Ann-Christina Lange

The noisy motions of instruments. The performative space of high-frequency trading

2017

https://doi.org/10.25969/mediarep/2106

Veröffentlichungsversion / published version
Sammelbandbeitrag / collection article

Empfohlene Zitierung / Suggested Citation:

Nutzungsbedingungen:
Dieser Text wird unter einer Creative Commons - Namensnennung - Nicht kommerziell - Keine Bearbeitungen 3.0 Lizenz zur Verfügung gestellt. Nähere Auskünfte zu dieser Lizenz finden Sie hier:
https://creativecommons.org/licenses/by-nc-nd/3.0

Terms of use:
This document is made available under a creative commons - Attribution - Non Commercial - No Derivatives 3.0 License. For more information see:
https://creativecommons.org/licenses/by-nc-nd/3.0
The noisy motions of instruments
The performative space of high-frequency trading

Ann-Christina Lange

In a recent article in London Review of Books, Donald Mackenzie (2014a) describes the construction of fiber-optic cables and how microwave technology is being developed at significant costs to facilitate high-speed financial trading activities. What is being built is communication lines between different financial exchanges to achieve the shortest distance and thereby transmission time, when issuing an order. The newest fiber-optic cable was finished last year and crosses the Atlantic Ocean to cut off a few milliseconds in the transmission of data from New York to London. The cables are a symptom of a fundamental digital transformation within the financial sector – one that might not best be described with reference to the notion of performativity as it has been presented from an actor-network theory/social studies of science field.¹

Michel Callon and Donald MacKenzie have been considered the founding fathers of this approach. They describe how models and calculative devices used by traders and financial experts to estimate the price of an instrument indeed co-produce the very price they attempt to measure. More generally this means that the theory or model do not only describe a given phenomenon in an objective fashion but in some cases helps to create it. MacKenzie empirically investigates “the incorporation of economics (a theory, model, concept, procedure, data set etc.) into the infrastructures of markets” (2006: 19). In An Engine, Not a Camera he studies how certain financial theories and models became authoritative and indeed shape the financial markets in quite fundamental ways. Economic models

¹ Research for this paper was supported by a ‘Crowd Dynamics in Financial Markets’ Sapere Aude Grant from the Danish Council for Independent Research (http://info.cbs.dk/crowds).
are, according to MacKenzie, an engine that produce empirical facts, rather than a camera representing such facts.

This has been a highly important and valuable approach to make sense of financial markets and what would at first sight seem to be an obvious choice when investigating digital finance and its performative aspects. Today around 30 per cent of the total trading volume is executed by high-speed algorithms in the UK and around 60 per cent in the US (cf. Foresight Final Report 2012: 19). One would expect to find that such material devices would have a range of assumptions written into them, which would affect how the financial markets function. In fact, MacKenzie states that the deepest kind of performativity is achieved when economics (concepts, models and assumptions) are incorporated into algorithms, procedures, routines, and material devices (cf. 2006: 19). However, as I will show in this chapter, the algorithms used to process the financial orders via the high-speed cables and micro wave connections is not to be considered faithful to such concepts, models or assumptions. Especially, in the sub-field formally known as high-frequency trading (HFT), algorithms are used to execute orders faster than human perception and seem to interact in quite unpredictable ways. Based on ethnographic observations and interviews inside the field of high-frequency trading and algorithmic trading, I aim to demonstrate the more ‘noisy’ motions that determine the performativity of digital finance. In order to do so, I turn my focus towards the relational interaction and spatial formations that at once condition and create digital finance.

