Joint Discussion Paper
Series in Economics

by the Universities of

Aachen · Gießen · Göttingen
Kassel · Marburg · Siegen

ISSN 1867-3678

No. 25-2016

Alexandru Mandes

Algorithmic and High-Frequency Trading Strategies: A Literature Review

This paper can be downloaded from
http://www.uni-marburg.de/fb02/makro/forschung/magkspapers

Coordination: Bernd Hayo • Philipps-University Marburg
School of Business and Economics • Universitätsstraße 24, D-35032 Marburg
Tel: +49-6421-2823091, Fax: +49-6421-2823088, e-mail: hayo@wiwi.uni-marburg.de

Algorithmic and High-Frequency Trading Strategies. A Literature Review

Alexandru Mandes
University of Giessen
Licher Str. 64
35394 Giessen, Germany
Alexandru.Mandes@wirtschaft.uni-giessen.de

Current draft: February 21, 2016

Abstract. The advances in computer and communication technologies have created new opportunities for improving, extending the application of or even developing new trading strategies. Transformations have been observed both at the level of investment decisions, as well as at the order execution layer. This review paper describes how traditional market participants, such as market-makers and order anticipators, have been reshaped and how new trading techniques relying on ultra-low-latency competitive advantage, such as electronic "front running", function. Also, the natural conflict between liquidity-consumers and liquidity-suppliers has been taken to another level, due to the proliferation of algorithmic trading and electronic liquidity provision strategies.

Key Words: algorithmic trading, high-frequency trading, electronic market making

JEL classification: C10, C61, C63, G19

1 Introduction

Market participants expose their buying and selling interests and finally trade with each other within an organised market system composed of several trading venues, e.g., regulated markets and multilateral trading facilities (MTF) – previously known as alternative trading systems (ATS). The most common mechanism of price discovery implemented by equity markets is the continuous double auction, where participants place their trading offer and trading demand as market or limit orders and where incoming orders are continuously matched against an order book formed of two queues of passive limit orders – one for buy (bid) and one for sell orders (ask) – sorted by price and time priority. In other words, the outstanding limit orders provide the liquidity necessary for the execution of aggressive, liquidity-taking market orders. From this point of view, the trading process, and the collateral price evolution, can be seen as an outcome of the interplay between order flow and persistent order book liquidity.

According to Harris (2002), market participants such as dealers and value traders are always passive liquidity suppliers, while other precommitted traders oscillate between liquidity-supplying and liquidity-consuming, depending on their current level of impatience. A passive execution has the advantage of a lower cost of trading as opposed to sending market orders. The latter is associated with an immediacy cost given by the bid-ask spread and, most important for large orders, a potential market impact if the order execution has to "walk up the book" in order to fill its entire size. Most informed traders, which have expectations about the future market direction, are strategically more aggressive since they need to open their positions as quickly as possible, before the market starts to move and their expected "alpha" is diminished.[1] They are also interested in minimizing transaction costs (implicit market impact) which directly reduces their overall profitability.

On the other side, while supplying liquidity, the uninformed dealers accumulate inventories which might lead to large losses in case the price should move against them, i.e. inventory or portfolio risk. If the price changes are independent of their positions – sometimes positive and sometimes negative, then the inventory risk is diversifiable (null on average). However, when being counterpart to informed traders, the order flow becomes unbalanced and the future price returns are usually inversely correlated with their current open positions, leading to trading losses, i.e. adverse selection or asymmetric information risk. Dealers can protect themselves from inventory and adverse selection risk only by quoting prices where the order flow is two-sided and the informed toxic flow is thus offset. Most of the time, this is consistent with updating the quotes in the market direction, which also takes away alpha from informed strategies ("hidden alpha"). This leads to a natural conflict between uninformed liquidity providers, which try to quickly

[1] In finance, the alpha coefficient stands for the relative return on an investment as compared to a market index benchmark.

detect market and liquidity shifts and update their quotes correspondingly, and informed traders which try to hide their intentions during large order executions ("stealth trading").

With the help of technology, both sides have upgraded their strategy implementations by replacing slow human operators with computer-based automated algorithms endowed with high computational power and fast reaction speeds. On one side, the uninformed liquidity providers have implemented automated market making programs with the goal of reducing their liquidity provision associated risk and of providing higher quality quotes – also known as Electronic Liquidity Providers (ELPs). On the other side, the informed traders rely on what is known as Algorithmic Trading (AT) strategies when executing their trades in order to minimize transaction costs. Additionally, order anticipators have incorporated high-frequency technology in order to decrease the time-frame of their microstructure trading strategies and improve their overall profitability, or even develop new strategies facilitated by their low-latency competitive advantage – commonly labeled as High-Frequency Trading (HFT).

Analyzing the recent and the future development of computer-based trading (CBT) strategies, as well as their impact on the overall market quality – captured along three distinct dimensions, i.e. liquidity, price efficiency and systemic risk – has been the subject of a large foresight study commissioned by the UK Government Office for Science (Government Office For Science, 2012). The empirical and theoretical scientific results show that the effect of CBT is controversial. On one side, market liquidity, transaction costs and price efficiency have been improved, but on the other side there is a greater risk of periodic illiquidity due to the new nature of liquidity providers, illustrated by the occurrence of several recent flash crashes. CBT may have altered the latent features of the market socio-technological system, such as self-reinforcing feedback loops, which under certain conditions exhibit chaotic properties, leading to significant financial instability.

Vuorenmaa (2012) reviews both the popularized media objections on this topic as well as the academic literature, and discusses the pros and cons of HFT and AT as seen from three different points of view, i.e. negative media writing, negative and positive academic research. Media allocated most of its publishing space to topics related to the Flash Crash of May 6, 2010 and to HFT being presented as "front runners" and predatory strategies taking advantage of slower market participants. Front running as well as manipulative techniques, such as quote stuffing, smoking, and spoofing, are claimed to rely on their "unfair" speed edge to take advantage of slower executions or to lure other traders into taking toxic positions. The negative academic research results point to increased correlations and complex interdependencies between the too homogeneous and overcrowded HFT strategies, as well as across assets and markets, contributing to higher market systemic risks and worse contagion effects – when various processes become coupled at very short time frames, volume feedback loops can be ignited leading to severe volatility extreme events. Finally, the positive academic research underlines the favorable market impact of HFT strategies, which due to their fast, predictive and accurate reactions are able to decrease bid-ask spreads and transaction costs, decrease volatility, increase liquidity both in normal and in times of high market distress, and contribute more to the efficient price discovery than slower (human) traders do.

As opposed to the previous type of reviews which focus on the aggregated market impact of automated trading, the goal of the current contribution is to review the literature which deals with the actual design of such computer based strategies. This is highly relevant from a developer's point of view, e.g., in the context of building simulations of nowadays financial markets. In Section 2, the algorithmic trading problem is defined and the two main subtypes of algorithmic trading strategies are presented. Similarly, Section 3 introduces a range of computer-based strategies, which can be applied by means of high-frequency trading. The first main HFT class – consisting in liquidity traders – is detailed in Subsection 3.1, while the second class – capturing orders anticipators – is described in Subsection 3.2. Section 4 concludes.

2 Algorithmic trading

Order size plays an important role in trading, since executing a large order is more difficult due to higher market impact and signaling risk (see Subsection 2.2 for more information on market impact). One way to overcome these issues is to slice large orders and spread their execution over time with the goal of minimising the associated implicit transaction costs. The automated programs implementing order executing strategies are widely known as Algorithmic Trading. In their definitions of AT, regulators underline the automated and computer-based decision process – with no human intervention – of determining the individual order trading parameters regarding timing, pricing and quantity setting, as well as the managing of orders after their submission.[2] For clarity, AT does not include any system which deals only with automated routing of orders or with confirming order execution, i.e. no actual determination of the trading parameters.

