Markets Committee

FX execution algorithms and market functioning

Report submitted by a Study Group established by the Markets Committee

The Group was chaired by
Andréa M Maechler (Swiss National Bank)

October 2020

JEL classification: F31, G14, G15

Keywords: FX market, price discovery, execution trilemma

BANK FOR INTERNATIONAL SETTLEMENTS

This publication is available on the BIS website (www.bis.org).

ISBN 978-92-9259-428-2 (online)

Preface

The FX market has undergone significant structural change in recent years. The proliferation of multiple trading venues has led to increased fragmentation, and trading has become more electronic and automated. This has fuelled an increase in the use of FX execution algorithms (EAs), including by bank and non-bank financial institutions, and certain non-financial corporates.

To understand the drivers and implications of the rising use of EAs in FX markets, the Markets Committee established in mid-2019 a Study Group chaired by Andréa M Maechler (Swiss National Bank). This report presents the Group's findings. In addition to data analysis and research, it draws on a unique survey among providers and users of EAs, as well as extensive industry-wide outreach.

The key takeaway of the report is that EAs support price discovery and market functioning in an increasingly fragmented market. However, they also contribute to the ongoing changes in market structure, and with increasing scale of use, give rise to new risks and challenges that warrant close monitoring. The report provides unique insights into central banks' use of EAs, and preliminary observations on the performance and use of EAs during a period of high volatility due to the COVID-19 crisis-induced market disruption in March 2020. While the primary focus is on the FX market, many of the report's findings also have broad relevance for other fast-paced electronic markets where similar trends in EA usage are observed.

Jacqueline Loh

Chair, Markets Committee
Deputy Managing Director, Monetary Authority of Singapore

Contents

Executive summary

The foreign exchange (FX) market has been undergoing rapid technological changes in recent years. These changes have led, among others, to the adoption of new tools, which have the potential to alter market dynamics. To gain greater insights about how technological innovation may affect market functioning in fast-paced markets, this report examines the role of execution algorithms (EAs) in the FX market.

EAs are automated trading programs designed to buy or sell a predefined amount of securities or FX according to a set of parameters and user instructions. In contrast to other common types of algorithms such as market-making or opportunistic algorithms, the sole purpose of EAs is to execute a trade as optimally as possible. They have become a well established means of FX execution over the last few years, reflecting in large part the rising electronification of the FX market, an exponential increase in computing power, and structural drivers such as the fragmentation of the fast-paced electronic FX market (Markets Committee (2018)).

To complement available data and research, the report draws on the findings of a unique survey among providers and users of EAs, as well as an extensive industry-wide outreach. While the report focuses on the FX market, its findings may be of broader relevance to all market participants and other actors, including central banks, actively monitoring or engaged in fast-paced electronic markets.

The report finds that the use of EAs is widespread but not dominant in the FX market. FX EAs came into use more than 10 years ago, and today account for an estimated 10–20% of global FX spot trading, or approximately USD 200–400 billion in turnover daily. During this time, FX EAs have also evolved from simple mechanical forms (eg programs that simply slice a large order into evenly sized smaller orders placed at a regular interval) to more sophisticated and adaptive types that respond to real-time changes in market conditions.

EAs have grown in usage as a response to the rapidly evolving and fragmented landscape of the FX market. EAs allow users to aggregate order book data across fragmented liquidity pools, to slice orders into smaller pieces and to distribute these pieces efficiently across liquidity pools. This helps users optimise their trade execution. The automated nature of EAs also helps increase operational efficiencies.

On a structural level, while EAs help improve market functioning, by changing the way market participants access the FX market and how orders are executed, they also introduce new risks. The report examines the impact of EAs on market functioning from three different perspectives:

Market microstructure perspective

- **EAs improve the price discovery and matching process in a highly fragmented market.** The ability of EAs to process fresh information quickly and to direct orders simultaneously across multiple trading venues helps market participants overcome hurdles associated with fragmented markets, such as information asymmetries and low visibility of market activity.
- **Market monitoring tools and activities need to account for the structural changes in the underlying microstructure of the FX market.** As orders are sliced into smaller pieces, market functioning depends less on the absolute depth

of the order book – as reflected in traditional measures of order book size – and more on how rapidly liquidity is replenished. If this trend gains in significance, novel liquidity indicators will need to adequately reflect these evolving dynamics.

- **Similarly, EAs facilitate the growing trend towards internalisation, which reduces the visibility of trades, and could eventually harm price discovery.** So far, market participants have been able to benefit from the ability of their liquidity providers to match trades internally. At the extreme, however, too many internalised trades could reduce the traded volume at primary venues to a point where it could jeopardise a sound price discovery process. Hence, as EAs continue to grow in relevance, further work will be needed to assess whether – and, if so, at which point – the growing share of "dark" trades associated with the use of EAs may start to negatively affect the price discovery process.

Market participants' perspective

- **Market fragmentation in the FX market poses challenges to trade execution: EAs help market participants optimise their execution but require adequate knowledge and information to ensure they are used effectively.** EAs can endow market participants with potentially better and more direct control over trade execution. However, they also imply greater challenges, as users of EAs carry market risk until completion of their trades. In particular, users of EAs must be aware that every execution strategy entails a trade-off – termed the "execution trilemma" in this report – between minimising the market impact, minimising the exposure to market risk and maximising the certainty of completing a trade.
- **Closing information gaps – in terms of both expertise and data – is central to ensuring greater transparency and a level playing field.** Market participants need to be able to assess the strength of their execution – pre-trade, in real time and post-trade. This, however, requires access to adequate data and information, which is typically costly and difficult to obtain. Issues that warrant further consideration include the provision of open access to a minimum set of data (akin to the "central tape" available in the equity markets), more uniform disclosures across the market, and a higher degree of standardisation with respect to the characterisation of EAs.

Market-wide perspective

- **While EAs contribute positively to market functioning in normal market conditions, the risk of self-reinforcing feedback loops triggering a sharp price move persists.** Initial observations after the outbreak of the Covid-19 pandemic suggest that the use of EAs was not impacted negatively by heightened volatility. On the contrary, the sharp increase in FX EA usage in March 2020, when FX market volatility reached multi-year highs, suggests that EAs remained a useful tool for users during this period. However, one cannot draw conclusions regarding EA performance in all market conditions from this single episode of heightened volatility alone.
- **Embedded controls, adequate education and consistent monitoring therefore remain key to helping reduce the risk of a local market disruption affecting market functioning in the broader FX market.** In particular, with no market-wide circuit breakers or kill switches in place in the FX spot market, the onus is on each provider and user to have adequate safeguards in place that

prevent the risk of unintentional trading behaviour from materialising. This area needs to stay in focus for providers, users, regulators and central banks.

These findings underline the fact that many questions remain open and warrant further analysis as the landscape of fast-paced electronic markets continues to evolve rapidly. The impact on market functioning will depend on many factors, including the direction in which EAs evolve, their market share, how well their risks are understood and managed, and how they interact with other developments in financial markets.

The ongoing three-year review of the FX Global Code will look into a number of the identified issues, particularly those pertaining to disclosure and algorithmic trading. As long as the FX market, and other markets such as fixed income, continue to evolve rapidly, fostering a better understanding of the role and ongoing evolution of EAs will remain of particular relevance to all actors actively engaged in or monitoring the FX market.

Central banks can also benefit from expanding their monitoring capabilities in this area, whether in the context of their market monitoring efforts, reserve management activities or monetary policy mandates. Indeed, as new tools, skills and access to data may often be required for comprehensive monitoring, central banks may want to consider creating a dedicated, fit-for-purpose platform to analyse relevant questions in the context of fast-paced markets. This may include, if required, pooling of resources to reduce costs for the central banking community as a whole, identifying relevant common issues of analysis, and fostering and disseminating knowledge, including through regular topical workshops.

Introduction

In this report, we examine the role of execution algorithms (EAs) in helping market participants navigate today's complex foreign exchange (FX) market structure, and their implications for market functioning.

The report draws on the findings of a unique survey among providers and users of execution algorithms as well as an extensive industry-wide outreach. Given the limited amount of readily available data and research on this topic, this report is of relevance to all market participants active in the FX market and to central banks in particular. A better understanding of FX EAs will help central banks optimise their market monitoring efforts, reserve management activities and the implementation of their monetary policy mandates. While the report focuses on FX, its findings may also be relevant for other over-the-counter (OTC) markets that evolve rapidly towards increased electronification. The findings may also be of use to the GFXC, particularly for its currently ongoing review of the FX Global Code.

The structure of the report is as follows: Section 1 takes stock of the evolution of FX EAs, their usage today, and relevant aspects from both users' and providers' perspectives. Section 2 provides a taxonomy for EAs to help understand the key trade-offs among different EAs and their decision logic. Section 3 provides new, important insights into the benefits and risks EAs pose for the functioning of the FX market as a whole – including initial findings covering the Covid-19-related market turmoil. Section 4 summarises the key takeaways. A comprehensive glossary explains technical terms used throughout the report in more detail.

1. Execution algorithms in the FX market: taking stock

1.1 Background to FX execution algorithms

For the purposes of this report, FX execution algorithms are defined as automated trading programs designed to buy or sell a predefined amount of FX according to a set of parameters and instructions, with the objective of filling the order. At their most basic level, EAs automate the process of splitting a larger order (eg USD 100 million), hereafter known as the "parent order", into multiple smaller orders (eg 100 transactions of USD 1 million each), known as "child orders", and executing them over a period of time separately rather than altogether. EAs seek to assist the user in entering into or closing a predefined position by either buying or selling a particular currency pair in one direction. In this way, they are distinct from other common types of algorithms used in the FX market which involve both buying and selling currencies. Examples of the latter include market-making algorithms, which typically seek to restore the liquidity provider's net aggregate position to a neutral or close-to-neutral value, and opportunistic algorithms, which are commonly used by principal trading firms and hedge funds to generate profit.

FX EAs allow users to navigate the fragmented FX market by aggregating liquidity and facilitating access to the various types of liquidity pools and trading venues, which would be difficult, if not impossible, manually. A significant implication is that they give market participants more direct control over how their transactions are executed. Prior to the emergence of FX EAs, this had been the domain of market-makers, hedge funds and other sophisticated financial institutions.

At the same time, in using FX EAs, market participants carry market risk until completion of the trade which they need to manage. In this way, FX EAs differ from other methods of trading such as "risk transfer" where market risk is swiftly shifted from end users (eg funds, corporates and small banks) to liquidity providers. When market participants execute a "risk transfer", they request a price from their liquidity provider (a request-for-quote (RFQ) or a request-for-stream (RFS)) and trade the full size of the ticket at the price received from their counterpart.[1] The direct cost of this immediate risk transfer, the bid-ask spread, is the compensation paid to the liquidity provider for taking on the market risk.[2] In this sense, the risk transfer price constitutes an almost instantaneous "all-in" price that typically depends on the transaction size, the prevailing liquidity conditions and market volatility. In contrast, when market participants execute via EAs, it is up to users to decide in what way and how fast to reduce market risk through the choice and parametrisation of algorithms – the result of which will be a trade-weighted average price known only at the end of execution plus an associated fee that EA providers typically charge for the usage of their EAs.

1.2 Adoption and evolution of FX execution algorithms

Drivers of FX execution algorithm adoption

EAs started to emerge in the FX market in the early-2000s, after having been first available in the equity market for a number of years. Based on information collected in the context of this report, usage of FX EAs has increased significantly in the past two decades, and is now estimated to account for 10–20% of daily spot FX volume in major currencies. According to the latest figures from the Bank for International Settlements (BIS) Triennial Central Bank Survey of Foreign Exchange and Over-the-counter Derivatives Markets, this equates to approximately $200–400 billion worth of FX spot traded via EAs each day globally.

Growing adoption of FX EAs has been driven in part by the rising electronification of the FX market. In spot FX, end users can now access liquidity via a range of electronic platforms. BIS FX Triennial Survey data (BIS (2010, 2019)) suggest that electronic execution of spot FX has increased from about 55% of total spot FX turnover globally in 2010 to about 70% in 2019. The commoditised nature of FX spot trading has lent itself particularly well to rapid and widespread adoption of EAs – in contrast to other markets such as corporate bonds or even FX options, which are much more diverse in terms of the products that are traded. Crucially, electronification has been supported by an exponential increase in computing power, lower data storage costs and the ability to transfer data almost instantly (Markets Committee (2018)). These technological advancements have enabled a dramatic

1 Traditionally, this RFQ was conducted via voice (ie over the phone) but is now often conducted electronically. Many participants also have access to live streaming prices (RFS prices) that can be transacted on immediately.

2 The bid-ask spread represents the basic transaction cost of trading. If a market participant wants to execute an order immediately, ie "aggressively", they need to "cross" the bid-ask spread: to buy, they have to pay the (higher) ask price; and to sell, they have to offer at the (lower) bid price. Whenever a market participant chooses to execute "passively", they place either a bid or an ask order in the order book and wait for another market participant to match them at that rate – thereby earning or "capturing" the spread, ie in this case they buy at the bid price and sell at the ask price.

increase in both the frequency of trading activity and the speed of information flow in the FX market and have driven growing adoption of algorithms in FX, as they have in other financial markets.

FX EA adoption has been both a response and a contributor to continued and growing fragmentation across a wide array of electronic trading venues (Moore et al (2016)). This fragmentation has been driven by increasing competition among trading platforms and different market players – especially by liquidity providers and new trading venues seeking to gain market share by offering innovative ways to access liquidity, match counterparties or provide trade data. Whereas in the past FX trading predominantly took place at just a few electronic trading venues – currently known as the primary trading venues – the market today comprises a large number of trading venues, each with its own unique set of trading rules and idiosyncrasies.[3] FX EAs provide users with a means to monitor, navigate and execute in the fragmented FX market.

Regulatory changes have also contributed to increased EA adoption – raising both the demand from end users for EA services and the supply of such services by liquidity providers. Many users use FX EAs to demonstrate compliance with so-called "best execution" requirements. This concept has gained significant traction among market participants following the implementation of the European MiFID II regulation in equities and other markets. While FX spot trading is typically not subject to the same regulatory requirements,[4] the practices have likewise become commonplace even in the FX market. This rise in demand on the end user side for more automated and traceable execution has coincided with a notable reduction in risk appetite and principal risk warehousing on the liquidity providers' side, as bank balance sheets have become more constrained and costly to deploy in the aftermath of the global financial crisis (Debelle (2018)). This has further facilitated the proliferation of EAs, as they rely less on liquidity providers' capacity to absorb risk.