This approach poses a methodological challenge as how to study the interaction among algorithms without reference to a conscious human subject. In order to deal with this challenge, I propose to look at the topological formations at play as they have been defined by Lury et al. from a media studies perspective. The first section defines the notion of topology and explains its suitability to the field of HFT. Secondly, I describe the market microstructure and regulatory changes that gave rise to the development towards HFT. The third section investigates three features that condition the spatial relations of digital finance, namely, the exploitation of time-delay, the interaction order between algorithms and the use of special order-types (i.e. how orders are executed). The chapter ends with a brief conclusion.
NOTES ON A TOPOLOGICAL APPROACH TO HIGH-FREQUENCY TRADING

The financial technologies and infrastructures used to send orders and receive them, have been widely studied from an ANT/Callon-inspired approach of technical devices. Ethnographies have been conducted where scholars followed and described these devices, their history, and the institutions in which they are embedded (cf. Muniesa 2008; Preda 2009; Lenglet 2011). However, the rise of algorithms as an interacting agent in financial trading has implications for how to study their embeddedness. MacKenzie et al. (2012) describe trading strategies designed to identify and exploit other traders’ algorithms (algo-sniffing). As a consequence, sophisticated algorithms are designed to hide their intentions from the market. The performativity thesis does not suffice to explain the spatial relations that now perform or shape the interaction that plays out between adaptive algorithms. The human traders and their infrastructure that used to be the object of ANT-oriented research and which might be said to be embedded within a specific spatial setting, has disappeared. MacKenzie explains himself:

“Clearly, Latour and Callon’s ‘actor network theory’ (e.g. Latour 2005) and Callon’s actor-network economic sociology (e.g. Çalışkan and Callon 2009 and 2010) are pertinent when most market participants are algorithms. Actor-network theory is prepared to use the term ‘actor’ to refer to non-human entities such as algorithms. While this usage remains controversial, it would plainly be a mistake to treat trading algorithms simply as the faithful delegates of human beings. As Adrian Mackenzie notes, ‘[a]n algorithm selects and reinforces one ordering at the expense of others’ (2006: 44), but that ordering may not be the one its human programmers intended.” (MacKenzie 2014b: 3)

This means that a study of the performativity of digital finance cannot be limited to a single-sided field study – like observing the behavior inside the trading room only. The spatial setting might simply not be taken for granted. Law (1999) has developed a topological approach to space, which he defines as post-ANT arguing that objects cannot be studied without taking into account the production of the spaces in which these objects circulate. Celia and Moor (2010) developed a topological approach focusing on media-related issues. What these approaches share is the focus on spatial formations that go beyond networks and differ from what could be imagined as a place or physical (often urban) site (like the open outcry trading pit). However, what is specific about the approach developed by Lury and Moor is that it allows for what Hansen (2015: 34) refers to as an “operationality of media culture”, which he further defines as “the capacity of today’s
media machines to generate appearances of worldly sensibilities, to directly manifest the world independently of any synthetic operation of a subject or a consciousness”.

The present chapter extends and builds upon this inspiration with the aim of reapplying the concept in order to study spaces topologically different from that of the open outcry pit (i.e. the space that once condition and is conditioned by the interaction between high-speed algorithms). Michael and Rosengarten (2012) identify a topological space as when points (entities or events) that are distant can also be proximal. Dispersed links might be drawn together (by contraction) as if they were in one place. Knorr Cetina and Bruegger have shown that “the screen brings that which is geographically distant and invisible near to participants, thus rendering it interactionally present […]” (2002: 909). The performative space of digital finance is defined by external parameters (such as the physical condition of locating computer servers close to or inside the exchange to minimize transaction time and to access data faster than other market participants) which gives rise to internally generated spatial relations between different kinds of financial actors (such as high-speed algorithms trading in front of slower market participants).