At one extreme, an order can be executed at once entirely – using a market order – with a high trading cost, i.e. worse execution price due to market impact. On the other side, the order can be equally split and scheduled at a constant execution rate over the entire trading period (see Figure 1) which is associated with the lowest impact, but with an intrinsic price risk, i.e. the difference between the effective execution price and the arrival price bench-

[2]Federal Financial Supervisory Authority (BaFin), Securities Trading Act (Wertpapierhandelsgesetz - WpHG); European Comission, MiFid 2 – Directive 2014/65/EU of the European Parliament and of the Council of 15 May 2014 on markets in financial instruments and amending Directive 2002/92/EC and Directive 2011/61/EU

mark due to random price movements (shortfall). The optimal scheduling lies in-between this range bounded by the minimum variance strategy, at one side, and the minimum impact strategy, at the other side. As expressed in Kissell and Malamut (2005), a trader faces the trade-off/dilemma of trading too quickly (aggressively) and trading too slowly (passively).

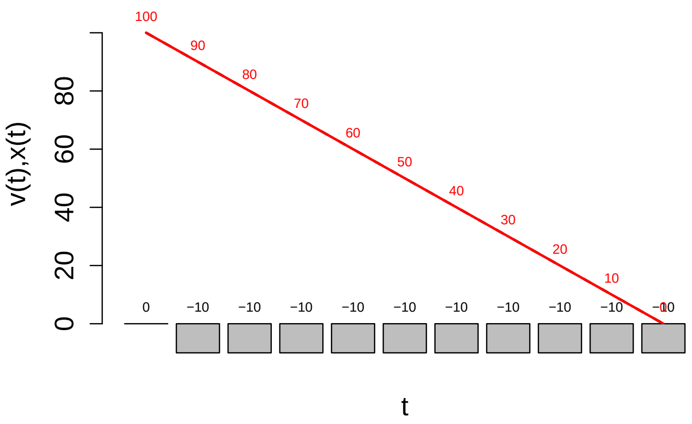

Figure 1: A linear execution schedule – time weighted average price (TWAP)

The line plot represents the trajectory $x(t)$, i.e. the remaining order size to be executed at time t. The bar chart illustrates the trading rate $\nu(t)$, the slice of the entire order executed at time t.

A formalization of the order scheduling problem is provided in Almgren and Chriss (2001). The objective is to compute a trajectory function $x(t)$ representing the number of units remaining to be traded at time t, with the initial target $x(0) = X_0$ and the completely executed order at the end of the execution time $x(T) = 0$. The corresponding trading rate, i.e. the number of units to be traded in one time interval, is denoted $\nu(t) = x(t-1) - x(t)$ (see an example of a constant trading rate in Figure 1). The optimization problem consists in finding the optimal balance between the expected trading costs and the uncertainty of these costs due to the exposure to timing risk:

$$\min_x \left(\mathtt{E}[C(x)] + \lambda \, \mathtt{Var}[C(x)] \right), \tag{2.1}$$

where:
- $C(x)$ cost of deviating from the benchmark
- $\mathtt{Var}[C(x)]$ variance of the execution cost as a proxy for risk
- $\lambda \geq 0$ agent's level of risk aversion which penalizes the variance relative to the expected cost

Johnson (2010) provides a classification of AT based on its target objectives into impact-driven, cost-centric and opportunistic algorithms. The first category is schedule-based and seeks to minimize the overall market impact costs by splitting larger orders over time. E.g., the Time Weighted Average Price (TWAP) strategy – also illustrated in Figure 1 – slices the original order into equal parts which are spread over fixed and equal time intervals throughout the day. Another impact-driven strategy, Volume Weighted Average Price (VWAP), tracks statically created trajectories based on historical volume profiles (see Subsection 2.1 for more details). Dynamic tracking algorithms such as Percentage of Volume (POV) – also known as Volume Inline – try to participate in the market at a given rate in proportion with the market's actual volume. Cost-centric algorithms try to reduce transaction costs by balancing implicit costs, such as market impact, and the exposure to timing risk by taking into account the investor's level of urgency or risk aversion. E.g., Implementation Shortfall (IS) – also known as Arrival Price – determines the optimal trade horizon, as well as the appropriate trading schedule, by applying various cost and market models which take into account factors such as order size, available time, expected price change (alpha), expected liquidity and volatility (see Subsection 2.2 for more details). Finally, opportunistic algorithms, such as Price Inline (PI), are variants of the impact-driven algorithms which are sensitive and self-adjusting to the current market conditions.

Fabozzi, Focardi, Kolm et al. (2010) identify IS and VWAP as the two most popular execution strategies. While VWAP's execution costs are assessed relative to the benchmark with the same name (see Subsection 2.1) and the algorithm aims at reducing the absolute market impact costs, IS is benchmarked to the arrival price – the price prevailing at the beginning of the execution period – and is optimized towards minimizing the overall potential risk-adjusted costs, with respect to a predefined coefficient of risk aversion. In the following two subsections, selected variants of these two algorithmic trading strategies are presented.

2.1 Volume Weighted Average Price

Volume-Weighted Average Price is an automated participation strategy which tries to achieve an execution price close to the VWAP benchmark. The latter corresponds to the overall turnover divided by the total volume $VWAP = \sum_n v_n \, p_n / \sum_n v_n$.[3] According to Madhavan (2002), the main reason for choosing this strategy is due to the benchmark choice, i.e. the criteria used for measuring the execution performance. Thus, VWAP is an ideal option for passive traders with no alpha and no need for urgency. The actual implementation consists in splitting the initial order over the entire trading period – usually the entire or a large part of a trading day – based on a model of the historical fractional daily volume pattern (see Figure 2). Since the actual intraday volume pattern for any single day will be different from the historical average, the VWAP is not guaranteed and will mostly randomly miss

[3] If a large order represents too much ($> 30\%$) of the average daily volume, VWAP has practically no meaning.

the actual VWAP, especially on days with highly unusual volume patterns. However, with the help of automation, the order can be split in a very large number of smaller orders which can be spread over finer time grids and thus reduce to some extent the problems associated with the model's noise.

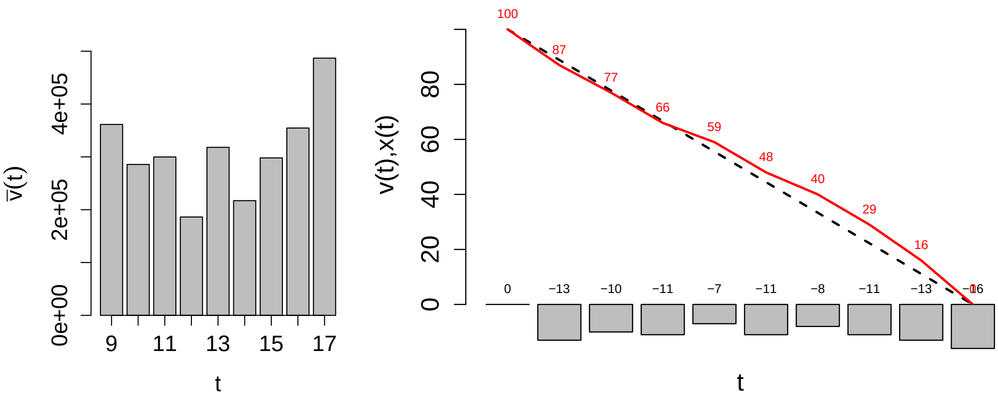

Figure 2: VWAP

The left plot presents the hourly average values of the trading volume for Siemens (SIE), traded at Xetra Frankfurt during the entire February 2012 – the notorious U-shape or the volume "smile" can be roughly observed. The right plot describes the evolution of the trading schedule and the hourly order sizes, corresponding to the historical volume pattern, i.e. the nine one hour bins in the left plot.

