidity Providers is the widest among dealing capacities—their algorithms are spread over the three clusters, where around one-third are clustered as market making algorithms. Additional contributions are (i) to provide a comprehensive list of feature importance when predicting direction, price, and volume, (ii) to discuss the implications of our models for regulators, and (iii) to show how to use our models to build agent-based market simulators.

2 Data Description

We use data from Euronext Amsterdam spanning the period 11 October 2021 to 30 January 2022 (80 trading days in 16 weeks) in four shares with tickers ASML, INGA, AD, TOM2—we refer to these shares as ASML, ING, AHOLD, and TOMTOM respectively. The data is labeled with algorithm identification and member identification, and contains transactions, orders, cancellations, and amendments processed by Euronext Amsterdam over the relevant period. Orders are matched via price-time-priority and are displayed in a central LOB.[8] The data are timestamped with nanosecond accuracy. The AFM receives the data in real time from the data systems in Euronext Amsterdam.

2.1 Assets Studied

In our study, the companies represent different sectors in the capital markets (technology, finance, and consumer goods) and their market capitalization ranges from large to small. The shares differ in the number of trading venues on which they trade, the number of derivatives written on the share, and the number of indices in which the share is a constituent. Euronext Amsterdam is the principal market in which the four shares trade. Specifically,

i) ASML Holding (ticker ASML) specializes in the development and manufacturing of machines that produce computer chips, its market capitalization is approximately €200B, it is traded on various trading venues, and it is a constituent in a number of indices.[9] We focus on ASML in the main text of the article; results for the other shares are in the Appendix.
ii) ING Groep N.V. (ticker INGA) is a Dutch multinational banking and financial services corporation with headquarters in Amsterdam, its market capitalization is approximately €50B, it is traded on various trading venues, and it is a constituent in a number of indices.[10]

[8] See (Euronext, 2023d, p. 43) for information about the main trading session and p. 44–46 for market mechanisms.

[9] Traded on more than 100 trading venues in 2022 according to AFM MIFID-II transaction data, where 47% of transactions in the period of our study are on Euronext Amsterdam, 21% on Cboe Europe (DXE order books), etc.; see Euronext (2023b) for more information on ASML.

[10] Traded on more than 100 trading venues in 2022 according to AFM MIFID-II transaction data, where 49% of transactions in the period of our study are on Euronext Amsterdam, 19% on Cboe Europe (DXE order books), etc.; see Euronext (2023c) for more information on ING.

Downloaded from https://academic.oup.com/jfec/article/23/2/nbae025/7826742 by guest on 20 June 2025

Table 1. Statistics for ASML, ING, AHOLD, and TOMTOM between 11 October 2021 and 30 January 2022

	Avg daily tradecount	Avg daily traded volume €	Lowest price period €	Highest price period €	Market share (Euronext) %
ASML	41,847	543,088,447	566.10	770.50	47
ING	21,469	186,226,078	11.58	13.62	49
AHOLD	10,535	78,665,021	27.54	31.35	44
TOMTOM	1385	4,329,991	6.00	9.23	55

iii) Koninklijke Ahold Delhaize N.V. (ticker AD) is a Dutch multinational retail and wholesaling company, its market capitalization is approximately €30B, it is traded on various trading venues, and it is a constituent in a number of indices.[11]
iv) TomTom (ticker TOM2) is a Dutch multinational developer and creator of location technology and consumer electronics, its market capitalization is approximately €1B, it is traded on fewer trading venues, and it is a constituent in a few indices.[12]

Table 1 presents summary statistics of the four shares between 11 October 2021 and 30 January 2022.

The tick size of each share traded on Euronext Amsterdam depends on the price range in which the share trades. These rules are unique for each share. ASML can trade in nineteen possible tick sizes; for example, tick size is €0.05 if the share trades between €200 and €500, and it is €0.1 if it trades between €500 and €1000.

The tick sizes of the shares do not change between 11 October 2021 and 30 January 2022. During this period, the tick size of ASML is €0.1, the tick size of ING is €0.002, the tick size of AHOLD is €0.005, and that of TOMTOM is €0.005. We observe that the tick sizes of ASML, ING, and AHOLD are similar tick sizes as a percentage of their share price (between 1.3 and 1.8 basis points), while the tick size of TOMTOM is larger as a percentage of its share price.[13]

2.2 Euronext: Members and Algorithms

In Euronext's rule book (Euronext 2023d), a member is any individual, corporation, partnership, association, trust, or entity who has been admitted to Euronext Securities Membership or Euronext Derivatives Membership and whose membership has not been terminated. Only members can trade on Euronext Amsterdam. A member can trade on its own account or for clients. The latter includes retail-brokers who process so-called "retail orders." A member can trade under one or multiple dealing capacities.

One of the dealing capacities is "retail liquidity provider." Orders of members acting as retail liquidity providers can be matched only with orders submitted by "retail" members—see Euronext (2023d, p. 14). Other dealing capacities (each with its own set of obligations) are (i) Liquidity Provider, (ii) House, and (iii) Client. This classification is determined by Euronext, not the regulator.

In this study, an "agent" is the concatenation of the membership, the dealing capacity, and the "entity" that sends the order. We provide a stylized example: suppose company

[11] Traded on more than 100 trading venues in 2022 according to AFM MIFID-II transaction data, where 44% of transactions in the period of our study are on Euronext Amsterdam, 23% on Cboe Europe (DXE order books), etc.; see Euronext (2023a) for more information on AHOLD.

[12] Traded on more than 50 trading venues in 2022 according to AFM MIFID-II transaction data, where 55% of transactions in the period of our study are on Euronext Amsterdam, 16% on Cboe Europe (DXE order books), etc.; see Euronext (2023e) for more information on TOMTOM.

[13] There is an extensive literature on the effect of tick sizes on liquidity; see Verousis, Perotti, and Sermpinis (2018) or Penalva and Tapia (2021) for relevant references.

Downloaded from https://academic.oup.com/jfec/article/23/2/nbae025/7826742 by guest on 20 June 2025

ABC is a member of the trading venue and trades in two dealing capacities: Liquidity Provider and House.

When company ABC trades as a Liquidity Provider, it employs either: algorithm X, or algorithm Y. When it trades as a House, it employs algorithm Z. The agents within company ABC are (i) ABC-LP-X, which refers to algorithm X within the Liquidity Provider dealing capacity of ABC, (ii) ABC-LP-Y, which refers to algorithm Y within the Liquidity Provider dealing capacity of ABC, and (iii) ABC-H-Z, which refers to algorithm Z within the House dealing capacity of ABC.

Euronext data shows which entity within the member executes a given order, but not if this entity is (i) a human trader, or (ii) a trading algorithm. However, by cross-checking with regulatory data we know that all Liquidity Provider agents, and most of the agents within House and Client, are trading algorithms. Therefore, we use the terms "agent" and "algorithm" interchangeably.

Figure 1 shows the number of algorithms associated with each of the members trading in ASML, grouped by dealing capacity (Liquidity Provider, House, and Client). We show algorithms that were active each week throughout the 16 weeks of data. We order members according to the number of orders they sent to Euronext; highest to lowest order count appears from top to bottom.

We label algorithms with the identifier reported by the firm using the algorithm. This is equivalent to the so-called "execution within firm" field in the MIFID-II reported transactions—field 59 in EU (2016a). In our dataset, from one month to the next, there are on average 7.5 new algorithms sending orders, and on average, 7.7 algorithms stop sending orders.[14] Although these numbers might appear high, their effect on order submission is small. New algorithms account, on average, for 0.31% of all orders sent per month.

Roughly, the algorithms that send the highest number of orders to Euronext are those with a Liquidity Provider dealing capacity. Out of the twelve members that have algorithms registered as Liquidity Providers, ten of these members trade exclusively with the Liquidity Provider dealing capacity.

2.3 Data Filters and the Orders We Study

In this study, we apply a number of filters to the data. First, we focus on messages entered between 09:05:00 and 17:25:00. This excludes the first and last five minutes of the main trading session and also omits messages sent during the pre-opening phase and closing phase.

Second, we ignore any undisclosed volume in the LOB when constructing the features of a given algorithm. This consists of volumes in the order book that—similar to iceberg orders—are not visible to market participants. We note, however, that under 0.1% of all orders in any of our shares include undisclosed volume.

Third, in creating our features, we ignore all orders, deletions, and amendments by agents classified as "retail liquidity providers" (retail LPs). This specific dealing capacity has the unusual behavior that their orders, deletions, and amendments (including the resulting transactions and volumes in the order book) are visible only to "retail" members. Retail clients are able to trade with both retail LPs and the rest of the order book. In effect, this creates two LOBs, one for retail members and one for all other agents. In our dataset, we exclude retail LPs' orders, and all retail member behavior which matches against these orders, because we are not concerned with retail trading behavior. The remaining orders of retail members are included in our dataset, for the purposes of determining the information available to the market, but we will not attempt to fit models for the behavior of retail members.

[14] For ASML there are ninety-six algorithms active each week during the 16 weeks of data.

Downloaded from https://academic.oup.com/jfec/article/23/2/nbae025/7826742 by guest on 20 June 2025

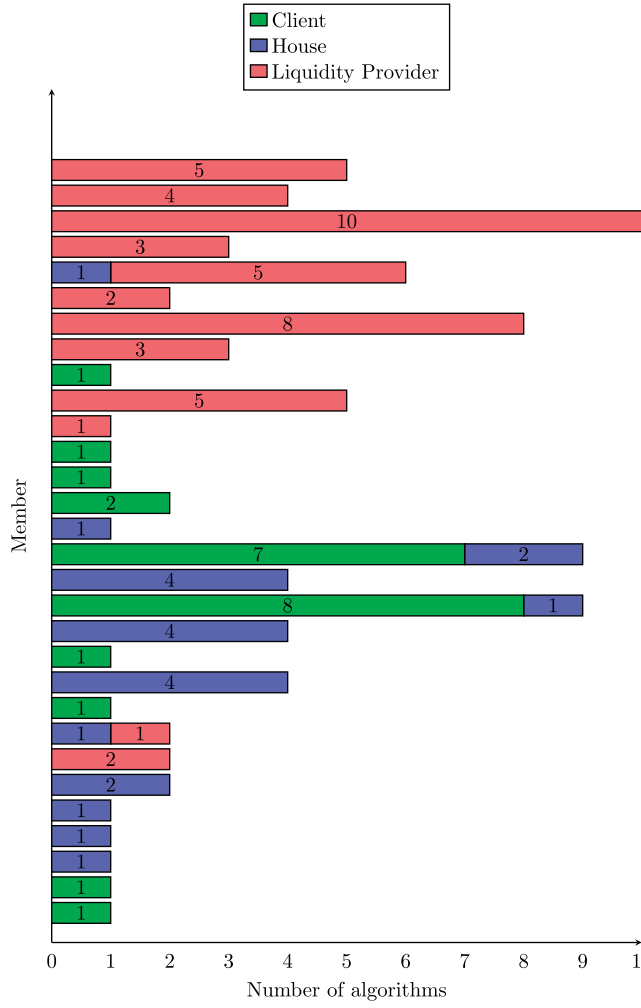


Figure 1. Number of algorithms and dealing capacities per member on Euronext Amsterdam in ASML. The members are ordered from top to bottom using order count. We show ninety-six algorithms active each week for the 16 weeks of data.

2.3.1 Order types

The order types available in Euronext are (i) limit orders, (ii) market orders, (iii) stop orders, (iv) pegged orders, and (v) mid-point orders. Table 2 shows the number of trades in the first 4 weeks of data for ASML for the following combinations: (i) the buy order is a limit order and the sell order is a limit order, (ii) the buy order is a market order and the sell order is a limit order, (iii) the buy order is a limit order and the sell order is a market order, and (iv) all other combinations.

The algorithms in our dataset mostly use limit orders, instead of market orders, to take liquidity. Approximately 97% of all transactions result from matching two limit orders. Specifically, in 97% of transactions one leg is a limit order that rests in the LOB and the other leg is an incoming limit order with a limit price generous enough to match the first leg.

Downloaded from https://academic.oup.com/jfec/article/23/2/nbae025/7826742 by guest on 20 June 2025

Table 2. Order types of transactions in ASML between 11 October 2021 and 7 November 2022

Order type		Trade count	
Buy side	**Sell side**	**Number**	**Percentage (%)**
Limit order	Limit order	752,922	97
Market order	Limit order	6686	1
Limit order	Market order	4974	1
All other combinations		10,446	1

Table 3. Validity types of transactions in ASML between 11 October 2021 and 7 November 2022

Validity parameter		Trade count	
Buy side	**Sell side**	**Number**	**Percentage (%)**
IoC	DAY	283,545	37
DAY	IoC	280,615	36
DAY	DAY	195,691	25
All other combinations		15,177	2

2.3.2 Time-in-force

Orders submitted to Euronext have a validity parameter—also known as the time-in-force of the order. Table 3 shows the trade count grouped by the validity parameters of the buy and sell orders that were matched. More precisely, we show the number of transactions in the first 4 weeks of data for ASML for the following combinations: (i) the buy order has validity parameter immediate-or-cancel ("IoC") and the sell order has validity parameter good-for-day ("DAY"), (ii) the buy order has validity parameter DAY and the sell order has validity parameter IoC, (iii) the buy order has validity parameter DAY and the sell order has validity parameter DAY, and (iv) all other combinations (for example, fill-or-kill).

In the first two rows of Table 3, liquidity is consumed with an IoC order. In the third row, liquidity is taken with a DAY order. IoC and DAY orders account for 98% of all transactions. Naturally, in Table 3, it is always the case that one of the two legs has DAY as a time-in-force, which represents the liquidity resting in the LOB waiting to be matched; the other leg is IoC roughly two-thirds of the time and DAY roughly one-third of the time.

2.3.3 The orders that we model

In this article, we model limit orders (including all possible validity parameters). As shown in the two tables above, these orders account for almost all trades in Euronext.

Table 4 shows the order and validity types of all limit orders sent during the first 4 weeks of data for ASML. This constitutes a four-week snapshot of the orders that we model in this article.

Limit orders with validity type DAY are 97% of all limit orders sent, and orders with validity type IoC account for only 3%. We also observe that IoC orders are often filled by more than one DAY order, which explains the difference between the figures in Tables 3 and 4.

Table 5 uses the first 4 weeks of data for ASML and reports the (contribution to) order count, trade count, and volume traded of groups of algorithms. More precisely, the table summarizes the aggregate contribution to (i) order count, (ii) trade count, and (iii) volume traded of the top five, top ten, top twenty, top fifty, and all algorithms. The ranking of algorithms is based on the order count of algorithms during the first 4 weeks of data.

Downloaded from https://academic.oup.com/jfec/article/23/2/nbae025/7826742 by guest on 20 June 2025

Table 4. Order and validity types of orders in ASML between 11 October 2021 and 7 November 2022

Order type	Validity type	Order count	Percentage (%)
Limit order	DAY	7,350,535	97
	IoC	252,744	3
	Other	10,210	0

Table 5. Descriptive statistics for algorithms trading in ASML between 11 October 2021 and 7 November 2022

	Order count	Order count Percentage of all %	Trade count	Trade count Percentage of all %	Volume €	Volume Percentage of all %
Top 5	2,612,307	34	382,040	25	2,332,746,616	19
Top 10	4,055,626	53	478,985	31	2,972,717,749	25
Top 20	5,652,783	74	679,310	44	4,431,976,153	37
Top 50	7,275,195	96	1,047,174	68	8,086,989,709	67
All	7,613,489	100	1,550,056	100	11,998,022,616	100

We note that each trade requires the involvement of two participants, leading to a double-counting of the number of trades in Table 5.

The top ten algorithms account for more than 50% of the orders, are involved in more than 30% of the trades, and are involved in more than 20% of the total traded volume. While algorithms below the top 50 provide only 4% of the orders, they are involved in 32% of all trades (and 33% of all traded volume by value). In the Appendix, we report the above statistics for ING, AHOLD, and TOMTOM in Appendix Tables A.1, A.2, and A.3, respectively.

3 Modeling Approach

3.1 Features

In this section, we describe the variables we use in the study. Brogaard, Hendershott, and Riordan (2014) find evidence that public information influences the direction of the orders submitted by HFTs, for example, news announcements, price movements, and LOB imbalances. Cont et al. (2023) suggest one should also consider the time of day and other market conditions such as momentum and volatility. Aït-Sahalia et al. (2022) use three clocks to construct predictor variables over non-overlapping lookback windows. AFM (2023) indicates that proprietary trading firms report that their trading algorithms use features such as order book imbalance, volume in the order book, and price trends, all measured over various time horizons. Furthermore, firms report that the most recent data has the most predictive power—for example, messages in the last few milliseconds have more predictive power than older messages.

In this article, we use fifty-three features that reflect (i) the state of the LOB, (ii) recent activity in the LOB, (iii) the current presence of the algorithm in the LOB, (iv) the current presence of the member in the LOB, (v) the cash and inventory of the algorithm, and (vi) the cash and inventory of the member. By inventory we mean a transformation $f(x) = \text{sign}(x)\log(1+|x|)$, where $|\cdot|$ is the absolute value, of the intraday accumulated position measured in number of shares, and the starting value of the accumulated position at the beginning of each trading day is zero by definition. Similarly, cash is the accumulated

Downloaded from https://academic.oup.com/jfec/article/23/2/nbae025/7826742 by guest on 20 June 2025

expense (in EUR) of purchases and sales of inventory; we work with expenses (so purchasing inventory increases the cash variable), to ensure that inventory and cash typically have the same sign. We also apply the above transformation f to the cash variable. Naturally, there is a strong positive relationship between these two variables. Given our data, we only consider transactions which occur on Euronext Amsterdam. All features are measured at the time just before the order enters the market.

To describe our features more precisely, let $\mathcal{A} = \{a_1, a_2, \ldots, a_M\}$ be the identifiers for the algorithms on Euronext, and $\mathcal{B} = \{b_1, b_2, \ldots, b_N\}$ be the identifiers for the trading members of Euronext—we have that $N \leq M$, that is, a trading member can trade with multiple trading algorithms. The features we use for algorithm $a_i \in \mathcal{A}$ from trading member $b_i \in \mathcal{B}$ are: (i) volumes in the LOB excluding those from trading algorithm a_i, (ii) volumes in the LOB posted by trading algorithm a_i, (iii) imbalance in the LOB excluding volumes posted by algorithm a_i, (iv) imbalance in the LOB of volume posted by algorithm a_i, (v) logarithm of the volume at the best bid and the best ask, (vi) bid–ask spread, (vii) returns of transaction prices over various time windows in the past, (viii) volatility of transaction prices over various time windows in the past, (ix) number of messages over various time windows in the past, (x) number of aggressive buy orders, number of aggressive sell orders, and net aggressive order flow over various time windows in the past,[15] (xi) logarithm of volume of last transaction, (xii) inventory of trading algorithm a_i, (xiii) inventory of trading member b_j, (xiv) cash of trading algorithm a_i, and (xv) cash of trading member b_j. In Appendix B, we provide formulae for each of the above features. The categories (i) to (xv) include between one and six variables each and account for a total of fifty-three features.

The features we employ are measurable to regulators and trading venues. Even if variables such as inventory of trading member (i.e., intraday accumulated inventory of trading member) are not used in the decision making process of an algorithm, it is nonetheless a variable, visible to the regulator or trading venue, that can be used to predict an algorithm's trading behavior. Effectively, our variables can be seen as proxies for some of the private signals used by trading firms.

There is correlation between many of the above features; for example, it is well-known that there is positive correlation between the bid–ask spread and any of the measures of volatility we consider; see for example, Grossman and Miller (1988), Glosten and Milgrom (1985), O'Hara (1998). For each share in our study, we use the first 4 weeks of data to fit a principal component analysis (PCA) transformation and project the fifty-three features down to thirty features, see Figure 2. This captures more than 90% of the variation in the standardized features for all shares.[16] For each share, we apply the transformation obtained for the first 4 weeks to the remaining 12 weeks. We refer to the features after the PCA transformation as the PCA-transformed features and we refer to the features before the PCA transformation as the original features. Unless stated otherwise, we work with the PCA-transformed features.

3.2 Output Variable

Consider an algorithm which sends a limit order at time t. We model the choice of (i) *direction*, (ii) *limit price*, and (iii) *volume* of the limit order sent to the market. Recall that Table 2 shows that around 97% of all trading activity employs limit orders. Our models accommodate both provision of liquidity and taking of liquidity. We return to this point in Section 4.5.8 when we consider a conditional model for liquidity taking activity.

We consider all limit orders regardless of their time-in-force. To formalize the problem, let $(D_t, P_t, V_t; a_i, b_j)$ be a limit order sent by algorithm $a_i \in \mathcal{A}$ and trading member $b_j \in \mathcal{B}$

[15] We use the terms "aggressive" and "liquidity taking" interchangeably.

[16] The exact percentages of the variation captured with 30 features are: 93% for ASML, 92% for ING, 93% for AHOLD, and 93% for TOMTOM.

Downloaded from https://academic.oup.com/jfec/article/23/2/nbae025/7826742 by guest on 20 June 2025

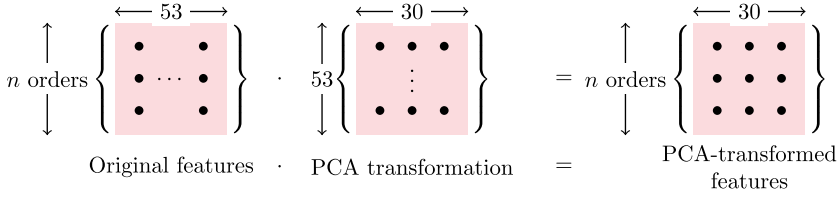


Figure 2. Transforming the original features to PCA-transformed features.

arriving at the exchange at time t, where $D_t \in \{-1, 1\}$ is the direction of the order ($D_t = 1$ if the order is to buy and $D_t = -1$ if the order is to sell), $P_t \in \mathbb{R}$ is the price limit attached to the order and $V_t \in \mathbb{R}^+$ its volume.[17] The signed difference from best quote is given by

$$\mathcal{P}_t = \begin{cases} P_t - S_t^b & \text{if } D_t = 1, \\ S_t^a - P_t & \text{if } D_t = -1, \end{cases} \tag{1}$$

where S_t^a and S_t^b are the best ask price and best bid price in Euronext, respectively, and

$$\mathcal{V}_t = \log(1 + V_t), \tag{2}$$

is the log-volume of the order. For an order to buy ($D_t = 1$) we have that $\mathcal{P}_t = P_t - S_t^b$, which is: (i) greater than zero if the order has a limit price that is more generous (higher than) than the best bid price, is (ii) equal to zero if the order has a limit price that is as generous as (equal to) the best bid price, and (iii) less than zero if the order has a limit price that is less generous (lower than) than the best bid price. Similarly, for an order to sell ($D_t = -1$) we have that $\mathcal{P}_t = S_t^a - P_t$, which is: (i) greater than zero if the order has a limit price that is more generous (lower than) than the best ask price, is (ii) equal to zero if the order has a limit price that is as generous as (equal to) the best ask price, and (iii) less than zero if the order has a limit price that is less generous (higher than) than the best ask price.

We study three regimes of the variable $\mathcal{P}_t$ and use them to define the price bucket variable $\mathfrak{P}_t$ as follows:

$$\mathfrak{P}_t = \begin{cases} 1 & \text{if } \mathcal{P}_t \in (0, \infty) \quad \text{(eager-to-trade orders – more generous than current best limit price)}, \\ 2 & \text{if } \mathcal{P}_t \in \{0\} \quad \text{(at-the-touch orders – at the current best limit price)}, \\ 3 & \text{if } \mathcal{P}_t \in (-\infty, 0) \quad \text{(orders deeper in the LOB)}. \end{cases} \tag{3}$$

For ASML, $\mathfrak{P}_t = 1$ for 9.6% of orders, $\mathfrak{P}_t = 2$ for 42.8% of orders, and $\mathfrak{P}_t = 3$ for 47.6% of orders.; see Table 6.

With 4 weeks of data from 11 October 2021 to 5 November 2021, we compute the (up to nine) unique population deciles of the variable $\mathcal{V}_t$, and use these population deciles to define the (up to ten) buckets associated with this variable. We denote by υ^c the number of buckets for the log-volume. The bucket version of $\mathcal{V}_t$ is denoted by $\mathfrak{V}_t \in \{1, \ldots, \upsilon^c\}$.

[17] Accounting for latency is out of the scope of this analysis and is left for future research. However, we mitigate some of the effects of latency in the features that agents observe, by measuring features over various time-intervals.

Downloaded from https://academic.oup.com/jfec/article/23/2/nbae025/7826742 by guest on 20 June 2025

Table 6. Order count per price bucket $\mathfrak{P}_t$ and cluster using first 4 weeks of training data for ASML

	Cluster 1 (Directional traders)	Cluster 2 (Opportunistic traders)	Cluster 3 (Market makers)	Total
$\mathfrak{P}_t = 1$	178,333	128,174	391,016	697,523
$\mathfrak{P}_t = 2$	386,678	456,828	2,257,863	3,101,369
$\mathfrak{P}_t = 3$	1,074,256	1,316,237	1,063,476	3,453,969
Total	1,639,267	1,901,239	3,712,355	7,252,861

We apply this bucketing to each of the twelve subsequent weeks of data. The variables we model are

$$(D_t, \mathfrak{P}_t, \mathfrak{V}_t), \tag{4}$$

and we call them (i) direction of the order, (ii) price bucket, and (iii) volume bucket, respectively.