As I mentioned above, and as Ignacio Farias and Anders Blok (2016: 12) also points out, investigating topological formations involves a methodological challenge: a study of the performativity of digital finance cannot be limited to one single-sided field site. In order to not only operationalize but also test the application and value of a topological approach to the study of finance, I use a combination of methods which compose a “multi-method” (Law 2004; Holmes/Marcus 2006). The methods include: qualitative interviews, observations and content analysis of documents. The data I draw upon here consist of ethnographic observations and interviews conducted inside a New York-based HFT firm near Wall Street. This data is supplemented with 50 interviews with a broad range of actors involved with HFT, including programmers, software developers, broker-dealers, exchange officials, investment bankers, and regulators (conducted in Copenhagen, London and New York since October 2013). The ethnographic work focused on the daily practices and conversations amongst HF traders, including how traders and programmers trade at their desk while monitoring preprogrammed algorithms, but the ethnographic work also followed their activities around designing and building high-frequency trading algorithms. The data offer insights into the ways in which traders reflect upon their own trading behavior and that of participants of other markets.
THE RISE OF HIGH-FREQUENCY TRADING

The transformation of the space of finance into fully automated systems started when the exchanges became electronic. Most literature dates the technological development toward fully automated trading back to the 1970s (cf. McGowan 2010; Hanson/Hall 2012). In 1971, the NASDAQ became electronic and introduced an electronic quotation system via which competing market makers could trade securities. In 1976, the New York Stock Exchange (NYSE) introduced its Designated Order Turnaround system, allowing for the electronic transmission of orders to buy and sell securities (cf. Burr 2014). This gave rise to what is called programme trading, which exploited the spread (the difference between the best offer to sell and the best bid to buy) between S&P 500 equity shares and the futures market. In the 1990s, with the introduction of Electronic Communications Networks (ECNs), this practice became widespread across different financial markets. The ECNs provided direct market access and eliminated the need for brokerage firms to facilitate trading inside the pit. In 1998, the SEC introduced the Regulation Alternative Trading Systems, which authorized ECNs. The intention was to restrict the monopoly that the NYSE and NASDAQ had gained by automating their order-matching systems. As a result, more computer systems were developed to facilitate the entry and execution of orders electronically via the use of algorithms.

However, HFT evolved more specifically as a response to both technological developments and regulatory changes. McGowan (2010), for instance, sees the rise of HFT as a direct result of the enactment of a set of US rules known as Regulation National Market System (Reg NMS). These were passed by the US Securities and Exchange Commission (SEC) in 2005 and fully enacted in 2007 in order to strengthen the US equity markets. In part, Reg NMS was a direct response to a problematization of the behavior of specialists and locals, who used to serve as market makers (meaning that if there are insufficient buyers or sellers, they maintain order flow by trading with their own capital). In 2004, however, a group of NYSE specialists were accused of not maintaining a fair market. Against this backdrop, Reg NMS aimed to secure fair competition and decrease the discretionary power of specialists (cf. Lewis 2014: 96). Among other things, this resulted in an updated rule prohibiting “trade-throughs”, i.e. the execution of trades at prices outside of the national best bid and offer (NBBO). By emphasizing the need for immediate and automatic order execution at the NBBO, Reg NMS not only targeted the discretionary power of specialists; in effect, it enabled ultra-fast market participants to exploit price discrepancies (caused by a time delay) between different exchanges.
More recent factors that have buttressed the rise of HFT include the narrowing of spreads. In 2001, US stock exchanges were permitted to quote prices in decimals instead of fractions in order to increase liquidity. This move is known as decimalization, and is widely acknowledged to have affected the overall functioning of all financial markets, as it reduced the minimum tick size or spread from one-eighth of a dollar to one cent (cf. Jennings 2001; Chen/Chou/Chung 2009). This further decreased the importance of specialists on the exchanges and eventually led to a vast increase in algorithmic trading. In this new and more liquid market structure, the institutional traders were splitting up orders executed by algorithms in order to reduce their market impact and to execute trades faster and at better prices (cf. Burr 2014). These changes all acted as catalysts for the increase of very fast, ultra-low-latency techniques, such as the use of high-speed computer programs for the execution of orders with a high level of frequency. The increased use of high-speed algorithms and the trading strategies used has led the regulators to define this as a practice with its own definition. The US Securities and Exchange Commission (SEC) defines HF traders as “professional traders acting in a proprietary capacity that engage in strategies that generate a large number of trades on a daily basis” (Securities and Exchange Commission 2010: 45). A working group under another US regulatory body, the Commodity Futures Trading Commission (CFTC), has proposed a broader definition that focuses more on the trading activity itself than on those engaged in it:

High frequency trading is a form of automated trading that employs:

(a) algorithms for decision making, order initiation, generation, routing, or execution, for each individual transaction without human direction;
(b) low-latency technology that is designed to minimize response times, including proximity and colocation services;
(c) high speed connections to markets for order entry; and
(d) high rates of orders or quotes submitted. (CFTC 2012)

HFT is both defined as a specific organizational practice, proprietary trading and as a specific use of technological tools to execute trading strategies. The later aspect is of great importance for the present chapter- low latency technology means that algorithms are designed as rather dumb and simple entities that are supposed to read the market in real-time. In order to be fast, they can only process very limited information. This point supports the argument that the assump-
tions, theories and concepts written into such devices might have a limited im-

pact on the financial markets in general.²

**EXPLOITATION OF TIME-DELAY**

The data I have collected demonstrate some common traits defining HFT strate-
gies. One key factor is that HF traders are able to exploit the price differences
between exchanges. One HF trader explained that “we profit from correlation
and hedge ourselves. We exploit securities that move in sync due to them being
tightly hedged”. This means that the traders issue orders and when that order is
“filled” – if it bought what it was asked – some of the traders’ other algorithms
would react to the price information. Similarly, a programmer from a research
firm specializing in HFT stated that “what [HF traders] do is to empirically
measure the correlation between securities. Virtually every pair of securities in
the market has a positive correlation”.

So, in most cases HF traders speculate on the correlation between different
financial products, which means that if the price of one stock moves up or down
it is very likely that another stock will do the same. It might be that they are
traded with the same index and have the same probability of following the price
moves of the whole index or that they are dependent upon the same factor, such
as oil prices or political initiatives etc. HF traders speculate on being faster than
the price move between two highly correlated financial instruments. In the words
of a CEO of a small HFT firm in New Jersey:

> “People are in the business of propagating that price impact to other securities […] So
what we are doing, basically, is transferring the price impact of one security to a large set
of other securities. That’s where liquidity comes from, we’re sourcing liquidity from other
securities and transferring them to a specific future contract and then we’re taking the
price impact from that future and spreading it to other securities.”

What the CEO characterizes as spreading is exactly this idea about profiting
from the time delay between different exchanges. This may materialize in vari-
ous ways. The algorithms used by HFT firms can be divided into three basic
types.

² For a more detailed description of the transition from the open outcry trading pit to
high-frequency trading see Borch et al. (2015) and Borch and Lange (2016).
The first type is called a spreader. It buys one instrument and sells another with as little internal latency as possible. For instance, the algorithm buys shares traded at the NYSE and futures traded at the Security Futures Exchange in Chicago (OCX). There is a 13-millisecond delay in the transmission of data from New York City to Chicago. This delay creates arbitrage opportunities of exploiting the price discrepancies between Shares traded at the NYSE and futures traded at the OCX. When the price of a share on the NYSE and its corresponding futures contract at OCX are out of sync, the algorithm would buy the less expensive one and sell it on the more expensive market.

The second type of algorithm is a scalper. This type of algorithm earns minimum incremental profits in a single instrument by buying and selling that same instrument many times a day across different trading venues.

This type of strategy is described by a trader who designs his algorithms to exploit slower market actors:

“What you do [in one HFT strategy] is making markets. So you are offering and bidding competitively on one exchange. That way when someone pays the spread, when someone buys the offer or sells the bid, they are first to know because they got filled. If they are part of that sell or buy, they find out immediately and that gives them the time-jump to go on to the next exchange and if they sold they can buy on that exchange and make profit on the difference.”