The VWAP can be improved by replacing the standard historical pattern with other models of the intraday volume dynamics. E.g., Białkowski, Darolles and Le Fol (2008) suggest a dynamical volume model, which decomposes the traded volume in two parts: one reflecting the common and seasonal market evolution – modeled by an extension of CAPM with factors estimated by principal component analysis – and a second one capturing the intraday specific volume dynamics by means of an ARMA(1,1) or a SETAR model. As a difference form the statical VWAP, the trading schedule cannot be determined in advance at the start of the trading period, but needs to be updated step-by-step, based on the one-step ahead prediction of the dynamic component which takes into consideration also the previous intraday period realization.

2.2 Implementation Shortfall

As its name suggests, Implementation Shortfall[4] is benchmarked to the arrival price – the price prevailing at the beginning of the execution period – and is optimized towards minimizing the overall potential risk-adjusted costs with respect to a predefined coefficient of risk aversion. According to Fabozzi et al. (2010), this strategy is especially appropriate for market participants

[4]The shortfall is the difference between the arrival and the effective execution price.

who know their risk-aversion profile and who have a strong belief about the future returns.

The total cost of trading $C(x)$ – in this case the shortfall – is the difference between the initial market value and the effective execution value, as defined in Equation 2.2. The price per share achieved on each trade $\tilde{P}_t$ is influenced by the temporary market impact which affects only the execution price $\tilde{P}_t$ and not the market equilibrium.[5] This liquidity premium can be reduced by sending smaller orders – corresponding to smaller $\nu(t)$ – while the permanent market impact is independent of the trading strategy and sums up to the same result, linear in total order size X_0.

$$C(x) = X_0\, P_0 - \sum_{t=1}^{T} \nu(t)\, \tilde{P}_t \tag{2.2}$$

A model for the trading schedule x_t, similar to the one proposed by Almgren and Lorenz (2006), can be assumed to be the sum of two distinct trajectories: (i) a linear execution (neutral/ TWAP) and (ii) a deviation – a quadratic function with the roots at $t = 0$ and $t = T$, which corresponds to a given "speed of execution" κ optimized by the proprietary trading strategy:

$$x(t) = X_0\, \frac{T-t}{T} - \kappa\, t\,(T-t), \tag{2.3}$$

where $T - t$ is the remaining time, and consequently:

$$\nu(t) = \frac{1}{T}\, X_0 + \kappa\,(T+1) - 2\,\kappa\, t. \tag{2.4}$$

The tilting factor κ is independent of the order size and reflects the timing cost (price drift and volatility risk), as well as the risk aversion / sense of urgency (see Equation 2.5). The trading schedule is front-weighted (aggressive), as in the left plot of Figure 3, when the agent's risk aversion or the expected risk are higher, the expected temporary market impact is lower and/or alpha is positive. In this case, the agent is willing to pay some premium in the form of higher temporary market impact to reduce the risk – at extreme, when $\kappa \to \infty$ there is a single execution of the entire quantity at the

[5] Over a longer period of time, two components of the total market impact can be distinguished. As described in Johnson (2010), (i) the permanent market impact leads to a new price equilibrium (an information-based effect equal to the difference between the pre-trade equilibrium and the post-trade equilibrium), while (ii) the temporary market impact represents the cost of immediacy and corresponds to the difference between the post-trade equilibrium and the trade print – the least favorable trade price obtained while executing the market order. Some time after the trade, the order book is assumed to be partially replenished.

start of the trading period. Conversely, when alpha is negative or when the agent is risk loving (negative λ), the trading schedule becomes back-weighted (passive), as in the right plot of Figure 3. Finally, when $\kappa = 0$, i.e. no significant drift is expected (zero or small alpha) and/or the expected temporary market impact is higher, the execution is neutral – constant trading rate.[6]

$$\kappa = \frac{\lambda\, \alpha\, \sigma}{\eta}, \tag{2.5}$$

where:
- $\hat{\alpha}$ price change expectation[7]
- $\hat{\sigma}$ expected stock volatility
- $\hat{\eta}$ expected temporary market impact cost[8]

Under certain assumptions, e.g., Glosten-Milgrom-Harris framework (Glosten and Milgrom (1985), Glosten and Harris (1988)) and a model for the temporary market impact, an analytically derivation of the optimal function κ^* is possible, but is beyond the goal of this paper. However, critical values for κ can be computed: if $\kappa > 0$ is too large, the trader will finish the execution earlier than T – execution time is shortened $T^* < T$ – while if κ is too small and negative, the trader will start its trading schedule later $t^* > 0$ – execution is delayed and time is also shortened. In the first case, the following holds at the end of the trading schedule $x(T^*) = 0$ with two solutions: $T^* = T$ and $T^* = \frac{X_0}{\kappa T}$. The critical value for $\kappa^*_+ = X_0/T^2 > 0$ is obtained by imposing the top of the parabola $x'(T^*) = 0$ to be at the end of the program $T^* = T$. In the second situation, the condition at the start of the program is $x(t^*) = X_0$, with two solutions $t^* = 0$ and $t^* = T + \frac{X_0}{\kappa T}$. The critical value is obtained by imposing $x'(t^*) = 0$ and $t^* = 0$ with the solution $\kappa^*_- = -X_0/T^2 < 0$. Various examples for different values of κ are presented in Figure 4.

3 High-frequency trading

High-frequency trading is not a strategy *per se*, but a technology which allows for the automation of a wide spectrum of trading strategies, propelled by the ongoing advances in computer technology. Schwartz (2010) underlines the

[6] Alternatively, the risk adjustment could also be applied to VWAP trading schedule, resulting into a tilted VWAP.

[7] A positive alpha indicates an unfavorable price change expectation, i.e. the expectation of prices moving lower while executing a sell order or moving higher for buy orders.

[8] The classic market impact function is considered to be a concave power function $h(\nu(t)) = \eta\, \nu(t)^\gamma$, where $\gamma < 1$ (**?**, Plerou, Gopikrishnan, Gabaix and Stanley (2002)). However, it is questionable if this functional form holds also for individual, unconditional orders (for a discussion, see Weber and Rosenow (2005)).

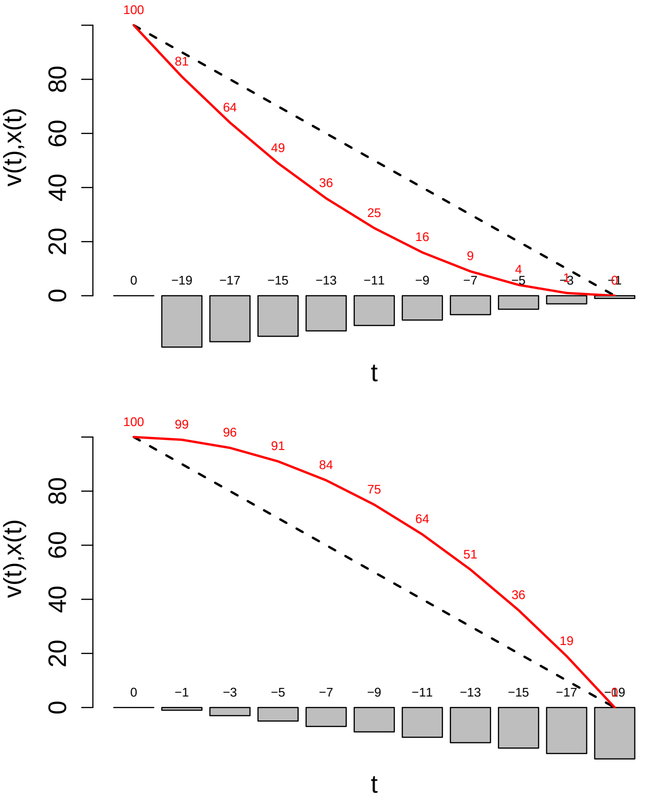

Figure 3: Front- and back-weighted trading schedules

The solid line plots show the trajectories for front-weighted (top) and back-weighted (bottom) trading schedules versus a neutral trading schedule plotted by means of a dashed line. The bar charts below each of the two plots represent the associated trading rate $\nu(t)$ at each time step t.