Finally, it should be noted that significant increases in the adoption of FX EAs have taken place over the last decade, which has been a period of historically subdued volatility in the FX market. One view is that this low volatility environment may have been supportive of user adoption, as the cost to users of taking on market risk during such periods is generally lower. That said, the sharp increase in FX EA usage in March 2020, when FX market volatility reached multi-year highs due to the Covid-19 pandemic, suggests that EAs remain a useful tool for users even during periods of increased volatility (Box E).

Evolution of FX execution algorithms

FX EAs have evolved in sophistication since their initial development. For instance, order slicing techniques have evolved from very simplistic approaches (based on time or number of slices) to more sophisticated approaches, with the latest generation able to dynamically adapt to market conditions. Broadly, there are three generations of algorithms:

[3] According to Sinclair (2018), market participants can connect to more than 75 different FX trading venues. The different types of venues are discussed in Box D.

[4] An FX spot transaction may fall under MiFID II best execution rules if it is provided for a client in connection with a derivatives order which itself falls under MiFID II.

- **First-generation algorithms:** The pioneering EAs generally had simple mechanical rules and were modelled after early algorithms in the equity market. The earliest FX EAs sought mainly to automate traders' practice of splitting parent orders into child orders, and followed strictly predetermined execution schedules. Their lack of sophistication generated distinct trading patterns that were easy for other market participants to detect.
- **Second-generation algorithms:** In subsequent iterations of FX EAs, providers strove to develop EAs that reduce market impact and avoid leaving distinct trading patterns by introducing some randomisation in the size and timing of child orders. Nevertheless, these algorithms remained essentially on statically defined schedules, and continued to be susceptible to detection through the use of more sophisticated forecasting and pattern recognition techniques.
- **Third-generation algorithms:** By the mid-2010s, FX EAs started to use complex statistical models to drive algorithmic decisions and react more dynamically to changes in market conditions, with the aim of further reducing market impact and signalling. These EAs leveraged the increasing availability of real-time market data and computing power to assess market conditions (eg order books) in near-real time, and to inform subsequent execution decisions.

Section 2 contains a detailed typology of common variants of EAs that are in use in markets today.

1.3 Prevalence of FX execution algorithms

To gain insight into the usage and provision of FX EAs, the study group conducted a survey of large buy-side and sell-side market participants and a series of interviews with sophisticated market participants on preferences, offerings, motivations and challenges related to EA usage and provision.[5] Among these large global institutions, roughly half were already using FX EAs, which is relatively high compared with other metrics on EAs' FX market share,[6] but similar to results from a survey conducted by the GFXC, where 40% of respondents indicated that they use algorithms in their trading (GFXC (2020)). Meanwhile, 55% of the potential providers (ie banks and principal trading firms) in this group of large institutions were already offering FX EAs to their clients, and half of the remaining 45% noted they were considering offering these tools.

The survey also found that respondents were predominantly using EAs to trade FX spot, while EA usage for FX derivatives such as non-deliverable forwards (NDFs) – where electronification did not gain traction until 2008 – was nascent.[7] EA usage was most prevalent in liquid advanced economy currencies and used less for emerging market economy (EME) currencies. Users' average ticket sizes for advanced economy currencies were also about 50–100% larger than in EME currencies (Graph 1).

5 The present survey covered 70 market participants from around the world and was accompanied by a series of interviews with a broad range of market participants. All providers invited to participate in the survey have a known electronic FX franchise. These surveys and interviews took place between October and November 2019., with a targeted follow-up in June 2020

6 As the study group focused on franchises with a known electronic FX offering and more sophisticated buy-side clients, the share of respondents using EAs is likely to have been biased upwards.

7 See Section 1.8 for a more detailed discussion of the outlook regarding FX EAs in NDF markets.

1.4 Users and usage of FX execution algorithms

Rationales for using FX execution algorithms

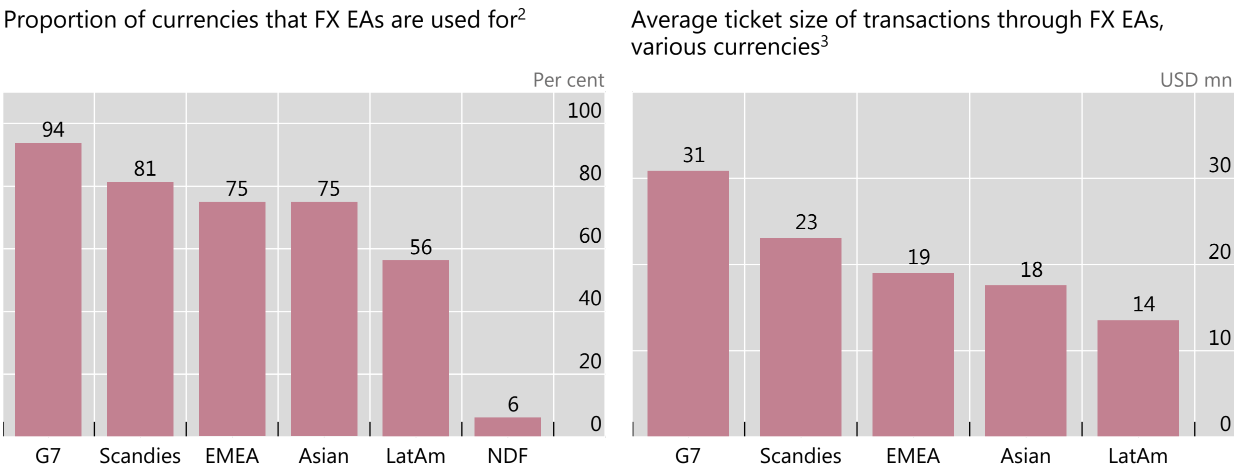
Survey respondents indicated that they use EAs primarily to reduce trading costs or market impact (Graph 2). Others may use EAs to ensure execution within some allocated period of time, or to trade close to or better than particular benchmarks.[8] Finally, EAs are often used by market participants for greater operational efficiencies related to trade entry, trade monitoring and trade settlement. EAs automate what is otherwise a manual process. EA transactions are generally settled as one single ticket regardless of the number of individual child orders, significantly reducing operational costs.

Customer groups of FX execution algorithms

The study group's survey and interviews with EA providers suggest that institutional investors, such as asset managers and hedge funds, are among the more common users of EAs. These market participants have the ability to accept some market risk in order to reduce execution costs by minimising the bid-ask spread paid, and reducing the market impact of their trades, especially for large-sized orders.

Banks – typically small to mid-size regional banks that do not offer EAs – were also found to be users of EAs. Yet penetration rates and volumes transacted were smaller than those of the buy side. These entities generally used EAs alongside other modes of execution, and mainly targeted swift execution. Many provider banks also allow EA access inside their own organisation by desks that need to transact FX – so-called

Usage of FX execution algorithms across currencies[1] and average trade sizes — Graph 1



[1] G7 currencies: US dollar (USD), euro (EUR), Japanese yen (JPY), pound sterling (GBP), Australian dollar (AUD), Canadian dollar (CAD) and Swiss franc (CHF). [2] Average of providers' responses to the question "Which currencies do clients typically use EAs for?" [3] Average of providers' responses to a question on the average size of clients' tickets in the respective currencies. Ticket sizes for NDFs excluded due to small sample size.

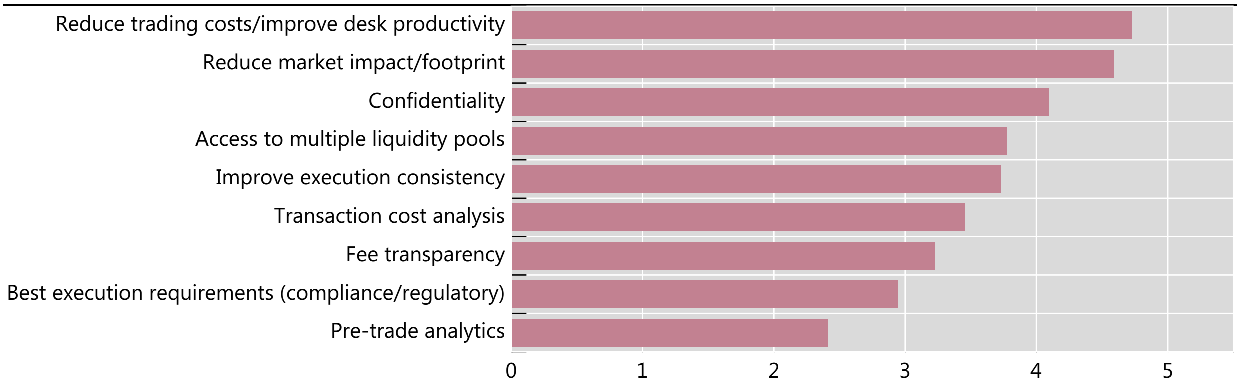
Sources: Markets Committee survey among execution algorithm users and providers; study group calculations.

[8] For example, in order to attain the 4 pm WM/Reuters FX spot benchmark price, an EA may be deployed during the five-minute calculation window, in a way that seeks to minimise the difference between the eventual benchmark fixing rate and the actual rate achieved through the execution.

What are the main reasons why users choose execution algorithms?

1 = not important at all; 5 = very important

Graph 2



Sources: Markets Committee survey among execution algorithm users and providers; study group calculations.

in-house users. Notably, feedback from providers suggests that in-house users typically have access to the same suite of EAs as are offered to clients, suggesting that tools for strictly proprietary use do not seem widespread among providers.

Non-financial corporations represented the smallest segment of EA users among FX market participants, in line with their generally smaller FX footprint (Schrimpf and Sushko (2019a)). Some corporations deliberately stay away from EAs, which require relatively deep market expertise and understanding of technical aspects of FX transactions. These institutions prefer traditional execution styles such as voice trading for risk transfer or use of fixing orders. Others, however, find they have the required level of sophistication in their organisation. These included several large multinational corporations for which EAs accounted for about 20–25% of their FX transaction volumes.

Finally, the survey also covered central banks' use of EAs, particularly 15 central banks from Asia-Pacific, Europe, North America and South America (Box A). Several central banks use EAs for some activities such as reserve portfolio rebalancing, but such usage is relatively limited. Two central banks indicated possible future use of EAs, while more than half of those surveyed noted that EAs are not necessary for their trading activities, which is likely to be due to their relatively low frequency of trades. A few central banks suggested that they were not able to use EAs because of internal restrictions.

Preparing, monitoring and assessing the use of FX execution algorithms

Most users apply EAs alongside other methods of execution. According to the survey results, EAs accounted for about 32% of their overall FX transaction volumes. And when determining whether to use an EA as opposed to other available means of execution, factors such as the size, currency pair, urgency and timing of the trade, and current liquidity conditions were typically considered.

Following a decision to use FX EAs, users also have to consider the choice of EA provider(s), the type of EA, the types of liquidity pool(s) / trading venues to execute in, and the parameterisation of the EA – all of which can affect the quality of execution. Notably, 64% of survey respondents who were using EAs had access to five to ten EA providers for their execution, implying a broad range of diverse EA offerings and specifications they had to choose from.

EA users also have to actively manage the trade-offs inherent in FX execution, eg balancing market impact, market risk and opportunity cost. For instance, a more passive style of execution (involving waiting longer to finalise a trade) can lower the average price of execution in comparison with an aggressive trade (aimed at executing a trade quickly and/or involving crossing the bid-ask spread). Moreover, with the trade execution being distributed over a longer period of time and the individual child orders being smaller, market impact is also expected to be diminished. However, the more time a trade takes to be executed, the longer a market participant is exposed to market risk, which leads to higher disparity of actual outcomes. Such

Box A

Central bank usage of FX execution algorithms

Among the 15 central banks that were surveyed, five central banks reported using EAs, with the extent of usage varying widely. One central bank indicated that it used EAs for almost 90% of overall volumes, two used EAs for about 25–30% of overall volumes, and the other two central banks used EAs for less than 10% of their overall FX volumes. Central banks' use of EAs for FX trading was concentrated in developed market currencies. More than half of them used time-sliced algorithms for more than 70% of their transaction volume. Most central banks indicated that they have three or more FX EA providers.

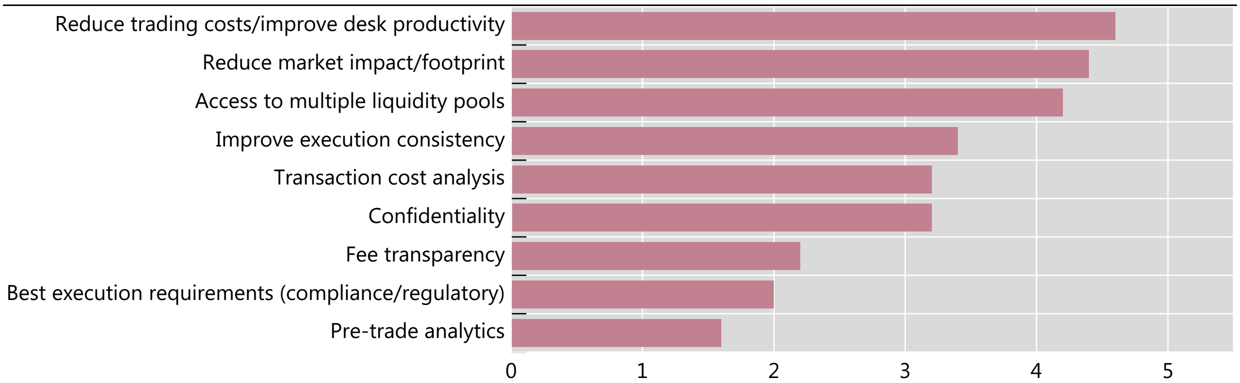
Ten of the 15 central banks reported that they were not currently using EAs. Of these, two are currently considering the future use of EAs. Six highlighted that they did not consider the use of EAs necessary due to the relatively low volumes of their activities in the FX market, while two highlighted internal restrictions on the use of such systems. Overall, central banks were relatively cautious when using EAs, and would not use them outside their main trading hours. When using EAs, all of them had people overseeing the execution. Most of the users attached limits to the orders, mainly limit price controls.

For central banks that have adopted EAs, the main motivations were to reduce trading costs, improve desk productivity, and access multiple liquidity pools in order to reduce market impact or footprint. Some central banks also cited improving execution consistency, transaction cost analysis and confidentiality as reasons for using EAs. None of the central banks highlighted pre-trade analytics or best execution requirements as key motivations for their use of EAs.

Central banks' reasons for using EAs[1]

1 = not important at all; 5 = very important

Graph A



[1] The findings in this graph are based on a relatively small sample of central banks, which also varied significantly in the extent of their use of EAs.