As discussed above, our models capture both aggressive and passive behavior depending on the price bucket and the time-in-force of the order. For example, note that all aggressive behavior falls within $\mathfrak{P}_t = 1$. More precisely, $\mathfrak{P}_t = 1$ includes orders that (i) cross the spread and trade against the opposite side of the LOB (aggressive orders), (ii) post liquidity inside the spread (generous liquidity provision), and (iii) are canceled by the exchange upon entry (missed attempts) because their price limit is not generous enough to trade and their time-in-force precludes them from resting in the LOB; we return to this point below in Section 4.5.8 where we study a conditional model for liquidity taking activity.

3.3 Regression Models

For each share c we employ $W = 16$ datasets $\{\mathcal{D}_k^c\}_{k=1}^W$, each corresponding to a given week of orders in the period 11 October 2021 to 30 January 2022. For each week k, the dataset is given by $\mathcal{D}_k^c = \{(\boldsymbol{x}_t^k, \boldsymbol{y}_t^k)\}_{t=T_1}^{T_k}$, where, $\boldsymbol{x}_t^k \in \mathbb{R}^{\tilde{K}}$ are the features, and $\boldsymbol{y}_t^k \in \{-1, 1\} \times \{1, 2, 3\} \times \{1, \ldots, v^c\}$ are the order details, given by $\boldsymbol{y}_t^k = (D_t^k, \mathfrak{P}_t^k, \mathfrak{V}_t^k)$. The vector $\boldsymbol{x}_t^k \in \mathbb{R}^{\tilde{K}}$ contains the features we use for the regressions $\boldsymbol{x}_t^k = (x_1^{k,1}, \ldots, x_1^{k,\tilde{K}})$. Here, $\tilde{K} = 30$ when we use PCA-transformed features. Note that the features $\boldsymbol{x}_t^k$ implicitly depend on the algorithm placing the order. We use logistic regressions to model the decisions of agents, that is

$$\text{(R1)} \quad \mathbb{P}(D_t^k = 1) = \frac{1}{1 + \exp\left(\beta^{D,k} + \boldsymbol{\beta}^{D,k} \cdot \boldsymbol{x}_t^k\right)}, \quad \text{on } \mathcal{D}_k^c, \tag{5}$$

$$\text{(R}\star\text{)} \quad \mathbb{P}(Y_t^k = y) = \frac{\exp\left(\beta_y^{Y,k} + \boldsymbol{\beta}_y^{Y,k} \cdot \boldsymbol{x}_t^k\right)}{1 + \sum_{j=1}^{y^c - 1} \exp\left(\beta_j^{Y,k} + \boldsymbol{\beta}_j^{Y,k} \cdot \boldsymbol{x}_t^k\right)}, \quad \text{on } \mathcal{D}_k^c, \tag{6}$$

where $Y \in \{\mathfrak{P}, \mathfrak{V}\}$ and the operator $\cdot$ denotes the dot product between two vectors. When Y is $\mathfrak{P}$ then $\text{R}\star = \text{R2}$, $y \in \{1, 2, 3\}$, and $y^c = 3$. Similarly, when Y is $\mathfrak{V}$ then $\text{R}\star = \text{R3}$, $y \in \{1, \ldots, v^c\}$, and $y^c = v^c$.

Downloaded from https://academic.oup.com/jfec/article/23/2/nbae025/7826742 by guest on 20 June 2025

We model $(D_t^k, \mathfrak{P}_t^k, \mathfrak{V}_t^k)$ with the vector of features $\boldsymbol{x}_t^k$. We also model price bucket and volume bucket conditional on the direction D_t^k; for this, we perform logistic regressions

$$(\mathrm{R}*) \quad \mathbb{P}(Y_t^k = y | D_t^k = d) = \frac{\exp\left(\beta_y^{Y,k,d} + \boldsymbol{\beta}_y^{Y,k,d} \cdot \boldsymbol{x}_t^k\right)}{1 + \sum_{j=1}^{\nu^c - 1} \exp\left(\beta_j^{Y,k,d} + \boldsymbol{\beta}_j^{Y,k,d} \cdot \boldsymbol{x}_t^k\right)}, \quad \text{on } \mathcal{D}_k^{c,d}, \tag{7}$$

where we have the following four combinations: (i) when Y is $\mathfrak{P}$ and $d = +1$ then $\mathrm{R}* = \mathrm{R4}$, $\nu^c = 3$, $\mathcal{D}_k^{c,d} = \mathcal{D}_k^{c,+1}$, and $y \in \{1,2,3\}$; (ii) when Y is $\mathfrak{P}$ and $d = -1$ then $\mathrm{R}* = \mathrm{R5}$, $\nu^c = 3$, $\mathcal{D}_k^{c,d} = \mathcal{D}_k^{c,-1}$, and $y \in \{1,2,3\}$; (iii) when Y is $\mathfrak{V}$ and $d = +1$ then $\mathrm{R}* = \mathrm{R6}$, $\nu^c = \nu^c$, $\mathcal{D}_k^{c,d} = \mathcal{D}_k^{c,+1}$, and $y \in \{1, \ldots, \nu^c\}$; (iv) when Y is $\mathfrak{V}$ and $d = -1$ then $\mathrm{R}* = \mathrm{R7}$, $\nu^c = \nu^c$, $\mathcal{D}_k^{c,d} = \mathcal{D}_k^{c,-1}$, and $y \in \{1, \ldots, \nu^c\}$. We also have that $\boldsymbol{\beta}_3^{\mathfrak{P},k,+1}, \beta_3^{\mathfrak{P},k,+1}, \boldsymbol{\beta}_3^{\mathfrak{P},k,-1}, \beta_3^{\mathfrak{P},k,-1}, \boldsymbol{\beta}_{\nu^c}^{\mathfrak{V},k,+1}, \beta_{\nu^c}^{\mathfrak{V},k,+1}, \boldsymbol{\beta}_{\nu^c}^{\mathfrak{V},k,-1}, \beta_{\nu^c}^{\mathfrak{V},k,-1}$ are all zero. Here we use

$$\mathcal{D}_k^{c,+1} = \{(\boldsymbol{x}_t^k, \boldsymbol{y}_t^k) \in \mathcal{D}_k^c : D_t = 1\}, \tag{8}$$

$$\mathcal{D}_k^{c,-1} = \{(\boldsymbol{x}_t^k, \boldsymbol{y}_t^k) \in \mathcal{D}_k^c : D_t = -1\}. \tag{9}$$

For each algorithm, we fit the above models. Section 4.2 reports the out-of-sample accuracies and outperformances over alternative methods for ASML. The results for the other three tickers are reported in Appendix C.1.

3.4 Machine Learning Models

3.4.1 Random forests

We also use random forests—introduced by Breiman (2001)—as a benchmark to assess the performance of the logistic regressions. Random forests go beyond linear models but lack explainability. We return to this point below when we discuss data privacy concerns.

We train random forests with fifty trees in the forest and a maximum depth of five for a given tree. To avoid overfitting and to obtain these parameter values, we tune hyperparameters to balance the in-sample and out-of-sample accuracies. We employ `sklearn.ensemble` in Python, and the same features as in our logistic regressions.

3.4.2 Algorithm clusters

The regression coefficients from our logistic regressions describe trading behavior. Indeed, the value of the regression coefficients models the probability with which algorithms choose direction, price, and volume. In this article, we use the regression coefficients of the models for direction (D_t) to cluster trading behavior. The clustering exercise employs hierarchical agglomerative clustering; see Section 21.2 in Murphy (2022). In particular, we use complete linkage, cosine affinity, and a target of three clusters, to compare with the dealing capacities of algorithms—recall that we study three dealing capacities in this article: (i) Liquidity Provider, (ii) House, and (iii) Client. We employ `AgglomerativeClustering` from `sklearn.cluster`.

3.5 Limitations

The models we discuss below have some limitations. First, our data covers trading in Euronext; thus, the models that follow do not account for trading in other venues, so we do not have information about the inventory of the members and their activity in other markets (e.g., stock market and derivatives market).[18] Also, the features we employ are

[18] For similar issues see Hansch, Naik, and Viswanathan (1998) who examine specialist inventory dynamics in shares where inventory is computed on the primary exchange only.

Downloaded from https://academic.oup.com/jfec/article/23/2/nbae025/7826742 by guest on 20 June 2025

ticker-specific. For instance, when modeling ASML, we do not use data from any of the other tickers in our Euronext dataset.

In addition to incomplete information, we do not study cancellation or modification of limit orders submitted to Euronext. Modifications are, according to Euronext, cancellations followed by a new order submission that keeps the same order ID as the original order. In total we discard ~ 1.1 million modifications of orders.[19]

Lastly, there are a number of features mentioned in the literature that we do not include. For example, the queue size at best quotes,[20] proxies for the latency of algorithms, and the latency of the Euronext's matching engine.[21]

In principle, if we had the trading activity of members in other venues, the accuracies that we report should improve. However, we remark that with the non-exhaustive feature list we employ, the models obtain reliable out-of-sample accuracies when predicting the direction, the price, and the volume of a new order. We think of these accuracies as a lower bound of what could be achieved in future research.

4 Model Fit

4.1 Train and Deploy Structure

As described in Section 2, we use 16 weeks of data in the study. The models are trained over four consecutive weeks of data and applied to the week after.[22] For example, we use weeks 1–4 (11 October 2021 to 5 November 2021) to train the models, then obtain out-of-sample predictions for week 5 (8 November 2021 to 12 November 2021), train models from weeks 2–5 (18 October 2021 to 12 November 2021), then obtain out-of-sample predictions for week 6 (15 November 2021 to 19 November 2021), and so on. We refer to each such run as a "train-and-deploy" exercise—see Figure 3.

With the 16 weeks of data, we perform twelve train-and-deploy exercises. Thus, there are 12 weeks of data for which we perform out-of-sample predictions. For each train-and-deploy exercise we obtain seven accuracies per algorithm (one for each model in R1 to R7; see Section 3.3). Next, we compute the order-weighted accuracy of the predictions per algorithm over the 12 weeks of data and report the results in Table 7. With these accuracies, we compute (i) the average, plus and minus the standard deviation, of the accuracies across groups of algorithms and we report it in the first five rows, and (ii) the order-weighted accuracy for all algorithms and report it in the bottom row.

4.2 Out-of-Sample Accuracy of Predictions

Table 7 reports out-of-sample accuracies of the predictions made by the logistic regression models for twelve train-and-deploy exercises on ASML. The direction of order D_t can take two values, the price bucket $\mathfrak{P}_t$ can take three values, and the volume bucket $\mathfrak{V}_t$ can take nine values.

The ordering of algorithms in Table 7 is based on the order count over all deploy-weeks of the train-and-deploy exercises. For example, the "Top 5" row refers to the top five

[19] The article of Bouchaud, Mézard, and Potters (2002) provides statistics on cancellations. For example, they state that roughly 10% of the orders are cancelled or modified before being executed. For ASML, 1.1 million orders is roughly 12.6% of the grand total of 8.7 (1.1 plus 7.6) million orders. Future avenues of research include the modeling of cancellations or modifications, together with the resting time of limit orders.

[20] For ASML, there are on average 4.19 orders on the best bid price level, and on average 4.25 orders on the best ask price level. This average is computed by sampling the number of orders in the queue at best quotes every minute for the complete time frame of the dataset we employ.

[21] See the work of Yao and Ye (2018) where the authors show the extent to which the queue is important for market makers, and the work of Aquilina, Budish, and O'neill (2022) and Cartea and Sánchez-Betancourt (2021) for the latency aspect; for example, Cartea and Sánchez-Betancourt (2021) shows how to use marketable limit orders sent by traders to compute the implied latency of their trading activity.

[22] The accuracies that we obtain are robust to increasing the training periods in this modeling choice.

Downloaded from https://academic.oup.com/jfec/article/23/2/nbae025/7826742 by guest on 20 June 2025

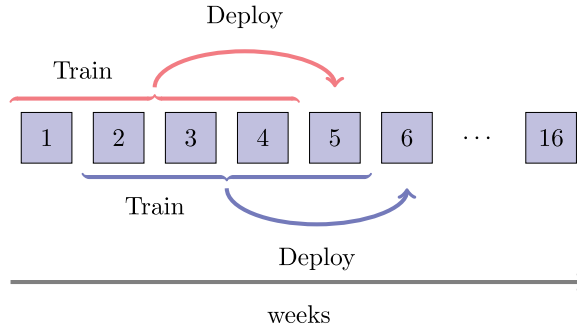


Figure 3. The models are trained on four consecutive weeks of data and applied to the week after.

Table 7. Accuracies of the logistic regression models for ASML, calculated over twelve train-and-deploy exercises. Accuracies are reported in % with ± the standard deviation. Here, D_t is the direction of the order, $\mathfrak{P}_t$ the price bucket of the order, and $\mathfrak{V}_t$ the volume bucket of the order.

	R1 D_t	R2 $\mathfrak{P}_t$	R3 $\mathfrak{V}_t$	R4 $\mathfrak{P}_t \vert D_t = 1$	R5 $\mathfrak{P}_t \vert D_t = -1$	R6 $\mathfrak{V}_t \vert D_t = 1$	R7 $\mathfrak{V}_t \vert D_t = -1$
Top 5	65 ± 7	80 ± 9	60 ± 23	85 ± 6	84 ± 6	61 ± 21	60 ± 22
Top 10	66 ± 7	86 ± 9	68 ± 22	89 ± 6	89 ± 6	69 ± 21	69 ± 22
Top 20	68 ± 9	87 ± 12	62 ± 23	89 ± 10	89 ± 10	63 ± 22	63 ± 22
Top 50	73 ± 12	86 ± 15	60 ± 25	86 ± 18	87 ± 18	59 ± 26	60 ± 25
All	77 ± 15	80 ± 19	48 ± 25	78 ± 24	77 ± 25	46 ± 25	46 ± 26
Order-weighted average	70	85	62	88	88	62	62

algorithms that sent most orders during the twelve deploy-weeks. The same ordering method is applied to Table 8 and the tables for other tickers.

Accuracy of predictions is high and stable over time for all algorithms.[23] The accuracies of the logistic regression models are stable for all shares—for example, when predicting the direction of the order D_t, accuracies range between 67% and 70% according to the order-weighted accuracy for all algorithms; see Appendix C.1. The order-weighted average of the regressions for $\mathfrak{P}_t$ that condition on direction of order perform better than regression R2. This is not the case for $\mathfrak{V}_t$ where we have a similar performance for R2, R6, and R7.

We compare the accuracies of the logistic regression models against a benchmark where the predicted bucket is the bucket most frequently used by the algorithm in the training data. Table 8 reports the outperformance of the logistic regression models over this benchmark.

On average, the logistic regression models outperform the benchmark. The outperformance is largest for predicting the direction of an order, at 18% on average. When predicting volume, we observe the smallest outperformance over the benchmark. See Appendix C.1 for comparable results for other tickers.

[23] These accuracies are for the methodology outlined in Figure 3. We tested the robustness of these accuracies when one increases the length of the training window. We find that the results in Table 7 are robust to the modeling choice in Figure 3. In particular, if we consider training on "expanding windows," that is, for deployment in a given week, we train the models on the data from all the weeks before, we find that the order-weighted average accuracies of the models are similar to those in Table 7. More precisely, we obtain an order-weighted accuracy of 70% for D_t, 84% for $\mathfrak{P}_t$, and 65% for $\mathfrak{V}_t$ when training on expanding windows, which is close to the ones reported in Table 7.

Downloaded from https://academic.oup.com/jfec/article/23/2/nbae025/7826742 by guest on 20 June 2025

Table 8. Outperformance over benchmark for ASML, calculated over twelve train-and-deploy exercises. Here, D_t is the direction of the order, $\mathfrak{P}_t$ the price bucket of the order, and $\mathfrak{V}_t$ the volume bucket of the order.

	D_t	$\mathfrak{P}_t$	$\mathfrak{V}_t$	$\mathfrak{P}_t \vert D_t = 1$	$\mathfrak{P}_t \vert D_t = -1$	$\mathfrak{V}_t \vert D_t = 1$	$\mathfrak{V}_t \vert D_t = -1$
Top 5	15 ± 7	8 ± 10	5 ± 6	14 ± 14	13 ± 14	6 ± 8	5 ± 7
Top 10	15 ± 7	4 ± 9	2 ± 5	7 ± 12	7 ± 12	3 ± 6	2 ± 6
Top 20	17 ± 10	4 ± 7	3 ± 7	7 ± 10	6 ± 10	4 ± 8	3 ± 7
Top 50	20 ± 13	3 ± 6	3 ± 6	5 ± 8	4 ± 8	3 ± 7	3 ± 7
All	21 ± 17	3 ± 10	0 ± 8	3 ± 10	4 ± 11	1 ± 10	0 ± 11
Order-weighted average	18	4	3	8	7	3	3

Table 9. Outperformance of random forests over benchmark and over logistic regressions for ASML, calculated over twelve train-and-deploy exercises. Here, D_t is the direction of the order, $\mathfrak{P}_t$ the price bucket of the order, and $\mathfrak{V}_t$ the volume bucket of the order.

	Outperformance					
	Over benchmark			Over logistic		
	D_t	$\mathfrak{P}_t$	$\mathfrak{V}_t$	D_t	$\mathfrak{P}_t$	$\mathfrak{V}_t$
Top 5	21 ± 7	7 ± 8	9 ± 10	6 ± 7	−1 ± 3	4 ± 4
Top 10	19 ± 7	4 ± 7	5 ± 8	4 ± 6	−1 ± 2	3 ± 3
Top 20	23 ± 11	3 ± 6	6 ± 8	6 ± 7	−1 ± 2	2 ± 5
Top 50	24 ± 13	2 ± 5	4 ± 7	4 ± 6	−1 ± 2	1 ± 4
All	22 ± 17	3 ± 7	2 ± 7	1 ± 7	0 ± 6	3 ± 5
Order-weighted average	23	4	5	5	0	2

Table 9 reports the outperformances of random forests when compared with the same benchmark as that used in Table 8, and compared with the logistic regressions in Table 7.

Overall, random forests outperform the benchmark by 23% when predicting whether algorithms send a buy or a sell order—this is 5% higher than the outperformance achieved by the logistic regression models. The outperformance over the benchmark is stable across shares; in all cases, random forests outperform logistic regression models. When predicting the price bucket, random forests outperform the benchmark by 4%, on average, which is the same outperformance attained by the multinomial logistic regression models. Lastly, when predicting the volume bucket, the outperformance of random forests over the benchmark is 5%, which is 2% higher than the outperformance of the logistic regression over the benchmark. On average, random forests tend to slightly outperform logistic regressions.

Going beyond logistic regressions comes at a cost. Machine learning models are known to have the potential to leak parts of the training data when one discloses the model (weights or coefficients obtained) as investigated in Shokri et al. (2017). That is, some models are vulnerable to privacy attacks—see Rigaki and Garcia (2020) for a comprehensive survey. For many applications of our study this suggests it is more appropriate to use logistic regressions because of their interpretability, their simplicity, and their privacy features. We therefore focus on logistic models in the remainder of this study.

4.3 Out-of-Sample Accuracy of Predictions: Blind to Member and Algorithm Identification

In this section, we show the out-of-sample accuracy of the predictions when we train the models with features that are built without using the identification of the member and of the algorithm.

Downloaded from https://academic.oup.com/jfec/article/23/2/nbae025/7826742 by guest on 20 June 2025

Table 10. Out-of-sample accuracies of the logistic regressions trained with features that are built without the member identification and without algorithm identification, and performance against logistic regression trained on the original, agent-labeled features. Here, D_t is the direction of the order, $\mathfrak{P}_t$ the price bucket of the order, and $\mathfrak{V}_t$ the volume bucket of the order. The logistic regressions using the agent-labelled features outperform the models that do not use the identity of member and identity of the algorithm by 16% for predicting direction, 32% for price, and 38% for volume.

	Blinded features					
	Accuracies			Decrease		
	D_t	$\mathfrak{P}_t$	$\mathfrak{V}_t$	D_t	$\mathfrak{P}_t$	$\mathfrak{V}_t$
Order-weighted average	54	53	24	16	32	38

Therefore, here we exclude features (i)–(v) in Appendix B, and replace them by features built without member and algorithm identification. For example, we replace available volumes in the LOB excluding trading algorithm a_i and available volumes in the LOB by trading algorithm a_i, with one feature that measures the total available volumes in the LOB. We do the same for features that measure the imbalance, and features that measure the change in imbalance. We also exclude inventory features.

The train-and-deploy structure is the same as that described in Section 4.1. Table 10 shows the out-of-sample accuracies of our models for the three target variables. We train and deploy only one model for all agents per target variable because we exclude any agent-specific features and, for this exercise, one cannot distinguish between algorithms.

Recall that the direction of order D_t can take two values, the price bucket $\mathfrak{P}_t$ can take three values, and the volume bucket $\mathfrak{V}_t$ can take nine values. The out-of-sample accuracy for predicting D_t is 54%, which is 16% lower than the accuracies obtained using the logistic regressions on the agent-labeled data. The out-of-sample accuracy for predicting $\mathfrak{P}_t$ is 53%, which is 32% lower than the accuracies obtained using the logistic regressions on the agent-labeled data, and the out-of-sample accuracy for predicting $\mathfrak{V}_t$ is 24%, which is 38% lower.

Therefore the added value of using agent-labeled data in the predictions is large; much larger (one order of magnitude higher) than the outperformance gained from going beyond linear models.

4.4 Feature Importance

Feature importance encompasses a set of machine learning techniques that orders features according to their contribution in explaining a given output variable. In this section, we assess the level of importance of each of the features we employ to predict D_t, $\mathfrak{P}_t$, and $\mathfrak{V}_t$—see Section 3.1 for the complete list of features used in our study.

We employ the "permutation feature importance" approach in Breiman (2001). For a given feature, this model-agnostic technique randomly permutes the values of the feature in the training set and computes the change in model score by taking the difference between (i) the accuracy using the original data and (ii) the accuracy with the permuted values of the feature. The changes in score indicate the sensitivity of the model to the feature. The permutation of values is random, so we repeat this process ten times for each feature.

Permutation feature importance attempts to quantify the contribution of an individual feature to the predictive model. This implies that if a variable is highly correlated with others, it will typically be assigned a low importance, as the model is able to use other variables to obtain the same information.

Downloaded from https://academic.oup.com/jfec/article/23/2/nbae025/7826742 by guest on 20 June 2025

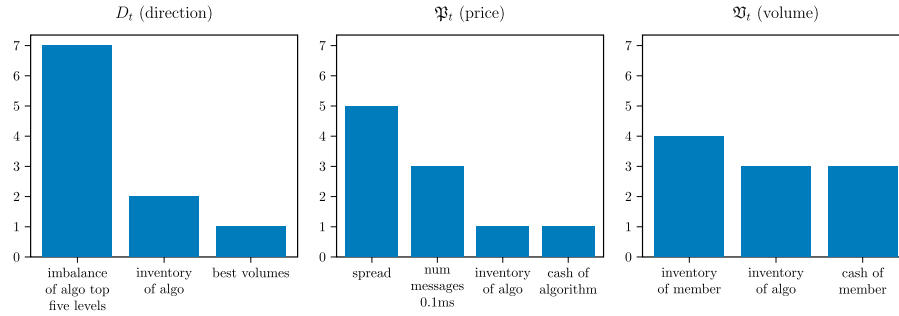


Figure 4. Most important feature for the top ten algorithms in ASML.

4.4.1 Logistic regressions

We repeat the train-and-deploy exercises as described in Section 4.1 on the original features described in Section 3.1 instead of the PCA-transformed features. The use of the original features as opposed to the PCA-transformed features in the logistic regressions is restricted to the feature importance results.[24] The results are reported below for each of the three regressions (R1, R2, and R3).

First, we look at the most important feature to predict D_t, $\mathfrak{P}_t$, $\mathfrak{V}_t$ for each of the top ten algorithms. The ranking of the top ten algorithms is based on the order count over all deploy-weeks of the train-and-deploy exercises. The most important feature is the feature with the highest median importance across the twelve train-and-deploy exercises, the ten algorithms, and the ten random shuffles per feature (giving us $12 \times 10 \times 10 = 1,200$ data points per feature describing their importance). See Figure 4 for the results.[25]

Although there are differences in the importance of features across predictions (D_t, $\mathfrak{P}_t$, and $\mathfrak{V}_t$), there is significant overlap on the most important feature within each of the predictions. For example, the imbalance of the volumes posted by the algorithm in the first five levels of the LOB is the most relevant feature for seven out of the top ten algorithms when predicting the direction of the order D_t. This is similar to other empirical results in the literature; see Cartea, Donnelly, and Jaimungal (2018) where the authors find that the imbalance of volumes resting in the LOB is a good predictor of the sign of the next liquidity taking order sent to the exchange, that is, whether the next liquidity taking order will be to buy or to sell.[26] We find that, for most algorithms, when deciding on the direction of the order D_t, the imbalance of the volumes posted by themselves is more important than the imbalance of the volumes posted by other algorithms.