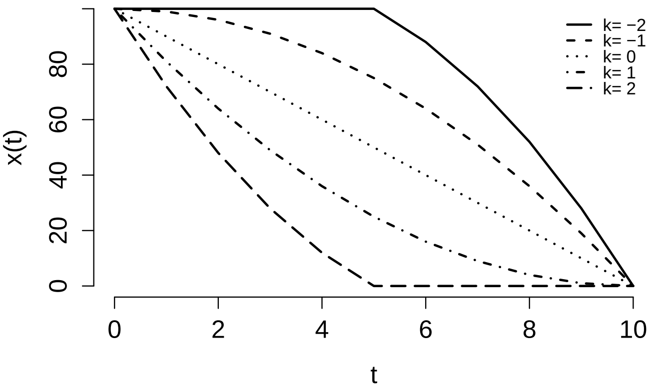

Figure 4: Five trading trajectories corresponding to different values of κ

The three middle trading schedules $\kappa \in \{-1, 0, 1\}$ correspond to the already presented back-weighted, neutral and front-weighted scenarios. When $\kappa = -2$ (upper, solid line), the trading schedule is not only back-weighted, but the execution is delayed and the trading period is shortened. Similarly, when $\kappa = 2$ (bottom, long-dashed line), the execution is front-weighted and the trader will finish the trading plan earlier.

same idea – most of the strategies assimilated with HFT are nothing new, they are only implemented using a modern and better technology which allows for faster execution and application at higher-frequencies (shorter time-frames).

Researchers (see, e.g., Aldridge (2009), Brogaard (2010)) as well as regulators across the world have tried to specify what are the attributes of HFT-based strategies.[9] Their findings could be summarised as follows: (1) computer-based non-discretionary / automated strategies (no human intervention in order initiation, generation, routing or execution), (2) proprietary trading (as opposed to agency activity), (3) use of low latency HFT technologies (e.g., co-location services, proximity hosting, direct market access, individual data feeds offered by exchanges), (4) real-time tick-by-tick data processing, (5) high amount of intraday messages (orders, quotes or cancellations) and trades, (6) small margins per trade ("scalping"), (7) high capital turnover, and (8) flat over-night positions (no positions or fully hedged).

A comparison to lower-frequency strategies (LFT) shows that HFT-based strategies address very short time-frames for generating entry and exit signals (associated with small expected returns), hold positions for very short periods of time (in terms of seconds or milliseconds),[10] trade very often and need a smaller risk capital, which they are able to relocate several times on a short-time basis, leading to their overall profitability. It has to be noted that profit opportunities at these short time-frames are only temporary and the available liquidity for entry/exit positions is limited, thus only the fastest competitors are able to trade. On the other side, in order to avoid losses, the liquidity-supplying strategies need to be able to quickly re-quote before others can take advantage of their outdated bids/asks (adverse selection risk). What these strategies have in common is that they are highly sensitive to their relative-to-competition reaction speed which continuously pushes them to minimising latencies of all kind.

According to EUREX, the main HFT-based strategies are liquidity provision, (statistical) arbitrage, short term momentum and liquidity detection.[11] HF liquidity provision traders are quasi market makers which try to cash in the bid-ask spread when the order flow is balanced. Pure and statistical arbi-

[9] BaFin, German HFT Bill – Hochfrequenzhandelsgesetz; European Comission, MiFid 2 – Directive 2014/65/EU of the European Parliament and of the Council of 15 May 2014 on markets in financial instruments and amending Directive 2002/92/EC and Directive 2011/61/EU; European Securities and Markets Authority (ESMA), Consultation paper "Guidelines on systems and controls in a highly automated trading environment for trading platforms, investment firms and competent authorities", Reference ESMA/2011/224; Securities and Exchange Commission (SEC), "Concept release on equity market structure", Release no. 34-61458, File no. S7-02-10, p. 45

[10] As a consequence, it is not unusual for HFT-based strategies "to switch between long and short net positions several times throughout the day", as opposed to their LFT-correspondents, as underlined in Brogaard (2010).

[11] "High-frequency trading – a discussion of relevant issues", London, 8 May 2013, `http://www.eurexchange.com/blob/exchange-en/455384/490346/6/data/presentation_hft_media_workshop_lon_en.pdf`

trageurs compare prices of the same assets across different markets or across economically similar (correlated) products and take advantage of any market inefficiencies. Short term momentum and liquidity detection strategies try to anticipate the short-term market direction due to new information or just due to participants' order flow.

3.1 Passive HFT - liquidity traders

Traditionally, on-exchange designated market makers (DMM) or specialists have played the role of providing liquidity by continuously double quoting bid and ask prices at which they are committed to buy and sell specific asset quantities. Trading is thus facilitated as other market participants are provided with immediacy, i.e. the ability to trade quickly and with reduced transaction costs. On the other side, when completing a round-trip, i.e. the buy and sell of the same quantity, without any price change, the market maker is compensated with the bid-ask spread differential. In case of an unfavourable price shift, the realised profit becomes smaller than the quoted spread or can even turn into a loss. Besides exchanges that employ a monopolistic market maker (e.g., NYSE), other exchanges (e.g., NASDAQ) rely on a different strategy for liquidity provision by allowing for multiple market-makers, where the overall performance comes from competition, rather than from strict individual obligations. In the case of exchanges which implement a maker-taker pricing model (e.g., NYSE Euronext's Arca Options platform, NASDAQ's NOM platform, BATS Global Markets options exchange), market-makers have an extra revenue source consisting in the rebates paid by the exchange for passively providing limit orders, which are matched by liquidity-taking market orders – as opposed to standard equal pricing.

According to Harris (2002), vanilla market making strategies – which do not also speculate, hedge, or invest – are uninformed with respect to the fair value of the underlying assets and do not form expectations about the future market direction. Therefore, in order to avoid any inventory risk, market makers try to keep their inventory under control by quoting prices which produce two-sided order flows. When their inventory deviates from the target levels, market makers have to adjust their quotes (bid/ask prices, spread width, bid/ask quantities) in order to stimulate order flow which will offset their imbalance. According to Aldridge (2009), another risk of market making, which is compensated by the bid-ask spread, is the time-horizon risk – the risk of an adverse market move decreases proportionally with the remaining trading time.[12]

Eventually, the specific levels of inventory targets, maximum allowed imbalances and holding time before restoring target inventories depend to a great extent on the degree of information about the market equilibrium price – the larger the confidence in the estimated future price dynamics, the lower

[12]In other words, the spreads at the beginning of the trading session are larger than near its end.

the perceived adverse selection risk and the higher the willingness to take large positions and hold them for a longer period of time in order to increase realised spreads and profitability, accordingly.

Passive HFTs apply similar market making strategies, which are also known as electronic liquidity provision (ELP), seeking to capture both the bid-ask spread and the rebates paid by the trading venues as incentives for posting liquidity. They are very poorly informed about fundamental values and therefore try to minimize the risk of liquidity provision by quickly offsetting even small inventory imbalances or by trying to identify toxic order flow, which ultimately leads to opening positions against the future market move (for more details on toxic order flow, see Easley, de Prado and O'Hara (2012)). Their round-trip gains and inventory sizes are smaller than the ones of better informed market makers, but their overall profitability comes from a larger number of round-trips due to quoting more often at the best bid and best ask, leading to a higher cumulative spread. HFT market makers compete to be the first counterparts in providing liquidity, as well as to profitably close their positions, which makes it vital being able to quickly update their outstanding orders in face of a changing market state.