Sources: Markets Committee survey among execution algorithm users and providers; study group calculations.

trade-offs need to be addressed explicitly, as illustrated by the "Execution Algorithm Trilemma" presented in Section 2. Unsurprisingly, the recent GFXC survey found that lack of experience and understanding were common reasons why users choose *not* to make use of EAs at all.[9]

Finally, when using EAs, users also highlighted the need to closely monitor market conditions, as well as to invest in people skills and technology to assess the performance of an execution – ideally before, during and after a trade. For this, some rely on built-in execution analytics on EA platforms to monitor each trade throughout its life cycle. *Pre-trade analytics* give users information on the prevailing market conditions and can be used in the selection of the specific execution algorithm and its appropriate parameters (eg over which time frame an order should be conducted). *Real-time statistics* may inform traders on the percentage of their order filled and whether execution is proceeding as scheduled, and help determine if a change in trading strategy or parameter is required. Finally, *post-trade analytics* such as transaction cost analysis (TCA) are used to assess the difference between the average cost of execution and a relevant benchmark (Box B). TCAs can help users better assess the extent to which they are adequately compensated for carrying the market risk of the execution.

Box B

Transaction cost analysis

Transaction cost analysis (TCA) is aimed at evaluating the quality of trade execution by comparing the final traded price of the execution against a benchmark. TCA can take many forms and will depend largely on the type, size and frequency of FX activity being undertaken. For an active market participant, TCA typically forms part of the broader process of proving and documenting best execution, which has become a common practice in recent years. In particular, regulations such as MiFID II have placed increased demands on traders to demonstrate that they have taken "sufficient" steps to obtain the best execution possible on a trade. Even though spot FX is not within the scope of these regulations, some participants apply similar best execution criteria across all products. Best execution regulatory requirements have subsequently encouraged the adoption of electronic trading more broadly and algorithmic trading more specifically, since electronic execution provides increased transparency around pricing and execution. TCA also helps the buy side manage relationships with their counterparties more effectively by allowing the discussion of execution quality to be based on quantitative metrics.

TCA depends on a number of key metrics. Detailed TCA requires accurate time stamps throughout the trade life cycle. These can then be measured against granular, high-frequency price data, and further analysis of price slippage, market impact and the ratio of rejected trades. Other metrics can include comparisons against the market at different times, such as when the order arrives at the trading desk, during the time of the execution and after the end of the execution.

TCA can be processed by liquidity providers, trading platforms, third-party TCA providers or the end user. Even though the calculations themselves are typically not too complicated, accessing and processing the required data as well as choosing an adequate benchmark can be a challenge in the absence of a complete consolidated tape in the FX market (see Section 3.3 for a more detailed discussion of aspects related to data access).

9 GFXC (2020).

1.5 Providers of execution algorithms

Institutions providing FX execution algorithms and their offerings

The main providers of EAs tend to be large international banks, many of which initially built the algorithmic technology for their own trading activity. In response to end-user demand, they then adapted their technology to offer their clients a suite of EAs.

A small number of non-bank liquidity providers and specialised independent firms also provide client-facing FX EAs. Non-bank liquidity providers chiefly refer to principal trading firms (PTFs) which developed highly sophisticated algorithms for their own trading activities. While the provision of client-facing FX EAs by PTFs is still nascent, it is expected to grow in the coming quarters. Finally, there are also a number of independent providers. These are technology firms that operate only as vendors (not as broker or dealer), providing clients with software and technical support to execute trades algorithmically.

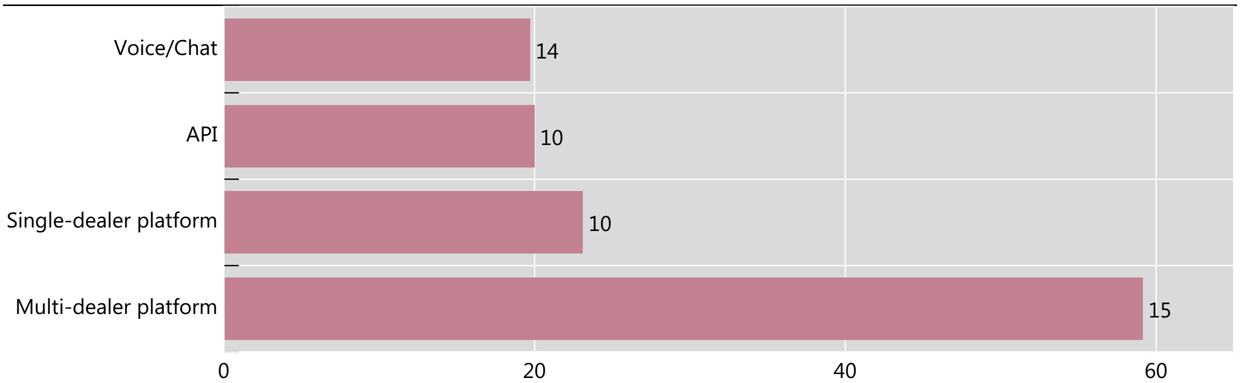
A few small to mid-size banks, which generally lack the resources to independently develop EAs, offer their clients "white-labelled" EAs acquired from other providers, particularly technology firms. This approach allows these smaller banks to provide a much demanded EA offering to their existing client base, while limiting their own development costs.

Distribution channels and tailoring of FX execution algorithms

Providers make their algorithms available through multiple channels (Graph 3). Generally, the most sophisticated clients access EAs via Application Programming Interfaces (APIs) that allow connections to different data feeds or trading tools, while voice/chats provide relatively quick access for clients that do not want to set up systems or deal directly with the EA software. By volume of transactions, EAs are most frequently accessed via multi-dealer platforms, which allow users a single point of access to their various providers' EAs. These include platforms provided by third parties that "host" the EA providers' algorithms.

Through which channels do providers distribute execution algorithms?

In per cent (number of responses for each category indicated next to the bars)[1] Graph 3



[1] Calculated as the FX EA volume for each selected channel as share of overall FX EA volume.

Sources: Markets Committee survey among execution algorithm users and providers; study group calculations.

Beyond the channels outlined above, independent "technology players" often offer applications that can be positioned within the client's workflow – with the incorporated EA tailored to suit the user's very specific needs. Buying such products can often be an intermediate solution between the cost of developing an algorithm suite in-house and relying on third-party providers, such as banks.

1.6 Provider-user relationship

How orders are managed and executed depends on the role a liquidity provider assumes when transacting with a client. The FX Global Code states that market participants should understand and clearly communicate their respective roles and capacities when trading with one another. In this context, three kinds of models can be distinguished.

Traditionally, FX transactions between providers and users have been conducted on either an agency or a principal basis. If liquidity providers act as an *agent*, they do not take on credit, settlement or market risk themselves but execute orders only for the account and on behalf of the client – thereby acting as a conduit for liquidity and charging an agreed fee for doing so. In contrast, if providers act as a *principal*, they trade on their own account and on their own behalf and thus take on not only credit and settlement risk, but also market risk, which they can either warehouse or pass on via another trade.

In the context of FX EAs, market participants sometimes describe a third, hybrid form of relationship called "riskless principal". In this arrangement, the EA provider is acting as an *agent* with respect to market risk, but remains a *principal* in terms of counterparty and settlement risk exposure. This process effectively allows the EA provider to face the client in a principal capacity, except in the case of market risk, to which the EA provider has no exposure and is "riskless". The term "riskless principal" can carry different interpretations, and it is therefore important that participants are clear about the way in which it is applied.

From an economic perspective, a key element is how risks are shared. In most cases of EA usage, users seem to take on the market risk, whereas providers often cover credit risk and the operational risk of an error in the algorithm or missing controls. A crucial takeaway from this report is the need for users to seek clarity on the specific risks and liabilities they bear when using an EA. While the FX Global Code states that market participants should understand and clearly communicate their roles and capacities in managing orders or executing transactions, it is not clear that this allocation of role and responsibilities is spelt out clearly in the EA user agreements and disclosures.

1.7 Algorithm safety

This section describes the key risks associated with the use of EAs. While operational risk stems from potential failures of algorithms, IT systems and processes as well as human errors, market risk arises through potential losses from adverse market moves.

These risks are not unique to FX EAs. Regulated securities markets are required to have controls in place to avoid systemic failures. In the FX market, however, due to their unregulated nature, the onus is on each EA provider, user and trading platform to provide its own safeguards against the risks of automated order execution.

In general, controls can be categorised by the phase of order execution in which they are used – pre-trade, in-flight and post-trade.

Pre-trade controls serve as a final check on market orders before they are transmitted to the trading venues. They are required by regulation in the major currency jurisdictions (Box C).[10] They allow for automatically blocking, halting or cancelling orders as soon as trades occur outside defined price thresholds, surpass a maximum size, or post an excess amount of orders automatically. The most common examples of controls include defining the maximum order size, checks of market data reasonability, price tolerance limits, or restricting the use of algorithms to trading hours where the market is most liquid.[11] Another key control for avoiding erroneous behaviour is to use restrictions on participation ratios, ie limit the share of executed volume at any point in time to a certain percentage of total traded volume. These pre-trade controls become particularly important during times of heightened volatility, as has been emphasised by some providers in reflecting on their Covid-19 crisis experience when liquidity conditions worsened.

In-flight controls allow users or providers to adjust execution parameters during an execution, often when market conditions change, or the algorithm behaves in an undesirable or unexpected way. This is important, for example, in instances of particularly low liquidity when algorithms could dominate trading volume or stop trading altogether. Some EA providers have taken to alerting users mid-way during their transactions when market conditions change significantly. This allows users to decide if they should accelerate, pause or abort a trade in view of changing market conditions. EA providers may also use a built-in "kill switch" – a mechanism designed to automatically pause or halt execution instantaneously in case of malfunction, lack of liquidity, unexplained volatility or unusual levels of trading activity. Unlike circuit breakers and pauses applied by exchanges, kill switches for execution algorithms are implemented and triggered by individual institutions. Potential market-wide implications are discussed in Section 3.4.

Post-trade controls involve continued monitoring of intraday market and credit exposures against limits, carrying trades with counterparties and taking mitigating actions when limits are breached. Transaction logs are reviewed to identify errors or potential conduct issues, while post-trade and event analysis is used by both users and providers to understand algorithm behaviour under particular scenarios and improve execution strategies and risk controls.

1.8 Outlook for the use of execution algorithms in the FX market

Looking ahead, the outlook for further adoption of FX EAs will depend in large part on their ability to deliver higher execution effectiveness to users. This relies on EA providers' ability to use "best in class" technology to help users compete effectively in the FX market. The future use and evolution of FX EAs is likely to be shaped by a number of developments:

[10] For example, Bank of England Prudential Regulatory Authority Supervisory Statement SS5/19 on Algorithmic Trading outlined risk management and other system and control functions.

[11] For example, the US Futures Industry Association (FIA) Guide to the Development and Operation of Automated Trading Systems outlines a list of pre-trade controls for automated trading systems.

Box C

Overview of relevant regulations, codes and standards

Although spot FX is exempted from best execution requirement under MiFID II, the latter has had a strong impact on practices in FX spot. Indeed, MiFID II, which came into effect in January 2018, is arguably the most comprehensive regulation on algorithmic trading.① Thus, MiFID II and its related regulations and administrative provisions by member states have had a great impact on the use of EAs, via both the obligation to achieve best execution, and the explicit requirements for governance oversight and risk controls.

Best execution requirements have facilitated the adoption of EAs. MiFID II requires that firms take "all sufficient steps" to achieve the best possible results for their clients. It further increased transparency requirements, with firms and execution venues required to produce periodic reports demonstrating compliance. A survey by Greenwich Associates finds that the best execution requirement has made the use of FX EAs essential, since, in addition to reducing execution costs, they provide a more efficient way to prove best execution to internal and external parties, including auditors and regulators.

MiFID II also specifies requirements on governance, oversight and risk controls in place for firms engaged in algorithmic trading (including EAs).② It specifies types of risk limits specific to algorithmic trading that are monitored in real time, and requires that firms have in place a "kill" functionality to completely stop transactions. Such requirements also raise the barrier to entry for smaller banks with less financial and technological capabilities, which may rather choose to rely on third-party providers to offer algorithmic execution to their clients. MiFID II is applicable only in the European Union.

In the United States, similar expectations are in place for equity trading broker dealers on risk assessment of trading algorithms and software testing and validation, as specified in regulatory guidance issued by FINRA.③ But trading in FX spot is not specifically addressed. FX futures traded on the Chicago Mercantile Exchange (CME) are subject to self-regulatory organisation (SRO) rules as specified under Commodity Futures Trading Commission (CFTC) guidance.④ Algorithmic trading guidelines and best practice standards are also produced by industry bodies such as the "FIA.⑤ While FIA rules do not directly apply to spot FX, they can serve as guidelines.⑥

The voluntary FX Global Code ("Code") is more directly applicable to FX.⑦ It recommends the provision of adequate disclosures to users regarding the operation of trading algorithms. A subsequent effect of the Code has been the development of curated Code liquidity pools, which comprise only liquidity providers that have signed up to the Code. Further, some industry associations and third-party providers have developed training and education programmes around the Code. The GFXC recently agreed that the increasing usage of algorithmic execution in the FX market warranted a review of the Code's existing guidance in this area.⑧

More recently, the Fixed Income, Currencies and Commodities (FICC) Markets Standards Board (FMSB) highlighted risks from the use of EAs, and identified gaps in regulatory coverage for commodities and spot FX, for both of which algorithmic trading is widely used. At the time of writing, it is inviting comments on a proposal for a Statement of Good Practice intended for the use of algorithms in all FICC businesses.⑨ Like the Code, these are broad standards for best practice, and not regulations.

Regulation and standards will need to continue to adapt to rapidly evolving changes in the FX market structure, including increased use of EAs by the buy side, fragmentation and new entrants. Additionally, machine learning and artificial intelligence are just emerging in the field of algorithmic execution, and the use of such technologies will bring new risks and challenges, to which regulations and standards will need to continue to adapt.

① See MIFID II Articles 27,28. ② MIFID II RTS6, "Specifying the organisational requirements of investment firms engaged in algorithmic trading", 2016. ③ Guidance on Effective Supervision and Control Practices for Firms engaging in Algorithmic Trading Activities, FINRA Regulatory notice 15-09. ④ https://www.cmegroup.com/trading/why-futures/welcome-to-cme-fx-futures.html. Additional references: https://www.cftc.gov/sites/default/files/idc/groups/public/@aboutcftc/documents/file/tacpresentation032912_cme.pdf; https://www.cftc.gov/About/CFTCCommittees/TechnologyAdvisory/index.htm; https://www.cftc.gov/About/CFTCCommittees/TechnologyAdvisory/tac_meetings.html. ⑤ See eg FIA best practices for exchange risk controls, October 2019. ⑥ FX derivatives and futures are included in MiFID II and exchange regulations. ⑦ FX Global Code, August 2018, https://www.globalfxc.org/fx_global_code.htm ⑧ GFXC press release, December 2019, https://www.globalfxc.org/press/p191204.htm. ⑨ Algorithmic Trading in FICC Markets & Statement of Good Practice for FICC Market Participants, draft, July 2018.