So, here, HF traders act upon a specific price move and at the same time participate in the resulting price move. They do so by constantly issuing and cancelling orders to be in front of the price move that they aim to profit from (cf. Lange 2016). Another trader, also acting CEO of a major HFT firm in Chicago, described a similar strategy:

“The fact that I am participating on the market gives me time to speed-jump because the information was a fill and that preempts market data significantly […] and when you receive that fill, that’s what triggers your hedge orders essentially, to these other exchanges.”

As one trader explained:

“We take advantage of the noisy motions on instruments where you’ll have price fluctuations that are not linked to any meaningful information, and in that case you know you can profit from that noise”. This involves reading the depth of the order book (that display the
bid- and ask-prices) and taking advantage of the probability that “there’s a large resting size at a certain level.”

The third type of algorithm is a market maker which seeks to quote bids and offers in the same instrument and makes the market buy and sell according to certain basic rules to control the risk in the same way that a scalper seeks to take advantage of noise in a single instrument. This type is explained to be a rather passive strategy since the algorithms are in fact doing nothing but waiting for the order to come in and act upon that information. In this case the HF trader does not act as a buyer or seller but acts more like a middle man that makes the buyer and seller meet.

**INTERACTING ALGORITHMS**

However, the story of HFT is more complicated than the exploitation of time differences between exchanges. As I explained previously, order execution works as messaging to the market and algorithms are designed to detect and counteract other order executions – so for example, if the algorithm puts in an order it would immediately react to the information that it gives to the market. However, if the market moves up, it waits instead of automatically executing a buy order not to get “spooked”. Similarly, larger investors with what the traders defined as “real money” (i.e. institutional investors, banks and pension funds) have developed randomization tools to hide their buy or sell intentions from the market and thereby prevent being read/predicted by HFTs. Normally they slice the order size in a way so only one-third of the order size would be revealed to the market every other second. This type of executing is done with the use of what is called an iceberg order (cf. Lenglet 2011; Lange 2016).

As a consequence, sophisticated tools are built by HF traders to detect such market moves initiated by larger investors in order to act upon or counteract expected price moves. A programmer explained his activity as “seeing if there are other people obscuring the signal, i.e. the edge that you are trying to capture, and part of that is doing constant market recognos [i.e. pattern recognition]”. Another trader offered a specific example of this kind of market recognos, the purpose of which is to detect the rhythms in buying and selling interests that the rest of the market is not aware of or does not know about:
“The shop that I started trading at, first thing they did – you know, I came from an automation background – was that they introduced me to markets and they immediately said, ‘we know that banks are using iceberg orders’, you know, hidden size, and they wanted to be able to detect the hidden size, because they are market makers and hidden size changes the typology of the market in ways that they can’t readily identify. So the first thing I did when I entered this business was to build an iceberg detector. And that is very much that kind of recognos where you’re looking for patterns that indicate other high-frequency or micro-structure activity and base decisions on that.”

What this means is that algorithms are designed to detect patterns of other algorithms with the purpose of trading in front of them. Thus a hierarchy exists between different ‘species’ of trading algorithms – between the slower and the faster ones. Iceberg orders are a device that is both conditioned and conditions how financial interaction plays out in space and time.

**The influence of order-types**

The last aspect that characterizes the transformation of the performance of financial space is the use of special order types. How traders send a message to the market (execute orders) is determined by which order types they use. The different order types are offered by the exchanges. Up and until the implementation of rule mentioned above, the Reg NMS only limit orders and existing market orders. A market order is an order that the trader uses to buy or sell an asset immediately at the best available and current price. A market order is set to execute a trade immediately with no time restrictions or the price range within which the order can be executed. The risk is that the bid and ask prices are a lot higher or lower than the current price at which the order is executed because of the time delay. A limit order on the other hand is an order used by the trader to buy or sell a set number of financial instruments at a specified price range. This means that if the price range for the specific asset (the difference between the bid and ask price is too big) the order will not be executed – it will be cancelled (rejected). If it is executed within the price range the order “got filled”. Limit orders are also used to limit the length of time an order can be outstanding before being cancelled.