To predict the price bucket and volume bucket of orders, the bid–ask spread and intraday accumulated inventory are respectively among the most important variables. Intuitively, as the spread widens (tightens), the algorithm is more likely to post orders further from (closer to) the midprice. However, as seen earlier, volumes of orders typically vary little per algorithm. We do not observe "directional" variables, for example, an algorithm's current exposure in the LOB, contributing to the price model, but we see that these play a role in the price conditioned on direction model; see Section 4.5.7.

[24] See Appendix C.2 for a robustness check where we compute the feature importance with the PCA-transformed features and we show that the results are consistent.

[25] In the figures that follow, we abbreviate the feature names for displaying purposes: "excl" is "excluding," "chg" is "change," "m" is "minutes," "s" is "seconds," "ms" is "milliseconds," "quad var" is "quadratic variation," "num" is "number," "agg" is "aggressive," and "algo" is "algorithm."

[26] Cartea, Jaimungal, and Wang (2020) show that volume imbalance helps to predict the direction of the next (non-aggressive) limit order.

Downloaded from https://academic.oup.com/jfec/article/23/2/nbae025/7826742 by guest on 20 June 2025

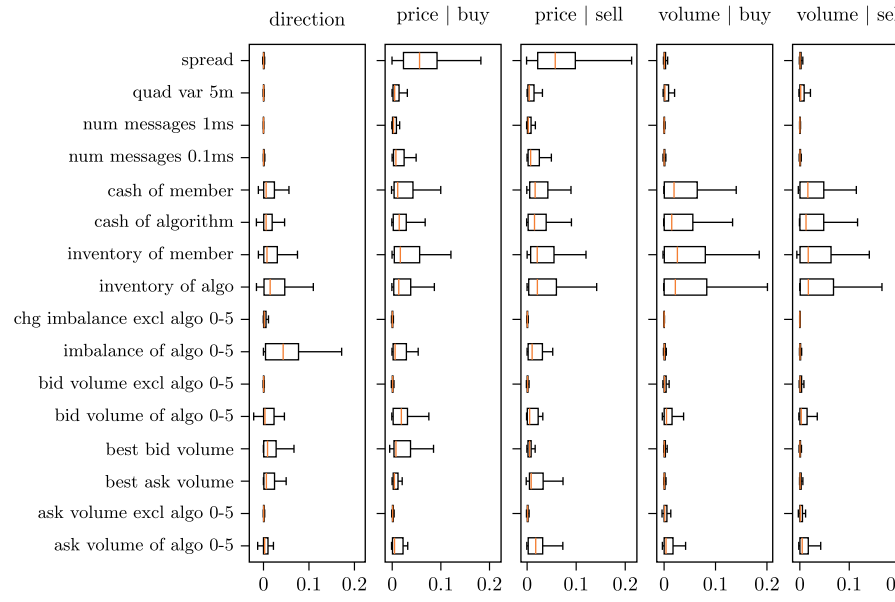


Figure 5. Importance of features to explain the direction D_t of an order, the price bucket $\mathfrak{P}_t$ of an order conditional on the direction D_t, and the volume bucket $\mathfrak{V}_t$ of an order conditional on the direction D_t, for ASML. We use permutation importance and logistic regressions, and only show the features that are in the top ten most important features for at least one target variable.

Next, Figure 5 shows a box-plot of the distribution of feature importance for the top ten algorithms computed over twelve train-and-deploy exercises, and ten repetitions of random permutations. We sort features according to their median importance and plot the top features—each feature contains 1200 data points corresponding to 12 sets of four consecutive weeks, 10 algorithms, and 10 repetitions.[27] The results in the box-plots differ from the results in Figure 4 in that Figure 5 shows the distribution of feature importance between algorithms instead of the most important feature for a given algorithm.

The left column in Figure 5 shows the results for ASML when predicting the direction of the order D_t. Again, the imbalance of volumes posted by the algorithm in the first five levels of the LOB is the most important variable when predicting whether the algorithm will send a buy order or a sell order, followed by the intraday accumulated inventory of the algorithm and the intraday accumulated inventory of the trading member, and the best volumes posted in the LOB. This is similar for all shares; the same set of features appears in each of the top ten features reported. This shows that inventory-related features, volumes posted on best quotes, and imbalances by the algorithm near the top of the LOB are important features to predict the direction of the order.

The second and third panels in Figure 5 show the most important features to predict the price bucket $\mathfrak{P}_t$ conditional on the direction of the order, that is, respectively $D_t = 1$ (buy), and $D_t = -1$ (sell). Spread is the most important feature to predict the price bucket $\mathfrak{P}_t$, which is in line with the results in Figure 4. There is consistency across shares—five of the top ten features in ASML also feature in the top ten of the three other shares. The order of importance is consistent: spread is the most important predictor for three of our four shares, and intraday accumulated inventory and cash of algorithm appear in the top three

[27] The results we obtain are robust to increasing the number of repetitions. In particular, the ordering does not change when increasing the number of repetitions to 15.

Downloaded from https://academic.oup.com/jfec/article/23/2/nbae025/7826742 by guest on 20 June 2025

for ASML, ING, and AHOLD. Other important features are the number of messages sent in the last 100 microseconds (i.e., 0.1 millisecond) and the volume posted by the algorithm on the first five levels of the LOB—both bid and ask side. Also, the best bid volume is more important than the best ask volume to explain the price bucket of buy orders, while for sell orders the best ask volume is more important than the best bid volume.

We draw attention to the differences and similarities between the top features that predict the direction of the order and the price bucket $\mathfrak{P}_t$. The main difference is the importance of spread, which is important to predict the price of an order, yet does not appear in the top ten features to predict the direction of the order.[28] Another difference is that the number of messages sent by all market participants in the last 100 microseconds is not important for predicting the direction of the order but it is important to predict price bucket $\mathfrak{P}_t$.

Lastly, the fourth and fifth panels in Figure 5 report the important features to predict the volume bucket $\mathfrak{V}_t$ conditional on the direction of the order, that is, respectively $D_t = 1$ (buy), and $D_t = -1$ (sell). We see little difference in feature importance between buy orders and sell orders. For all shares, the four inventory variables are the most important features to predict the volume of an order. Also, the volume quoted by the algorithm on the first five levels, on both the buy side and the sell side, is an important feature for all shares. We see that the number of messages in the last second is an important feature for ING, AHOLD, and TOMTOM. Contrary to the other output variables, we observe that quadratic variation—measured over various intervals, is an important feature for ASML, ING, and AHOLD.

Inventory related features are important for predicting the three variables $(D_t, \mathfrak{P}_t, \mathfrak{V}_t)$. Also, among order book features, those that describe behavior closer to the top of the LOB have more predictive power than features that describe behavior deeper in the LOB. Similarly, the number of messages sent in the recent past (e.g., in the last 100 microseconds) tends to be more important than the number of messages contained in longer horizons. This is consistent with the information that trading firms report to the AFM; see AFM (2023).

We also observe that features describing the algorithm's presence in the LOB tend to be more predictive than features describing the presence of other participants. For example, imbalances of the algorithm and volumes of the algorithm tend to be among the most important features, yet imbalances and volumes of other agents are not—this is apart from best bid volume and best ask volume.

4.5 Clustering of Algorithms

The model coefficients from predicting (i) direction of the order D_t with a logistic regression, (ii) price bucket $\mathfrak{P}_t$ with a multinomial logistic regression, and (iii) volume bucket $\mathfrak{V}_t$ with a multinomial logistic regression, capture information about the trading behavior of an algorithm. More precisely, the coefficients show how the features affect the probability that an agent selects a given value for the variables $(D_t, \mathfrak{P}_t, \mathfrak{V}_t)$. Here, we use these coefficients to cluster algorithms.

We revert to using the logistic regression models on the PCA-transformed features because of their higher predicting power. In particular, we employ the model parameters from the logistic regressions to predict trade direction D_t. We focus on D_t because results are easy to interpret and because D_t is the output variable where outperformances are highest. Recall that for D_t we have only one coefficient per feature, instead of one coefficient per bucket per feature when predicting $\mathfrak{P}_t$ and $\mathfrak{V}_t$.

Arguably, the clusters that arise within the regression coefficients are clusters that identify similarity in trading behavior. In Section 4.5.2, we compare these clusters of trading

[28] The spread is the cost of immediacy in the LOB and it is the same cost for buy and sell aggressive orders.

Downloaded from https://academic.oup.com/jfec/article/23/2/nbae025/7826742 by guest on 20 June 2025

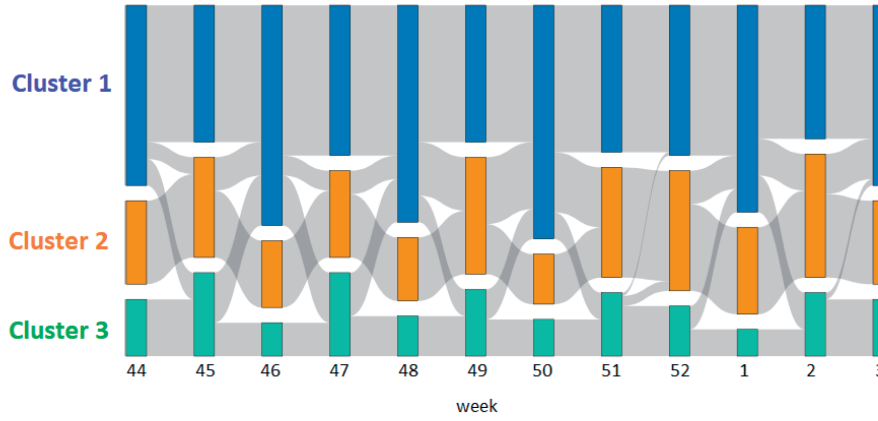


Figure 6. Clusters of trading behavior for ASML for the twelve clustering exercises. The first column represents the clustering exercise on weeks 41–44 of 2021, the second on weeks 42–45, etc. Cluster 1 is at the top (blue), cluster 2 is in the middle (orange), and cluster 3 is at the bottom (green). The size of the bars corresponds to the number of algorithms in the clusters—they add up to 96 algorithms in each of the twelve clustering exercises. The gray areas connecting consecutive clustering exercises represent the transition of algorithms from a cluster in one clustering exercise to a cluster in the next clustering exercise.

behavior with the dealing capacity of algorithms in Euronext Amsterdam—recall that the dealing capacities are (i) Liquidity Provider, (ii) House, and (iii) Client and are determined by the exchange, not the regulator.

We employ the clustering technique described in Section 3.4 with the coefficients of the regressions applied to the PCA-transformed features and a target of three clusters.[29] We do this over twelve sets of four consecutive weeks of data each and refer to this as clustering exercises. Each clustering exercise includes only the ninety-six algorithms present in all of the twelve clustering exercises we perform.

Figure 6 shows the size of the clusters we obtain in each clustering exercise. We include the transitions of algorithms from a given cluster in one clustering exercise to another cluster in the next clustering exercise.

The number of algorithms in each of the three clusters varies for each of the twelve clustering exercises. Given that the clusters aim to capture fundamental trading behavior, it is crucial that they exhibit a degree of consistency over time. Below, we study this point in more detail.

4.5.1 Stability through time

We explore the stability of the clusters across time. For this, given two consecutive clustering exercises A and B where the last week in B is one week after the last week in A, we compute the probability that two randomly chosen algorithms exhibit one of the following two properties: (i) they belong to the same cluster in exercise A and belong to the same cluster in exercise B, or (ii) they do not belong to the same cluster in exercise A and do not belong to the same cluster in exercise B. The blue line in Figure 7 shows the probability of (i) or (ii) along the twelve clustering exercises we perform across time. That is, for two consecutive clustering exercises, we count the pairs of algorithms that were clustered together and

[29] We choose a target of three clusters as it agrees with the number of primary dealing capacities in Euronext and it provides a trade-off between stability of clusters over time and the discretization of trader behavior. Using a larger number of clusters does not give significantly better separation between groups—we believe that this clustering should be better interpreted as a quantization of the distribution of trader behavior, rather than identifying truly discrete trader types.

Downloaded from https://academic.oup.com/jfec/article/23/2/nbae025/7826742 by guest on 20 June 2025

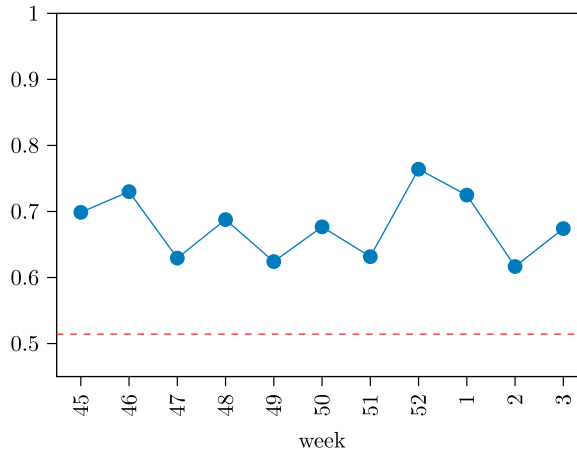


Figure 7. Stability of clusters for ASML across the twelve clustering exercises. The *y*-axis shows the probability of two randomly selected agents in two consecutive clustering exercises A and B (i) being members of the same cluster in A and B or (ii) different clusters in A and B. The *x*-axis shows the last week of the clustering exercise B.

remained clustered together, and those that were clustered differently and remained in different clusters, over all pairs of algorithms. This calculation is a probability, so it lies in $[0, 1]$. The red dash-line is the theoretical value of the probability when algorithms are randomly shuffled and reassigned into the clusters while keeping the number of algorithms in each cluster fixed as those we find for the first 4 weeks of data.

Intuitively, if the blue line is close to 1, then the clustering does not change over time, so the clusters are stable. On the other hand, if the blue line is close to the red dash-line, then the clustering resembles that of random reshuffling of algorithms.

The blue line lies above the red dotted line, which indicates stability of algorithm clusters. However, there is a still a fair amount on instability in the clusters over time, as indicated by the gap between the blue line and 1. Part of the instability might be due to algorithms updating their model parameters frequently, as shown in the results of the survey (AFM 2023). Recall that our clustering exercises include only the ninety-six algorithms present in all clustering exercises. Indeed, if any of these algorithms updates the model parameters they use to send orders, then we expect that the parameters we obtain will change too, resulting in unstable clusters.

We also study the link between the level of activity of the algorithms and the stability of their regression coefficients. We use the logistic regression coefficients of the fifty-three original features described in Section 3.1, and assign each algorithm according to the clustering obtained on the coefficients of the logistic regressions on the PCA-transformed features. The first column of Table 11 lists the most important features to predict the direction D_t of an order. The second and third columns report the temporal deviation, which for a given feature is defined as the standard deviation of the regression coefficients over time. We report the average temporal deviation for the top ten algorithms (second column) and for the bottom ten algorithms (third column)—the ranking is according to the number of orders sent during the first 4 weeks of data.

Given the large number of data points used to fit the models of the top 10 algorithms, we expect that the observed temporal deviation is principally due to changes in the algorithms themselves, for example due to re-calibration of parameters. Conversely, for the bottom 10 algorithms, the number of observed actions is relatively small, so the

Downloaded from https://academic.oup.com/jfec/article/23/2/nbae025/7826742 by guest on 20 June 2025

Table 11. Average temporal deviation of coefficients. The temporal deviation is the standard deviation of the regression coefficients over twelve train-and-deploy exercises when predicting the direction D_t of an order.

	Top 10 Algorithms	Bottom 10 Algorithms
Imbalance of algo top five levels	0.27	0.33
Inventory of algo	0.25	0.84
Ask volume of algo top five levels	0.19	0.36
Bid volume of algo top five levels	0.15	0.26
Best bid volume	0.04	0.21
Best ask volume	0.04	0.18
Chg in imbalance excl algo 0–5	0.03	0.14
Spread	0.03	0.22
Net agg buy–sell last 1 s	0.02	0.15
Number of messages 1 ms	0.02	0.21
Number of messages 0.1 ms	0.02	0.22

systematically higher temporal deviation for these algorithms may be due to statistical estimation error, rather than systematic differences between the top and bottom algorithms. Regardless of the reason for the large temporal deviation of the coefficients for the bottom algorithms, Table 11 lends support to the claim that the instability of our clusters is due primarily to variation in the estimated coefficients of the smaller algorithms.

4.5.2 Types of market participants

Here, we explore how the clusters relate to the dealing capacity of market participants. Recall that we study three dealing capacities: (i) Liquidity Providers (LP), (ii) House, and (iii) Client. Figure 8 shows the confusion matrix between dealing capacity and clusters.

The L-shaped confusion matrix in Figure 8 lends support to the claim that Liquidity Providers exhibit the highest variability in trading behavior.[30]

Cluster 1 consists of an almost equal number of Liquidity Providers, Houses, and Clients. Furthermore, the majority of algorithms with type House or Client are in cluster 1. Therefore, one would expect this cluster to show trading behavior associated with directional trading firms, such as hedge funds or pension funds—see Section 4.5.5. Clusters 2 and 3 consist mostly of Liquidity Providers. Hence, one would expect these clusters to exhibit behavior most often associated with traditional market makers. The statements above are speculative and comment on what (arguably) one would expect. In what follows, we do not inform our understanding of trading behavior with the dealing capacity in our dataset. We remark that our understanding of trading behavior comes from the coefficients of the regressions we build and the clustering exercises we perform.

4.5.3 Clusters and members

Figure 9 shows the number of algorithms that each of the members have, grouped by cluster. The sorting of members is similar to that of Figure 1 (by order count); we reproduce Figure 1 on the left-hand side for comparison and convenience.

Figure 9 adds an extra dimension (member identification) to the confusion matrix in Figure 8. The figure shows the variability of trading behavior among members. Only three members (the first, third, and fourth from top to bottom) have algorithms in each of the

[30] Here, L-shaped refers to the higher values in Figure 8 being arranged along the left column and bottom row.

Downloaded from https://academic.oup.com/jfec/article/23/2/nbae025/7826742 by guest on 20 June 2025

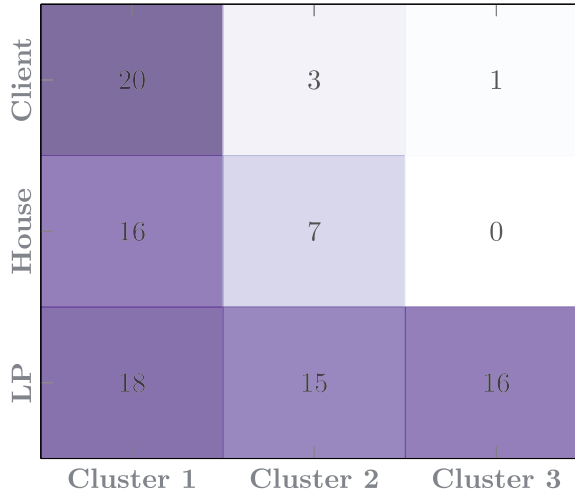


Figure 8. Confusion matrix between the clusters obtained with the parameters from the first 4 weeks of data and the dealing capacity for ASML.

three clusters, five members have algorithms in two of the three clusters, and the remaining twenty-two members have algorithms in only one cluster.[31] Comparing the left-hand side to the right-hand side of Figure 9 we see that the three members with algorithms in each of the three clusters are members with algorithms that are all registered as Liquidity Providers in Euronext.

4.5.4 Summary of results about trading behavior

Consistently, across exercises in the following subsections, algorithms in cluster 3 show behavior associated with inventory-averse market makers. We describe the algorithms in this cluster as "market makers." These algorithms are keen to maintain a balanced presence in the LOB (e.g., they are more likely to post an order on the bid side of the LOB if their volumes on the ask side of the LOB are higher than those of the bid side of the LOB), they revert their inventories to zero, improve the spread, post liquidity at-the-touch, and do not exhibit aggressive behavior unless their inventory is large.

In contrast, algorithms in cluster 2 do not show inventory aversion, they send orders in the direction of the imbalance of the volumes posted at the best quotes (send sell orders if the volume in the bid is much smaller than the volume in the ask), have the lowest percentage (across clusters) of eager-to-trade orders that provide liquidity inside the spread and have the highest percentage of eager-to-trade orders that traded aggressively. We describe the algorithms in cluster 2 as "opportunistic traders."

Lastly, algorithms in cluster 1 have position-building behavior in terms of their own imbalances and the way they choose direction as a function of its inventory and its imbalance in the LOB. These algorithms have propensity to send eager-to-trade orders and at-the-touch orders. Around 82% of their eager-to-trade orders trade aggressively. We describe the algorithms in cluster 1 as "directional traders."

It should be recognized that this is only a description of these algorithms' behavior on this particular exchange—some of them may be taking an opposite directional position in a different market, for which we do not have data.

[31] Out of the twenty-two members with only one cluster of trading behavior, twelve have only one algorithm.

Downloaded from https://academic.oup.com/jfec/article/23/2/nbae025/7826742 by guest on 20 June 2025

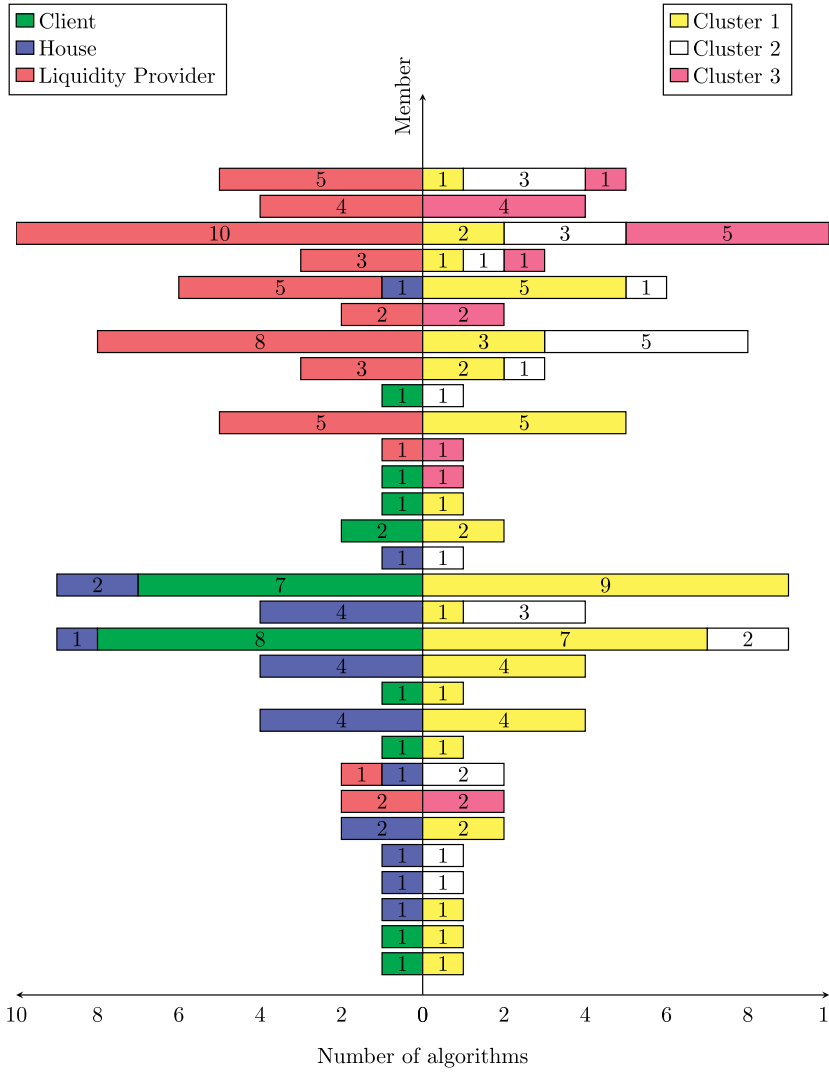


Figure 9. Number of algorithms per dealing capacity (left) and cluster (right) per member in ASML. The members are ordered from top to bottom using order count. We show ninety-six algorithms active each week for the 16 weeks of data.

4.5.5 Clustered market behavior

Table 12 provides summary statistics about the trading activity of the algorithms within the clusters.[32] To facilitate the reading of the tables that follow, we highlight in bold the numbers that are used for the discussion in the main text.

Although algorithms in cluster 3 send the most orders (51%), they do fewer transactions (33% compared to 42% for cluster 1 and 25% for cluster 2). This results in a higher order

[32] Recall that a single aggressive order may result in multiple trades (depending on the number of limit orders it executes against). The number of orders in Table 4 is larger than the number of orders in Table 12 because the later only includes the algorithms present in all of the twelve clustering exercises we perform.

Downloaded from https://academic.oup.com/jfec/article/23/2/nbae025/7826742 by guest on 20 June 2025

Table 12. Statistics for algorithms in clusters 1, 2, and 3 for ASML. First 4 weeks of data. Numbers in bold are discussed in the main text.