The strategies employed by passive HFTs are sometimes referred to as quasi market making, because HFTs can suspend their activity whenever the market state would lead to their unprofitability, e.g., due to high volatility or trending, as they face no legal obligation to maintain quotes and guarantee market liquidity. Therefore, there is a general critique that the liquidity provided by HFT is often illusory, since HFTs stay in the market only when they are confident to make profits and retreat or even turn into liquidity-consumers under adverse conditions. In Government Office For Science (2012), this phenomena is named periodic illiquidity and is also explained by the opportunistic style and tight risk management of HFT market makers. Moreover, a particular type of ELP, known as rebate arbitrage, tries to profit only from collecting liquidity rebates paid by the trade venue by opening and closing positions at the same price on different exchanges (no spread capture), without really contributing to liquidity. Menkveld (2011) states that HFT "cream skimming" strategies, requiring little capital, have driven out traditional market makers with large capital and inventories, leading to a drastic change in the structure of liquidity provision sources, with large potential implications for the market dynamics in times of stress. On the other side, the Eurex Exchange finds evidence against this critique, by analyzing the behavior of HFTs on the 25th of August, 2011 – a high volatile day for the DAX Futures (FDAX).[13] The report concludes that HFTs actually contribute to liquidity and prevent fast price movements during periods of high volatility or strong directional trading.

Nevertheless, some ELP strategies are considered to be used as signal detectors of large institutional block orders, which launch active HFT strategies

[13] "High-frequency trading in volatile markets – an examination", Eurex Exchange, November 2011.

seeking to profit on this new information, and ultimately lead to amplifying investors' market impact (more details in Subsection 3.2).[14]

In a literature review, Aldridge (2013) identifies two main types of automated market making models: inventory models and information-based models.[15] The first class is concerned only with the effective management of inventory, without any opinion on the drift or any autocorrelation structure. A representative model is described in Avellaneda and Stoikov (2008). The optimal bid and ask quotes in the Avellaneda-Stoikov model are identified following a two-step procedure. First, assuming a standard Brownian motion with constant volatility and no drift, a reservation price (a personal indifference valuation for the stock) is computed as a deviation from the mid-price m_t, given the current inventory imbalance I_t (see Equation (3.1)).

$$r_t(I_t) = m_t - \gamma\, I_t\, \sigma^2\, (T - t), \tag{3.1}$$

where γ is the risk aversion coefficient, σ^2 is the mid-price variance, $(T - t)$ is the remaining time.

In the second step, the optimal spread size (s_t) around the reservation price r_t is determined (see Equation (3.2)). The execution probabilities of the quoted bid and ask are assumed to be given by a Poissonian process, where the arrival rates are functions of the relative distance to the mid-price. Moreover, the size of market orders is considered to follow a power law distribution ($f^Q(x) \sim x^{-1-\alpha}$) and the market impact function to have a logarithmic shape ($\Delta p \sim K\, ln(Q)$).

$$s_t = \gamma\, \sigma^2\, (T - t) + (2/\gamma)\, \ln(1 + \gamma/k) \tag{3.2}$$

The final solution of s_t depends on the mid-price variance σ^2, remaining time $(T - t)$, risk aversion γ and a parameter $k = \alpha\, K$, which captures the overall market environment as a combination between exponent α and coefficient K. The optimal quotes are set at $q_t^{a,b} = r_t(I_t) \pm s_t/2$. A numerical simulation shows that the Avellaneda-Stoikov strategy has a smaller average return, but also a smaller variance, compared to a naïve fixed offset and symmetric around the mid-price strategy.

Simplified versions of the Avellaneda-Stoikov model have been proposed in the literature. In Aldridge (2009), the arrival rates are not functions of the relative limit distance any more, but just per minute basis counters of the best bid changes λ^b, respectively best ask changes λ^a. Every new minute, the optimal bid and ask prices are computed as in Equation (3.4), relative

[14] "The hidden alpha in equity trading", Oliver Wyman Consulting, 2013.

[15] As explained later, information-based models do not rely on external fundamental information, but try to identify the true value from the market itself.

to the buy and sell reservation prices defined in Equation (3.3) or directly to the outstanding bid and ask quotes.

$$r_t^{b,a}(I_t) = m_t - \gamma\,(I_t \pm 1/2)\,\sigma^2\,(T-t) \tag{3.3}$$

$$p_t^b = r_t^b - \frac{1}{\gamma}\,\ln\left(1 - \gamma\,\frac{\lambda_t^b}{\lambda_t^b - \lambda_{t-1}^b}\right);\, p_t^a = r_t^a + \frac{1}{\gamma}\,\ln\left(1 - \gamma\,\frac{\lambda_t^a}{\lambda_t^a - \lambda_{t-1}^a}\right) \tag{3.4}$$

In case the optimal bid price p^b crosses the current ask quote, a unitary market buy order is initiated or, *vice versa*, a market sell order. Alternatively, a market buy order is triggered if the optimum bid price is closer to the current bid than the difference between the optimum ask price and the current ask or, *vice versa*, a market sell order. At the end of the day, the entire net position is offset at market prices.

The second class of market making strategies tries to extract the information which other market participants may possess, by analysing the order flow (buying and selling pressure) and/or the shape of the order book. The theoretical model proposed in Glosten and Milgrom (1985) relies on Bayesian learning to combine new information into the market maker's prior beliefs about the true market value of the traded asset (V) – the general form of the Bayes rule is:

$$Pr(h|D) = \frac{Pr(h)\,Pr(D|h)}{Pr(D)}, \tag{3.5}$$

where $Pr(h|D)$ is the conditional posterior belief, $Pr(h)$ the prior belief, $Pr(D|h)$ the degree to which the hypothesis predicts the observed data (likelihood), $Pr(D)$ the prior probability of the data.

In a market making setting, the observed data (D) can take the form either of a buy or of a sell market order. The market maker assumes two mutually exclusive hypothesis: the true market value deviates from ($h : V \lesseqgtr m$) or is equal to the current mid-price ($\bar{h} : V = m$).[16] The marginal likelihood of the data $Pr(D)$ can be independently computed using the theorem of total probability: $Pr(D|h)\,Pr(h) + Pr(D|\bar{h})\,Pr(\bar{h})$. Let the probability of informed trading $Pr(\texttt{informed}) = \alpha$.[17] If the market maker cannot distinguish between informed and uninformed trading ($\alpha = 0.5$), then the probability of receiving a market order given that the true market value is

[16] Equivalently, the two hypothesis can be considered the true market value is outside ($h : V < P_b, V > P_a$) or in-between the current market maker spread ($\bar{h} : P_b \leq V \leq P_a$).

[17] An estimate for informed trading could be given by the volume imbalance and trade intensity indicator (VPIN toxicity metric) proposed in Easley et al. (2012).

lower/higher is given by $Pr(D|h) = 0.75$.[18] Complementary, the probability of a market order under the assumption of an "equal" market value is $Pr(D|\bar{h}) = 1 - Pr(D|h) = 0.25$. If the market maker is uninformed, then the unconditional probabilities of the two hypothesis are $Pr(h) = Pr(\bar{h}) = 0.5$. Finally, the optimal bid and ask quotes, which include the adverse trading risk and assume zero profit (perfect competition), are computed as expectations of the true market value given the two possible market events:

$$P_b = E[V|\texttt{sell}]\,; P_a = E[V|\texttt{buy}] \tag{3.6}$$

The Das model in Das (2005, 2008) extends the theoretical model of Glosten and Milgrom by introducing a method to explicitly compute the solutions to the quote-setting Equation (3.6), based on an online probabilistic estimate of the true underlying value of the asset. This density estimate takes the form of a discrete distribution $Pr(V = x)$, computationally represented as a vector with a range from $\mu_V - 4\,\sigma_V$ to $\mu_V + 4\,\sigma_V$, where μ_V and σ_V^2 are the mean and variance of the true value. The vector is initialized with a normal pdf $N(\mu_V, \sigma_V)$ and is continuously updated by means of a nonparametric density estimation technique. There are three possible market events which lead to updating the density estimate: receiving a buy, a sell or no market order. E.g., in the case of a buy order, standard Bayesian updating, as described in Equation (3.7), is applied over the entire vector. The buy prior $Pr(\texttt{buy})$ in the denominator is the same for all x and can be ignored during the update, before renormalizing.

$$Pr(V = x|\texttt{buy}) = \frac{Pr(V = x)\,Pr(\texttt{buy}|V = x)}{Pr(\texttt{buy})}, \tag{3.7}$$

where $Pr(V = x)$ is the prior density estimate.