New preferences: user interactivity. One trend is an emerging bifurcation between providers that seek to give users more vs less direct control over their execution. Among those giving more control, this entails giving users the ability to alter execution strategies mid-way and to refine execution parameters (eg speed of execution, liquidity pools), particularly as users become more informed through better pre-, post- and real-time trade analytics. Notably, these providers use dashboards to equip users with an understanding of market developments during the execution. In contrast, other EA providers have reported a shift towards algorithms, where decisions on the type and time of execution are delegated to the providers. One provider is considering offering only one algorithm, where the user only gets to define the broad objective (eg size of transaction, urgency) and all of the other execution decisions are determined by the provider.

New algorithm capabilities: algorithm wheels and basket algorithms. A further evolution within the EA landscape are so-called algorithm wheels that allocate transactions across various algorithms and providers. The algorithm wheel is a third-party algorithm and utilises granular data on execution performance to automate routing decisions. Algorithm wheels have become more widely used in equity markets and have the potential for wider adoption in the FX market.[12] Basket algorithms are EAs that allow users to group together transactions in a few currencies, with the aim of reducing overall execution costs in comparison with executing each currency individually. The cost savings derive from increased opportunities for matching opposing order flow directly and netting across the various legs of the trade.

New markets: non-deliverable forward (NDF) markets. EAs for use in NDF markets are currently nascent, as electronic trading of NDFs did not start until 2008. Only a small number of providers (three large tier-one banks among those engaged) currently offer EAs for NDFs. Buy-side participants have also observed that the technology for algorithmic execution of FX derivatives (including NDFs) is still relatively nascent. However, several providers have identified such algorithms as a growth area for prioritisation in development plans. These developments are supported by the increase of market-makers and electronic communication networks (ECNs) that stream prices for NDFs, and the shift to central clearing making trading more efficient (Schrimpf and Sushko (2019b)). The expectation is that institutions will eventually provide the same suite of algorithms for NDFs as they currently have for FX spot.

New technologies: artificial intelligence and machine learning. Another emerging area of development is the adoption of artificial intelligence (AI) and machine learning (ML) techniques in the development and design of FX EAs. The findings of this report suggest that the adoption of AI and ML in the FX EA context remains limited. These technologies are mostly confined to being used "offline" in the calibration of algorithm parameters, eg for improving an algorithm's decision logic, rather than being applied "live" during execution, eg allowing an EA to autonomously and dynamically adjust its own decision logic while trading. Moreover, they are often restricted to a finite set of actions, such as preventing the erratic behaviour of an EA. For example, a user could define the broad execution schedule of a trade, while the ML-enabled algorithm could be given flexibility in deciding how to deploy child orders, eg whether to post them passively or more aggressively by allowing to cross the bid-ask spread. The restrictive use of AI and ML in EAs may reflect the difficulty of explaining and reproducing the decision logic of these techniques. This makes robust and comprehensive testing, as well as proper documentation, very challenging.

[12] https://fxalgonews.com/articles/the-rise-of-the-fx-algo-wheel.html.

Finally, ML-based algorithms have not yet demonstrated their ability to substantially outperform more traditional EAs. A wider adoption of ML/AI techniques in the development and design of EAs remains to be seen.

2. The design and application of FX execution algorithms: what they are, what they do and how they work

2.1 Types of FX execution algorithms and key trade-offs

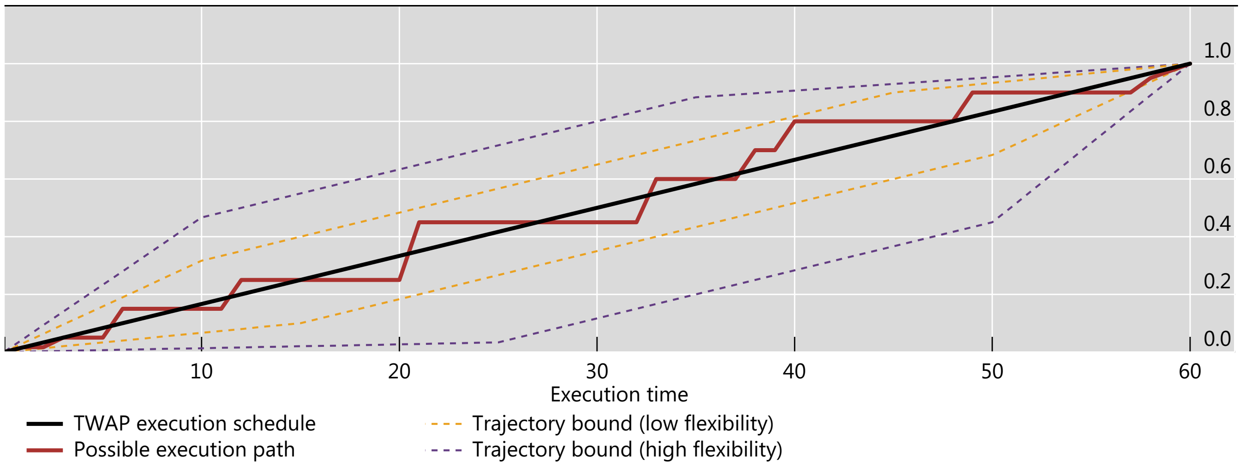
Typology of execution algorithms

There are a large number of FX EAs in use today. Notwithstanding the wide range of names for FX EAs, they are minor variations or hybrids of **six basic archetypes**:

- **Time-sliced:** algorithms that split parent orders into small child orders evenly over time. These algorithms are also known as time-weighted average price (TWAP) algorithms. In the example in Graph 4, an order is placed to sell USD 60 million USD/JPY through a TWAP algorithm with an execution time of 60 minutes. Based on this input, the broad schedule of execution would imply USD 1 million being executed every minute (black line). Current versions of TWAP algorithms typically provide for some randomisation in the timing of execution to reduce the predictability and signalling from orders (red line). The algorithms would also typically have some flexibility to opportunistically diverge from the broad execution schedule, to obtain a better transacted price (lower and upper trajectory bounds).
- **Historical volume-sliced:** algorithms that split parent orders into small child orders, scheduled according to historical measures of traded volume. Also known as historical volume-weighted average price (VWAP) algorithms.
- **Percentage-of-volume (POV):** algorithms that target a level of participation by current estimates of market volume in the particular currency pair.

Illustrative example of a TWAP algorithm

Fraction of order completed — Graph 4



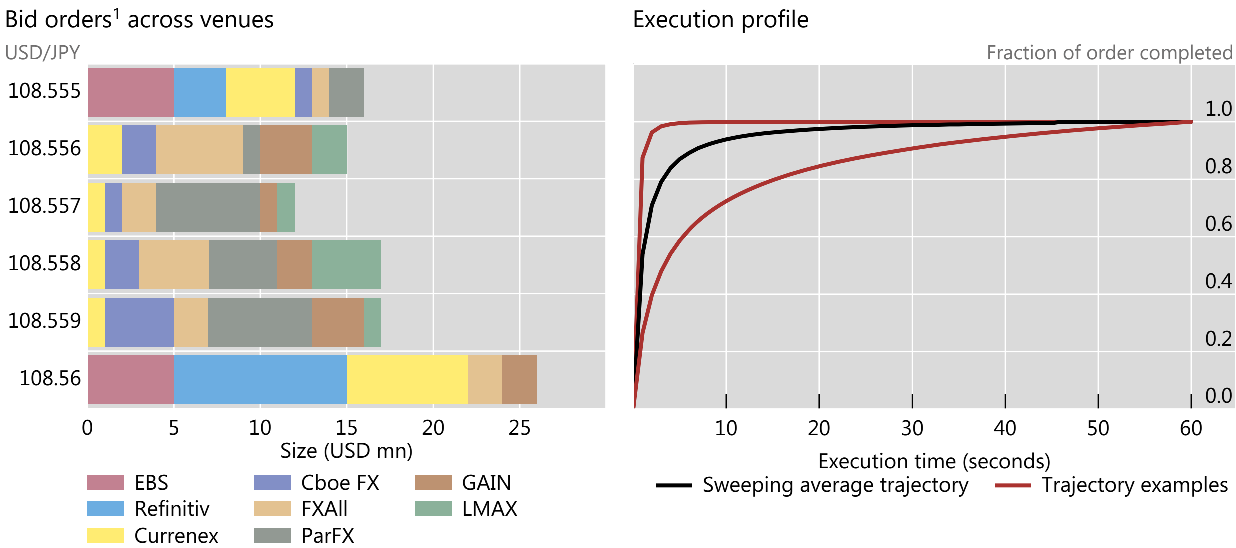
Source: Markets Committee study group calculations.

- **Pegged/tracker:** algorithms that place orders "tracking" the market. The aim of the pegged/tracker algorithms is to execute the orders at levels better than the prevailing mid-price. To that end, the orders can be placed at the prevailing bid (if buying) or ask (if selling) price, or at levels just slightly above or below these. These algorithms are by nature the most passive, since they follow the market and generally do not trade at prices worse than mid-price.
- **Implementation shortfall:** algorithms that seek to minimise slippage, defined as the difference between the average price achieved by the EA throughout the entire execution and the arrival price, ie the mid-price at the start of the transaction. Transactions are typically completed within a relatively short time frame but are also aimed at reducing market impact through dynamically adapting the aggressiveness of execution to market conditions.
- **Limit-based/sweeping:** algorithms that aim for immediate or rapid execution by targeting all liquidity at various venues better than a user-specified limit price. In the example in Graph 5, an order is placed to sell USD 60 million USD/JPY through a limit-based algorithm which will generally first take stock of available liquidity at the various trading venues. In case of sufficient liquidity, the algorithm will instantaneously consume all available liquidity up to the order amount. In case of insufficient liquidity, the algorithm will have several options to complete the remaining part of the order, depending in large part on the user's preferences regarding the urgency of execution.

For more illustrative examples of all six archetypes of FX EAs, see the Annex.

The TWAP – the most simplistic EA – is the most widely used type of algorithm. This may in part be due to the fact that time-sliced EAs are the most commonly offered, featuring in most algorithm providers' suite of offerings (Graph 6). Given

Illustrative example of a limit-based/sweeping algorithm — Graph 5

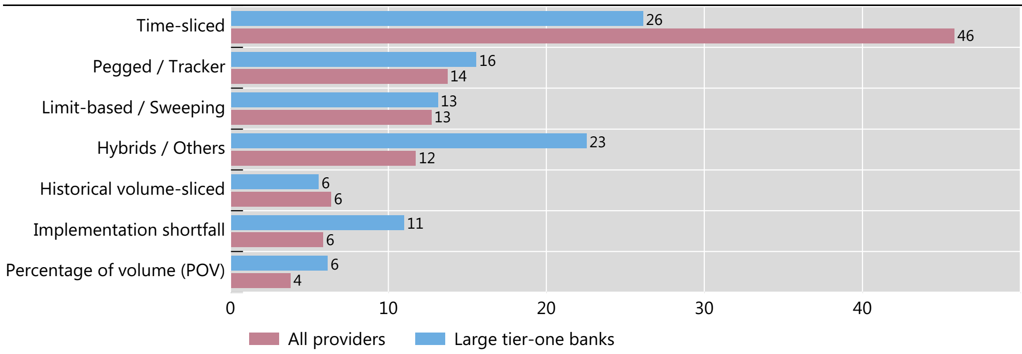


[1] Limit/passive orders to buy USD/JPY

Source: Markets Committee study group calculations.

Usage of providers' EA offerings in 2018, by type of algorithm

In per cent[1] Graph 6



[1] Percentage of volume for each algorithm type as a share of providers' overall execution algorithm volume.

Sources: Markets Committee survey among execution algorithm users and providers; study group calculations.

the relatively nascent development of more dynamic execution algorithms (eg implementation shortfall, pegged/tracker algorithms, percent-of-volume (POV) and hybrids of such), their use in the FX market is still relatively limited. However, among large tier-one banks that typically do offer such dynamic EAs, these kinds of algorithms account for a large proportion of EA volumes. Moreover, anecdotal feedback from market participants suggests that more dynamic algorithms have performed quite well even during the Covid-19-related phase of elevated volatility in March 2020, which may imply that both offerings and usage of dynamic EAs are likely to rise further in the future.

Key trade-offs between execution objectives: the execution algorithm trilemma

The overarching goal of EAs is to achieve optimal execution. However, this concept comprises several competing dimensions and will be defined differently for different users depending on their individual objective functions. While the various types of EAs can accommodate several preferences with respect to execution, no single execution algorithm can optimise all aspects simultaneously. This is depicted by the execution algorithm trilemma, which is explained in greater detail below.

Perold (1988) defines optimal execution as the goal of minimising "implementation shortfall" in terms of (i) market impact and spread costs, (ii) market risk and (iii) opportunity costs or execution uncertainty:[13]

- **Market impact** is the difference between a fair price benchmark prior to execution and the actual execution price. This difference incorporates both a spread and any market movement between these two moments in time (Collins and Fabozzi (1991)). Market impact arises for two reasons: (i) the liquidity absorption of orders, ie when an order consumes available liquidity in the opposing direction; and (ii) the information content of orders.

[13] Other definitions of best execution can be found, for example, in Wagner and Edwards (1993), Kissell (2006) or Kissell (2014).