	Cluster 1 (Directional traders)	Cluster 2 (Opportunistic traders)	Cluster 3 (Market makers)
Number of algorithms	54	25	17
Order count	1,639,267	1,901,239	3,712,355
Trade count	573,417	343,874	461,874
Order to trade ratio	**2.86**	**5.53**	**8.04**
Order count %	23	26	**51**
Trade count %	**42**	**25**	**33**
Liquidity taking trade count	291,545	251,785	120,353
Liquidity taking trade count %	**51**	**73**	**26**
Average algo volume in ask side of LOB 0–5	**1.43**	**2.07**	**2.47**
Average algo volume in bid side of LOB 0–5	**1.49**	**2.04**	**2.48**
Average absolute imbalance of algo 0–5	**1.93**	**1.12**	**0.77**
Average ask volume of algo 11–20	0.43	1.69	1.07
Average bid volume of algo 11–20	0.44	1.68	1.08
Average daily directionality	**0.53**	**0.46**	**0.14**
Median resting time milliseconds	**2,923**	**2,013**	**691**

to trade ratio for cluster 3 (8.04) than for clusters 1 and 2 (2.86 and 5.53, respectively). Furthermore, algorithms in cluster 3 have a relatively low fraction of liquidity taking transactions; 26% of their transactions are liquidity taking, versus 51% and 73% for clusters 1 and 2. This suggests that algorithms in cluster 3 behave as traditional market makers. Algorithms in clusters 2 and 3 post more volume (on average) on the first five levels of the LOB than algorithms in cluster 1. Similarly, algorithms in clusters 2 and 3 have more balanced volumes posted in the LOB when compared to algorithms in cluster 1.[33] The second to last entry of Table 12 shows the average directionality per day by algorithms of the relevant clusters. We use the definition of directionality in Van Kervel and Menkveld (2019), that is, the absolute net volume divided by total volume traded by the algorithm. Algorithms in cluster 3 are the least directional and algorithms in cluster 1 are the most directional. This reinforces the understanding that algorithms in cluster 3 behave as traditional market makers, and algorithms in cluster 1 behave as directional traders. The last entry of Table 12 shows the median resting time of orders per cluster.[34] Algorithms in cluster 3 have the lowest median resting time with 691 milliseconds, versus 2013 milliseconds for cluster 2, and 2923 milliseconds for cluster 3.

In the next sections, we investigate how the various features we consider affect the average behavior of algorithms in each of these clusters. We focus on the impact of features on direction and price of orders, as the model for volume does not display much benefit over a naive model (as seen in Table 8). More precisely, in Section 4.5.6, we study the average coefficients associated with the regressions for predicting the direction D_t, then, Section 4.5.7 discusses the average coefficients for predicting the price bucket $\mathfrak{P}_t$ conditional on direction, and Section 4.5.8 explores liquidity taking activity and liquidity provision in more detail.

4.5.6 Direction: average behavior

We show the average regression coefficients per cluster for the most important features to explain the direction of the order D_t. Table 13 reports the average coefficients per feature

[33] Recall that if the imbalance of the volumes posted by the algorithm is close to zero then the volume posted by the algorithm on the bid side is close to the volume posted by the algorithm on the ask side.

[34] Resting time is the difference between time of cancellation and time of entry of limit orders.

Downloaded from https://academic.oup.com/jfec/article/23/2/nbae025/7826742 by guest on 20 June 2025

Table 13. Average regression coefficients per cluster on first 4 weeks of training data for ASML when predicting direction D_t. Numbers in bold are discussed in the main text.

	Cluster 1 (Directional traders)	Cluster 2 (Opportunistic traders)	Cluster 3 (Market makers)
Imbalance of algo 0–5	**1.48**	**−1.04**	**−0.42**
Imbalance of algo 11-20	0.21	**−0.97**	0.81
Imbalance of algo 6-10	0.48	−1.00	−0.09
Cash of algorithm	**0.70**	**0.13**	**−0.30**
Inventory of algo	**0.70**	**0.12**	**−0.30**
Best ask volume	−0.26	−0.20	0.23
Best bid volume	0.24	0.19	−0.24
Volume of algo 11–20	0.16	**−0.05**	0.40
Volume of algo 0–5	−0.07	−0.04	−0.28
Chg imbalance excl algo 0–5	0.04	0.16	0.05
Net agg buy–sell last 1 s	−0.03	0.07	0.10
Return 1 s	0.01	0.08	0.10
Volume of algo 6–10	0.01	**−0.11**	0.04
Chg imbalance excl algo 6–10	0.00	0.03	0.10

computed on the first 4 weeks of data. For each of the clusters, the average is taken over all algorithms in the cluster. To obtain the coefficients of the fifty-three features, for each of the ninety-six algorithms, we perform a matrix multiplication of the coefficients of the logistic regressions on the PCA-transformed features (96×30 matrix) and the principal components (PCs) matrix (30×53 matrix), both computed in Section 4.1. We represent the effects of the volume in the book with the total volume and the imbalance (rather than the separate volumes on the bid and ask side).

In the tables which follow, we exclude features where the magnitude of the average coefficients is smaller than or equal to 0.1 for all clusters. We order features by their average absolute coefficient size.

We recall the formulation for the logistic regression in Equation (5), where $D_t = 1$ means that the order is to buy and $D_t = -1$ means that the order is to sell. Thus, given that all features are normalized and centered around zero, all else being equal, if an average coefficient is positive (negative) it means that positive values of the relevant feature increase (decrease) the probability of a given order being a buy order; and vice-versa for negative values of the relevant feature.

Table 13 provides a number of stylized facts about the difference among clusters to understand differences in trading behavior.

(D-i) Cluster 1 is the only cluster with an average positive value for the coefficient for the imbalance of the algorithm in the first five levels—recall that when imbalance is positive the algorithm has more volume posted in the bid than in the ask. Thus, algorithms in cluster 1 are inclined to send a buy order if they have already posted more volume in the bid than in the ask (and similarly with the roles of bid and ask, and buy and sell reversed). Conversely, algorithms in clusters 2 and 3 are more likely to revert their imbalance back to a level where the volume provided in the ask side of the LOB is close to that provided in the bid side of the LOB. This suggests that clusters 2 and 3 behave in some ways like traditional market makers, that is, trading algorithms that provide liquidity in both sides of the LOB in a balanced way. In contrast, algorithms in cluster 1 tend to post liquidity with a preferred direction, which is consistent with most algorithms of House and Client type being in cluster 1.

Downloaded from https://academic.oup.com/jfec/article/23/2/nbae025/7826742 by guest on 20 June 2025

(D-ii) Cluster 3 exhibits mean-reversion to zero in inventories at the algorithm level, that is, when the intraday accumulated inventory is positive they are more likely to send sell orders, while the inventories of algorithms in cluster 2 are not mean-reverting to zero.[35] Conversely, algorithms in cluster 1 have a strong preference to send orders in the direction of their inventory (buy orders if inventory is positive and sell orders if inventory is negative).

(D-iii) The estimation we perform through the PCA coefficients leads to an "averaging" effect over coefficients for similar features. This is seen in the coefficients associated with the inventory and cash of the algorithm which are similar.

In summary, algorithms in cluster 3 exhibit trading behaviors that the literature describes as market markers. For example, algorithms in cluster 3 seem keen to maintain a balanced provision of liquidity and take actions to revert their inventories to zero. On the other side of the spectrum we have the algorithms in cluster 1. These algorithms do not balance the liquidity posted in the LOB, and their orders are more likely to be in the same direction as their inventory (i.e., position building and directional trading). Algorithms in cluster 2 lie somewhere in the middle of the behavior observed by algorithms in clusters 1 and 3. More precisely, out of the coefficients shown in Table 13, there are three instances in which the sign of the coefficient associated with cluster 2 is different from both of the signs of the coefficients for clusters 1 and 3. The three instances involve volumes deep in the LOB.

4.5.7 Price limit: average behavior

To gain further insights into the trading behavior of the algorithms in the above clusters we study how they choose their price limits as measured through the price bucket $\mathfrak{P}_t$.

Recall that by the definition of $\mathfrak{P}_t$ in Equation (3), the first price bucket contains orders whose price limit is more generous than the best quotes; that is, the limit price is lower than the best ask price for a sell order or the limit price is greater than the best bid price for a buy order. The second price bucket contains orders whose limit price equals that of the best quotes; that is, the limit price is the best ask price if the order is to sell or the limit price is the best bid price if the order is to buy. The third price bucket contains orders deeper in the LOB.

Thus, given that orders in the first price bucket either (i) trade with the opposite side of the LOB, (ii) improve the spread, or (iii) get canceled upon entry because their time-in-force precludes them from resting in the LOB and their price limit is better than best quotes, we refer to them as orders that show "eagerness to trade." Similarly, recall that we call price bucket one the "eager-to-trade" bucket. Next, we study the average regression coefficients associated with orders in the "eager-to-trade" bucket ($\mathfrak{P}_t = 1$).

We observe the following stylized features of trading algorithms in each cluster:

(P-i) According to the intercept of the regressions, algorithms in cluster 3 are noticeably less likely than those in clusters 1 and 2 to place orders that are eager to trade.

(P-ii) For algorithms in all clusters, the higher the imbalance of the volumes (by all algorithms) posted at the best bid and ask, the more likely a buy order is eager to trade. This effect is most pronounced for algorithms in cluster 3, where the bid volume has a much greater impact on orders to buy than the ask volume; conversely for orders to sell.

(P-iii) Algorithms in cluster 1 seem indifferent to their own imbalance in the book when determining eagerness to trade, whereas imbalance of their own volume indicates a

[35] See Chapter 10 in Cartea, Jaimungal, and Penalva (2015) for a mathematical model of the market making problem illustrating how risk aversion plays a role in how market makers provide liquidity as a function of their inventory.

Downloaded from https://academic.oup.com/jfec/article/23/2/nbae025/7826742 by guest on 20 June 2025

moderate eagerness to trade for cluster 2 algorithms, and a strong eagerness for cluster 3. The direction of the order for cluster 3 is in the direction of the imbalance (so if an algorithm has a higher bid than ask volume posted, which means they have a positive imbalance, its buy orders are typically more eager to trade and its sell orders are typically less eager to trade).

(P-iv) Regardless of the direction, the wider the bid–ask spread the more likely algorithms are to send orders that are eager to trade.[36] This is particularly pronounced for cluster 3 (below we see that this cluster often refills the book when the spread is above one tick).

(P-v) For algorithms in clusters 1 and 2, the higher their own posted volumes at levels 0–5 on either the bid or ask side, the less likely they are to send an order (either to buy or to sell) that is eager to trade. Algorithms in cluster 3 exhibit the opposite behavior; already having volume at the first five levels of the LOB makes it more likely for them to send eager-to-trade orders. For algorithms in clusters 2 and 3, having volume posted deeper in the book decreases the likelihood of sending eager-to-trade orders.

(P-vi) For algorithms in clusters 1 and 2, the higher the number of recent messages, the more likely an order is eager to trade, in either direction.

(P-vii) For all algorithms, there is a reasonably clear symmetry when the roles of bid and ask, and buying and selling, are reversed.

We also look at the second price bucket ($\mathfrak{P}_t = 2$), which describes orders adding volume to the LOB at the current best bid or best ask (also known as posting "at-the-touch"), and at the third ($\mathfrak{P}_t = 3$, whose coefficients are simply the negative sum of those for $\mathfrak{P}_t = 1, 2$), which describes orders adding volume deeper within the LOB.

By looking at Table 14 for $\mathfrak{P} = 2$ and $\mathfrak{P} = 3$, we observe the following stylized facts for the typical behavior of each cluster. It should be emphasized that these observations are for a given direction of order, and so the conclusions should be read in the light of Table 13.

(P-viii) Cluster 3 algorithms are much more likely to place their orders at-the-touch than either of the other clusters (see the intercept of Table 14 for $\mathfrak{P} = 2$). Cluster 2 is less likely to place its orders at-the-touch than cluster 1 (all other things being equal).

(P-ix) Algorithms in cluster 2 have the highest propensity to place orders deeper in the book.

(P-x) For all algorithms, the higher the volume at the best ask price and the lower the volume at the best bid (i.e., the lower the imbalance of the best volumes in the LOB) the more likely that a buy order will be placed deeper in the LOB; conversely for sell orders. This indicates that algorithms pre-empt that when a large volume builds at the best ask (bid) price, then prices will likely move down (up), hence why they choose to place their buy (sell) orders deeper in the LOB.

(P-xi) The likelihood that an algorithm in cluster 2 posts at-the-touch is affected by its own imbalance near the touch. The likelihood is impacted in the direction of the imbalance (so if an algorithm has a higher bid than ask volume posted, which corresponds to a positive imbalance, it is more likely that the algorithm posts buy orders at the best bid price). A weaker effect is observed for cluster 3, but little effect for cluster 1.

(P-xii) For all algorithms, an increase in the recent number of messages reduces the probability of posting at-the-touch in either direction.

In summary, for algorithms in the first two clusters, the higher the number of recent messages (sent to the exchange by all market participants) the more likely they are to send an

[36] We refine this result in Section 4.5.8. In particular, we find that the larger the bid-ask spread, the more likely all clusters are to provide liquidity inside the spread, and the less likely they are to cross the spread.

Downloaded from https://academic.oup.com/jfec/article/23/2/nbae025/7826742 by guest on 20 June 2025

Table 14. Average regression coefficients for price bucket describing eager-to-trade orders ($\mathfrak{P}_t = 1$), at-the-touch orders ($\mathfrak{P}_t = 2$), and orders deeper in the LOB ($\mathfrak{P}_t = 3$) per cluster, on first 4 weeks of training data for ASML, conditioning on the direction of order D_t. Numbers in bold are discussed in the main text.

		$D_t = 1$ (Buy)			$D_t = -1$ (Sell)		
		Cluster 1	**Cluster 2**	**Cluster 3**	**Cluster 1**	**Cluster 2**	**Cluster 3**
$\mathfrak{P}_t = 1$	Intercept	**−0.21**	**−0.46**	**−0.74**	**−0.17**	**−0.48**	**−0.77**
	Best bid volume	**0.35**	**0.30**	**1.09**	**−0.21**	**−0.17**	**−0.26**
	Best ask volume	**−0.20**	**−0.16**	**−0.30**	**0.35**	**0.30**	**1.05**
	Imbalance of algo 0–5	**−0.00**	**0.29**	**0.68**	**0.08**	**−0.32**	**−0.75**
	Spread	**0.11**	**0.07**	**0.56**	**0.10**	**0.07**	**0.54**
	Volume of algo 0–5	**−0.25**	**−0.12**	**0.40**	**−0.29**	**−0.08**	**0.30**
	Volume of algo 11–20	**0.04**	**−0.32**	**−0.22**	**0.07**	**−0.26**	**−0.21**
	Num messages 1 ms	**0.17**	**0.20**	**−0.08**	**0.21**	**0.15**	**−0.06**
	Volume of algo 6–10	−0.13	−0.23	0.11	−0.13	−0.18	0.04
	Imbalance of algo 6–10	−0.07	0.04	−0.20	0.10	−0.08	−0.28
	Num messages 0.1 ms	**0.12**	**0.19**	**−0.08**	**0.15**	**0.13**	**−0.06**
	Num messages 0.1 s	**0.13**	**0.04**	**0.07**	**0.15**	**0.09**	**0.07**
	Imbalance of algo 11–20	−0.02	0.11	−0.15	0.02	−0.09	−0.05
		$D_t = 1$ (Buy)			$D_t = -1$ (Sell)		
		Cluster 1	**Cluster 2**	**Cluster 3**	**Cluster 1**	**Cluster 2**	**Cluster 3**
$\mathfrak{P}_t = 2$	Intercept	**−0.15**	**−0.58**	**1.25**	**−0.13**	**−0.68**	**0.94**
	Volume of algo 0–5	0.13	0.52	0.16	0.24	0.52	0.21
	Volume of algo 6–10	−0.03	0.38	0.19	0.00	0.33	0.17
	Volume of algo 11–20	−0.19	0.06	0.14	−0.27	−0.01	0.06
	Imbalance of algo 0–5	**−0.02**	**0.22**	**0.09**	**0.01**	**−0.21**	**−0.15**
	Imbalance of algo 11–20	−0.05	0.09	0.07	0.06	−0.07	0.13
	Best ask volume	0.08	0.03	0.04	−0.06	0.08	−0.12
	Imbalance of algo 6–10	−0.04	0.09	0.18	0.02	−0.05	−0.04
	Num messages 1 ms	**−0.09**	**−0.01**	**−0.04**	**−0.11**	**−0.02**	**−0.05**
	Num messages 0.1 s	**−0.10**	**−0.02**	**−0.01**	**−0.12**	**−0.03**	**−0.03**
	Spread	−0.00	0.04	−0.10	−0.01	0.04	−0.09
		$D_t = 1$ (Buy)			$D_t = -1$ (Sell)		
		Cluster 1	**Cluster 2**	**Cluster 3**	**Cluster 1**	**Cluster 2**	**Cluster 3**
$\mathfrak{P}_t = 3$	Intercept	0.36	**1.04**	−0.51	0.30	**1.16**	−0.17
	Imbalance of algo 0–5	0.02	−0.51	−0.76	−0.09	0.52	0.90
	Best bid volume	**−0.31**	**−0.36**	**−1.00**	**0.13**	**0.15**	**0.22**
	Best ask volume	**0.11**	**0.13**	**0.26**	**−0.29**	**−0.38**	**−0.92**
	Volume of algo 0–5	0.11	−0.40	−0.55	0.05	−0.44	−0.51
	Spread	−0.11	−0.11	−0.46	−0.09	−0.11	−0.45
	Volume of algo 11–20	0.14	0.26	0.09	0.19	0.26	0.15
	Volume of algo 6–10	0.16	−0.15	−0.30	0.13	−0.16	−0.21
	Imbalance of algo 6–10	0.11	−0.13	0.02	−0.11	0.14	0.31
	Num messages 0.1 ms	−0.07	−0.23	0.11	−0.09	−0.16	0.09
	Num messages 1 ms	−0.09	−0.20	0.12	−0.10	−0.13	0.11
	Imbalance of algo 11–20	0.07	−0.20	0.08	−0.08	0.16	−0.08

eager-to-trade order (in either direction), similarly, for algorithms in all clusters, the higher the spread the less likely an algorithm is to place eager-to-trade or at-the-touch orders. Again, algorithms in cluster 1 seem indifferent to their own imbalance in the book when deciding to send an eager-to-trade order. Algorithms in cluster 2 have a moderate eagerness

Downloaded from https://academic.oup.com/jfec/article/23/2/nbae025/7826742 by guest on 20 June 2025

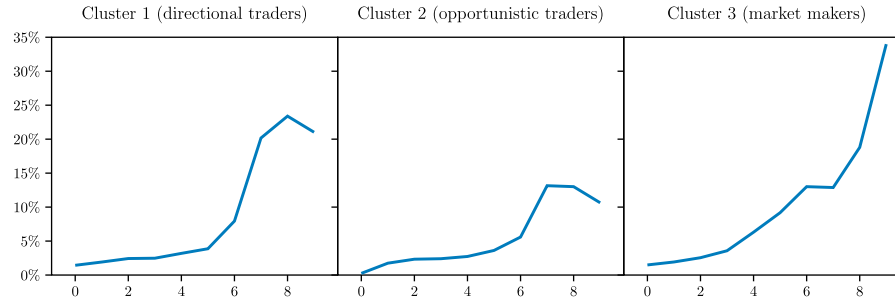


Figure 10. Percentage of eager-to-trade orders as a function of the absolute value of the transformed intraday accumulated inventory of the algorithm. See Appendix B for the formula of the log-inventory of the algorithm.

to trade according to their imbalance in the LOB (similar for posting at-the-touch). Algorithms in cluster 3 have a strong eagerness to trade according to their imbalance in the LOB, and they also have the highest baseline probability of posting at-the-touch. These results are consistent with our previous findings about cluster 3 exhibiting behavior associated with market makers.

The low value of the coefficients for the inventory of the algorithm (this predictor is absent from Table 14, indicating that all clusters had an average coefficient below 0.10 in magnitude) may suggest that this has limited effect on the eagerness to trade. However, this may be due to the limitation of our statistical model to linear-logistic relationships. The plots in Figure 10 show the percentage of orders which were eager to trade, as a function of (transformed, intraday accumulated) inventory. In the rare cases where transformed inventory exceeds 6 (in absolute value), there is a noticeable increase in the eagerness to trade for all algorithms. For algorithms in cluster 3, this effect is smoother as inventory changes, and larger than in clusters 1 or 2.

4.5.8 Liquidity taking activity

We now focus on orders in price bucket $\mathfrak{P}_t = 1$, that is, orders with prices better than the current best bid or ask price. When such an order is sent, there are three possible scenarios, which we label by a variable A_t describing whether it takes liquidity, provides liquidity, or is a missed attempt.

i) The case $A_t = 0$ encompasses aggressive trading, that is, orders that cross the spread. These are buy orders with limit price greater than or equal to the best ask price, or sell orders with limit price less than or equal to the best bid price. Almost all orders within this category are executed upon entry (or partially executed) against liquidity on the other side of the LOB depending on the time-in-force, the limit price of the order, the volume of the order, and the liquidity available in the LOB.[37] From Table 3, we know that orders with FoK time-in-force are negligible so for all practical purposes $A_t = 1$ is aggressive behavior.

ii) The case $A_t = 1$ covers generous provision of liquidity. These are orders that provide liquidity inside the spread. The time-in-force of the orders is DAY and their limit price is higher than the best bid price and lower than the best ask price if the order is to buy; conversely for sell orders.

[37] We provide an example. Assume the best ask price in the LOB is 100 and the volume available is 50 units. An IoC order arrives willing to buy with price limit of 100 and volume of 100, then, 50 units will get executed against the available liquidity and the remaining 50 will be cancelled. Alternatively, if the order had FoK time-in-force then it is cancelled upon arrival.

Downloaded from https://academic.oup.com/jfec/article/23/2/nbae025/7826742 by guest on 20 June 2025

Table 15. Percentage of orders sent with $A_t = 0$ (aggressive trading), $A_t = 1$ (generous liquidity), and $A_t = 2$ (missed attempts) out of all eager-to-trade orders, that is, orders with $\mathfrak{P}_t = 1$.

	Cluster 1 (Directional traders)	Cluster 2 (Opportunistic traders)	Cluster 3 (Market makers)	Total
Number of orders $\mathfrak{P}_t = 1$	178,333	128,174	391,016	697,523
IoC orders	68%	84%	4%	245,988
$A_t = 0$ (aggressive trading)	82%	89%	18%	333,073
$A_t = 1$ (generous liquidity)	17%	10%	81%	364,650
$A_t = 2$ (missed attempt)	1%	1%	1%	2800

iii) The case $A_t = 2$ describes missed attempts. These are orders that are canceled upon entry because their price limit is not generous enough to trade with the available liquidity in the LOB and their time-in-force precludes them from resting in the LOB. This category covers (i) orders to buy with limit price less than the best ask price and IoC time-in-force; it also covers (ii) sell orders with limit price greater than the best bid price and IoC time-in-force. Missed attempts are likely aimed at (a) liquidity that disappeared over the latency period, (b) hidden liquidity, or (c) new liquidity being added over the latency period; see Cartea and Sánchez-Betancourt (2023).

For example, during the first 4 weeks of data for ASML we find that algorithms in cluster 1 (respectively, cluster 2 and cluster 3) sent 178,333 orders (respectively, 128,174 and 391,016 orders) to buy or to sell that were eager to trade; around 82% of these orders (respectively, 89% and 18%) traded against the opposite side of the LOB, that is, show aggressive behavior, similarly, 17% of the orders (respectively, 10% and 81%) provided liquidity inside the spread, that is, generous liquidity provision, and 1% of the orders (respectively, 1% and 1%) were canceled by the exchange upon entry because of their lack of generosity and time-in-force, that is, missed attempts. These results are summarized in Table 15 together with the usage of IoC associated with each of the clusters.

These percentages support our findings that algorithms in cluster 3 make markets for ASML. Algorithms in cluster 3 have the lowest percentage of aggressive behavior, and they have the highest proportion of provision inside the spread. On the other hand, algorithms in clusters 1 and 2 render similar proportions for aggressive behavior and missed aggressive behavior, but they differ in their proportions for providing liquidity inside the spread. Here, algorithms in cluster 1 provide liquidity inside the spread more often than algorithms in cluster 2. For both clusters 1 and 2, the vast majority of IoC orders successfully trade, and only a small fraction are canceled.

Next, we perform a regression similar to those above to predict if the order will take liquidity, conditioned on direction and conditioned on the order being in the first price bucket. We recall that if an order sent at time t consumes liquidity we write $A_t = 0$, if the order provides liquidity inside the spread we write $A_t = 1$, and we write $A_t = 2$ otherwise. Table 16 shows the average coefficients for the regressions associated with $A_t = 0$ (aggressive trading), $A_t = 1$ (generous provision of liquidity), and $A_t = 2$ (missed attempts).

We observe the following stylized features of trading algorithms in each cluster:

(LT-i) According to the intercept of the above tables, algorithms in clusters 1 and 2 are more likely to send aggressive orders that consume liquidity from the opposite side of the LOB. Algorithms in cluster 3 have the highest propensity to send orders that provide liquidity inside the spread, whereas algorithms in cluster 2 are

Downloaded from https://academic.oup.com/jfec/article/23/2/nbae025/7826742 by guest on 20 June 2025

Table 16. Average regression coefficients of orders that (i) take liquidity upon arrival in the exchange ($A_t = 0$), (ii) provide liquidity inside the spread upon arrival in the exchange ($A_t = 1$), and (iii) are immediately canceled by the exchange upon arrival ($A_t = 2$), conditioned on orders being eager to trade and conditioned on the direction D_t of orders. Numbers in bold are discussed in the main text.