According to Equation (3.6), the market maker's optimal quotes are equal to the conditional expected value given that a particular type of order is received. E.g., an approximation of the ask price $P_a = E[V|\texttt{buy}]$ can be computed as in Equation (3.8).

$$P_a = \sum_{x_i = V_{min}}^{V_{max}} x_i\,Pr(V = x_i|\texttt{buy}) = \frac{\sum_{x_i = V_{min}}^{V_{max}} x_i\,Pr(V = x_i)\,Pr(\texttt{buy}|V = x_i)}{Pr(\texttt{buy})} \tag{3.8}$$

[18] The market order can be generated either by an informed or uninformed trader. If the trader were informed, the probability of placing exactly an order of the respective sign is 100%, while if uninformed, the probability of placing a buy or sell order is equal to 50%. Therefore, the conditional probability $Pr(D|h) = \alpha + 0.5\,(1 - \alpha)$.

Such quotes correspond to a zero-profit situation and represent the maximum bid price and the minimum ask price the market maker is willing to quote in order to cover the adverse selection risk.

Das (2005) also proposes a basic profit-making strategy, achieved by widening the spread around the break-even bid-ask spread with a fixed amount. The market maker faces a trade-off between large spreads (higher profits) with few round-trips and small spreads with many trades. The optimal solution of this trade-off depends on the market informativeness degree, as well as on the level of competition. When the competition between market makers is high, Das (2008) suggests placing bids and asks just inside the current spread as long as they are associated with a non-negative expected profit, i.e. the Glosten and Milgrom condition. It is to be noted that the only signals about the true market value come from direct trading, so that quote placement plays an important role in sampling the distribution on trades induced by the true value.

The prior of a buy order is:

$$Pr(\texttt{buy}) = \sum_{x_i=V_{min}}^{V_{max}} Pr(V = x_i)\, Pr(\texttt{buy}|V = x_i) \tag{3.9}$$

The computation of the above conditional probabilities depend on assumptions about the trader population. Das (2008) describes a market maker trading against informed agents, who only know the true value under the form of a noisy signal $W = V + N(0, \sigma_W)$.[19] In other words, the noisy informed trader sends a buy market order if $W > P_a$, a sell order if $W < P_b$ and no order if $P_b \leq W \leq P_a$.[20] Under these conditions, the conditional probability of a buy order is:

$$Pr(\texttt{buy}|V = x) = Pr(W > P_a) = Pr(N(0, \sigma_W) > P_a - x) \tag{3.10}$$

The optimum ask and bid quotes are the solutions of the fixed point equations given by substituting Equation (3.10) into Equation (3.8). Das (2005) also finds out that, following a fundamental shock, the market-maker's spread increases, reflecting its uncertainty about the underlying true value. While the probability mass of the density estimate becomes more concentrated, spreads also get lower. The drawback is that the pdf values in the outer regions are also getting very small and hinder the estimate update when future fundamental shocks occur. A solution proposed by Das (2005) is to recenter the density estimate around the current expected value and reinitialize it with a normal distribution after a fundamental jump has happened – such

[19] The σ_W parameter can control for the level of order flow toxicity (informativeness).

[20] The case of a mixed populations of uninformed and (perfectly or noisy) informed traders is analysed in Das (2005) and described in Appendix A, B.

a moment could be identified by an order imbalance-based classifier which simply counts the difference between the number of buy and sell orders (or bought and sold volumes) over a certain amount of time. Another extension, which takes into account also the portfolio risk, is presented in Das (2005). The proposed technique consists in shifting the pre-determined bid and ask quotes with a linear function of the current inventory I: $f(I) = -\gamma I$, where γ is a risk-aversion coefficient.[21]

Another market making model based on the Bayesian framework is introduced in Lin (2006). This time, the true value belief is updated by means of a discrete Kalman filter, where the measurement is given by the net order flow observed over some period of time. A different framework used to derive market making algorithms is reinforcement learning. Chan and Shelton (2001) train a market-maker strategy which observes three (discrete) state variables: own inventory, order flow imbalance, bid-ask spread. The market maker's set of actions include changing of the bid and/or ask prices, while the quoted sizes remain fixed. Finally, the reward function takes into consideration multiple objectives by means of a weighted-linear combination between profit and bid-ask spread (as a measure of market quality).

3.2 Active HFT - order anticipators

As opposed to their passive counterpart, active HFTs act as liquidity takers, by trading with aggressive, liquidity-consuming market-orders. There are two main lines of active HFT development: the first one seeks to predict market momentum and incoming order-flow and makes profits from short-term market shifts, while the second one tries to exploit the technological structure of the trading network system and superior speed advantage in order to detect and "front run" distributed executions of large orders.

Many order anticipator strategies have existed before the advent of HFT and have only been adapted and applied at higher-frequencies, where human traders are not able to react. Traditionally, according to Harris (2002), order anticipators try to profit from information about third parties' trading intentions, rather than from own fundamental information regarding the traded asset. This type of profit-motivated speculators, also described as "parasitic traders" or "predatory", can be further subdivided between front runners, sentiment-oriented technical traders and squeezers. While front-running and squeezing are considered to be illegal strategies, due to exploiting confidential information or being guilty of market manipulation, the sentiment-oriented technical traders (technical analysts or chartists) use public available information in order to predict future price returns. Depending on the underlying time-frame, various sources of information can be used by these technical strategies. For example, end-of-day strategies process aggregated information related to the daily price (open, high, low, close, average, median) or

[21] Alternatively, a sigmoid function allows for a gradual increase, as well as for limiting the inventory control impact.

to the daily turnover in order to generate profitable trading signals. On the other side, intraday trading relies on additional information available at the market microstructure level, such as order book liquidity, order flow, trade and tick history. The main idea is to identify persistent relationships between various market indicators and future short-term market moves or incoming customer order flow, upon which profitable trading strategies can be built.

In a high-frequency setting, the analysed time-series are usually in a raw format (tick-by-tick), i.e. without any aggregation, and the information processing as well as trading are conducted in a non-discretionary and automated fashion by means of computer programs (no human intervention). The actual trading rules can either be specified by a human specialist or can be "learned" from past data through various quantitative and computational methods (e.g., machine learning, artificial intelligence, evolutionary techniques). A related general concern, conformed by Brogaard (2010), is that HFT-based strategies are not very diversified or, equivalently, their trading signals are highly correlated, leading to an exacerbation of market movements.

One simple HFT strategy proposed by Aldridge (2009) is based on order flow short-term autocorrelation, which consists in opening position in the direction of the order flow imbalance, i.e. the difference between the cumulative number/volume of buy and sell orders, or buyer- and seller-initiated trades. Another strategy relies on mimicking aggressive trading, as a proxy for informed traders' expectations. An indicator measuring the orders' aggressiveness can be computed as the percentage of market as opposed to limit orders. Besides the share of market orders, where limit orders are placed, i.e. the shape of the order book, can also reveal the future expectations of market participants. For example, the order book imbalance, quantified as the difference between the cumulated volumes up to a certain depth level on each side of the order book, can influence the aggressiveness of incoming orders in two ways.[22] When their side of the book is thicker and crowded, traders price their orders more aggressively in order to increase their order execution probability (competition effect). Conversely, traders become less aggressive when the opposite side is deeper, forecasting a favorable short-term order flow (strategic effect). Order-flow related are also the inter-trade and price/volume durations, which can indicate changes in liquidity and volatility due to new information.[23] Short-term momentum/reversal can also be detected by means of econometric models taking into consideration phenomenological observations such as autocorrelation of price returns (serial dependence for small intraday scales), fat tails and volatility clustering.