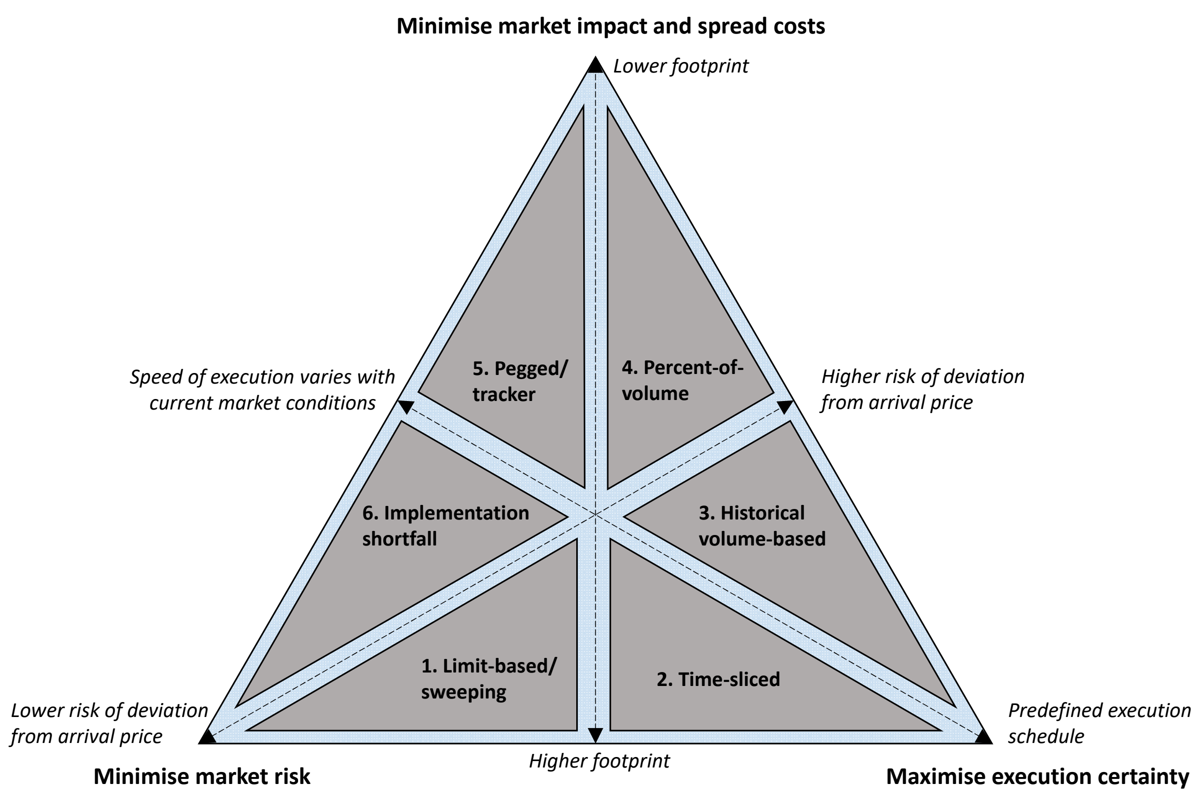
- **Market risk** is the risk arising from fluctuations in market prices over the duration of the execution covering fluctuations in market prices unrelated to the actions of the EA.
- **Opportunity cost** is a measure of the forgone profit or avoided loss of not being able to transact the entire order within the allocated time period.

When executing, market participants have to determine the appropriate trade-off between these three aspects of best execution. For instance, an EA that minimises market risk by executing as swiftly and aggressively as possible will not be able to minimise market impact simultaneously, as the latter would require a much more passive and thus slower execution. Similarly, an EA that seeks to respond dynamically to changing market conditions to optimise the trade-off between market risk and market impact may fail to complete if liquidity conditions are thinner than usual.

Importantly, no single EA can optimise all three components of implementation shortfall at the same time but will inevitably need to strike a balance between the three. Graph 7 displays this execution algorithm trilemma, of which the two main trade-offs are presented below.

Market impact vs market risk. The main trade-off exists between market impact and market risk (Graph 7, top and lower left-hand corner).

The execution algorithm trilemma — Graph 7



- This trade-off is exemplified by **pegged/tracker algorithms** (upper left), which minimise market impact by transacting almost exclusively with passive orders, but are exposed to substantial market risk as execution tends to be slow and market prices can drift substantially over the life of the execution. Conversely, **limit-based/sweeping algorithms** (lower left) minimise market risk by transacting very rapidly, but can have significant impact on market prices up to the limit price. **Implementation shortfall algorithms** (middle left) seek to internalise the trade-off between market impact and market risk in a single algorithm, by using proprietary market impact and market volatility risk models to determine an optimal approach to the transaction.
- Most algorithms include user-specified settings on the level of aggression or urgency for an execution, reflecting the importance of user choice in the determination of the trade-off between market impact and market risk. This choice might in turn be shaped by the user's view of the market. For instance, a user may prefer lower (higher) market risk if s/he expects prices to shift unfavourably (favourably) against his transaction, and a lower (higher) tolerance for market impact risk if markets are expected to be illiquid (deep).

Trading costs vs opportunity costs. Another important trade-off is between trading costs (comprising market impact and market risk) and opportunity costs (Graph 7, lower right-hand corner).

- The trade-off is exemplified by the **TWAP and historical VWAP algorithms**, which provide users with a high degree of execution certainty, but are unable to optimally manage trading costs as they do not respond dynamically to market conditions. For instance, a TWAP could be set to transact gradually, but will suffer from high exposure to market risk. Conversely, the same algorithm may be set to transact quickly to minimise market risk, but will result in significant market impact if the availability of liquidity in the market is lower than expected.
- **POV algorithms** are better able to manage the trade-off between market risk and market impact, as they will dynamically adjust to the availability of liquidity in the market, but they risk opportunity costs as they cannot guarantee completion within a fixed period.

Many providers increasingly recognise that users typically have preferences that do not lie on the corner solutions of the trilemma. Users ideally want to reduce market risk, reduce market impact and have a degree of execution certainty. **Hybrid algorithms** attempt to bridge this gap with logic to switch between two or more types of algorithms in response to changes in market conditions. One example of such a hybrid algorithm can switch between a TWAP and a **limit-based/sweeping algorithm** when prices shift from one zone to another. The algorithm may start as a TWAP but will transition to a sweeping algorithm when prices move to a more favourable level.

For some users, the way they **benchmark the execution performance** will determine their choice of algorithm. This is especially evident for users who undertake large-sized transactions that can be spread out over the course of a day. Widely used benchmarks in such circumstances include VWAP or TWAP benchmarks. Other users that may have requirements to complete their orders in a relatively shorter time could use a point-in-time benchmark like the arrival price. Such users may prefer to use implementation shortfall algorithms since such algorithms seek to minimise slippage relative to the arrival price.

2.2 Application and usefulness of EAs in a fragmented FX market

The FX market is fragmented, and there is no consolidated tape providing data on transacted volume or price quotes. However, EAs can help users overcome these challenges by combining market data into an aggregate order book and navigating orders efficiently across different liquidity pools (Box D) and trading venues.

Constructing the aggregate order book

A key function of EAs is to help users reduce the FX market's opaqueness. They do so by taking information from their own transactions and market data from various trading venues, stitching them together and creating an aggregated picture of prevailing market conditions that informs execution decisions. Thanks to this aggregated view – also referred to as the aggregate order book – EAs provide estimates on transaction volumes in near real-time that otherwise would only become available with long lags. Different types of market data that feed into this aggregated view of the order book and market conditions include:

Quotes and order book data: Market data in near-real time are generally available in the form of quotes from trading venues (including from the primary FX trading venues such as EBS Market, Refinitiv Spot Matching, and CME for some currency pairs), providers' internal principal trading desk, and a variety of secondary ECNs. How close market quotes can get to real time depends on the frequency of quote updates, which can vary among trading venues.[14]

- *Quotes as a measure of liquidity:* Quotes represent the bids and offers from various market participants in the trading venue. Collectively, quotes on any given trading venue form the order book, or measure the liquidity available at the trading venue at any given point in time. However, quotes represent only the visible liquidity, as market participants may use EAs (such as "iceberg" orders) where only a fraction of the entire order may be visible to other market participants. Some venues, including the primary trading venues, may provide only a snapshot of the order book up to a few pips away from the best bid and best ask. Others may provide the full order book data.

- *Primary market quotes as reference prices*: Quotes from the primary trading venues are generally used by most EAs as their key reference for "where the market is". Indeed, some algorithms may stop functioning if primary venue pricing is shut off.

- **Executed transactions.** To enhance the picture of market conditions, many providers – and particularly large dealers – rely substantially on their own execution "experience". Key measures include fill ratios – which measure the traded amount as a share of the order amount submitted to a trading venue – and market impact – which is a provider's own measure of the impact of its transactions in a trading venue (or relative to particular counterparties where identifiers are available).

[14] EBS Market currently offers real-time prices at 5 millisecond time slices for customers meeting certain participation criteria requirements (volume and maker/taker share). The fastest data feed of Refinitiv Spot Matching at the time of the analysis had 25 millisecond time slices for the most actively traded currency pairs (Refinitiv has since launched a version of the feed with 5 millisecond time slices).

- **Transaction data.** The use of transaction data (price, size and direction) for real-time decision-making of EA execution is usually quite limited, as most trading venues provide them only on a lagged basis. Primary trading venues provide deal indicators for buy/paid and sell/given transactions in real time, but more detailed transaction volume information is typically provided only on an end-of-day basis.[15] Some other venues provide transaction data with shorter time delay, though generally not at a frequency sufficient for market participants to leverage in their real-time decisions.

Box D

Liquidity pools accessed by FX execution algorithms

As noted before, EAs help users navigate the fragmented FX market by providing access to different liquidity pools. While primary electronic venues remain important sources of liquidity for price discovery in the FX market, their market share has dwindled in recent years, and many alternative ECNs and liquidity pools have emerged.① These various types of liquidity pools can be characterised along two broad dimensions: their degree of anonymity and their use of principal or franchise liquidity.

Anonymous vs disclosed: Anonymous liquidity pools based on a central limit order book (CLOB) do not allow apportioning of liquidity or price discrimination via the use of customised tags or other means to identify a counterparty pre-trade. They typically rely on "firm liquidity", ie trades are matched without the optionality for rejection.② In contrast, disclosed liquidity pools or alternative venues allow for liquidity partitioning via the use of customised tags. They typically include "non-firm liquidity", ie the option for counterparties at the venue to reject orders within a specific time frame ("last look").③

Principal vs franchise liquidity: In principal liquidity pools, the liquidity provider assumes the risk of the position itself, ie transactions are executed against the market-making desk of the liquidity provider. In franchise liquidity pools, the liquidity provider uses flow from clients to hedge offsetting flows from other clients, ie by internalising rather than hedging on external markets.④ This would include less lit forms of trading such as dark pools, where there may be relatively little information provided to participants on prices, depth of market, counterparties and other such information which is typically available at other venues on either a pre-trade or a post-trade basis. The main aim of this less lit form of trading is to reduce market impact, with high levels of internalisation expected to improve execution performance. The largest dealers report internalisation ratios of as high as 90%.⑤ However, the true market share of internalisation is hard to gauge, as there are multiple interpretations of the term "internalisation".⑥ As a result, the reported degree of internalisation can vary from one provider to the other. For instance, internalisation at one extreme could include only volumes transacted between users of an EA platform. This would tend to be quite low. At the other extreme, internalisation could include the amount of EA volumes routed to a provider's market-making desk, or even include those externalised into mid-matching venues.

① See Markets Committee, "Monitoring of fast-paced electronic markets", Section 3.2, September 2018. ② This would include the primary venues, as well as other markets such as FXall Midbook. ③ This would include venues such as Currenex, EBS Direct and 360T. Some ECNs, like Cboe FX (formerly HotspotFX) and Euronext FX (formerly FastMatch), offer both firm and non-firm liquidity. ④ Internalisation refers to the matching of offsetting client flows of a dealer without the need to revert to the inter-dealer market. When intermediating such opposite interests in a short period of time, the market-maker avoids paying spread and any brokerage fees that would arise when dealing externally. Additionally, the market-maker avoids unnecessary signalling to the market: a trade in the external market has the potential to reveal the trader's intentions, and so carries the risk of market impact and information leakage, which could have an impact on subsequent prices. ⑤ See eg Box C in Moore et al (2016). ⑥ In the reporting guidelines for the BIS Triennial Survey, internalisation is defined as the ratio of total customer spot turnover that is not hedged on external markets (BIS (2019)). Customer FX spot turnover hedged on external markets is further defined as trading on anonymous venues.

[15] An exception is the EBS Spectrum feed which provides near-real-time volume information.

Derived data. Execution venues typically use the above market data and combinations thereof to compute statistical measures, including volatility indicators, cross currency/asset correlation and trends to further enrich the basis for execution decision-making.

Smart order routing

The FX market comprises a plethora of trading platforms and liquidity pools, which differ in terms of transparency, anonymity, internalisation practices and firmness of liquidity. EAs enable users to navigate this complex market structure by using a technology known as smart order routing (SOR).

SOR constitutes a powerful tool that entails key execution decisions: it governs which liquidity is accessed in which pool, at what time and at which price. It is the gateway to FX liquidity – the last interface between the order and the market. All subsequent actions of execution relate to the actual placement of EA orders in the market – the execution scheduling, which is discussed in the following section.

While the "smart" component of SOR concerns the decision logic related to the choice of trading venue, the degree of sophistication of smart order routers varies. In the more basic variations, order routing logic itself is rather simple, eg seeking the best available price. More recent improvements in SOR take into account other factors, including quoted size, queue length,[16] and venue efficacy measures (eg fill ratio and fill probability) to determine the allocation of orders across trading venues. This is enabled by the availability of more real-time data and rising IT capacity/infrastructure investment. Order routing has also been one of the first areas where machine learning techniques have been employed for the development of more flexible EAs, which include self-learning techniques that enhance their ability to respond to market conditions in real time.

Differences in price impact profiles are also important when routing orders to different liquidity pools. Routing rules generally assume a lower price impact for fully internalised orders – that is, for client orders that a liquidity provider can offset internally with orders from other clients, without having to resort to an external trading venue, where the trade becomes visible. In contrast, transactions taking place at external trading venues are assumed to result in a larger price impact because of the information leakage associated with a "lit" trade. These assumptions are regularly verified through transaction cost analysis.

2.3 Execution scheduling

Another key function of FX EAs is to slice users' orders into smaller ones in order to limit market impact and to manage the trade-offs outlined in the trilemma above. The decision logic for pricing and scheduling of child orders is an important aspect. Indeed, it is typically the design of this decision logic that determines the typology of an FX EA. Further detail on the decision logic and execution profiles of the six archetypes of FX EAs can be found in Annex.

[16] Active orders reside in a queue inside the order book until they are either cancelled or executed against an order of opposite type.

Use of child orders and replenishment rates

By design, EAs slice the user's transaction or parent order into smaller-sized orders called child orders. The size of these child orders is often close to the minimum transaction size on trading venues, which facilitates the absorption of the order and thereby limits market impact.

In order to achieve the required transaction turnover, passive EAs repeatedly place new individual child orders close to or at the prevailing best bid or offer prices. Hence, in contrast to single block trades that are placed inside the order book in full size at once, EAs constantly "replenish" their quoted liquidity. The rate at which individual orders are being renewed is known as the "replenishment rate".