One aspect of the Reg MNS was that every order had to go to the exchange offering the best price. This effected a proliferation of more or less advanced order types (execution commands) reaching far beyond basic market and limit orders. More than 200 different order types now exist. Exchanges imitate and in-
vent new and different order types in order to differentiate themselves and to serve and attract HF traders to trade at their exchange. HF traders are considered as the liquidity-providers and makes sure that the exchange always have someone to be on the other side of a given bid or offer (which means that the trade will be executed at their venue and they will earn the transaction fee).

Direct Edge’s “hide-not-slide” order type is a good example. The basic principle at US stock exchanges is that the trader who places the order first at the best current price is the one being allowed to execute his trade first. But in some cases the rule is difficult to maintain for instance in a situation where no seller is there to fill the buy order. To avoid this situation the order should be routed to the next exchange with a matching sell order. However, traders can place an order to be executed only at one specific exchange, so in the case of no matching sell order, the offered price will slide to a lower level until it gets filled. The hide-not-slide order type offer traders to issue a trade that is not displayed in the order book so that the price will not slide, but it will wait until a matching sell order comes in and only then will it be displayed. This means that the hidden order has a time advantage over other traders as that order will be executed before new incoming orders. HFTs can actually jump the queue. Apart from such special order types, rebates are offered to HF traders by most exchanges – a reduction in the fees they would normally have to pay per order executed – a feature the HF traders are highly dependent upon as they execute a high level of orders every second.

What is established here looks like a rather complex feedback structure where high-speed trading algorithms condition a specific market microstructure, without which it cannot exist. The trading algorithms that have been presented here in fact condition and shape the social structures in which they are also embedded.

**Conclusion**

The models and theories that lay the ground for the performativity thesis presented by Callon and MacKenzie were designed with the purpose of representing, predicting or even forecasting the movements of the market. High-speed algorithms on the other hand, are pre-programmed to read price moves directly as they appear in the order book. Based on a dynamic interaction with other financial actors, high-speed algorithms work by issuing an order to see how other actors (human traders and other algorithms) react to that order and is pre-programmed to issue another quote based on that information. This means that
financial devices are not performative in the sense described by Mackenzie and Callon, as they do not work to increase its ‘predictive fit’ (cf. Stark 2011).

Instead complex spatial relations are constituted where each financial actor (HFT-algorithms, execution algorithms, the exchanges matching engine etc.) exploit the inefficiencies of other species and in the process, creates new inefficiencies, again exploited by yet another kind of actor. A financial algorithm is not simply an automated rule but a device that is both conditioned and conditions, which exploits inefficiencies and at the same time, creates the interdependence of the different kinds of algorithms. It is from this background that the performative space of HFT is probably best understood as a topological one composed by interacting agents. Thus the feedback relation between interactive algorithms does not only apply to structurally correlated instruments but also to the interactional algorithmic responses between the exchanges’ matching algorithms (and order types), institutional investors’ and broker-dealers’ executing algorithms and other HFT algorithms.

What makes topology a distinctive approach to the study of social dynamics vis-à-vis other approaches to the study of finance is that it provides tools for the understanding of the financial markets that reach beyond the study of its actor-networks or agent-based interaction within the trading room. What is offered in this chapter is a presentation of how the rise and function of algorithmic trading strategies and execution technologies contribute to the making and reshaping of financial markets. Algorithms both act within and outside the market; they are both ahead of the price move they aim to profit from while also creating it. Such deformation comes with and enacts a particular spatialization of finance, in which distances and temporalities are continuously redrawn or folded into each other, complicating notions of inside and outside, distance and proximity. The elementary component of physics and non-linear algebra might inspire the analysis of how dispersed actors create an economic pattern opposed to an Euclidean geometry, to which the prevailing economic system aspires (Deleuze/Guattari 1980; Delanda 2002).

REFERENCES


Borch, Christian/Lange, Ann-Christina (2016): “High-frequency trader subjec-