		$\mathfrak{P}_t = 1, D_t = 1$ (Buy)			$\mathfrak{P}_t = 1, D_t = -1$ (Sell)		
		Cluster 1	**Cluster 2**	**Cluster 3**	**Cluster 1**	**Cluster 2**	**Cluster 3**
$A_t = 0$	Intercept	**0.73**	**0.97**	**−0.00**	**0.82**	**1.08**	**0.10**
	Spread	**−1.09**	**−0.95**	**−1.37**	**−1.29**	**−0.88**	**−1.35**
	Best bid volume	**−0.36**	**−0.35**	**−0.50**	**−0.93**	**−0.65**	**−0.78**
	Best ask volume	**−0.77**	**−0.65**	**−0.77**	**−0.44**	**−0.26**	**−0.47**
	Volume of algo 0–5	−0.49	−0.84	0.08	−0.43	−0.77	−0.02
	Volume excl algo 0–5	0.30	0.31	0.45	0.30	0.29	0.42
	Volume of algo 6–10	−0.39	−0.58	0.10	−0.30	−0.58	0.03
	Quad var 5 m	**−0.20**	**−0.13**	**−0.34**	**−0.27**	**−0.16**	**−0.36**
	Quad var 1 m	**−0.17**	**−0.12**	**−0.28**	**−0.21**	**−0.12**	**−0.30**
	Quad var 60 m	**0.14**	**0.25**	**0.16**	**0.16**	**0.19**	**0.17**
	Quad var 15 m	**−0.14**	**−0.04**	**−0.25**	**−0.19**	**−0.09**	**−0.26**
	Last volume transacted	−0.08	−0.12	−0.15	−0.10	−0.12	−0.12
	Imbalance of algo 0–5	0.12	−0.02	0.21	−0.12	−0.00	−0.20
	Num messages 0.1 ms	−0.21	−0.12	0.01	−0.16	−0.14	0.02
	Volume of algo 11–20	−0.18	−0.09	−0.00	−0.14	−0.19	−0.01
	Net agg buy−sell last 1 s	−0.09	−0.12	−0.05	0.13	0.10	0.10
	Imbalance excl algo 0–5	−0.09	−0.11	−0.01	0.11	0.17	0.08
	Num messages 0.1 s	0.07	0.11	0.08	0.11	0.09	0.08
	Agg sell last 360 s	0.10	0.09	0.10	0.08	0.05	0.06
	Agg sell last 60 s	0.09	0.06	0.13	0.07	0.05	0.09
	Agg sell last 1 s	0.03	0.06	0.02	−0.10	−0.10	−0.09
	Num messages 1 ms	−0.15	−0.03	0.01	−0.10	−0.05	0.03
	Agg buy last 60 s	0.04	0.04	0.06	0.10	0.07	0.06
	Return 300 s	0.04	0.04	0.12	0.02	0.01	0.08
		$\mathfrak{P}_t = 1, D_t = 1$ (Buy)			$\mathfrak{P}_t = 1, D_t = -1$ (Sell)		
		Cluster 1	**Cluster 2**	**Cluster 3**	**Cluster 1**	**Cluster 2**	**Cluster 3**
$A_t = 1$	Intercept	**0.05**	**−0.25**	**0.94**	**0.10**	**−0.24**	**0.83**
	Spread	**0.85**	**0.57**	**1.25**	**0.95**	**0.55**	**1.25**
	Best bid volume	**0.33**	**0.28**	**0.50**	**0.71**	**0.39**	**0.71**
	Best ask volume	**0.62**	**0.37**	**0.70**	**0.36**	**0.23**	**0.49**
	Volume of algo 0–5	0.21	0.68	−0.17	0.17	0.55	−0.15
	Volume of algo 6–10	0.09	0.47	−0.16	0.02	0.40	−0.13
	Volume excl algo 0–5	−0.12	−0.15	−0.36	−0.13	−0.15	−0.34
	Quad var 5 m	**0.17**	**0.08**	**0.32**	**0.15**	**0.04**	**0.29**
	Quad var 60 m	**−0.14**	**−0.16**	**−0.14**	**−0.20**	**−0.19**	**−0.18**
	Quad var 1 m	**0.13**	**0.07**	**0.25**	**0.11**	**0.05**	**0.25**
	Imbalance excl algo 0–5	−0.14	−0.11	−0.08	0.20	0.12	0.11
	Quad var 15 m	**0.11**	**0.02**	**0.23**	**0.07**	**−0.02**	**0.20**
	Num messages 1 ms	−0.17	−0.09	0.02	−0.22	−0.07	0.05
	Num messages 0.1 ms	−0.17	−0.04	0.01	−0.22	−0.05	0.04
	Inventory of algo	0.12	0.11	−0.03	0.05	0.11	0.05
	Cash of algorithm	0.12	0.11	−0.03	0.05	0.11	0.05
	Agg sell last 60 s	−0.07	−0.03	−0.08	−0.11	−0.04	−0.10
	Imbalance of algo 0–5	−0.06	−0.08	−0.12	0.02	0.00	0.12
	Net agg buy–sell last 1 s	−0.10	−0.07	−0.03	0.08	0.07	−0.03
	Volume excl algo 6–10	0.08	0.02	0.06	0.05	−0.02	0.13
	Volume of algo 11–20	0.02	0.11	−0.01	−0.04	0.12	0.04
	Net agg buy–sell last 360 s	−0.06	−0.03	0.03	0.11	0.07	0.06

(continued)

Downloaded from https://academic.oup.com/jfec/article/23/2/nbae025/7826742 by guest on 20 June 2025

Table 16. Continued

		$\mathfrak{P}_t = 1, D_t = 1$ (Buy)			$\mathfrak{P}_t = 1, D_t = -1$ (Sell)		
		Cluster 1	Cluster 2	Cluster 3	Cluster 1	Cluster 2	Cluster 3
$A_t = 2$	Intercept	−0.78	−0.72	−0.94	−0.92	−0.83	−0.93
	Spread	**0.24**	**0.39**	**0.12**	**0.35**	**0.33**	**0.11**
	Imbalance excl algo 0–5	0.23	0.23	0.09	−0.31	−0.29	−0.19
	Volume of algo 0–5	0.28	0.16	0.09	0.26	0.22	0.17
	Num messages 0.1 ms	0.38	0.17	−0.01	0.38	0.18	−0.06
	Volume of algo 6–10	0.30	0.11	0.06	0.28	0.17	0.10
	Num messages 1 ms	0.32	0.12	−0.03	0.32	0.12	−0.08
	Net agg buy–sell last 1 s	0.19	0.20	0.07	−0.20	−0.17	−0.07
	Volume excl algo 0–5	−0.18	−0.16	−0.10	−0.17	−0.13	−0.08
	Last volume transacted	0.12	0.11	0.09	0.16	0.13	0.06
	Best bid volume	**0.03**	**0.07**	**−0.00**	**0.21**	**0.25**	**0.07**
	Best ask volume	**0.15**	**0.28**	**0.07**	**0.08**	**0.03**	**−0.02**
	Agg sell last 1 s	−0.10	−0.13	−0.07	0.15	0.12	0.03

unlikely to provide liquidity inside the spread. This is consistent with the proportions reported at the beginning of this subsection, where algorithms in cluster 2 exhibit the lowest percentage of liquidity provision inside the spread (10%).

(LT-ii) Once again, algorithms across clusters agree on their behavior according to spread. Given they post eager-to-trade orders, algorithms are more likely to provide liquidity inside the spread when the bid–ask spread is wider. This is consistent with the literature, in that high-frequency traders are more likely to add limit orders to the book when the spread is wide; see Carrion (2013), Hagströmer, Nordén, and Zhang (2014), and Jarnecic and Snape (2014). Similarly, algorithms are more likely to have missed attempts when the spread is wider.

On the other hand, the wider the spread is, the less likely algorithms are to consume liquidity from the other side of the LOB. This is consistent with Hendershott et al. (2009), who show that algorithms are quick to post limit orders when spreads are wide, but are more likely to take liquidity when spreads are narrow. This observation is unsurprising, as for ASML the spread is usually equal to one tick, in which case it is neither possible to provide liquidity inside the spread nor to miss a trade attempt, making crossing the spread relatively more likely. This should also be compared with observation (P-iv); the sums of the spread coefficients in Table 14 with $\mathfrak{P}_t = 1$ and Table 16 with $A_t = 0$ show that the likelihood of an order crossing the spread is reduced when the spread is larger, unconditionally on whether the order is eager to trade or not.

(LT-iii) The results we discuss for spread also hold for some of the variables describing recent activity (as one would expect). For example, for most horizons of the quadratic variation, the larger the quadratic variation, the less likely algorithms will show aggressive behavior, and the more likely it is that they will post liquidity inside the spread. For both cases, the quadratic variation over the last trading hour has an inverse effect.

(LT-iv) The best available volumes in the LOB (best bid volume and best ask volume) also have a similar consistent effect across algorithms in all three clusters. The higher the available volumes in the market, the less likely algorithms are to consume liquidity from the opposite side of the LOB, the more likely they are to provide liquidity inside the spread, and the less likely they are to miss aggressive attempts. For liquidity taking, we observe that the magnitude of coefficients for

Downloaded from https://academic.oup.com/jfec/article/23/2/nbae025/7826742 by guest on 20 June 2025

the best ask volume is roughly twice the magnitude of coefficients for the best bid volume when describing the aggressiveness of orders to buy; conversely for orders to sell. This is consistent with the findings in (P-x); indeed, if the market builds volume in the best ask price for example, then everything else being equal, one expects prices to move down, this in turn lowers the probability of an aggressive order to buy as a consequence of the expected price move; conversely for the case of sell orders.

(LT-v) For liquidity provision inside the spread, we observe that the coefficients for best bid volume and best ask volume swap their importance when considering buys or sells. In particular, the best buy volume is more important to determine if a sell order will be sent inside the spread; conversely, for the best ask volume and buy orders.
These results are consistent with what one expects from the usual models of trading activity.

Our work has focused on ASML, which is a liquid asset. Similar clustering results hold for ING, see Appendix C.3. For shares with lower volumes traded (here AHOLD and TOMTOM), the results are less clear, possibly due to lower levels of trading in these shares. See Appendix Tables A.13–A.18 for the clustered coefficients for other shares.

5 Implications For Supervision

The results above provide a starting point to formalize and to quantify the links between a number of market features and the behavior of trading algorithms. Our findings provide statistical evidence for many microstructural stylized facts discussed in the literature, and our study is the first to confirm these facts with a unique data set that contains both member and algorithm identification. One key contribution is to show how algorithms make trading decisions as a function of fifty-three market features, some of which are visible to all market participants and the others are idiosyncratic features of the algorithms and of each market participant.

We believe the findings in the above sections are of special relevance to regulators and supervisors.

5.1 Surveillance

Given the vast number of messages that are processed every day by exchanges in electronic markets, it is difficult to spot behavior that intentionally or inadvertently may harm the integrity of markets. Either way, the regulator must develop sophisticated tools to monitor the market and understand the impact of individual algorithms on metrics of market quality.

Currently, financial regulators and supervisors use algorithms on transaction and order data to detect practices that may harm the integrity of the market. For example, in electronic trading, the Dutch Authority for the Financial Markets (AFM) relies on a combination of Suspicious Transaction and Order Reports from market participants and custom made algorithms on transaction and order data to prevent and detect market manipulation, see AFM (2021b). Similarly, the Financial Industry Regulatory Authority in the USA uses the SONAR system to monitor suspicious trading activity such as well-timed trades occurring just before public announcements. These surveillance algorithms mine data to search for unusual patterns, some of which could be designed to manipulate markets (e.g., pump and dump), to mislead the information in the LOB (e.g., spoofing), or to trade with privileged information (e.g., front running, insider trading).

Most detection algorithms tend to look for "misleading" signals in the market.[38] To learn what is misleading, one should first know which features agents take into account

[38] See Article 12 "Market Manipulation" in EU (2014).

Downloaded from https://academic.oup.com/jfec/article/23/2/nbae025/7826742 by guest on 20 June 2025

when making a trading decision. As a substitute for misleading signals, supervisors often look for outliers or unusual trading behavior. But this ignores the fundamental issue at stake: is the unusual trading behavior misleading market participants?

One should not mistake features that predict algorithmic trading behavior for features that algorithms use when making trading decisions. Anecdotally, trading firms often use a variety of non-market information sources when making decisions, while our data is restricted to quantities which are based in the market (in particular, which are observable to a regulator). This suggests that some of our conclusions will be affected by confounding factors. For example, a firm with an off-market signal which encourages them to take a directional position may often trade in a consistent direction—in this case, their intraday accumulated inventory may be a good predictor of their trading behavior in the near future, as it acts as a proxy for their unobserved signal. This distinction is important when drawing causal conclusions from our results.

Our work has implications for both (i) data-driven detection of market manipulation and (ii) supporting claims about misleading behavior in cases of market manipulation. With regards to (i), our findings lend support to the claim that supervisors should look at "unusual" behavior in the most important features as listed in Section 4.4. Regarding (ii), supervisors have difficulties backing up why some trading behavior could be considered misleading. It is often difficult to show that some behavior (e.g., creating an imbalance in the LOB), causes a reaction in other agents. We believe our findings show how changes in the most important features affect trading decisions, hence could be informative in cases of market manipulation. More precisely, our models can be used to simulate the sequence of events that follow a case of potential market manipulation. By comparing simulations including and excluding the potential manipulation, one could estimate if and how market participants reacted to the potential manipulation, thereby quantifying the effect of the potential manipulation.

5.2 Testing of Trading Algorithms

Our results provide a unique starting point to build more sophisticated trading models than those used in the extant literature. These models will be useful for market participants to build simulators to test strategies and will help financial regulators to understand market dynamics, individual behavior, and the impact of trading algorithms and strategies on the integrity of markets.

Trading firms are required to test their trading algorithms to make sure they do not behave in an unintended manner or contribute to disorderly trading conditions, see EU (2016c).[39] To do so, firms can use their own testing environment or one provided by the trading venue. Trading venues are required to provide members with simulation facilities which reproduce as realistically as possible the production (i.e., the real) environment. The simulations should allow members to test a range of scenarios that they consider relevant to their activity, and realistically reproduce disorderly trading conditions, see EU (2016b).[40]

[39] Article 5(4): The methodologies referred to in paragraph 1 shall ensure that the algorithmic trading system, trading algorithm or algorithmic trading strategy:
a) does not behave in an unintended manner;
b) complies with the investment firm's obligations under this Regulation;
c) complies with the rules and systems of the trading venues accessed by the investment firm;
d) does not contribute to disorderly trading conditions, continues to work effectively in stressed market conditions and, where necessary under those conditions, allows for the switching off of the algorithmic trading system or trading algorithm.

[40] Article 10(2)(a) Trading venues shall provide their members with access to a testing environment which shall consist of any of the following:
a) simulation facilities which reproduce as realistically as possible the production environment, including disorderly trading conditions, and which provide the functionalities, protocols and structure that allow members to test a range of scenarios that they consider relevant to their activity;
b) testing symbols as defined and maintained by the trading venue.

Downloaded from https://academic.oup.com/jfec/article/23/2/nbae025/7826742 by guest on 20 June 2025

AFM (2021a) notes that there is room for improvement for both trading firms and trading venues in testing trading algorithms. Current testing environments tend to range from one-off order books without new orders entering the market over time, generated by the trading venue at the request of a trading firm, to markets generating orders with a predetermined frequency and set parameters in real-time. Overall, AFM (2021a) notes that trading venues have difficulty creating sufficiently realistic simulation facilities. Furthermore, this report stresses that testing against disorderly trading conditions should be designed with a view to addressing the reaction of the algorithm or strategy to conditions that may create a disorderly market. Agent-based modeling is not currently being used widely in simulation facilities, nor agents trained on real data. AFM (2021a) encourages trading venues to explore innovative ways to make their simulation environments more realistic so that they are in line with the requirements.

This study is a starting point to build a simulator where market dynamics can be replicated at a granular level to understand how each message to the order book may affect individual and therefore affect collective dynamics. In particular, our statistical framework is devised to capture the fine microstructural facts that drive the trading decisions of individual algorithms, which is key to understanding how trading algorithms react to new market information and to messages to the LOB. With such a market simulator, firms will be able to test their trading algorithms and regulators will be able to study the effect of new trading algorithms on the quality of the market. Similarly, financial authorities will have a tool to build counterfactual trading tapes to gain insights into market behavior in the absence of certain trading algorithms or with potential new entrants in the market.

5.3 Clustering

In terms of the clustering exercise, there are a number of relevant insights for supervisors. Irrespective of the trading venue, a supervisor could impose its own criteria on what behavior would be expected of any Liquidity Provider, House, or Client. Then, our clustering methodology can be used to highlight trading firms whose algorithms might show behavior contrary to what one would reasonable expect based on their dealing capacity.

6 Conclusions

We built models to predict how individual agents choose the direction, the price, and the volume of orders they send to Euronext Amsterdam. These models achieved high and reliable out-of-sample accuracies and were employed to cluster trading behavior. We drew insights from the models that we fitted to data and showed (i) the value of using member and algorithm identifications, (ii) the trading behavior of the various dealing capacities, and (iii) the feature importance in the prediction tasks. We discussed how regulators can use the results of our models. We detailed the next steps to build agent-based market simulators that can test counterfactuals, which may be used in cases of market manipulation.

Future work will address the limitations we discussed. In particular, we envisage the construction of models that incorporate a large variety of features that capture events happening across markets and instruments. Lastly, future work will explore other potential implications of the results we obtain. For example, the consequences for market quality of a hypothetical update of members' algorithms according to how the remaining algorithms choose their direction, price, and volumes.

Author Contributions

The principal author is L.S.B.; the authors contributed to the article as follows: S.C. and R. G. initiated the collaboration; Á.C., S.C., R.G., and L.S.B. developed the modeling

Downloaded from https://academic.oup.com/jfec/article/23/2/nbae025/7826742 by guest on 20 June 2025

framework, with input from S.L. and L.V.; R.G. and L.S.B. implemented the primary modeling code, with input from S.C., S.L., and L.V.; Á.C., S.C., R.G., and L.S.B. wrote the article.

Funding

This work was supported by The Alan Turing Institute.

Disclaimer

The views expressed in this article are those of the authors, and not necessarily those of the Autoriteit Financiële Markten.

Appendix

Appendix A: Description of Shares

Similar to Table 5, we compute the order count, trade count, and volume traded of groups of algorithms during the first 4 weeks of training data for ING, AHOLD, and TOMTOM. Table A.1 displays the results for ING, Table A.2 for AHOLD, and Table A.3 for TOMTOM.

Table A.1. Descriptive statistics for algorithms trading in ING between 11 October 2021 and 7 November 2022

	Order count	Order count Percentage of all %	Trade count	Trade count Percentage of all %	Volume €	Volume Percentage of all %
Top 5	2,326,236	42	190,042	24	970,242,121	19
Top 10	3,542,047	64	268,601	35	1,338,267,032	27
Top 20	4,495,680	81	328,265	42	1,675,327,497	34
Top 50	5,388,559	97	600,684	77	3,631,692,876	73
All	5,555,240	100	777,720	100	4,986,039,098	100

Table A.2. Descriptive statistics for algorithms trading in AHOLD between 11 October 2021 and 7 November 2022

	Order count	Order count Percentage of all %	Trade count	Trade count Percentage of all %	Volume €	Volume Percentage of all %
Top 5	925,564	49	68,711	24	294,230,009	25
Top 10	1,250,004	66	97,494	34	378,566,543	33
Top 20	1,590,981	84	121,131	42	429,239,323	37
Top 50	1,837,989	97	228,241	80	918,281,638	79
All	1,891,709	100	285,440	100	1,163,250,600	100

Downloaded from https://academic.oup.com/jfec/article/23/2/nbae025/7826742 by guest on 20 June 2025

Table A.3. Descriptive statistics for algorithms trading in TOMTOM between 11 October 2021 and 7 November 2022

	Order count	Order count Percentage of all %	Trade count	Trade count Percentage of all %	Volume €	Volume Percentage of all %
Top 5	151,036	59	6,266	17	12,802,205	14
Top 10	193,241	75	9,173	25	24,460,828	28
Top 20	230,321	89	18,410	49	41,265,509	48
Top 50	253,284	98	32,989	88	75,158,861	87
All	257,902	100	37,284	100	86,580,402	100

Appendix B: Feature Specifications

In this section, we provide a detailed description of the variables we use in the study; these were introduced in Section 3.1.

We denote time by t and let the trading horizon be $[0, T]$ on any given day. At any time $t \in [0, T]$, let $\tilde{S}_t$ be price of the last transaction that took place before time t; if the trade involved multiple orders, $\tilde{S}_t$ is the volume weighted average price of the transaction. Let S^a_t be the best ask price available in the LOB at time t and V^a_t the aggregated volume at price level S^a_t. Similarly, let S^b_t be the best bid price available in the LOB at time t and V^b_t the aggregated volume at price level S^b_t. Thus, the spread at time t and the midprice at time t are given by

$$\mathfrak{S}_t = S^a_t - S^b_t, \quad S_t = \frac{S^a_t + S^b_t}{2}, \tag{10}$$

respectively. For a given number of levels $n \in \mathbb{N}$ denote by $V^{a,n}_t$ the total volume available in the first n levels of the LOB on the ask side; similarly, denote by $V^{b,n}_t$ the total volume available in the first n levels of the LOB on the bid side. Recall that $\mathcal{A} = \{a_1, a_2, \ldots, a_M\}$ are the identifiers for the algorithms on Euronext, and $\mathcal{B} = \{b_1, b_2, \ldots, b_N\}$ are the identifiers for the trading members of Euronext with $N \leq M$ because a trading member can have multiple trading algorithms. For $a_i \in \mathcal{A}$ (similar for $b_i \in \mathcal{B}$) let $V^{a,n,-a_i}_t$ be the volume available in the first n levels of the LOB on the ask side when we *exclude* any volume posted by algorithm a_i (similar definition for $V^{b,n,-a_i}_t$). Likewise, we use V^{a,n,a_i}_t and V^{b,n,a_i}_t for the volume posted by algorithm $a_i \in \mathcal{A}$. We define the imbalance between positive volumes V^a and V^b by

$$\mathfrak{I}(V^b, V^a) = \log(1 + V^b) - \log(1 + V^a), . \tag{11}$$

For a time window $\delta > 0$, we define the volatility of the transaction prices at time t of period δ by

$$\mathfrak{V}^\delta_t = \sqrt{\sum_{\Delta \log \tilde{S}_u \neq 0; u \in [t-\delta, t)} |\Delta \log \tilde{S}_u|^2}, \tag{12}$$

where

$$\Delta \log \tilde{S}_u = \log \tilde{S}_u - \log \tilde{S}_{u^-}, \quad \text{and } \tilde{S}_{u^-} = \lim_{v \nearrow u} \tilde{S}_v. \tag{13}$$

Downloaded from https://academic.oup.com/jfec/article/23/2/nbae025/7826742 by guest on 20 June 2025

The above definition extends to other stochastic processes in the article, that is $Y_{u^-} = \lim_{v \nearrow u} Y_v$ for any càdlàg stochastic process Y. The return at time t of period δ is given by

$$\mathfrak{R}_t^{\delta} = \log\left(\frac{S_{t^-}}{S_{t-\delta}}\right). \tag{14}$$

The number of messages at time t and period δ is denoted by $\mathfrak{U}_t^{\delta}$ and records the number of messages in the LOB during the window $[t-\delta, t)$. Here, a "message" is an order entry, an order deletion, or an order amendment.

Lastly, for algorithm $a_i \in \mathcal{A}$ and the associated trading member $b_j \in \mathcal{B}$, we let $\mathfrak{Q}_t^{a_i}$ and $\mathfrak{Q}_t^{b_j}$ denote the intraday accumulated inventory up to time t with the assumption that $\mathfrak{Q}_0^{a_i} = 0$ and $\mathfrak{Q}_0^{b_j} = 0$, and we let $\mathfrak{C}_t^{a_i}$ and $\mathfrak{C}_t^{b_j}$ be the accumulated expense (in EUR) of purchases of inventory with the assumption that $\mathfrak{C}_0^{a_i} = 0$ and $\mathfrak{C}_0^{b_j} = 0$. Note that for $x \in \{a_i, b_j\}$, the variables $\mathfrak{C}_t^{x}$ and $\mathfrak{Q}_t^{x}$ change in the same direction, that is, if inventory goes up then cash goes up and vice-versa. Cash and inventory variables are computed for trading in Euronext Amsterdam.