EAs are also able to place orders quite close to the best market bid and offer as they seek to respond dynamically to price developments. In contrast, manual traders that have slower responses to market developments may tend to place resting orders along various levels in the order book to ensure their orders are transacted.

As a consequence, electronic order books with significant participation by EAs tend to look shallower in terms of visible liquidity at any point in time, as compared with markets in which there is more participation by manual traders. However, this may not imply poorer liquidity if replenishment of consumed liquidity is sufficiently quick and continuous.[17]

The pricing of child orders, ie the level of aggression of an algorithm, is mainly determined by the type of algorithm, and to a certain extent by user-defined parameters. While pegged/tracker algorithms give rise almost exclusively to passive orders, a limit-based or sweeping algorithm will rely almost exclusively on the use of aggressive orders. Most of the other algorithms will use a mix of passive and aggressive orders, and face choices in each time step related to the optimal degree of aggressiveness of the orders.

Such micro-decisions related to execution are usually taken based on internally calibrated static values that determine specific parameters, such as the amount of spread an algorithm is willing to cross to achieve a fill under given conditions. This applies especially to EAs that aim to minimise market impact and spread cost (eg implementation shortfall algorithms), as it is essential for these types of algorithms to consider how their orders affect the market.

In this regard, the modelling of market impact becomes a crucial ingredient for the timing of individual child orders. One approach to market impact modelling assumes that market impact decays very slowly towards zero in a pattern usually referred to as the "mark-out pattern" or "price signature" (Oomen (2019)).[18] This means that every child order implies an opportunity cost for later child orders:

[17] See Section 3.2 for a detailed discussion.

[18] The slow price impact decay can be seen as the reflection of long memory in order flow. Empirically, order flow (ie the difference between buy and sell orders) has a long memory, meaning that it is highly autocorrelated over long time horizons. Prices, however, should be hard to predict in an efficient market – hence, autocorrelation of returns should be low. For the long memory of order flow to be compatible with low autocorrelation of returns, the price impact profile has to show a slow decay. If the impact decay were fast, the autocorrelation in order flow would induce predictability in prices. Interestingly, the long memory in order flow is likely to result from order splitting, either by EAs or by human traders. For more details, see eg part V of Bouchaud et al (2018).

because of the lingering price impact of the current order, a less favourable price has to be paid for subsequent orders. Other, more recent approaches to market impact modelling may also take into account the implications of individual orders on aggregate order flow and imbalances.[19]

3. Implications for market functioning

This section aims to assess the benefits and risks from the increased use of execution algorithms from a market functioning perspective and, where relevant, highlights areas for further research and possible policy recommendations.

Definition of market functioning

Market functioning is a difficult concept to capture in a simple definition. For the purpose of this report, we build on the general definition provided in another recent Markets Committee (2019) study and apply it to the analysis of the use of EAs in the FX market. According to that definition, a well functioning market is a **market that allows timely and efficient access for participants who wish to trade and that creates price signals that reflect the underlying fundamentals of the relevant currency markets**.

Hence, an effectively functioning market must fulfil two fundamental functions:

1. **Provide adequate access and matching:** allow participants with diverse trading interests to be brought together in an efficient and cost-effective manner in order to adjust and redistribute their financial exposures.
2. **Allow proper price discovery:** incorporate all relevant publicly available information in an appropriate, prompt and reliable manner, in turn generating meaningful price signals that allow an efficient allocation of resources.

Based on this broad definition, the study group identified three main issues relevant for market functioning that may arise from the increasing use of EAs in the FX market: (i) changes in the microstructure of the FX market; (ii) considerations for the effective use of EAs from a market participants' perspective; and (iii) market-wide implications of the use of EAs.

3.1 Market microstructure changes and implications

A key finding of this report is that EAs are both a response and a contributor to the changing microstructure of the FX market. From a market functioning perspective, three drivers that both affect and are affected by the use of EAs are particularly relevant: (i) the increasing fragmentation of the FX market; (ii) price discovery and the role of internalisation; and (iii) the changing patterns of liquidity dynamics.

Market fragmentation

EAs help market participants overcome hurdles associated with market fragmentation. The FX market is complex, with trading taking place simultaneously on a bilateral basis and at many different trading venues. This reduces the visibility of

19 *Risk.net* (2017).

market transactions, complicates the analysis of market conditions and makes price discovery difficult.

Assessing liquidity in a fragmented market is very challenging. As trading interest may be posted simultaneously across multiple trading venues, this can create a so-called "liquidity mirage",[20] ie an illusion of deeper liquidity than is actually available for transacting.[21] Once trading interest is matched at one venue, it could be withdrawn immediately from other venues in the market. Alternatively, hidden trading interest can create the perception of lower liquidity than is available. This so-called latent liquidity can emerge when traders opt to hide the actual size of their trading interest.[22] Many providers try to differentiate their EA offering by, among other things, the quality of the various liquidity pools that their EAs can route orders to, and the logic applied in doing so.

In this fragmented environment, an important benefit of EAs is their ability to provide access to liquidity across multiple trading venues. By routing orders to the best available source of liquidity, EAs are an effective tool to help match diverse trading interests. Most providers surveyed indicate that their algorithms access more than 10 liquidity pools. As EAs express trading interests across different venues, market participants are more likely to find a matching interest and execute their trades at the venue that provides the best execution outcome.

Price discovery and the role of internalisation

EAs contribute to improving the price discovery process. By slicing larger orders into smaller pieces, EAs enable the market to absorb a large order at or close to the best bid or offer available without creating a large implementation shortfall. This allows for smoother price dynamics. By routing orders to a multitude of venues, EAs also help incorporate price signals across the fragmented market. In this way, they support an efficient market-wide price discovery process.

EAs facilitate the internalisation of customer flows. As orders are split into smaller pieces and spaced out over time, there is a higher probability that these smaller pieces can be offset internally with trades from other customers (ie without having to trade on external trading venues).

Internalisation can be beneficial to both customers and dealers. Dealers benefit from internalisation by avoiding intermediation costs (eg the costs associated with trading on external venues). Similarly, internalisation can benefit users by reducing the information leakage – and hence market impact – typically associated with conducting trades on a visible ("lit") trading venue, such as a primary trading venue. This is why many EAs can be configured to route child orders to internal liquidity pools. That said, some market participants point out that certain dealers may try to skew their prices up- or downwards, depending on their interest to buy or sell a particular position. This practice may negate some of the market impact reduction initially sought by customers, as skewing prices may reveal information to the broader market.

[20] For an empirical illustration of a liquidity mirage, see Dobrev and Schaumburg (2015).

[21] Liquidity mirage is likely to be exacerbated if the practice of last look is permitted, as is the case in the FX market.

[22] A classic order type used to hide the true size of a particular trading interest is known as "iceberg order". As the name suggests, an iceberg order is an order type where only a small fraction of the total order size is visible in the order book.

From a market functioning perspective, a trade-off can emerge between price discovery and internalisation. What is optimal for an individual market participant may not be optimal for broader market functioning. From a market's perspective, "lit trading", eg submitting an order to an external trading venue where the order book is displayed is typically preferable, as market prices can more fully reflect available economic information to all market participants. From an individual market participant's perspective, however, "taking trades into the dark" through internalisation may be preferable, as it generally reduces the information leakage and, hence, the market impact of a trade. In other words, the individual incentives for dark trading to limit market impact contrast with the market-wide desirability of lit trading to ensure a smooth price discovery process.

The impact of internalisation on price discovery warrants greater analysis and close monitoring over time. Current levels of internalisation or other forms of "dark" trading are not seen as negatively impacting the price discovery process at the moment, or at least not substantially. However, market participants agree that there is likely to be some minimum – yet still unknown – threshold of necessary trading volume taking place on primary venues to maintain the integrity of the price discovery process. This is because prices from primary trading venues are used as reference prices for other trading venues (lit and dark) as well as for bilateral trading. Thus, understanding how much "lit" vs "dark" trading is needed for primary trading venues to remain reliable sources for reference prices will help better assess the role of EAs in market functioning.

More generally, deepening central banks' understanding of the precise role of primary vs "dark" trading venues will be useful in monitoring the changing nature of price formation dynamics in highly automated and fragmented markets. While the present report focuses on the FX market, this area of analysis could be relevant for other markets facing rapid electronification, such as fixed income markets. Hence, it could become an integral part of the ongoing efforts of central banks to adapt their market monitoring frameworks to increasingly fast and electronic financial markets.

New liquidity dynamics

The increasing use of EAs is contributing to the emergence of new liquidity dynamics in the FX market. Several market participants mentioned in interviews conducted by the study group that visible depth in public order books is lower today than it used to be. While this phenomenon is largely driven by the automation of FX trading, there are signs that EAs specifically are contributing to the thinning of the order book as they slice large orders into smaller ones and spread them over time.

However, as long as the order book is replenished fast enough, a thinner order book does not necessarily reduce market functioning. As velocity of trading increases with electronification, some market participants point out that liquidity replenishment, ie the rate at which new top-of-book liquidity is renewed, is what matters for a well functioning market. If market-makers know that liquidity takers demand liquidity in smaller amounts split over time, there is little need to provide a large amount of liquidity at any point in time, provided new liquidity is replenished at an adequate rate.[23] Historically, when trading was done manually, the market

[23] This development might be part of the evolution from a "fully automated stock exchange" (Black (1971)) towards "a fully continuous exchange" (Kyle and Lee (2017)). The latter propose a new order type: a "continuous scaled limit order". Such buy and sell orders represent continuous demand over

relied predominantly on the availability of large limit orders to buffer shocks in periods of stress. Therefore, a thinner order book would mean a lower ability of the market to absorb shocks, such as from unanticipated news or "fat finger" trades. Today, a thinner order book is not necessarily a sign of fragility, as long as the liquidity demanded by EAs can be matched, in small pieces, with liquidity provision at adequate replenishment rates.

Initial observations from the Covid-19 crisis support this conclusion. Most market participants interviewed by the study group stated that EAs were able to continuously execute transactions throughout this period of heightened price volatility, albeit at somewhat higher cost of execution (Box E). As market volatility was generally seen as high but not extreme, this recent episode should not be used to draw general conclusions regarding the robustness of order book dynamics and EA performance during shocks.

The use of EAs reinforces the changing pattern of liquidity dynamics. Liquidity providers will always require a premium to buy or sell large amounts instantaneously, because quoting large amounts at tight prices exposes them to adverse market developments and selection risk – the risk of being exploited by market participants with superior information or faster technology. The ability of EAs to split large orders into smaller pieces therefore provides an interesting alternative means of execution. Furthermore, the more EAs are being used – and the "thinner" the order book becomes – the more trades may need to be split into smaller pieces to reduce market impact, reinforcing the importance of EAs.

To adequately assess liquidity conditions in highly fragmented markets such as the FX market, novel liquidity indicators may be required to support market monitoring activities in fragmented and fast-paced markets. Immediately available and visible liquidity might be lower and more dispersed, but as long as algorithms are able to trade in a continuous way, market functioning is not necessarily impaired provided the order book can be replenished quickly. Traditional order book-based indicators may no longer be a good proxy to assess liquidity conditions. Instead, indicators that capture the rate at which liquidity is replenished may be more useful to the use of EAs. Provided the market relevance of EAs continues to grow, central banks could benefit from developing indicators that are specifically designed to account for the changes in the liquidity dynamics. Developing such indicators, however, may be costly and require novel technologies.

Overall, EAs appear to have a net positive impact on the FX market microstructure. In normal market conditions, the ability of EAs to generate quotes that reflect fresh information quickly, and to transmit this information simultaneously across multiple liquidity venues, improves the price discovery process and matching in an otherwise highly fragmented market.[24] Although visible liquidity is likely to be thinner due to the slicing of orders into small pieces, market resilience is not affected as long as the order book is replenished sufficiently fast. Over time, however, by reinforcing the trend towards smaller order sizes and internalisation, EAs could contribute to a reduction in visibility of depth and turnover on primary markets, which could, in the extreme, hamper price discovery and market functioning.

time rather than immediate purchases or sales. They can be considered as time-slicing algorithms taken to the extreme, ie continuously spreading the volume over time.

24 Chaboud et al (2014) suggest that algorithms overall, not exclusively execution algorithms, support price discovery through their increased ability to constantly update posted quotes.

3.2 Considerations for the effective use of FX execution algorithms

FX EAs endow market participants with more direct control over trade execution and can help them trade more effectively. However, with greater control also comes a greater responsibility for end users to understand, monitor and manage risks appropriately. First, users need to be aware of and actively manage the risks associated with the use of EAs and their inherent trade-offs. Second, they need to evaluate to what extent the superior execution quality compensates for the higher risk they bear. Third, and as a key prerequisite, they need access to relevant information and market data to understand, analyse and make effective use of FX EAs. All three aspects are central to ensuring an efficient redistribution of trading interests.

Managing execution risks and their trade-offs

Adequate controls are essential to mitigate operational risks of EAs. EAs allow embedding controls within the execution process. This is an improvement over manual trading, where the applicability and enforcement of controls are far more limited. However, risks may take a new form or a new order of magnitude due to the reliance of EAs on new technology, their higher information-processing speed and greater level of automation.

In contrast to trading at the risk transfer price, users executing via EAs carry market risk for an extended period of time. Furthermore, the use of EAs may alter users' trading habits, as automation enables trading over time zones outside their normal operating hours, which might introduce a new source of risk. In this context, users need to have adequate controls in place to address such risks (Section 1.7).

Users need to be aware of the trade-offs inherent in algorithmic execution and need to actively manage them. They face the trade-offs related to execution, as outlined by the Execution Algorithm Trilemma (Graph 7), and need to manage those themselves. Consequently, users' choices related to the style of execution need to weigh market risk against other potential costs such as the market impact or the risk of not fully completing a trade within some allocated time period.

Additionally, the resilience of the new distribution of risk is untested so far. The development of EA usage has in large part been concomitant with a global low volatility environment. As a result, there are concerns that, during a stress event, the market might not be able to withstand shocks as efficiently as in the past when banks were the ultimate risk bearers. Encouragingly, the most recent experience during the Covid-19 crisis has been positive in this regard (Box E). However, with the most recent bout in volatility being high but far from extreme, an ultimate test in the context of disorderly market conditions remains to be faced.

Evaluating execution quality

Evaluating the extent to which users are adequately compensated for the additional execution risk they bear is challenging – even if one had access to all existing market data. However, given the non-negligible market risk that users carry in the context of EAs, users should scrutinise thoroughly whether a particular EA does in fact deliver on its promise of superior execution quality in comparison with alternative means of execution and adjusted for the additional market risk incurred. In this context, performance analytics come into play.