A second class of active HFTs tries to identify patterns of execution algorithms (processing large institutional orders) and trade ahead the remain-

[22] Alternatively, the book imbalance and the short-term price return can be assessed by comparing the bid-ask midpoint with the weighted price at a given depth of the order book as in Cao, Hansch and Wang (2009).

[23] The time interval between subsequent volume changes of a specified magnitude is known as volume duration.

ing execution program ("electronic front-running"). For example, when the liquidity is fragmented over a number of trading venues, the execution algorithms route their orders across different locations, and in case their trading activity is exposed, ultra-HFTs can exploit their ultra-low-latency competitive advantage and profit from the "working" of the large order (latency-based arbitrage). Similarly, algorithmic traders can be taken advantage of if their order generating pattern is predictable from the observable order flow (order-flow detection). In the same framework of multiple exchanges, ultra-HFTs can profit from temporary market inefficiencies by picking off outdated orders belonging to market participants with a higher reaction latency (slow-market arbitrage). As a negative side effect, the institutional investor suffers from an increased market price impact and, consequently, a reduced "alpha". Retail investors are usually not affected because their orders are too small. First of all, it is questionable whether the term "front-running" is justified or not in this case, since the ultra-HFTs do not have direct access to the actual order to be executed, and this is only induced in a noisy manner from public trading data. On the execution side, smart routing strategies which are able to minimize the speed advantage of ultra-HFTs have been developed. E.g., Royal Bank of Canada's THOR® routing technology precomputes the routing latencies and sends the different slices of the total order in such a way so they reach the targeted trading venues at the same time and therefore eliminating any information leakage.

4 Conclusions

The current contribution describes a range of computer-based trading strategies which are widely applied by market participants, in the context of nowadays increased trading automation and market fragmentation. The strategies were classified both based on the profile of market participants and on the technology type. Informed traders use algorithmic trading strategies in order to reduce their transaction costs and smart order routing technologies in order to minimize information leakage. Based on their risk profile and on the performance benchmark for the execution, different algorithmic trading strategies are preferred. The classic optimisation problem of trading has been defined and both basic implementations, as well as extended versions of the two main algorithmic trading strategies, i.e. Volume Weighted Average Price and Implementation Shortfall, have been introduced. In the case of uninformed liquidity traders or market-makers, high-frequency trading technologies have been introduced in order to maintain the low-latency advantage as trading edge. Several models of market making strategies, which can be classified in two types based on their core objective, i.e. inventory- and information-based models, have been introduced. For the latter type, two different machine learning frameworks, i.e. Bayesian learning and Reinforcement learning, have been discussed. Finally, order anticipators make use of the increased computational power in order to process in real time a larger amount of trading information – including the rich microstrucutre data, e.g., order flow and order book data – and to develop new predictive models which can be applied at high-frequency time-series. Others simply rely on their ultra-low-latency advantage and speculate the fragmented structure of the trading network in order to detect and "front run" the execution of large orders.

References

Aldridge, I. (2009). *High-frequency trading: a practical guide to algorithmic strategies and trading systems*, Vol. 459, John Wiley & Sons, Hoboken, New Jersey.

Aldridge, I. (2013). *High-frequency trading: a practical guide to algorithmic strategies and trading systems, Second Edition*, John Wiley & Sons, Hoboken, New Jersey.

Almgren, R. and Chriss, N. (2001). Optimal execution of portfolio transactions, *Journal of Risk* **3**: 5–40.

Almgren, R. and Lorenz, J. (2006). Bayesian adaptive trading with a daily cycle, *The Journal of Trading* **1**(4): 38–46.

Avellaneda, M. and Stoikov, S. (2008). High-frequency trading in a limit order book, *Quantitative Finance* **8**(3): 217–224.

Białkowski, J., Darolles, S. and Le Fol, G. (2008). Improving vwap strategies: a dynamic volume approach, *Journal of Banking & Finance* **32**(9): 1709–1722.

Brogaard, J. (2010). High frequency trading and its impact on market quality, *Northwestern University Kellogg School of Management Working Paper* **66**.

Cao, C., Hansch, O. and Wang, X. (2009). The information content of an open limit-order book, *Journal of futures markets* **29**(1): 16–41.

Chan, N. T. and Shelton, C. (2001). An electronic market-maker, *Technical Report 195*, Massachusetts Institute of Technology – Artificial Intelligence Laboratory.

Das, S. (2005). A learning market-maker in the glosten–milgrom model, *Quantitative Finance* **5**(2): 169–180.

Das, S. (2008). The effects of market-making on price dynamics, *Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems-Volume 2*, International Foundation for Autonomous Agents and Multiagent Systems, pp. 887–894.

Easley, D., de Prado, M. and O'Hara, M. (2012). Flow toxicity and liquidity in a high-frequency world, *Review of Financial Studies* **25**(5): 1457–1493.

Fabozzi, F. J., Focardi, S. M., Kolm, P. N. et al. (2010). *Quantitative equity investing: Techniques and strategies*, John Wiley & Sons.

Glosten, L. R. and Harris, L. E. (1988). Estimating the components of the bid/ask spread, *Journal of financial Economics* **21**(1): 123–142.

Glosten, L. R. and Milgrom, P. R. (1985). Bid, ask and transaction prices in a specialist market with heterogeneously informed traders, *Journal of financial economics* **14**(1): 71–100.

Government Office For Science (2012). Foresight: The future of computer trading in financial markets, *Technical report*.

Harris, L. (2002). *Trading and exchanges: Market microstructure for practitioners*, Oxford University Press, USA.

Johnson, B. (2010). *Algorithmic Trading & DMA: An introduction to direct access trading strategies*, 4Myeloma Press, London.

Kissell, R. and Malamut, R. (2005). Understanding the profit and loss distribution of trading algorithms, *Trading* **2005**(1): 41–49.

Lillo, F., Farmer, J. and Mantegna, R. (2002). Single curve collapse of the price impact function for the new york stock exchange, *arXiv preprint cond-mat/0207428* .

Lillo, F., Farmer, J. and Mantegna, R. (2003). Master curve for price-impact function, *Nature* **421**(129): 176–190.

Lin, H. (2006). *Essays in market microstructure*, PhD thesis, University of Warwick.

Madhavan, A. N. (2002). Vwap strategies, *Trading* **2002**(1): 32–39.

Menkveld, A. (2011). Electronic trading and market structure, *UK Government Office for Science, Foresight Driver Review - The Future of Computer Trading in Financial Markets* .

Plerou, V., Gopikrishnan, P., Gabaix, X. and Stanley, H. E. (2002). Quantifying stock-price response to demand fluctuations, *Physical Review E* **66**(2): 027104.

Schwartz, R. (2010). Dark pools, fragmented markets, and the quality of price discovery, *The Journal of Trading* **5**(2): 17–22.

Vuorenmaa, T. (2012). The good, the bad, and the ugly of automated high-frequency trading, *Technical report*, Valo Research and Trading.

Weber, P. and Rosenow, B. (2005). Order book approach to price impact, *Quantitative Finance* **5**(4): 357–364.

A Das model with uninformed and perfectly informed traders

Assume the proportion of informed traders to be α. Uninformed traders send buy and sell orders with equal probability 0.5. The informed trader sends a sell market order with certainty if $V < P_b$, a buy market order if $V > P_a$ and no order otherwise ($P_b \leq V \leq P_a$).