We are now able to describe the variables we use as features to describe the way algorithms make decisions in the market. The set of variables that an algorithm $a_i \in \mathcal{A}$ (of trading member $b_j \in \mathcal{B}$) has at its disposal at time $t \in [0, T]$ is given by:

i) Available volumes in the LOB excluding trading algorithm a_i (six variables).
In particular, we use $\log(1+V)$ where V is one of the following variables: $V_{t^-}^{x,5,-a_i}$, $V_{t^-}^{x,10,-a_i} - V_{t^-}^{x,5,-a_i}$, $V_{t^-}^{x,20,-a_i} - V_{t^-}^{x,10,-a_i}$, for $x \in \{a, b\}$.

ii) Available volumes in the LOB by trading algorithm a_i (six variables).
In particular, we use $\log(1+V)$ where V is one of the following variables: $V_{t^-}^{x,5,a_i}$, $V_{t^-}^{x,10,a_i} - V_{t^-}^{x,5,a_i}$, $V_{t^-}^{x,20,a_i} - V_{t^-}^{x,10,a_i}$, for $x \in \{a, b\}$.

iii) Imbalance in the LOB excluding trading algorithm a_i (three variables).
In particular, we use: $\mathfrak{I}(V_{t^-}^{b,5,-a_i}, V_{t^-}^{a,5,-a_i})$, $\mathfrak{I}(V_{t^-}^{b,10,-a_i} - V_{t^-}^{b,5,-a_i}, V_{t^-}^{a,10,-a_i} - V_{t^-}^{a,5,-a_i})$, and $\mathfrak{I}(V_{t^-}^{b,20,-a_i} - V_{t^-}^{b,10,-a_i}, V_{t^-}^{a,20,-a_i} - V_{t^-}^{a,10,-a_i})$.

iv) Imbalance in the LOB of trading algorithm a_i (three variables).
In particular, we use: $\mathfrak{I}(V_{t^-}^{b,5,a_i}, V_{t^-}^{a,5,a_i})$, $\mathfrak{I}(V_{t^-}^{b,10,a_i} - V_{t^-}^{b,5,a_i}, V_{t^-}^{a,10,a_i} - V_{t^-}^{a,5,a_i})$, and $\mathfrak{I}(V_{t^-}^{b,20,a_i} - V_{t^-}^{b,10,a_i}, V_{t^-}^{a,20,a_i} - V_{t^-}^{a,10,a_i})$.

v) Change of imbalance in the LOB excluding trading algorithm a_i (three variables).
In particular, we use: $\mathfrak{I}(V_{t^-}^{b,5,-a_i}, V_{t^-}^{a,5,-a_i}) - \mathfrak{I}(V_{s^-}^{b,5,-a_i}, V_{s^-}^{a,5,-a_i})$, $\mathfrak{I}(V_{t^-}^{b,10,-a_i} - V_{t^-}^{b,5,-a_i}, V_{t^-}^{a,10,-a_i} - V_{t^-}^{a,5,-a_i}) - \mathfrak{I}(V_{s^-}^{b,10,-a_i} - V_{s^-}^{b,5,-a_i}, V_{s^-}^{a,10,-a_i} - V_{s^-}^{a,5,-a_i})$, and $\mathfrak{I}(V_{t^-}^{b,20,-a_i} - V_{t^-}^{b,10,-a_i}, V_{t^-}^{a,20,-a_i} - V_{t^-}^{a,10,-a_i}) - \mathfrak{I}(V_{s^-}^{b,20,-a_i} - V_{s^-}^{b,10,-a_i}, V_{s^-}^{a,20,-a_i} - V_{s^-}^{a,10,-a_i})$ where $s < t$ is the time of the last message before t and as usual s^- is the time just before s.

vi) Log volume of quantity at best bid and best offer (two variables).
In particular, we use $\log(1+V)$ where V is one of the following variables: $V_{t^-}^{x,1}$ for $x \in \{a, b\}$.

vii) Spread in basis points (one variable).
In particular, we use $\mathfrak{S}_{t^-}/S_{t^-} \times 10,000$.

viii) Returns over a number of periods (four variables).
In particular, we use: $\mathfrak{R}_t^{1s}$, $\mathfrak{R}_t^{5s}$, $\mathfrak{R}_t^{60s}$, and $\mathfrak{R}_t^{300s}$.

ix) Volatility over a number of periods (four variables).
In particular, we use: $\mathfrak{V}_t^{1m}$, $\mathfrak{V}_t^{5m}$, $\mathfrak{V}_t^{15m}$, and $\mathfrak{V}_t^{60m}$.

x) Number of messages over number of periods (four variables).
In particular, we count the number of messages sent to the given instrument in Euronext in the last: 1 s, 100 milliseconds, 1 millisecond, and 100 microseconds.

Downloaded from https://academic.oup.com/jfec/article/23/2/nbae025/7826742 by guest on 20 June 2025

xi) Aggressive buys minus aggressive sells over previous 1, 5, 60, and 360 seconds (four variables).
xii) Aggressive buys over previous 1, 5, 60, and 360 s (four variables).
xiii) Aggressive sells over previous 1, 5, 60, and 360 s (four variables).
xiv) Volume of last transaction (one variable).
In particular, if the volume of the last transaction is $V > 0$, we employ $\log(1+V)$.
xv) Inventory of trading algorithm $\mathfrak{Q}_{t^-}^{a_i}$ (one variable).
In particular, we employ

$$\text{sign}(\mathfrak{Q}_{t^-}^{a_i}) \times \log(1 + |\mathfrak{Q}_{t^-}^{a_i}|). \tag{15}$$

xvi) Inventory of trading member $\mathfrak{Q}_{t^-}^{b_j}$ (one variable).
In particular, we employ

$$\text{sign}(\mathfrak{Q}_{t^-}^{b_j}) \times \log(1 + |\mathfrak{Q}_{t^-}^{b_j}|). \tag{16}$$

xvii) Cash of trading algorithm $\mathfrak{C}_{t^-}^{a_i}$ (one variable)

$$\mathfrak{C}_t^{a_i} = \int_0^t \tilde{S}_u \mathrm{d}\mathfrak{Q}_u^{a_i}, \quad \mathfrak{C}_0^{a_i} = 0. \tag{17}$$

In particular, we employ

$$\text{sign}(\mathfrak{C}_{t^-}^{a_i}) \times \log(1 + |\mathfrak{C}_{t^-}^{a_i}|). \tag{18}$$

xviii) Cash of trading member $\mathfrak{C}_{t^-}^{b_j}$ (one variable)

$$\mathfrak{C}_t^{b_j} = \int_0^t \tilde{S}_u \mathrm{d}\mathfrak{Q}_u^{b_j}, \quad \mathfrak{C}_0^{b_j} = 0. \tag{19}$$

n particular, we employ

$$\text{sign}(\mathfrak{C}_{t^-}^{b_j}) \times \log(1 + |\mathfrak{C}_{t^-}^{b_j}|). \tag{20}$$

Appendix C: Model Fit Results for Other Shares

C.1 Model performance

Similar to Tables 6–8, we show (i) the accuracies of the logistic regressions, (ii) the outperformance of the logistic regressions over the most frequent bucket, and (iii) the outperformance of random forests over the most frequent bucket and over the logistic regression; we do this for the shares ING, AHOLD, and TOMTOM.

Table A.4 shows the accuracies for ING, Table A.5 shows the outperformance of the logistic regressions over the most frequent bucket, and Table A.6 shows the outperformance of random forests over the most frequent bucket and over logistic regression. Here, D_t has two buckets, the variable $\mathfrak{P}_t$ has three buckets and $\mathfrak{V}_t$ has nine buckets.

Downloaded from https://academic.oup.com/jfec/article/23/2/nbae025/7826742 by guest on 20 June 2025

Table A.4. Accuracies of the logistic regression models for ING, calculated over twelve train-and-deploy exercises

	R1 D_t	R2 $\mathfrak{P}_t$	R3 $\mathfrak{V}_t$	R4 $\mathfrak{P}_t \vert D_t = 1$	R5 $\mathfrak{P}_t \vert D_t = -1$	R6 $\mathfrak{V}_t \vert D_t = 1$	R7 $\mathfrak{V}_t \vert D_t = -1$
Top 5	64 ± 9	88 ± 9	55 ± 2	91 ± 6	90 ± 7	56 ± 2	56 ± 2
Top 10	65 ± 7	87 ± 12	64 ± 21	90 ± 10	90 ± 10	64 ± 21	64 ± 21
Top 20	71 ± 13	90 ± 12	63 ± 19	92 ± 11	92 ± 10	63 ± 19	63 ± 18
Top 50	75 ± 14	87 ± 15	55 ± 21	88 ± 15	89 ± 14	56 ± 21	55 ± 21
All	78 ± 15	78 ± 22	46 ± 23	75 ± 27	76 ± 28	43 ± 25	42 ± 25
Order-weighted average	69	87	60	90	90	61	61

Table A.5. Outperformance over benchmark for ING, calculated over twelve train-and-deploy exercises

	D_t	$\mathfrak{P}_t$	$\mathfrak{V}_t$	$\mathfrak{P}_t \vert D_t = 1$	$\mathfrak{P}_t \vert D_t = -1$	$\mathfrak{V}_t \vert D_t = 1$	$\mathfrak{V}_t \vert D_t = -1$
Top 5	13 ± 8	1 ± 1	0 ± 8	3 ± 4	2 ± 4	0 ± 9	1 ± 9
Top 10	14 ± 7	3 ± 7	1 ± 6	6 ± 9	6 ± 9	1 ± 7	1 ± 7
Top 20	20 ± 13	3 ± 5	3 ± 7	5 ± 7	5 ± 7	3 ± 9	2 ± 7
Top 50	23 ± 13	4 ± 8	1 ± 6	5 ± 10	5 ± 9	2 ± 8	2 ± 7
All	25 ± 17	1 ± 12	−3 ± 12	3 ± 11	3 ± 13	−2 ± 13	−1 ± 12
Order-weighted average	18	3	0	5	5	1	1

Table A.6. Outperformance of random forests over benchmark and logistic regression for ING, calculated over twelve train-and-deploy exercises

	Outperformance					
	Over benchmark			Over logistic		
	D_t	$\mathfrak{P}_t$	$\mathfrak{V}_t$	D_t	$\mathfrak{P}_t$	$\mathfrak{V}_t$
Top 5	22 ± 8	0 ± 1	2 ± 9	9 ± 9	0 ± 1	2 ± 1
Top 10	20 ± 8	2 ± 4	2 ± 8	6 ± 8	−1 ± 2	2 ± 3
Top 20	25 ± 12	2 ± 3	3 ± 9	4 ± 6	−1 ± 2	0 ± 4
Top 50	26 ± 13	2 ± 6	2 ± 7	3 ± 6	−1 ± 2	2 ± 4
All	25 ± 17	2 ± 7	1 ± 9	1 ± 8	1 ± 7	4 ± 8
Order-weighted average	23	2	2	5	−1	2

Table A.7 shows the accuracies for AHOLD, Table A.8 shows the outperformance of the logistic regressions over the most frequent bucket, and Table A.9 shows the outperformance of random forests over the most frequent bucket and over logistic regression. Here, D_t has two buckets, the variable $\mathfrak{P}_t$ has three buckets and $\mathfrak{V}_t$ has nine buckets.

Downloaded from https://academic.oup.com/jfec/article/23/2/nbae025/7826742 by guest on 20 June 2025

Table A.7. Accuracies of the logistic regression models for AHOLD, calculated over twelve train-and-deploy exercises

	R1 D_t	R2 $\mathfrak{P}_t$	R3 $\mathfrak{V}_t$	R4 $\mathfrak{P}_t \vert D_t = 1$	R5 $\mathfrak{P}_t \vert D_t = -1$	R6 $\mathfrak{V}_t \vert D_t = 1$	R7 $\mathfrak{V}_t \vert D_t = -1$
Top 5	63 ± 7	84 ± 8	58 ± 14	86 ± 6	87 ± 6	58 ± 13	58 ± 14
Top 10	66 ± 9	87 ± 10	66 ± 21	89 ± 9	89 ± 9	66 ± 21	66 ± 21
Top 20	71 ± 13	89 ± 11	67 ± 23	90 ± 10	91 ± 10	67 ± 23	66 ± 23
Top 50	74 ± 14	84 ± 16	53 ± 23	85 ± 15	84 ± 17	55 ± 22	53 ± 23
All	78 ± 16	77 ± 21	42 ± 24	73 ± 28	73 ± 28	39 ± 27	40 ± 25
Order-weighted average	69	86	59	87	88	60	59

Table A.8. Outperformance over benchmark for AHOLD, calculated over twelve train-and-deploy exercises

	D_t	$\mathfrak{P}_t$	$\mathfrak{V}_t$	$\mathfrak{P}_t \vert D_t = 1$	$\mathfrak{P}_t \vert D_t = -1$	$\mathfrak{V}_t \vert D_t = 1$	$\mathfrak{V}_t \vert D_t = -1$
Top 5	13 ± 7	7 ± 9	5 ± 6	9 ± 10	9 ± 10	6 ± 7	5 ± 7
Top 10	16 ± 7	5 ± 7	4 ± 8	6 ± 8	7 ± 8	4 ± 9	4 ± 9
Top 20	20 ± 14	3 ± 6	4 ± 7	5 ± 7	5 ± 7	5 ± 8	3 ± 8
Top 50	22 ± 14	3 ± 7	2 ± 6	5 ± 8	3 ± 11	2 ± 8	3 ± 8
All	22 ± 21	0 ± 12	−2 ± 12	3 ± 13	4 ± 12	−1 ± 11	−1 ± 11
Order-weighted average	18	4	4	6	6	5	4

Table A.9. Outperformance of random forests over benchmark and logistic regression for AHOLD, calculated over twelve train-and-deploy exercises

	Outperformance					
	Over benchmark			Over logistic		
	D_t	$\mathfrak{P}_t$	$\mathfrak{V}_t$	D_t	$\mathfrak{P}_t$	$\mathfrak{V}_t$
Top 5	20 ± 6	5 ± 7	3 ± 9	7 ± 8	−2 ± 3	−2 ± 4
Top 10	21 ± 7	3 ± 5	2 ± 9	5 ± 6	−2 ± 3	−2 ± 3
Top 20	25 ± 13	2 ± 4	4 ± 9	5 ± 5	−1 ± 2	0 ± 5
Top 50	25 ± 14	2 ± 5	3 ± 7	3 ± 5	−1 ± 3	1 ± 3
All	22 ± 19	2 ± 7	2 ± 10	1 ± 11	2 ± 8	3 ± 7
Order-weighted average	23	3	4	5	−1	0

Table A.10 shows the accuracies for TOMTOM, Table A.11 shows the outperformance of the logistic regressions over the most frequent bucket, and Table A.12 shows the outperformance of random forests over the most frequent bucket and over logistic regression. Here, D_t has two buckets, the variable $\mathfrak{P}_t$ has three buckets and $\mathfrak{V}_t$ has ten buckets.

Downloaded from https://academic.oup.com/jfec/article/23/2/nbae025/7826742 by guest on 20 June 2025

Table A.10. Accuracies of the logistic regression models for TOMTOM, calculated over twelve train-and-deploy exercises

	R1 D_t	R2 $\mathfrak{P}_t$	R3 $\mathfrak{V}_t$	R4 $\mathfrak{P}_t \vert D_t = 1$	R5 $\mathfrak{P}_t \vert D_t = -1$	R6 $\mathfrak{V}_t \vert D_t = 1$	R7 $\mathfrak{V}_t \vert D_t = -1$
Top 5	68 ± 15	88 ± 15	57 ± 27	89 ± 15	89 ± 15	58 ± 27	58 ± 26
Top 10	66 ± 13	90 ± 13	61 ± 23	90 ± 13	91 ± 12	61 ± 23	62 ± 23
Top 20	70 ± 15	88 ± 13	62 ± 26	89 ± 12	89 ± 12	62 ± 25	62 ± 26
Top 50	73 ± 13	78 ± 19	51 ± 24	79 ± 18	77 ± 20	50 ± 24	50 ± 24
All	71 ± 24	70 ± 23	44 ± 28	54 ± 38	54 ± 39	31 ± 30	31 ± 30
Order-weighted average	67	87	52	88	88	52	52

Table A.11. Outperformance over benchmark for TOMTOM, calculated over twelve train-and-deploy exercises

	D_t	$\mathfrak{P}_t$	$\mathfrak{V}_t$	$\mathfrak{P}_t \vert D_t = 1$	$\mathfrak{P}_t \vert D_t = -1$	$\mathfrak{V}_t \vert D_t = 1$	$\mathfrak{V}_t \vert D_t = -1$
Top 5	16 ± 15	2 ± 4	4 ± 9	3 ± 5	3 ± 4	5 ± 9	5 ± 9
Top 10	14 ± 13	1 ± 3	3 ± 7	2 ± 4	2 ± 3	2 ± 8	2 ± 7
Top 20	17 ± 16	0 ± 3	1 ± 6	1 ± 4	2 ± 3	1 ± 7	1 ± 6
Top 50	19 ± 13	1 ± 5	−2 ± 7	2 ± 7	1 ± 5	−3 ± 8	−2 ± 9
All	13 ± 19	−2 ± 15	−6 ± 15	−2 ± 16	−3 ± 16	−6 ± 17	−6 ± 20
Order-weighted average	15	0	3	2	2	3	3

Table A.12. Outperformance of random forests over benchmark and logistic regression for TOMTOM, calculated over twelve train-and-deploy exercises

	Outperformance					
	Over benchmark			Over logistic		
	D_t	$\mathfrak{P}_t$	$\mathfrak{V}_t$	D_t	$\mathfrak{P}_t$	$\mathfrak{V}_t$
Top 5	23 ± 16	2 ± 3	7 ± 10	8 ± 9	−1 ± 1	2 ± 2
Top 10	19 ± 15	1 ± 2	4 ± 8	5 ± 7	0 ± 1	1 ± 2
Top 20	22 ± 15	0 ± 2	2 ± 6	5 ± 9	0 ± 1	1 ± 3
Top 50	22 ± 14	1 ± 4	0 ± 6	3 ± 7	0 ± 3	3 ± 3
All	13 ± 19	0 ± 7	0 ± 8	0 ± 13	2 ± 13	6 ± 13
Order-weighted average	24	0	5	9	0	2

C.2 Feature importance

Below, we carry out a robustness check where we compute the feature importance with the PCA-transformed features instead of using the original features. Figure A.1 shows the importance of the PCs when predicting the direction D_t in ASML.

Downloaded from https://academic.oup.com/jfec/article/23/2/nbae025/7826742 by guest on 20 June 2025

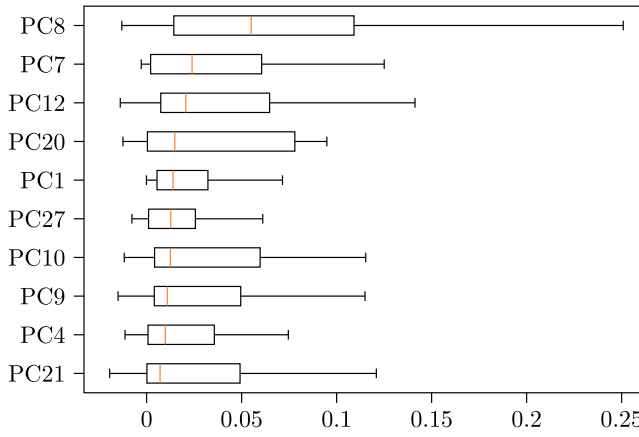


Figure A.1. Importance of PCs to explain the direction D_t of an order for ASML. We use permutation importance and logistic regressions, and only show the top ten most important PCs.

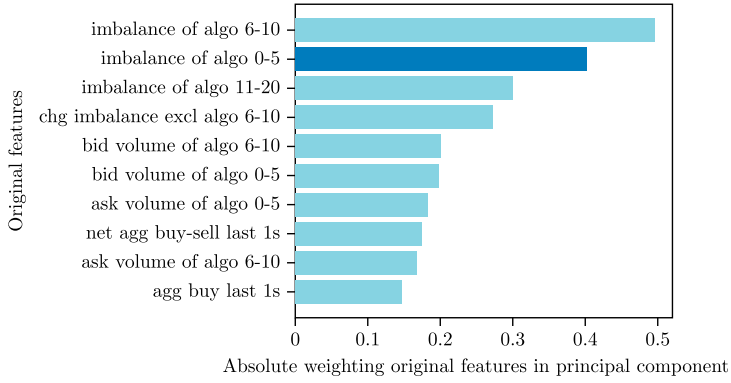


Figure A.2. Absolute weighting original features in PC for ASML. We only show the top 10 original features with the highest absolute weighting.

Next, in Figure A.2 we look at the original features that have a higher influence in PC number 8.

The results are consistent with the previous finding that pointed at the imbalance of algorithms being the most important to predict direction.

C.2.1 Feature importance: ING, AHOLD, and TOMTOM

We present the permutation feature importance using the logistic regressions. Similar to Figure 5—where we show the most important features for predicting D_t, $\mathfrak{P}$, and $\mathfrak{V}$—Figure A.3 reports the analogous for ING, Figure A.4 reports the analogous for AHOLD, and Figure A.5 reports the analogous for TOMTOM.

Unlike the case for ASML, inventory or cash-related features are the most important features when predicting D_t for ING, AHOLD, and TOMTOM. Spread retains its place as most important feature when predicting $\mathfrak{P}$, while inventory of algorithm and cash of algorithm are also important features for ING, and AHOLD. Consistently across ING,

Downloaded from https://academic.oup.com/jfec/article/23/2/nbae025/7826742 by guest on 20 June 2025

AHOLD, and TOMTOM, the four most important features to predict $\mathfrak{V}$ are either inventory variables (inventory of algorithm and inventory of member) or cash variables (cash of algorithm and cash of member).

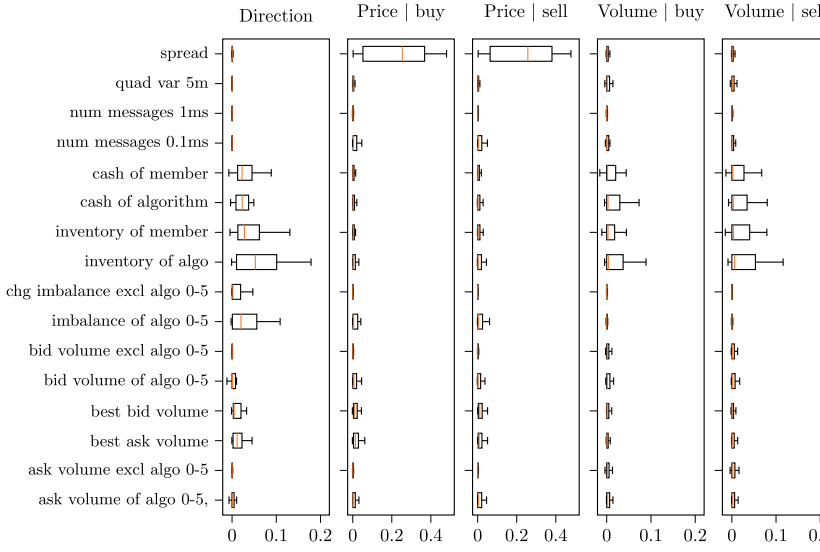


Figure A.3. Importance of features to explain the direction D_t of an order, the price bucket $\mathfrak{P}_t$ of an order conditional on the direction D_t of the order, and the volume bucket $\mathfrak{V}_t$ of an order conditional on the direction D_t of the order for ING. We use permutation importance and logistic regressions, and only show the features that are in the top ten most important features for at least one target variable.

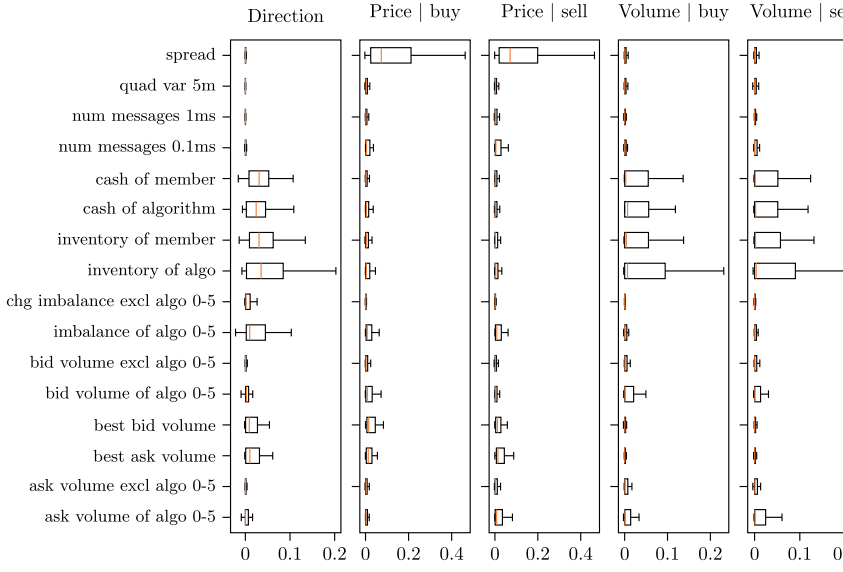


Figure A.4. Importance of features to explain the direction D_t of an order, the price bucket $\mathfrak{P}_t$ of an order conditional on the direction D_t of the order, and the volume bucket $\mathfrak{V}_t$ of an order conditional on the direction D_t of the order for AHOLD. We use permutation importance and logistic regressions, and only show the features that are in the top ten most important features for at least one target variable.