Ideally, proper performance analytics cover the entire life cycle of an execution. Pre-trade and real-time analytics help inform decisions such as the selection of the execution benchmark, the choice of liquidity provider(s), and the choice of EA itself and its parameterisation. Post-trade analytics help evaluate execution performance and can, thereby, serve to validate pre-trade and at-trade decisions. Over time, these performance analytics tools can help improve the execution handling.

When assessing EA performance, it is key to consider both the explicit and the implicit costs of execution. Looking at explicit headline costs – eg bid-ask spreads and fees – alone would be incomplete and potentially misleading. Rather, implicit trading costs such as market impact and opportunity costs also need to be taken into account. In addition, the soundness of performance analytics crucially depends on the ability to access independent, external sources of data rather than relying on – potentially subjective – data and metrics offered by liquidity providers. Such data also come at a high cost, which needs to be included in the overall performance analysis.

Testing the performance and impact of EAs is technologically challenging. Because many factors can influence the impact of the algorithm, many executions in similar market conditions are necessary to obtain a statistically robust metric that disentangles the performance of the algorithm itself from the underlying market dynamics. The fact that EAs both react to and affect the market constitutes a simultaneous causality problem that further complicates inference. Meaningful analysis may only be possible in a dedicated test environment or through pooling of execution data across many market participants. While the latter is being tested by the industry, it may face practical obstacles such as confidentiality concerns. Central banks could explore the benefits of pooling expertise and resources to conduct relevant analyses on a dedicated and technologically fit-for-purpose platform to reduce costs while expanding knowledge.

Access to data and information

As outlined above, the effective use of FX EAs requires expertise as well as adequate information and data. However, access to market data varies and can be costly. High-quality data on the FX market are not accessible to all end users and can depend on users' ability to pay, their business model and their degree of sophistication. Some market participants are able to pay for access to faster and more granular data – thereby further widening information gaps in the market. Some market data, however, may not be available to certain market participants, because they are restricted, eg dealers that act as a price-maker for a certain minimum amount of volume per day. Finally, access to data is also dependent on the number of liquidity providers a user has access to – with the most sophisticated user groups typically enjoying the broadest menu of data available. Even for this study group, collecting robust metadata in the context of EAs constituted an unsurmountable challenge. As a result, some empirical analysis could not be conducted, as the study group was unable to obtain sufficiently broad, independent trade data where EA-induced transactions were identifiable.

Transparency of trading activity in the FX market is limited by virtue of the market remaining primarily bilateral and OTC in nature, fostering a relatively high degree of opaqueness. In particular, the absence of a common master record of data ("central tape") on transactional prices and volumes makes any objective evaluation of reference prices, benchmark prices or execution performance a difficult task.

Box E

FX EA performance during the March 2020 market volatility

FX market volatility increased significantly in March 2020 as the Covid-19 outbreak evolved into a pandemic, putting to the test the effective functioning of the FX market. This box briefly examines the use of EAs and their role in FX markets during this period. It draws on market intelligence and discussions with market participants, supported by quantitative data on FX market conditions and FX EA usage, where available.

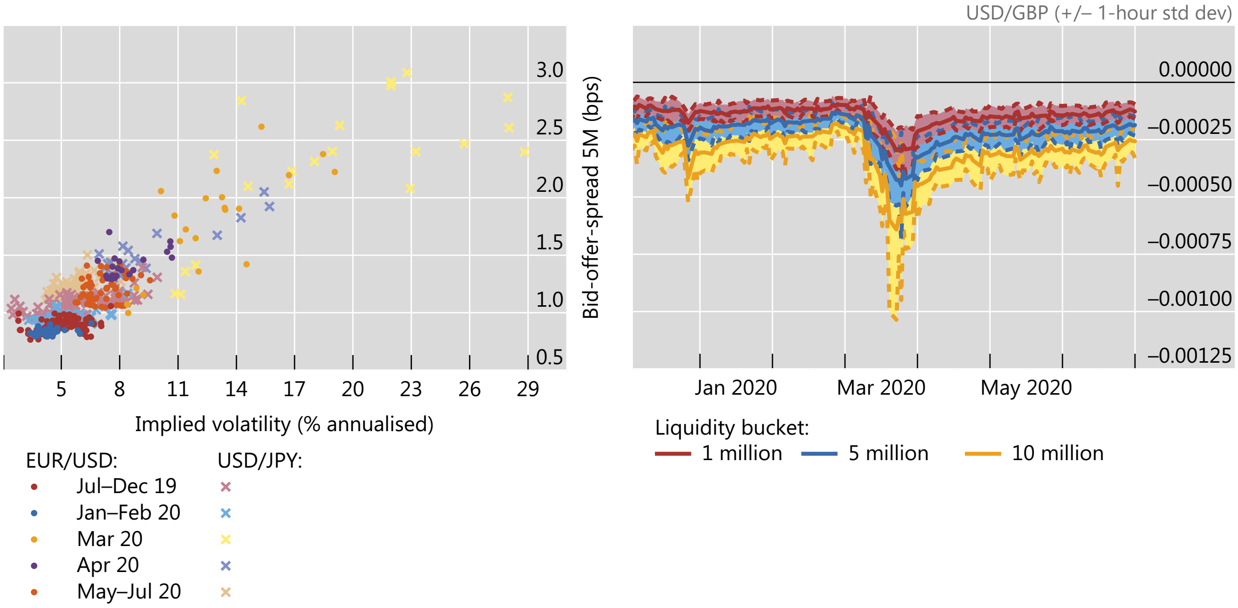
After several years of low FX market volatility, measures of volatility across most currencies more than quadrupled in March 2020 from earlier in the year (Graph E.1, left-hand panel). This heightened volatility was not as pronounced as in previous crises (eg the global financial crisis) or when compared with some fixed income and equity markets. However, bid-ask spreads deteriorated (right-hand panel), particularly for larger orders, and to a greater extent in some instances than in other asset classes (Dobrev and Meldrum (2020)). Meanwhile, volumes in FX markets rose significantly with relatively good two-way order flow, indicating that market functioning was not significantly impaired. Most providers reported more than a doubling of EA volumes relative to the average in usual conditions, approximately 30–50% more than the increase in FX turnover across all execution methods. The most significant driver of increased

Measures of FX liquidity conditions during Covid-19

Graph E.1

Spread vs volatility (EUR/USD and USD/JPY)

GBP/USD price dispersion between bid and mid-price

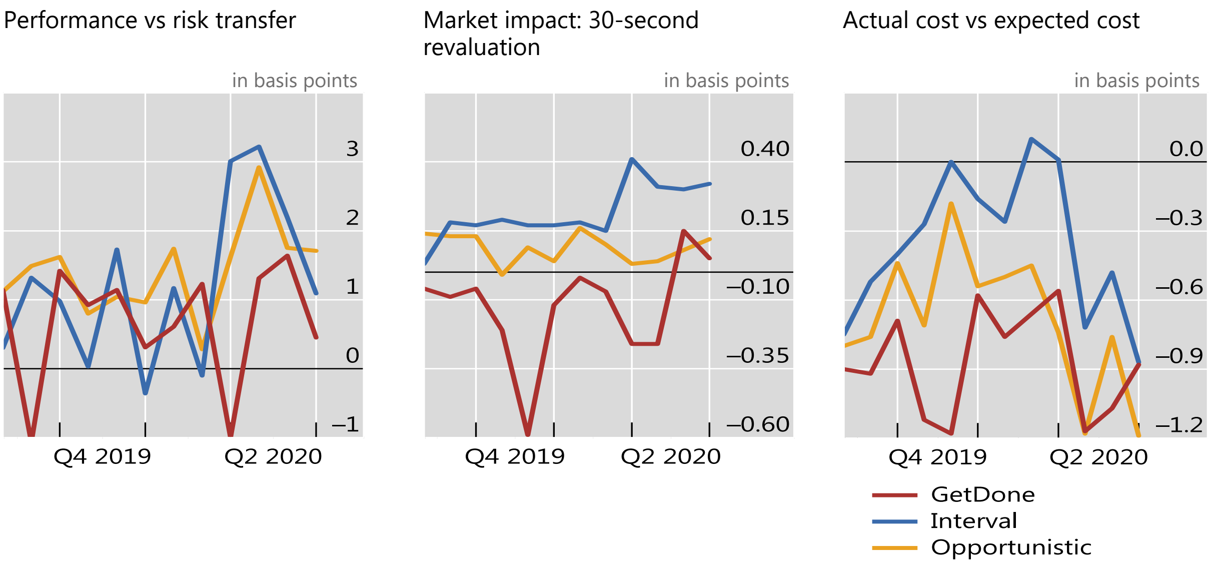


Sources: EBS; Refinitiv; study group calculations.

EA volumes was a relative deterioration in other methods of execution. In particular, risk transfer spreads widened significantly, especially for larger-sized orders, as market-makers were less inclined to take risk onto their balance sheets. Meanwhile, passive EAs allowed users to manage their orders over time at prices inside the bid-ask spreads. Moreover, passive child orders filled more quickly as FX volumes picked up. Consequently, FX EAs – particularly passive pegged/tracker (labelled as "Opportunistic") and time-sliced ("Interval") EAs – generally outperformed risk transfer during this period (Graph E.2, left-hand panel). They also had lower market impact compared with more aggressive limit-based/sweeping ("GetDone") EAs (centre panel). Work-from-home arrangements were cited as another possible contributor to increased EA volumes, as EAs readily facilitate record-keeping and audit trails for compliance purposes. However, this is likely to have been a less dominant driver, as EA volumes returned to more normal levels even as many traders continued to work from home in April and subsequent months.

Monthly performance by execution algorithm style

Graph E.2



Data include G10 and EME currency pairs. Left-hand panel: The chart compares the performance of limit-based/sweeping execution algorithms ("Get Done") and more passive pegged/tracker ("Opportunistic") and time-sliced ("Interval") execution algorithms with risk transfer pricing during the March–April 2020 bout of Covid-19 volatility. Higher positive numbers indicate relative outperformance. "Interval" EAs include some more sophisticated varieties and passive/neutral/aggressive TWAPs. Centre panel: Market impact is the volume-weighted average movement in the market 30 seconds after execution time. Higher positive numbers indicate lower market impact. Right-hand panel: Expected cost is the expected fair value cost of execution normalised for currency pair, size and market conditions. Higher negative numbers indicate better performance of actual vs expected costs.

Sources: BestX; study group calculations.

A further look into the use of FX EAs during this period reveals several unique insights. First, providers highlighted an increase in the use of more passive or hybrid algorithms relative to aggressive algorithms. Faced with the EA trilemma (Graph 7), users focused more on minimising market impact and spread costs (apex of the trilemma triangle), and surprisingly placed a lesser emphasis on market risk minimisation (bottom left-hand corner) despite the more volatile prices during this period. Increased usage of passive EAs may in turn have contributed positively to liquidity during the peak of this crisis by increasing the volume of passive orders sent to the market. Second, there was divergence across currencies, with turnover in the majors generally increasing significantly (more so for USD and GBP than for EUR and JPY), while the increase in turnover for Scandinavian and EME currencies was less significant. Finally, providers highlighted the outperformance of more dynamic algorithms relative to the earlier generation of EAs such as simple TWAPs that were less able to adapt to prevailing market conditions (Graph E.2, right-hand panel).[1] In fact, some of the dynamic algorithms also needed manual retuning in March, as they were not able to adapt quickly enough to the sudden change in conditions at the end of February.

Overall, FX EAs appear to have stood up well to the recent period of increased FX market volatility, and this could encourage increased adoption of EAs going forward. Indeed, while providers have attributed the higher volumes in March mainly to increased volumes from existing users, they have also noted more interest from new users, some of whom had started to test the use of EAs by placing orders indirectly through voice channels. Notwithstanding these positive outcomes for EAs, it is premature to conclude from this short episode that EAs are a panacea to execution in all market conditions. For instance, it is not yet clear what contributed to the more favourable market dynamics in this episode of high volatility, in comparison with, for example, recently observed flash events, which caused some electronic pricing and execution to stop. Further research in these areas is warranted.

① Dobrev and Meldrum (2020) also document a persistent wedge between volume-weighted and equally-weighted quoted spreads, consistent with the potential benefits from the use of more sophisticated algorithmic execution strategies capable of adapting in real time to rapidly changing liquidity conditions.

Disclosures related to EAs are typically high-level and non-standardised. Given the myriad of EAs on offer, making an informed decision about which one to choose requires detailed information on what an algorithm does and how it does it, ie its characteristics and decision logic. However, the extent to which such information can be provided is typically limited, as the inner workings of any EA form part of providers' intellectual property. In addition, disclosures surrounding controls embedded in EAs are typically vague and contractual liabilities in the event of EA malfunction are therefore often unclear. Disclosures on providers' choice of liquidity sources and related financial incentives for order routing decisions are typically scarce or non-existent. This information, however, is indispensable in order to detect potential conflicts of interest and biases in order routing that may run counter to the objective of improving execution. Disclosures are important in understanding how child orders are handled in different trading environments, such as those that allow for last-look or anonymous practices.

Efforts to address complexity and reduce opaqueness come at a cost. Market participants typically need to invest heavily to ensure that their staff have the appropriate knowledge and experience before using EAs. To help bridge the gap, liquidity providers offer extensive onboarding processes for their clients; are ready to provide ample advice before, during and after execution; and facilitate access to pre-trade, real-time and post-trade analytics for executed trades. It is important for end users to be able to assess the quality and effectiveness of their execution. To that end, third-party providers have entered the market to offer a more independent view on price and transaction cost analysis, and various initiatives based on proprietary and limited data sets enable EA users to share data among themselves. However, these services typically come at a non-negligible cost and hence may be available only to the most sophisticated users.

In sum, the bar for navigating today's FX market is high. Users who seek relevant information and data to understand and analyse the performance and risks of EAs continue to face great challenges. The uneven access to data and the presence of information asymmetry, the high degree of opaqueness related to transacted prices and volumes, and the lack of standardised disclosures constitute major hurdles. While some degree of information asymmetry may be warranted to ensure adequate incentives for liquidity providers to play their role as market-makers, concerns are – similarly to other markets – that certain industry practices may give a one-sided advantage to providers to the detriment of users.[25]

While some of these hurdles are inherent in the FX market, closing information gaps is central to ensuring a more level playing field. Enhancing transparency and encouraging common standards related to the necessary market expertise and disclosures is in everyone's interest. As addressing these issues requires the joint effort of the official and private sector, the GFXC's three-year review of the FX Global Code will include specific workstreams[26] to assess, among other things, where there is room for improvement in the area of algorithmic trading and disclosures. Beyond

[25] For instance, some market participants have highlighted that users who use simplistic time-weighted average price (TWAP) algorithms to execute against the 4 pm WM/R benchmark fixing can be systemically disadvantaged, even as "smart players are making money from the fix every day by trading the predictable moves" (*Financial Times* (2020)).