The conditional probabilities are:

$$
\begin{aligned}
&Pr(\texttt{buy}|V \leq P_a) = 0.5\,(1-\alpha)\\
&Pr(\texttt{buy}|V > P_a) = 0.5\,(1-\alpha) + \alpha = 0.5\,(1+\alpha)\\
&Pr(\texttt{sell}|V < P_b) = 0.5\,(1-\alpha) + \alpha = 0.5\,(1+\alpha)\\
&Pr(\texttt{sell}|V \geq P_b) = 0.5\,(1-\alpha)
\end{aligned}
$$

The priors of buy and sell orders are given by:

$$
\begin{aligned}
Pr(\texttt{buy}) &= \sum_{x_i=V_{min}}^{V_{max}} Pr(\texttt{buy}|V = x_i)\,Pr(V = x_i)\\
&= \sum_{x_i=V_{min}}^{P_a} Pr(\texttt{buy}|V = x_i, x_i \leq P_a)\,Pr(V = x_i)\\
&+ \sum_{x_i=P_a}^{V_{max}} Pr(\texttt{buy}|V = x_i, x_i > P_a)\,Pr(V = x_i)\\
&= 0.5\,(1-\alpha) \sum_{x_i=V_{min}}^{P_a} Pr(V = x_i) + 0.5\,(1+\alpha) \sum_{x_i=P_a+1}^{V_{max}} Pr(V = x_i)
\end{aligned}
$$

$$
\begin{aligned}
Pr(\texttt{sell}) &= \sum_{x_i=V_{min}}^{V_{max}} Pr(\texttt{sell}|V = x_i)\,Pr(V = x_i)\\
&= \sum_{x_i=V_{min}}^{P_b-1} Pr(\texttt{sell}|V = x_i, x_i < P_b)\,Pr(V = x_i)\\
&+ \sum_{x_i=P_b}^{V_{max}} Pr(\texttt{sell}|V = x_i, x_i \geq P_b)\,Pr(V = x_i)\\
&= 0.5\,(1+\alpha) \sum_{x_i=V_{min}}^{P_b-1} Pr(V = x_i) + 0.5\,(1-\alpha) \sum_{x_i=P_b}^{V_{max}} Pr(V = x_i)
\end{aligned}
$$

The ask and bid quote-setting equations (fixed point equations):

$$
\begin{aligned}
P_a = E[V|\texttt{buy}] &= \sum_{x_i=V_{min}}^{V_{max}} x_i\, Pr(V = x_i|\texttt{buy}) \\
&= \sum_{x_i=V_{min}}^{V_{max}} \frac{x_i\, Pr(\texttt{buy}|V = x_i)\, Pr(V = x_i)}{Pr(\texttt{buy})} \\
&= \frac{1}{Pr(\texttt{buy})} \sum_{x_i=V_{min}}^{P_a} x_i\, Pr(\texttt{buy}|V = x_i, x_i \leq P_a)\, Pr(V = x_i) \\
&+ \frac{1}{Pr(\texttt{buy})} \sum_{x_i=P_a+1}^{V_{max}} x_i\, Pr(\texttt{buy}|V = x_i, x_i > P_a)\, Pr(V = x_i) \\
&= \frac{0.5\,(1-\alpha)}{Pr(\texttt{buy})} \sum_{x_i=V_{min}}^{P_a} x_i\, Pr(V = x_i) + \frac{0.5\,(1+\alpha)}{Pr(\texttt{buy})} \sum_{x_i=P_a+1}^{V_{max}} x_i\, Pr(V = x_i)
\end{aligned}
$$

$$
\begin{aligned}
P_b = E[V|\texttt{sell}] &= \sum_{x_i=V_{min}}^{V_{max}} x_i\, Pr(V = x_i|\texttt{sell}) \\
&= \sum_{x_i=V_{min}}^{V_{max}} \frac{x_i\, Pr(\texttt{sell}|V = x_i)\, Pr(V = x_i)}{Pr(\texttt{sell})} \\
&= \frac{1}{Pr(\texttt{sell})} \sum_{x_i=V_{min}}^{P_b-1} x_i\, Pr(\texttt{sell}|V = x_i, x_i < P_b)\, Pr(V = x_i) \\
&+ \frac{1}{Pr(\texttt{sell})} \sum_{x_i=P_b}^{V_{max}} x_i\, Pr(\texttt{sell}|V = x_i, x_i \geq P_b)\, Pr(V = x_i) \\
&= \frac{0.5\,(1+\alpha)}{Pr(\texttt{sell})} \sum_{x_i=V_{min}}^{P_b-1} x_i\, Pr(V = x_i) + \frac{0.5\,(1-\alpha)}{Pr(\texttt{sell})} \sum_{x_i=P_b}^{V_{max}} x_i\, Pr(V = x_i)
\end{aligned}
$$

Update equations:

1. Receive a market buy order: multiply all probabilities for $V = x_i, x_i > P_a$ by $0.5\,(1+\alpha)$ and all the other by $1 - 0.5\,(1+\alpha)$, before renormalizing.

2. Receive a market sell order: multiply all probabilities for $V = x_i, x_i < P_b$ by $0.5\,(1+\alpha)$ and all the other by $1 - 0.5\,(1+\alpha)$, before renormalizing.

B Das model with uninformed and noisy informed traders

The conditional probabilities are:

$$\begin{aligned} Pr(\texttt{buy}|V = x_i, x_i \le P_a) &= (1-\alpha)\,0.5 + \alpha\,Pr(x_i + N(0,\sigma_W) > P_a) \\ &= (1-\alpha)\,0.5 + \alpha\,Pr(N(0,\sigma_W) > P_a - x_i) \end{aligned}$$

$$\begin{aligned} Pr(\texttt{buy}|V = x_i, x_i > P_a) &= (1-\alpha)\,0.5 + \alpha\,Pr(x_i + N(0,\sigma_W) > P_a) \\ &= (1-\alpha)\,0.5 + \alpha\,Pr(N(0,\sigma_W) < x_i - P_a) \end{aligned}$$

$$\begin{aligned} Pr(\texttt{sell}|V = x_i, x_i < P_b) &= (1-\alpha)\,0.5 + \alpha\,Pr(x_i + N(0,\sigma_W) < P_b) \\ &= (1-\alpha)\,0.5 + \alpha\,Pr(N(0,\sigma_W) < P_b - x_i) \end{aligned}$$

$$\begin{aligned} Pr(\texttt{sell}|V = x_i, x_i \ge P_b) &= (1-\alpha)\,0.5 + \alpha\,Pr(x_i + N(0,\sigma_W) < P_b) \\ &= (1-\alpha)\,0.5 + \alpha\,Pr(N(0,\sigma_W) > x_i - P_b) \end{aligned}$$

$$\begin{aligned} Pr(\texttt{no order}|V = x_i, x_i < P_b) &= \alpha\,Pr(x_i + N(0,\sigma_W) > P_b) \\ &= \alpha\,Pr(N(0,\sigma_W) > P_b - x_i) \end{aligned}$$

$$\begin{aligned} Pr(\texttt{no order}|V = x_i, P_b \le x_i \le P_a) &= \alpha\,(Pr(x_i + N(0,\sigma_W) > P_b) + \\ &\quad Pr(x_i + N(0,\sigma_W) < P_a)) = \alpha\,(Pr(N(0,\sigma_W) > P_b - x_i) + \\ &\quad Pr(N(0,\sigma_W) < P_a - x_i)) \end{aligned}$$

$$\begin{aligned} Pr(\texttt{no order}|V = x_i, x_i > P_a) &= \alpha\,Pr(x_i + N(0,\sigma_W) < P_a) \\ &= \alpha\,Pr(N(0,\sigma_W) > x_i - P_a) \end{aligned}$$

The priors of buy and sell orders, the optimal ask and bid quotes, as well as the updating equations can be derived by substituting the above conditional probabilities in the general forms 3.9, 3.8 and 3.7.