Downloaded from https://academic.oup.com/jfec/article/23/2/nbae025/7826742 by guest on 20 June 2025

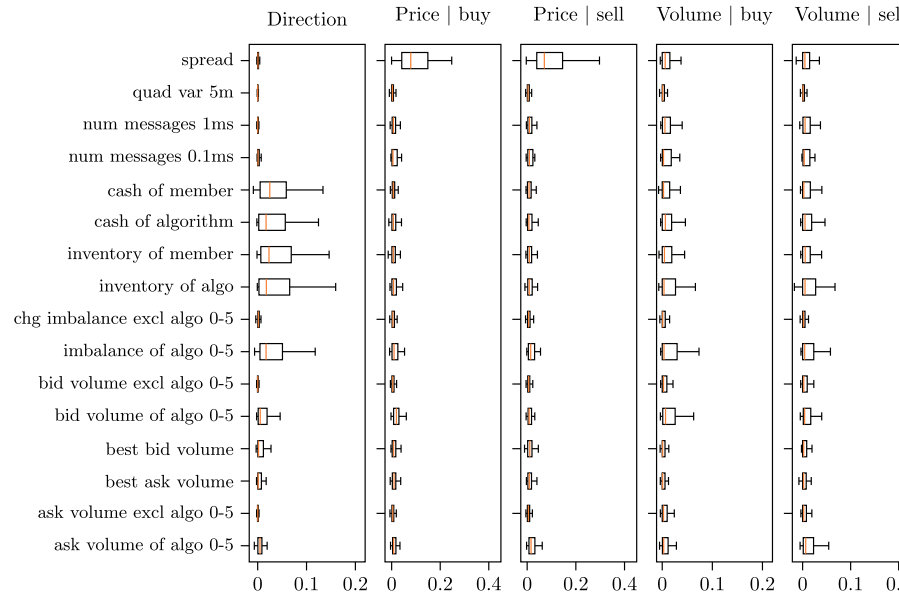


Figure A.5. Importance of features to explain the direction D_t of an order, the price bucket $\mathfrak{P}_t$ of an order conditional on the direction D_t of the order, and the volume bucket $\mathfrak{V}_t$ of an order conditional on the direction D_t of the order for TOMTOM. We use permutation importance and logistic regressions, and only show the features that are in the top ten most important features for at least one target variable.

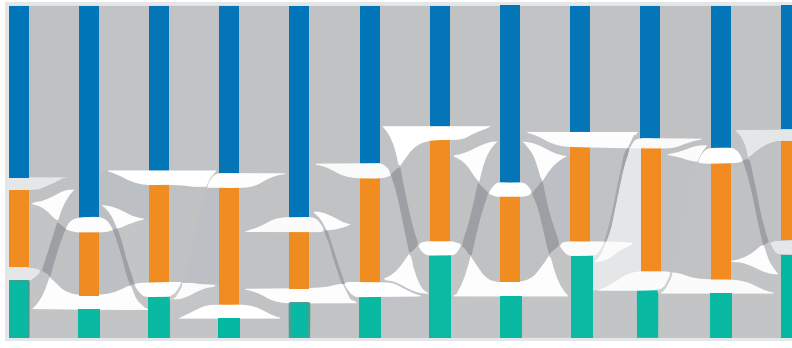


Figure A.6. Clusters of trading behavior for ING for the twelve clustering exercises. The first column represents the clustering exercise on weeks 41–44 of 2021, the second on weeks 42–45, etc. Cluster 1 is at the top (blue), cluster 2 is in the middle (orange), and cluster 3 is at the bottom (green). The size of the bars corresponds to the number of algorithms in the clusters—they add up to ninety-six algorithms in each of the twelve clustering exercises. The gray areas connecting consecutive clustering exercises represent the transition of algorithms from a cluster in one clustering exercise to a cluster in the next clustering exercise.

C.3 Clustering of agents

This section presents the clustering results for the shares ING, AHOLD, and TOMTOM according to the methodology explained in Section 4.5. First, we show the size of the clusters we obtain in each of the twelve clustering exercises. As before, for each of the clustering exercises, the first cluster has the most algorithms, and the third cluster has the least number of algorithms.

Downloaded from https://academic.oup.com/jfec/article/23/2/nbae025/7826742 by guest on 20 June 2025

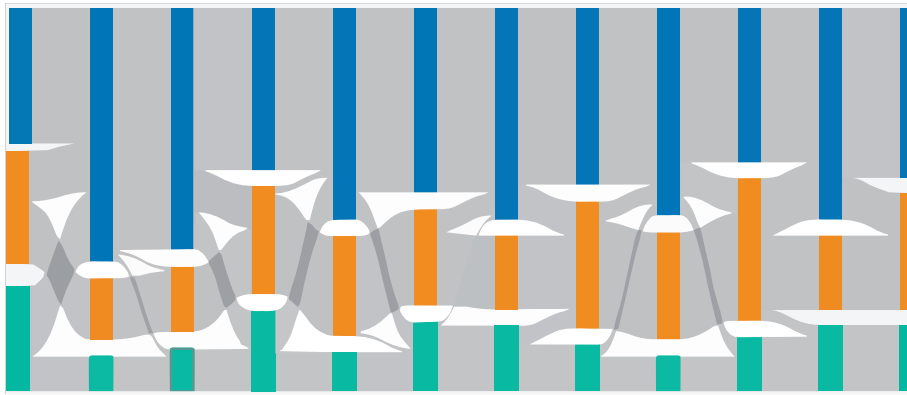


Figure A.7. Clusters of trading behavior for AHOLD for the twelve clustering exercises. The first column represents the clustering exercise on weeks 41–44 of 2021, the second on weeks 42–45, etc. Cluster 1 is at the top (blue), cluster 2 is in the middle (orange), and cluster 3 is at the bottom (green). The size of the bars corresponds to the number of algorithms in the clusters—they add up to ninety-one algorithms in each of the twelve clustering exercises. The gray areas connecting consecutive clustering exercises represent the transition of algorithms from a cluster in one clustering exercise to a cluster in the next clustering exercise.

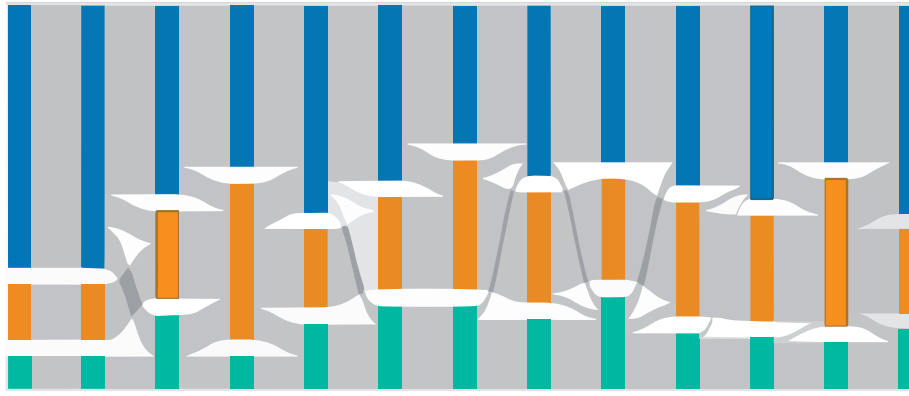


Figure A.8. Clusters of trading behavior for TOMTOM for the twelve clustering exercises. The first column represents the clustering exercise on weeks 41–44 of 2021, the second on weeks 42–45, etc. Cluster 1 is at the top (blue), cluster 2 is in the middle (orange), and cluster 3 is at the bottom (green). The size of the bars corresponds to the number of algorithms in the clusters—they add up to seventy-six algorithms in each of the twelve clustering exercises. The gray areas connecting consecutive clustering exercises represent the transition of algorithms from a cluster in one clustering exercise to a cluster in the next clustering exercise.

Second, we show the stability of clusters through time using the technique described in Section 4.5.1 to create Figure 7.

Downloaded from https://academic.oup.com/jfec/article/23/2/nbae025/7826742 by guest on 20 June 2025

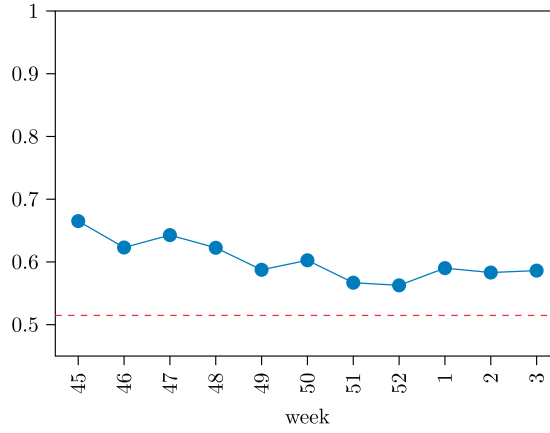


Figure A.9. Stability of clusters for ING across the twelve train exercises.

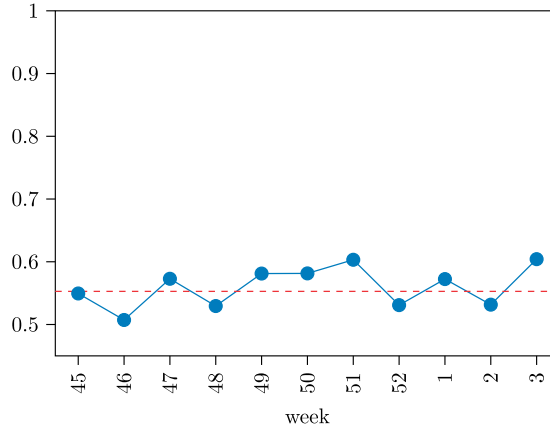


Figure A.10. Stability of clusters for AHOLD across the twelve train exercises.

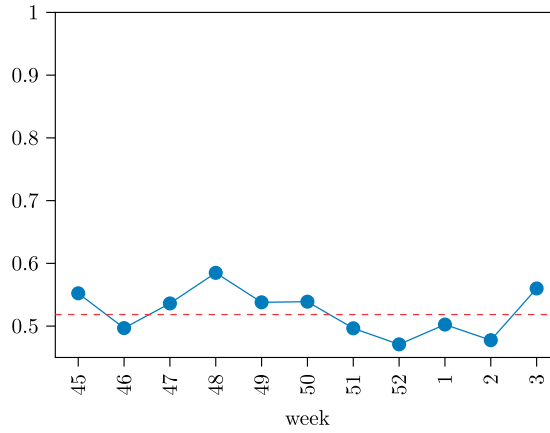


Figure A.11. Stability of clusters for TOMTOM across the twelve train exercises.

Downloaded from https://academic.oup.com/jfec/article/23/2/nbae025/7826742 by guest on 20 June 2025

Figures A.10 and A.11 show that the stability of the clusters we obtain for AHOLD and TOMTOM is not good. In particular, we see that random reshuffling of algorithms produces a similar stability score. Therefore we do not interpret the clustering results of these two shares. We believe the poor clustering might be due to the clusters on the first 4 weeks of data having substantially different sizes than the clusters obtained on other months of data of the same share (see Figures A.7 and A.8). If we were to take a clustering on the second month of data, for example, with clusters whose size is more representative of other months, then the stability metric would improve.

Figure A.12 shows an "L shape" similar to that in Figure 8. That is: the majority of algorithms are in cluster 1, with types Liquidity Provider, House, and Client all well represented. In cluster 2 we again observe mostly algorithms with type Liquidity Provider, and the same goes for cluster 3.

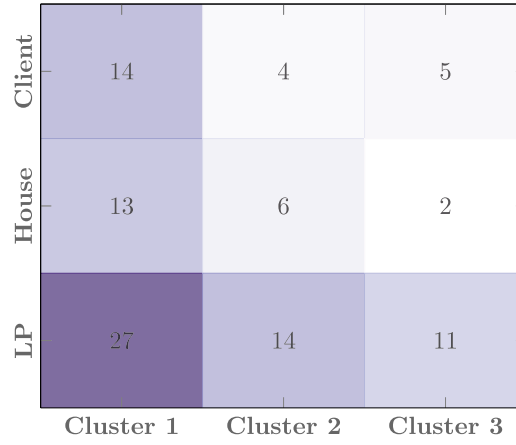


Figure A.12. Confusion matrix between dealing capacity and clusters obtained on the first 4 weeks of data for ING.

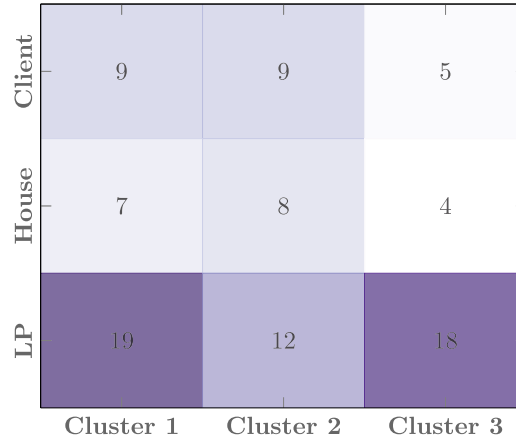


Figure A.13. Confusion matrix between dealing capacity and clusters obtained on the first 4 weeks of data for AHOLD.

Downloaded from https://academic.oup.com/jfec/article/23/2/nbae025/7826742 by guest on 20 June 2025

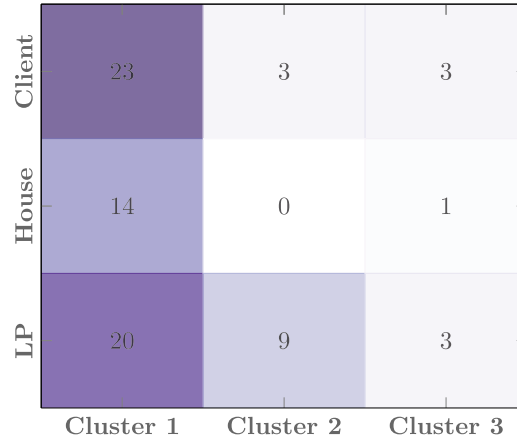


Figure A.14. Confusion matrix between dealing capacity and clusters obtained on the first 4 weeks of data for TOMTOM.

Table A.13. Average regression coefficients per cluster on first 4 weeks of training data for ING when predicting direction D_t. All excluded features have average coefficients with a magnitude smaller than 0.1 for all clusters.

	Cluster 1	Cluster 2	Cluster 3
Imbalance of algo 0–5	1.25	−1.43	0.26
Imbalance of algo 6–10	0.26	−0.55	0.98
Cash of algorithm	0.58	0.25	−0.38
Inventory of algo	0.58	0.25	−0.38
Volume of algo 11–20	0.30	−0.11	0.44
Best bid volume	0.41	−0.16	0.23
Best ask volume	−0.35	0.17	−0.24
Imbalance of algo 11–20	0.01	−0.47	−0.11
Volume of algo 6–10	0.08	−0.14	0.24
Imbalance excl algo 0–5	−0.10	−0.00	−0.22
Inventory of member	0.09	−0.09	0.15
Cash of member	0.09	−0.08	0.15
Num messages 0.1 ms	−0.02	−0.12	−0.10
Num messages 1 ms	−0.02	−0.10	−0.08
Return 1 s	0.07	0.00	0.12
Volume excl algo 0–5	0.10	0.01	−0.08
Net agg buy–sell last 1 s	0.03	−0.04	0.11
Quad var 60 m	−0.11	−0.01	0.02

C.4 Direction: Average behavior

Here we report the average regression coefficients for ING, AHOLD, and TOMTOM using the same techniques as those described in Section 4.5.6.

Downloaded from https://academic.oup.com/jfec/article/23/2/nbae025/7826742 by guest on 20 June 2025

There are similarities between ASML and ING. For imbalance of the algorithm on the first five levels, we observe a strong positive coefficient for cluster 1, a strong negative coefficient for cluster 2, and we see cluster 3 lie between the two. These results are consistent with what we observe in ASML. For inventory of algorithm, we observe a strong positive coefficient for cluster 1, a strong negative coefficient for cluster 3, and cluster 2 is somewhere in between—again consistent with the clustering obtained for ASML, including the signs of the coefficients. Lastly, we see that imbalances of the algorithm near the top of the LOB have the largest positive (resp. negative) coefficients for both ASML and ING.

Table A.14. Average regression coefficients per cluster on first 4 weeks of training data for AHOLD when predicting direction D_t. All excluded features have average coefficients with a magnitude smaller than 0.1 for all clusters.

	Cluster 1	Cluster 2	Cluster 3
Imbalance of algo 0–5	1.58	−0.24	−0.12
Cash of algorithm	0.50	0.43	0.56
Inventory of algo	0.50	0.43	0.55
Best bid volume	0.18	−0.08	0.81
Best ask volume	−0.16	0.11	−0.76
Imbalance of algo 11–20	−0.04	0.02	−0.93
Imbalance of algo 6–10	0.12	−0.16	−0.59
Volume of algo 11–20	−0.33	0.26	−0.18
Volume of algo 0–5	0.30	−0.08	0.24
Cash of member	0.03	0.31	−0.02
Inventory of member	0.03	0.31	−0.02
Quad var 60 m	−0.08	0.18	−0.01
Imbalance excl algo 11–20	0.09	0.12	−0.06
Chg imbalance excl algo 0–5	−0.01	−0.08	0.17
Return 1 s	0.10	0.01	0.15
Num messages 0.1 ms	0.03	−0.19	−0.03
Num messages 1 ms	−0.03	−0.17	−0.02
Chg imbalance excl algo 11–20	0.08	0.02	0.11
Imbalance excl algo 6–10	−0.00	0.06	−0.13
Return 5 s	0.05	0.03	0.11
Volume excl algo 11–20	−0.02	0.14	−0.02
Return 300 s	0.02	0.03	−0.10
Net agg buy–sell last 1 s	0.01	0.03	0.10
Imbalance excl algo 0–5	0.03	−0.00	−0.11
Agg sell last 1 s	−0.01	−0.02	−0.10
Quad var 15 m	0.01	0.10	−0.00

Downloaded from https://academic.oup.com/jfec/article/23/2/nbae025/7826742 by guest on 20 June 2025

Table A.15. Average regression coefficients per cluster on first 4 weeks of training data for TOMTOM when predicting direction D_t. All excluded features have average coefficients with a magnitude smaller than 0.1 for all clusters.

	Cluster 1	Cluster 2	Cluster 3
Imbalance of algo 0–5	0.58	−1.11	0.42
Imbalance of algo 6–10	−0.05	−0.76	−0.13
Cash of algorithm	0.44	0.11	−0.39
Inventory of algo	0.43	0.10	−0.39
Volume of algo 6–10	0.05	−0.26	−0.16
Num messages 1 ms	−0.02	0.04	−0.36
Imbalance of algo 11–20	−0.18	−0.06	−0.18
Best bid volume	0.30	−0.06	0.03
Num messages 0.1 ms	−0.03	0.05	−0.30
Best ask volume	−0.24	0.01	−0.12
Return 1 s	0.15	0.08	0.14
Return 5 s	0.13	0.09	0.12
Agg buy last 5 s	0.02	0.07	0.25
Return 300 s	0.10	0.05	−0.16
Volume of algo 0–5	0.08	−0.21	−0.02
Imbalance excl algo 11–20	0.19	0.05	−0.07
Num messages 0.1 s	0.03	−0.02	−0.26
Quad var 5 m	0.05	−0.04	−0.20
Chg imbalance excl algo 6–10	−0.02	0.15	−0.11
Agg buy last 1 s	0.04	0.06	0.16
Imbalance excl algo 6–10	0.04	0.09	−0.12
Agg sell last 5 s	−0.06	−0.02	0.17
Inventory of member	0.04	−0.03	0.18
Spread	−0.09	0.03	0.12
Cash of member	0.04	−0.03	0.18
Volume excl algo 6–10	0.06	0.03	−0.13
Volume of algo 11–20	−0.05	0.08	0.10
Volume excl algo 11–20	0.15	0.00	−0.05
Num messages 1 s	0.01	−0.03	0.13
Imbalance excl algo 0–5	−0.01	0.02	−0.12

C.5 Price limit: Average behavior

Here we report the results of Section 4.5.7 for ING, AHOLD, and TOMTOM. We do not comment on the coefficients of the tables below; the insights are similar to those discussed for ASML.

Downloaded from https://academic.oup.com/jfec/article/23/2/nbae025/7826742 by guest on 20 June 2025

Table A.16. Average regression coefficients for price bucket describing eager-to-trade orders ($\mathfrak{P}_t = 1$), at-the-touch orders ($\mathfrak{P}_t = 2$), and orders deeper in the LOB ($\mathfrak{P}_t = 3$) per cluster, on first 4 weeks of training data for ING, conditioning on the direction of order D_t.

		$D_t = 1$ (Buy)			$D_t = -1$ (Sell)		
		Cluster 1	**Cluster 2**	**Cluster 3**	**Cluster 1**	**Cluster 2**	**Cluster 3**
$\mathfrak{P}_t = 1$	Intercept	−0.09	−0.64	−0.20	−0.04	−0.64	−0.28
	Best bid volume	0.42	0.69	0.32	−0.22	−0.12	−0.16
	Best ask volume	−0.24	−0.12	−0.16	0.38	0.66	0.30
	Volume of algo 0–5	−0.26	0.20	−0.28	−0.26	0.28	−0.33
	Volume of algo 6–10	−0.20	0.19	−0.33	−0.22	0.25	−0.35
	Spread	0.11	0.39	0.12	0.13	0.38	0.15
	Imbalance of algo 0–5	0.05	0.43	0.01	−0.03	−0.58	0.09
	Num messages 1 ms	0.19	0.12	0.09	0.16	0.18	0.12
	Volume of algo 11–20	−0.00	0.07	−0.28	−0.06	0.05	−0.27
	Num messages 0.1 ms	0.16	0.10	0.06	0.13	0.17	0.09
	Num messages 0.1 s	0.05	0.12	0.07	0.06	0.12	0.01
	Imbalance of algo 6–10	−0.06	0.05	−0.02	0.05	−0.17	0.04
	Imbalance excl algo 6–10	−0.03	0.01	0.02	0.07	0.10	0.05
	Inventory of member	0.03	0.01	0.10	−0.05	0.01	0.08
	Cash of member	0.03	0.01	0.10	−0.05	0.01	0.08
		$D_t = 1$ (Buy)			$D_t = -1$ (Sell)		
		Cluster 1	**Cluster 2**	**Cluster 3**	**Cluster 1**	**Cluster 2**	**Cluster 3**
$\mathfrak{P}_t = 2$	Intercept	−0.07	0.08	0.09	−0.08	0.15	−0.02
	Volume of algo 0–5	0.17	0.26	0.27	0.25	0.27	0.34
	Volume of algo 11–20	−0.14	−0.23	−0.13	−0.21	−0.30	−0.10
	Imbalance of algo 0–5	0.01	0.53	0.02	0.02	−0.32	−0.03
	Volume of algo 6–10	0.05	0.10	0.12	0.09	0.06	0.19
	Spread	0.03	−0.16	0.08	0.05	−0.17	0.06
	Best ask volume	0.06	−0.02	0.11	0.04	−0.19	0.10
	Num messages 0.1 s	−0.07	−0.08	−0.08	−0.07	−0.10	−0.08
	Imbalance of algo 6–10	0.00	0.24	0.00	0.02	−0.11	−0.00
	Best bid volume	−0.01	−0.17	0.03	0.05	0.02	0.07
	Num messages 1 ms	−0.10	−0.04	−0.04	−0.10	−0.02	−0.03
		$D_t = 1$ (Buy)			$D_t = -1$ (Sell)		
		Cluster 1	**Cluster 2**	**Cluster 3**	**Cluster 1**	**Cluster 2**	**Cluster 3**
$\mathfrak{P}_t = 3$	Intercept	0.16	0.56	0.11	0.12	0.49	0.30
	Imbalance of algo 0–5	−0.06	−0.96	−0.03	0.01	0.90	−0.06
	Best ask volume	0.18	0.14	0.05	−0.41	−0.47	−0.40
	Best bid volume	−0.41	−0.52	−0.35	0.17	0.09	0.09
	Volume of algo 11–20	0.14	0.15	0.41	0.27	0.24	0.38
	Volume of algo 6–10	0.14	−0.29	0.21	0.13	−0.31	0.16
	Spread	−0.14	−0.23	−0.20	−0.17	−0.21	−0.22
	Volume of algo 0–5	0.09	−0.46	0.02	0.00	−0.55	−0.01
	Imbalance of algo 6–10	0.06	−0.28	0.02	−0.07	0.28	−0.03
	Num messages 0.1 ms	−0.08	−0.12	−0.07	−0.08	−0.21	−0.11
	Num messages 1 ms	−0.08	−0.08	−0.06	−0.07	−0.16	−0.08
	Imbalance of algo 11–20	0.06	−0.03	−0.04	−0.01	0.14	−0.04

Downloaded from https://academic.oup.com/jfec/article/23/2/nbae025/7826742 by guest on 20 June 2025

Table A.17. Average regression coefficients for price bucket describing eager-to-trade orders ($\mathfrak{P}_t = 1$), at-the-touch orders ($\mathfrak{P}_t = 2$), and orders deeper in the LOB ($\mathfrak{P}_t = 3$) per cluster, on first 4 weeks of training data for AHOLD, conditioning on the direction of order D_t.

		$D_t = 1$ (Buy)			$D_t = -1$ (Sell)		
		Cluster 1	**Cluster 2**	**Cluster 3**	**Cluster 1**	**Cluster 2**	**Cluster 3**
$\mathfrak{P}_t = 1$	Intercept	−0.30	−0.59	0.11	−0.41	−0.47	0.12
	Best bid volume	0.45	0.76	0.40	−0.19	−0.08	−0.34
	Best ask volume	−0.20	−0.10	−0.30	0.42	0.62	0.46
	Spread	0.20	0.51	0.17	0.16	0.43	0.18
	Volume of algo 11–20	−0.11	−0.08	−0.38	−0.14	−0.07	−0.39
	Volume of algo 6–10	−0.01	−0.08	−0.50	0.00	0.01	−0.49
	Volume of algo 0–5	0.07	−0.02	−0.38	0.10	0.07	−0.39
	Imbalance of algo 0–5	−0.06	0.20	0.26	0.09	−0.11	−0.21
	Num messages 1 ms	0.14	0.15	0.03	0.08	0.06	0.06
	Volume excl algo 0–5	−0.10	−0.08	−0.07	−0.02	−0.07	−0.06
	Num messages 0.1 ms	0.12	0.12	0.01	0.08	0.06	0.04
	Imbalance excl algo 11–20	−0.11	−0.06	−0.08	−0.03	0.05	0.06
		$D_t = 1$ (Buy)			$D_t = -1$ (Sell)		
		Cluster 1	**Cluster 2**	**Cluster 3**	**Cluster 1**	**Cluster 2**	**Cluster 3**
$\mathfrak{P}_t = 2$	Intercept	−0.22	0.10	0.11	−0.24	0.29	0.08
	Volume of algo 0–5	−0.03	0.18	0.33	−0.06	0.12	0.31
	Volume of algo 11–20	0.02	−0.09	−0.24	0.07	−0.11	−0.18
	Best bid volume	−0.03	−0.26	−0.02	0.08	0.03	0.11
	Num messages 0.1 s	−0.10	−0.19	0.01	−0.07	−0.15	0.01
	Spread	0.00	−0.23	−0.01	0.05	−0.20	−0.03
	Best ask volume	0.07	0.04	0.12	0.02	−0.23	−0.04
	Volume excl algo 0–5	0.10	0.15	0.02	0.09	0.09	0.03
	Volume of algo 6–10	−0.03	0.08	0.13	−0.03	0.02	0.14
	Num messages 1 ms	−0.13	−0.14	0.03	−0.06	−0.06	0.01
	Imbalance of algo 0–5	0.05	0.11	−0.07	−0.04	−0.09	0.02
	Num messages 0.1 ms	−0.11	−0.07	0.05	−0.04	−0.00	0.03
		$D_t = 1$ (Buy)			$D_t = -1$ (Sell)		
		Cluster 1	**Cluster 2**	**Cluster 3**	**Cluster 1**	**Cluster 2**	**Cluster 3**
$\mathfrak{P}_t = 3$	Intercept	0.52	0.50	−0.21	0.65	0.19	−0.20
	Volume of algo 11–20	0.08	0.17	0.62	0.08	0.19	0.56
	Best bid volume	−0.42	−0.49	−0.37	0.11	0.05	0.22
	Best ask volume	0.13	0.06	0.18	−0.43	−0.40	−0.42
	Spread	−0.21	−0.28	−0.16	−0.21	−0.23	−0.15
	Imbalance of algo 0–5	0.01	−0.31	−0.18	−0.06	0.20	0.19
	Volume of algo 6–10	0.04	−0.00	0.38	0.03	−0.04	0.35
	Volume of algo 0–5	−0.04	−0.16	0.05	−0.03	−0.19	0.08
	Imbalance of algo 6–10	0.15	−0.10	−0.12	−0.03	0.03	0.02
	Imbalance excl algo 11–20	0.14	0.08	0.11	0.01	−0.05	−0.06
	Num messages 0.1 s	0.01	0.12	−0.01	0.01	0.11	−0.02

Downloaded from https://academic.oup.com/jfec/article/23/2/nbae025/7826742 by guest on 20 June 2025

Table A.18. Average regression coefficients for price bucket describing eager-to-trade orders ($\mathfrak{P}_t = 1$), at-the-touch orders ($\mathfrak{P}_t = 2$), and orders deeper in the LOB ($\mathfrak{P}_t = 3$) per cluster, on first 4 weeks of training data for TOMTOM, conditioning on the direction of order D_t.

		$D_t = 1$ (Buy)			$D_t = -1$ (Sell)		
		Cluster 1	**Cluster 2**	**Cluster 3**	**Cluster 1**	**Cluster 2**	**Cluster 3**
$\mathfrak{P}_t = 1$	Intercept	−0.02	−0.71	−0.28	−0.04	−0.76	−0.31
	Volume of algo 0–5	−0.40	0.01	−0.55	−0.43	−0.15	−0.44
	Best bid volume	0.18	0.63	0.06	−0.15	−0.19	−0.07
	Best ask volume	−0.16	−0.15	−0.09	0.11	0.52	0.03
	Num messages 1 ms	0.14	0.19	0.15	0.12	0.04	0.14
	Imbalance of algo 0–5	−0.01	0.21	0.04	0.08	−0.30	0.01
	Volume excl algo 0–5	0.10	0.08	0.20	0.06	0.06	0.16
	Num messages 0.1 ms	0.11	0.20	0.11	0.09	0.02	0.12
	Spread	−0.03	0.22	0.01	0.02	0.18	−0.00
	Num messages 0.1 s	0.12	0.07	0.06	0.05	0.11	0.01
	Imbalance excl algo 6–10	−0.02	−0.02	0.18	0.04	0.05	0.11
	Volume of algo 6–10	−0.11	−0.02	−0.06	−0.07	0.09	−0.04
	Imbalance excl algo 11–20	−0.00	−0.04	−0.11	0.01	0.06	−0.12
	Imbalance of algo 11–20	0.04	0.12	−0.01	0.03	−0.10	0.03
	Last volume transacted	0.05	−0.01	−0.11	0.04	0.01	−0.07
	Volume of algo 11–20	−0.04	−0.03	0.07	0.00	0.10	0.05
	Imbalance of algo 6–10	−0.00	0.05	0.01	0.09	−0.10	−0.00
	Num messages 1 s	0.02	−0.00	−0.04	−0.01	0.10	−0.07
		$D_t = 1$ (Buy)			$D_t = -1$ (Sell)		
		Cluster 1	**Cluster 2**	**Cluster 3**	**Cluster 1**	**Cluster 2**	**Cluster 3**
$\mathfrak{P}_t = 2$	Intercept	0.08	−0.36	−0.06	−0.02	−0.25	−0.03
	Volume of algo 0–5	0.13	−0.01	0.13	0.22	0.04	0.18
	Best bid volume	−0.06	−0.26	−0.02	0.09	0.13	0.06
	Volume of algo 11–20	−0.14	0.09	−0.13	−0.10	−0.01	−0.14
	Best ask volume	0.11	0.12	0.03	0.01	−0.24	0.04
	Num messages 0.1 s	−0.05	−0.08	0.06	−0.06	−0.14	0.12
	Imbalance excl algo 0–5	0.04	0.06	−0.10	0.00	−0.08	−0.12
	Volume of algo 6–10	−0.04	−0.08	−0.07	−0.02	−0.13	0.06
	Imbalance excl algo 11–20	0.02	0.05	0.23	−0.02	−0.05	−0.01
	Volume excl algo 0–5	−0.00	−0.03	−0.15	0.01	0.08	−0.09
	Num messages 1 s	−0.01	−0.06	0.04	−0.02	−0.13	0.10
	Volume excl algo 11–20	−0.05	0.07	0.15	−0.04	0.01	−0.04
	Last volume transacted	−0.05	0.08	0.11	0.01	0.07	0.02
	Imbalance of algo 6–10	−0.03	0.14	−0.00	0.03	−0.11	−0.01
	Quad var 15 m	−0.06	−0.04	0.04	−0.02	−0.13	−0.02
	Imbalance of algo 0–5	0.01	0.09	−0.00	−0.02	−0.15	−0.04
	Num messages 1 ms	−0.08	−0.10	−0.00	−0.05	0.01	−0.06
	Quad var 60 m	−0.05	−0.04	0.03	−0.02	−0.14	0.00
	Volume excl algo 6–10	0.01	0.05	0.10	0.05	−0.00	−0.01
	Cash of member	−0.00	−0.00	−0.03	−0.02	0.00	0.12
	Inventory of member	0.00	−0.00	−0.03	−0.02	0.00	0.12

(continued)

Downloaded from https://academic.oup.com/jfec/article/23/2/nbae025/7826742 by guest on 20 June 2025

Table A.18. Continued

		$D_t = 1$ (Buy)			$D_t = -1$ (Sell)		
		Cluster 1	Cluster 2	Cluster 3	Cluster 1	Cluster 2	Cluster 3
$\mathfrak{P}_t = 3$	Intercept	−0.06	1.07	0.34	0.06	1.02	0.33
	Volume of algo 0–5	0.26	0.00	0.41	0.21	0.11	0.25
	Imbalance of algo 0–5	0.01	−0.30	−0.04	−0.06	0.46	0.03
	Best bid volume	−0.11	−0.37	−0.04	0.06	0.06	0.01
	Best ask volume	0.05	0.03	0.05	−0.12	−0.28	−0.07
	Volume of algo 11–20	0.18	−0.06	0.06	0.10	−0.09	0.09
	Imbalance of algo 6–10	0.03	−0.19	−0.01	−0.12	0.22	0.02
	Volume of algo 6–10	0.15	0.10	0.13	0.09	0.04	−0.02
	Num messages 1 ms	−0.07	−0.09	−0.15	−0.07	−0.04	−0.08
	Volume excl algo 0–5	−0.09	−0.05	−0.05	−0.07	−−0.14	−0.07
	Spread	−0.04	−0.15	−0.03	−0.07	−0.13	−0.04
	Imbalance excl algo 6–10	0.00	0.02	−0.23	−0.01	−0.10	−0.09
	Num messages 0.1 ms	−0.04	−0.11	−0.10	−0.07	−0.05	−0.03
	Imbalance of algo 11–20	−0.08	−0.13	−0.01	0.01	0.12	−0.03
	Num messages 0.1 s	−0.07	0.01	−0.12	0.02	0.02	−0.13
	Cash of algorithm	0.00	0.05	−0.02	0.05	−0.07	0.14
	Inventory of algo	0.00	0.05	−0.02	0.05	−0.07	0.14
	Volume excl algo 11–20	−0.04	−0.04	−0.11	−0.04	−0.02	0.06
	Imbalance excl algo 0–5	−0.08	0.02	0.04	0.02	0.11	0.03
	Imbalance excl algo 11–20	−0.02	−0.01	−0.12	0.01	−0.01	0.13
	Quad var 15 m	−0.01	0.12	0.03	0.00	0.10	0.03
	Quad var 60 m	−0.01	0.12	0.04	0.00	0.08	0.01
	Agg sell last 360 s	−0.04	0.01	−0.11	−0.04	−0.00	−0.06

References

Abergel, F., M. Anane, A. Chakraborti, A. Jedidi, and I. M. Toke. 2016. *Limit Order Books*. Cambridge: Cambridge University Press.

AFM. 2021a. "Algorithmic Trading—governance and Controls." Available at https://www.afm.nl/en/sector/actueel/2021/april/beheersing-controles-handelsalgoritmes. Accessed 1 May 2023.

AFM. 2021b. "Prevention and Detection of Market Abuse." Available at https://www.afm.nl/en/sector/themas/beurzen-en-effecten/afm-market-watch. Accessed 18 April 2023.

AFM. 2023. "Machine Learning in Trading Algorithms Application by Dutch Proprietary Trading Firms and Possible Risks." Available at https://www.afm.nl/en/sector/actueel/2023/maart/her-machine-learning. Accessed 7 March 2023.

Aït-Sahalia, Y, J. Fan, L. Xue, and Y. Zhou. 2022. "How and When are High-frequency Stock Returns Predictable? *SSRN 4095405*.

Aït-Sahalia, Y, and M. Saglam. 2013. *High Frequency Traders: Taking Advantage of Speed*. Technical Report, Cambridge, MA, National Bureau of Economic Research.

Amihud, Y., and H. Mendelson. 1980. Dealership Market: Market-Making with Inventory. *Journal of Financial Economics* 8: 31–53.

Aquilina, M., E. Budish, and P. O'neill. 2022. Quantifying the High-Frequency Trading "Arms Race". *The Quarterly Journal of Economics* 137: 493–564.

Assefa, S. A., D. Dervovic, M. Mahfouz, R. E. Tillman, P. Reddy, and M. Veloso. 2020. "Generating Synthetic Data in Finance: Opportunities, Challenges and Pitfalls." *Proceedings of the First ACM International Conference on AI in Finance*, 1–8. New York, NY: ACM.

Avellaneda, M., and S. Stoikov. 2008. High-Frequency Trading in a Limit Order Book. *Quantitative Finance* 8: 217–224.

Bouchaud, J. P., Y. Gefen, M. Potters, and M. Wyart. 2003. Fluctuations and Response in Financial Markets: The Subtle Nature of Random' Price Changes. *Quantitative Finance* 4: 176–190.

Downloaded from https://academic.oup.com/jfec/article/23/2/nbae025/7826742 by guest on 20 June 2025

Bouchaud, J. P., M. Mézard, and M. Potters. 2002. Statistical Properties of Stock Order Books: Empirical Results and Models. *Quantitative Finance* 2: 251–256.
Breiman, L. 2001. Random Forests. *Machine Learning* 45: 5–32.
Brogaard, J., T. Hendershott, and R. Riordan. 2014. High-Frequency Trading and Price Discovery. *Review of Financial Studies* 27: 2267–2306.
Brogaard, J., T. Hendershott, and R. Riordan. 2019. Price Discovery without Trading: Evidence from Limit Orders. *The Journal of Finance* 74: 1621–1658.
Byrd, D, M. Hybinette, and T. H. Balch. 2020. ABIDES: Towards High-Fidelity Multi-Agent Market Simulation. *Proceedings of the 2020 ACM SIGSIM Conference on Principles of Advanced Discrete Simulation*. New York, NY, USA: Association for Computing Machinery, 11–22.
Carrion, A. 2013. Very Fast Money: High-Frequency Trading on the NASDAQ. *Journal of Financial Markets* 16: 680–711.
Cartea, Á., R. Donnelly, and S. Jaimungal. 2018. Enhancing Trading Strategies with Order Book Signals. *Applied Mathematical Finance* 25: 1–35.
Cartea, Á., and S. Jaimungal. 2015. Optimal Execution with Limit and Market Orders. *Quantitative Finance* 15: 1279–1291.
Cartea, Á., S. Jaimungal, and J. Penalva. 2015. *Algorithmic and High-Frequency Trading*. Cambridge: Cambridge University Press.
Cartea, Á., S. Jaimungal, and Y. Wang. 2020. Spoofing and Price Manipulation in Order-Driven Markets. *Applied Mathematical Finance* 27: 67–98.
Cartea, Á., R. Payne, J. Penalva, and M. Tapia. 2019. Ultra-Fast Activity and Intraday Market Quality. *Journal of Banking & Finance* 99: 157–181.
Cartea, Á., and J. Penalva. 2012. Where is the Value in High Frequency Trading? *Quarterly Journal of Finance* 02: 1250014.
Cartea, Á., and L. Sánchez-Betancourt. 2021. The Shadow Price of Latency: Improving Intraday Fill Ratios in Foreign Exchange Markets. *SIAM Journal on Financial Mathematics* 12: 254–294.
Cartea, Á., and L. Sánchez-Betancourt. 2023. Optimal Execution with Stochastic Delay. *Finance and Stochastics* 27: 1–47.
Cohen, S. N., D. Snow, and L. Szpruch. 2023. Black-box Model Risk in Finance. *Machine Learning and Data Sciences for Financial Markets: A Guide to Contemporary Practices*. Cambridge: Cambridge University Press, 687–717.
Cohen, S. N., and L. Szpruch. 2012. A Limit Order Book Model for Latency Arbitrage. *Mathematics and Financial Economics* 6: 211–227.
Cont, R., M. Cucuringu, V. Glukhov, and F. Prenzel. 2023. Analysis and Modeling of Client Order Flow in Limit Order Markets. *Quantitative Finance* 23: 187–205.
Cont, R., S. Stoikov, and R. Talreja. 2010. A Stochastic Model for Order Book Dynamics. *Operations Research* 58: 549–563.
Dutta, C., K. Karpman, S. Basu, and N. Ravishanker. 2022. Review of Statistical Approaches for Modeling High-Frequency Trading Data. *Sankhya B* 85: 1–48.
EU. 2014. "Market Abuse Regulation." *Official Journal of the European Union*. Available at https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32014R0596&from=EN. Accessed 9 March 2023.
EU. 2016a. "Annexes to the Commission Delegated Regulation (EU) 600/2014." *Official Journal of the European Union*, 17. Available at https://ec.europa.eu/finance/securities/docs/isd/mifid/rts/160728-rts-22-annex_en.pdf. Accessed April 2024.
EU. 2016b. "Commission Delegated Regulation (EU) 2017/584." *Official Journal of the European Union*. Available at https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32017R0584&from=EN. Accessed 1 May 2023.
EU. 2016c. "Commission Delegated Regulation (EU) 2017/589." *Official Journal of the European Union*. Available at https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32017R0589&from=EN. Accessed 1 May 2023.
Euronext. 2023a. "AHOLD Asset Description." Available at https://live.euronext.com/en/product/equities/NL0011794037-XAMS. Accessed 7 March 2023.
Euronext. 2023b. "ASML Asset Description." Available at https://live.euronext.com/en/product/equities/NL0010273215-XAMS. Accessed 7 March 2023.

Downloaded from https://academic.oup.com/jfec/article/23/2/nbae025/7826742 by guest on 20 June 2025

Euronext. 2023c. "ING Asset Description." Available at https://live.euronext.com/en/product/equities/NL0011821202-XAMS. Accessed 7 March 2023.
Euronext. 2023d. "Rule Book." Available at https://www.euronext.com/en/media/1905. Accessed 7 March 2023.
Euronext. 2023e. "TOMTOM Asset Description." Available at https://live.euronext.com/en/product/equities/NL0013332471-XAMS. Accessed 7 March 2023.
Farmer, J. D., and D. Foley. 2009. The Economy Needs Agent-Based Modelling. *Nature* 460: 685–686.
Farmer, J. D., P. Patelli, and I. I. Zovko. 2005. The Predictive Power of Zero Intelligence in Financial Markets. *Proceedings of the National Academy of Sciences of the United States of America* 102: 2254–2259.
Glosten, L. R., and P. R. Milgrom. 1985. Bid, Ask and Transaction Prices in a Specialist Market with Heterogeneously Informed Traders. *Journal of Financial Economics* 14: 71–100.
Goldstein, M., A. Kwan, and R. Philip. 2023. High-Frequency Trading Strategies. *Management Science* 69: 4413–4434.
Gould, M. D., M. A. Porter, S. Williams, M. McDonald, D. J. Fenn, and S. D. Howison. 2013. Limit Order Books. *Quantitative Finance* 13: 1709–1742.
Grossman, S. J., and M. H. Miller. 1988. Liquidity and Market Structure. *The Journal of Finance* 43: 617–633.
Guéant, O. 2016. *The Financial Mathematics of Market Liquidity: From Optimal Execution to Market Making*, Vol. 33. Boca Raton, FL: CRC Press.
Guéant, O., C. A. Lehalle, and J. Fernandez-Tapia. 2013. Dealing with the Inventory Risk: A Solution to the Market Making Problem. *Mathematics and Financial Economics* 7: 477–507.
Hagströmer, B, and L. Nordén. 2013. The Diversity of High-Frequency Traders. *Journal of Financial Markets* 16: 741–770.
Hagströmer, B., L. Nordén, and D. Zhang. 2014. How Aggressive Are High-Frequency Traders? *Financial Review* 49: 395–419.
Hambly, B., J. Kalsi, and J. Newbury. 2020. Limit Order Books, Diffusion Approximations and Reflected Spdes: From Microscopic to Macroscopic Models. *Applied Mathematical Finance* 27: 132–170.
Hansch, O., N. Naik, and S. Viswanathan. 1998. Do Inventories Matter in Dealership Markets? evidence from the London Stock Exchange. *The Journal of Finance* 53: 1623–1656.
Hasbrouck, J. 2018. High-Frequency Quoting: Short-Term Volatility in Bids and Offers. *Journal of Financial and Quantitative Analysis* 53: 613–641.
Hendershott, T., C. M. Jones, and A. J. Menkveld. 2011. Does Algorithmic Trading Improve Liquidity? *The Journal of Finance* 66: 1–33.
Hendershott, T., and R. Riordan. 2009. *Algorithmic Trading and Information*. Manuscript. Berkeley, CA: University of California.
Ho, T., and H. R. Stoll. 1981. Optimal Dealer Pricing under Transactions and Return Uncertainty. *Journal of Financial Economics* 9: 47–73.
Hoffmann, P. 2014. A Dynamic Limit Order Market with Fast and Slow Traders. *Journal of Financial Economics* 113: 156–169.
Jarnecic, E., and M. Snape. 2014. The Provision of Liquidity by High-Frequency Participants. *Financial Review* 49: 371–394.
Kyle, A. S. 1985. Continuous Auctions and Insider Trading. *Econometrica* 53: 1315–1335.
Lehalle, C. A., O. Guéant, and J. Razafinimanana. 2011. "High-frequency Simulations of an Order Book: A Two-scale Approach." In F. Abergel, B. K. Chakrabarti, A. Chakraborti and M. Mitra (eds.), *Econophysics of Order-driven Markets. New Economic Windows*, Milano: Springer, 73–92. https://doi.org/10.1007/978-88-470-1766-5_6
Libman, D., S. Haber, and M. Schaps. 2021. Forecasting Quoted Depth with the Limit Order Book. *Frontiers in Artificial Intelligence* 4: 667780.
Mankad, S., G. Michailidis, and A. Kirilenko. 2013. Discovering the Ecosystem of an Electronic Financial Market with a Dynamic Machine-Learning Method. *Algorithmic Finance* 2: 151–165.
Megarbane, N., P. Saliba, C. A. Lehalle, and M. Rosenbaum. 2017. The Behavior of High-Frequency Traders Under Different Market Stress Scenarios. *Market Microstructure and Liquidity* 03: 1850005.
Menkveld, A. J. 2016. The Economics of High-Frequency Trading: Taking Stock. *Annual Review of Financial Economics* 8: 1–24.
Murphy, K. P. 2022. *Probabilistic Machine Learning: An Introduction*. Cambridge, MA: MIT Press.

Downloaded from https://academic.oup.com/jfec/article/23/2/nbae025/7826742 by guest on 20 June 2025

O'Hara, M. 1998. *Market Microstructure Theory*. Hoboken, NJ: John Wiley & Sons.
Penalva, J. S., and M. Tapia. 2021. Heterogeneity and Competition in Fragmented Markets: Fees vs Speed. *Applied Mathematical Finance* 28: 143–177.
Rigaki, M., and S. Garcia. 2023. A Survey of Privacy Attacks in Machine Learning. Vol. 56. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3624010.
Roşu, I. 2009. A Dynamic Model of the Limit Order Book. *Review of Financial Studies* 22: 4601–4641.
Ruan, R., E. Bacry, and J. F. Muzy. 2023. Agent Market Orders Representation Through a Contrastive Learning Approach. *arXiv preprint arXiv:2306.05987*.
Shokri, R., M. Stronati, C. Song, and V. Shmatikov. 2017. "Membership Inference Attacks against Machine Learning Models." *2017 IEEE Symposium on Security and Privacy (SP)*, 3–18. Piscataway, NJ: IEEE.
Sirignano, J., and R. Cont. 2019. Universal Features of Price Formation in Financial Markets: Perspectives from Deep Learning. *Quantitative Finance* 19: 1449–1459.
Tao, X., A. Day, L. Ling, and S. Drapeau. 2022. On Detecting Spoofing Strategies in High-Frequency Trading. *Quantitative Finance* 22: 1405–1425.
Van Kervel, V., and A. J. Menkveld. 2019. High-Frequency Trading around Large Institutional Orders. *The Journal of Finance* 74: 1091–1137.
Verousis, T., P. Perotti, and G. Sermpinis. 2018. One Size Fits All? High Frequency Trading, Tick Size Changes and the Implications for Exchanges: Market Quality and Market Structure Considerations. *Review of Quantitative Finance and Accounting* 50: 353–392.
Vyetrenko, S., D. Byrd, N. Petosa, M. Mahfouz, D. Dervovic, M. Veloso, and T. Balch. 2020. "Get Real: Realism Metrics for Robust Limit Order Book Market Simulations." *Proceedings of the First ACM International Conference on AI in Finance*, 1–8. New York, NY: ACM.
Wang, X., C. Hoang, Y. Vorobeychik, and M. P. Wellman. 2021. Spoofing the Limit Order Book: A Strategic Agent-Based Analysis. *Games* 12: 12.
Williams, B., and A. Skrzypacz. 2020. "Spoofing in Equilibrium." Stanford University Graduate School of Business Research Paper.
Wright, I. D., M. Reimherr, and J. Liechty. 2022. A Machine Learning Approach to Classification for Traders in Financial Markets. *Stat* 11: e465.
Yao, C., and M. Ye. 2018. Why Trading Speed Matters: A Tale of Queue Rationing Under Price Controls. *The Review of Financial Studies* 31: 2157–2183.

Downloaded from https://academic.oup.com/jfec/article/23/2/nbae025/7826742 by guest on 20 June 2025

© The Author(s) 2024. Published by Oxford University Press.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.
Journal of Financial Econometrics, 2024, 23, 1–64
https://doi.org/10.1093/jjfinec/nbae025
Article