[26] https://www.globalfxc.org/events/20191204_summary_3_year_review_feedback.pdf.

this, the central banking community could evaluate benefits of enhancing the exchange of market-specific knowledge, such as through regular topical workshops.

There are several areas where it may be worthwhile to explore further the benefits and costs of greater transparency. First, a stocktaking analysis of what kind of pricing and volume data on the FX market are available could help market participants identify the data gaps they face. This could also shed light on the usefulness of a central tape to help facilitate an objective evaluation of the costs and benefits of FX execution. Second, a more uniform disclosure of the available liquidity pools, their characteristics and potential conflicts of interest could help market participants better assess the risks associated with various sources of liquidity. Third, a higher degree of standardisation with respect to the information on EA characteristics, their decision logic and embedded controls would facilitate comparability and selection of EAs across liquidity providers.

3.3 Market-wide implications of FX execution algorithm usage

This section considers how the use of EAs could amplify sharp price dislocations and cause a wider market impact. It first considers mechanisms by which EAs potentially reinforce or trigger flash events and feedback loops, and then discusses potential implications of the lack of coordinated regulation in the FX market and the importance of controls to limit these risks.

The precise role of EAs in flash events is hard to ascertain due to the lack of sufficient data. Theoretically, it is conceivable that an overly aggressive EA that exhausts the risk-bearing capacity of intermediaries could trigger a flash event. There is evidence that, among other factors, a large automated selling program played an important role in the 2010 flash crash in equity markets (Kirilenko et al (2017)). However, analysis of recent flash events in FX suggests that usually a range of factors, such as a concentration of stop-loss orders or the presence of less experienced staff outside a currency's core time zone, contribute to such events.[27]

Embedded controls, such as throttles that limit or switch off an EA altogether, should prevent EAs from amplifying sharp price moves. Market intelligence gathered by the study group on recent flash events suggests that EAs may have actually supported market functioning in recent flash events on the margin. Although most market-making algorithms stopped providing liquidity, some EAs continued to transact throughout these events at a steady pace, potentially mitigating price movements and even supporting price retracement after such an event. In fact, the reappearance of bids early in the market rebound of the sterling flash crash suggests the presence of EAs and their calming effects on liquidity conditions (BIS (2017)).

To assess the role of EAs during flash events, it is important to distinguish between static and more adaptive EAs. Both static algorithms (such as the time-sliced and historical volume-based types in Graph 7) and adaptive EAs (such as POV, pegged/tracker or implementation shortfall types) pose potential risks. A static behaviour is possibly no longer appropriate when market conditions change significantly. For example, a fixed order size and schedule may work well in normal conditions, but can be too aggressive in the context of worsening liquidity conditions.

[27] See eg Markets Committee (2017). For a discussion of the role of high-frequency trading during such events, see Markets Committee (2011).

An adaptive behaviour, on the other hand, can be advantageous, but also problematic when trading takes place in a new and untested market environment.

Adaptive-type EAs can create potential feedback loops that exacerbate the price movements. Feedback loops occur when a small change in the market environment loops back to the algorithm as an input, in turn triggering bigger changes. The possibility of such feedback loops needs to be examined closely by central banks, as they could cause price distortions and affect market functioning (US CFTC-SEC (2010)). EAs have the potential to reinforce and even generate feedback loops under specific conditions, eg outsize orders, crowded trades or periods of thin liquidity.

The risk of feedback loops is higher for algorithms that are more reactive towards market conditions. Within the EAs categorised in Graph 7, the algorithms that adjust the pace of execution to market turnover (POV) or try to minimise implementation shortfall (IV) are probably more prone to reinforcing feedback loops compared with passive (pegged/tracker) or deterministic algorithms (TWAP). For example, a POV algorithm selling during a flash crash potentially increases the speed of execution to match the increased market turnover, thereby increasing the selling pressure in the market. On the other side, an EA buying in a flash crash can help support prices during the price drop and eventually initiate a rebound. Hence, whether an EA exacerbates or stabilises price swings depends on its trading direction.[28]

Correlated EA orders could generate another potential source of risk to market functioning. A large number of similar algorithms, otherwise designed to minimise price impact, may end up collectively reinforcing market impact by acting as if placing a single large order. This could occur if a large number of orders caused by algorithms with similar logic were to respond in the same way to an external event. In line with findings on algorithmic trading in general, the more reactive the EA is to market conditions (eg second- and third-generation EAs), the more they may be prone to exhibiting correlated behaviour (Chaboud et al (2014)). Data from the study group's survey show that, as of today, the majority of volume is executed through rather simple, deterministic algorithms, with TWAP being the most popular. As their trading schedule is predetermined, they are not expected to increase correlation risks. But with a shift to more adaptive algorithms, correlation risks need to be monitored.

The recent Covid-19 period with elevated volatility can be seen as a test case for the functioning of EAs and potential market-wide implications. There have been no indications of unexpected behaviour of EAs or the presence of any feedback effects or other market-wide implications. Market participants noted that the rise in FX volatility was not large enough to trigger the embedded controls of the algorithms that are geared more towards flash-type events. Of course, this does not necessarily mean that the risks are not present. In stress periods in particular, the deployment of adequate controls such as those discussed in Section 1.7 are crucial to mitigate the risk of an EA entering self-reinforcing feedback loops and to contain market and operational risks for the users. Some market participants interviewed by the study group stressed the importance of throttles that limit the participation of EAs in stress periods to avoid their actions becoming self-referential.[29]

[28] The fact that an EA can be either buying or selling during a flash event differentiates the EA from momentum-seeking strategies that follow price trends. The latter by their nature either sell in a flash crash or buy in a rally, and thus have the greatest potential to exacerbate flash events.

[29] For an analysis of the role of algorithmic trading in general during the Covid-19 period, see Fukuma and Kadogawa (2020).

Kill switches and other types of circuit breakers are applied at the user level with the aim of preventing unintentional trading behaviour. If they are triggered simultaneously, however, they could have an adverse market-wide impact on liquidity and short-term price discovery. For instance, if kill switches (or other homogeneous volatility controls embedded in the algorithm logic) are activated for many EAs concurrently, this may trigger a sharp movement in price by withdrawing liquidity simultaneously.[30] In these instances, asymmetrically applied price controls that stop the EA when chasing an adverse price movement, but allow the EA to continue to transact if the price moves favourably, could have positive implications for matching and price discovery.

Due to the lack of coordinated circuit breakers across venues, EAs can potentially cause spillovers of price events by routing liquidity between them. The significance of this risk is difficult to assess, but recent empirical studies on circuit breakers in equity markets suggest that liquidity moves to primary venues during stress (Guillaumie et al (2020)) just like it does in the FX market (Moore et al (2016)). Given the considerable fragmentation, the OTC nature of the FX market and its linkages to other markets, further work is necessary to understand the impact of kill switches and circuit breakers on liquidity resilience and market functioning.

Recent episodes of flash events in the FX market have raised concern that automation makes markets more prone to sudden bursts of volatility. Users of EAs need to be aware of these risks and able to manage them through embedded controls and monitoring. This will substantially reduce systemic risks. Increasing transparency and building the right knowledge of execution market dynamics and its interaction with other algorithms, particularly during extreme events, could form part of the monitoring frameworks of central banks and standard setters. Systemic risks such as feedback loops are complex to assess, both due to the multiple interactions between EAs and other algorithms or trading venues, and owing to the challenges in sourcing the required empirical data. Further collaboration among central banks and empirical analysis/information-sharing will enable a better understanding of the nature of these risks, particularly in OTC markets such as the FX or fixed income markets.

4. Conclusion

Execution algorithms have become a well established execution channel in the FX market in recent years, which was mostly characterised by subdued volatility. And while many observers and market participants expected them to be less widely used during a crisis, it appears that the opposite has happened: EAs look to have retained an important role during the turbulence caused by the Covid-19 pandemic. And they are likely to be here to stay.

This report discusses the proliferation of EAs and aspects pertaining to users and providers. It analyses the typology of EAs and the execution trilemma, and assesses the risks and benefits of EAs from a market functioning perspective. The **key takeaways** from this analysis are as follows:

30 It is possible that a highly uniform regulatory, accounting and supervisory landscape could lead to more synchronised actions during stress events, potentially causing feedback loops (Edmans et al (2015)).

EAs are a response to the rapidly evolving and fragmented landscape of the FX market. Growing adoption of EAs has taken place amid broad structural developments such as the increasing electronification of financial markets, the continually growing fragmentation of the FX market across a wide array of electronic venues, the emergence of new actors and a changing regulatory landscape. By slicing large orders into smaller ones and by distributing orders efficiently across different liquidity pools and trading venues, EAs help to optimise trade execution in this environment. Furthermore, the drive to automate processes and the effort to improve audit trails are other important factors that are contributing to the increased usage of EAs.

The use of EAs today is widespread but not dominant in the FX market. Based on our latest estimate, approximately 10–20% of FX spot trading volume is executed via EAs, with adoption greater in G10 currencies than in EME currencies. The group of EA users and their trading objectives are almost as diverse as the types of EAs, liquidity pools and trading platforms on offer. EAs have evolved in sophistication from the very simplistic approaches introduced in the late 1990s to the latest generation with self-learning capabilities. For the moment, the more simplistic algorithms remain the most widely used ones.

Several competing execution objectives lead to an inherent "execution algorithm trilemma". The concept of "best execution" comprises several competing dimensions particular to each user's case, balancing the trade-off between market risk, market impact and opportunity cost. While various types of EAs can accommodate different objectives, no single execution algorithm can optimise all aspects simultaneously. This is what gives rise to what we call the execution algorithm trilemma. Different types of EAs respond differently to the trade-offs and are situated accordingly within the trilemma.

This report approached the issues for market functioning posed by FX EAs from **three main perspectives**:

Market microstructure perspective

EAs appear to support price discovery and hence market functioning. EAs facilitate price discovery by slicing large orders and routing these smaller orders to fragmented liquidity pools efficiently. Anecdotal evidence suggests that this positive impact persisted during the market turbulence witnessed during the outbreak of Covid-19. In particular, adaptive EAs reportedly coped well with increased volatility and wider bid-ask spreads, and performed favourably relative to other means of execution. That said, more extreme market conditions may still reveal deficiencies in EAs. Robustness of algorithms and appropriate controls remain key considerations, which require ongoing scrutiny and evaluation by market participants as well as regulators and standard setters.

EAs have also accelerated the trend towards internalisation. The spreading of a large order over time can facilitate internalisation, as dealers no longer have to keep large trades in inventory to bridge the time mismatch between customer orders, but can more easily match the customer flow. While the topic of internalisation is much broader, EAs may at least indirectly contribute towards taking trades into the dark. A particular trend in recent years was the reduction in the proportion of traded volumes on primary venues, traditionally a key source of price discovery. Deepening central banks' understanding of the precise role of primary vs secondary and "lit" vs "dark" trading venues could contribute substantially to understanding the changing nature of price formation dynamics in highly automated and fragmented markets.

Novel liquidity indicators could support efforts to effectively monitor FX and other fast-paced electronic markets. Developing indicators designed to account for the identified changes in liquidity dynamics is likely to be costly and require technology-intensive tools and expertise. Hence, it could become an integral part of the ongoing efforts of central banks to adapt their market monitoring frameworks to increasingly fast and electronic financial markets.

Market participants' perspective

EAs shift the execution risk from the dealers to the users, which implies new challenges for users. EAs endow market participants with more direct control over trade execution by providing access to means of execution that had previously been available only to market-makers. However, with greater control also come greater challenges. First, participants need to actively manage and address related risks. Second, they need to define their own benchmarks and evaluate whether they are compensated for the shift in risk by superior execution quality. Third, and as a key prerequisite, users need access to relevant information and market data to understand, analyse and make effective use of algorithmic execution. In order to meet these challenges, users need to be well equipped in terms of access to market data and relevant information to ensure an efficient and cost-effective redistribution of their trading interests. The emergence of independent TCA providers has helped to address some of these issues, but does come at additional cost. Central banks could explore the benefits of pooling expertise and resources to conduct relevant analyses on a dedicated and technologically fit-for-purpose platform to reduce costs while expanding knowledge.

The bar for navigating today's FX market is high. Information in the highly fragmented FX market is unevenly distributed. Closing information gaps is therefore central to ensuring a more level playing field for all. Enhancing transparency and ensuring common standards related to training and disclosures are consistent with broad market fairness and effectiveness. Areas of further exploration include the identification of data gaps that market participants face, the need for more uniform disclosures related to liquidity pools and potential conflicts of interest, and a potentially higher degree of standardisation with respect to the information on EA characteristics. This could help market participants to better assess the strength of their pre-trade, real-time and post-trade capabilities and would also shed light on the usefulness of a common master record of data ("central tape"). Beyond this, the central banking community could evaluate benefits of enhancing the exchange of market-specific knowledge, such as through regular topical workshops.

Market-wide perspective

Controls are key in containing the operational risk. The onus is on each EA provider and user to have adequate safeguards in place that prevent unintentional trading behaviour from materialising, especially as no market-wide circuit breakers or kill switches exist in the FX market. Operational risks are either algorithm malfunctions or human errors, such as overly aggressive parameterisation of an EA. Embedded controls as well as education and constant monitoring are key to ultimately reducing systemic risk. This area needs to be kept in focus by providers, users, regulators and central banks.

Central banks can draw lessons from these insights for their market monitoring and FX trading activities, including for reserve management and monetary policy implementation. As addressing some of the identified issues requires the joint effort of the official and private sectors, the GFXC's three-year review of the FX Global Code has established a workstream to look at aspects of algorithmic trading. A second workstream will focus on disclosures, a very relevant area for EAs given the market-wide benefits of appropriate transparency and disclosure in the context of EAs.

While this report focuses on the FX market, many financial markets have undergone or are undergoing a similar transformation to become increasingly fast and electronic. While each market has its idiosyncrasies, findings of this report may shed some light on important aspects central banks and other market participants should consider when monitoring or actively engaging in fast-paced markets.

Glossary

Algorithm: Broadly refers to a step-by-step procedure used for calculation or analysis. A wide range of computer programs – not limited to automated trading systems – are often made up of many algorithmic steps, often shared across multiple programs within the same organisation. An algorithm used within an automated trading system defines a set of instructions on when and how to submit, revise or cancel an order.

Algorithm wheel: Third-party algorithms used to allocate transactions between various execution algorithms.

API: *Application Programming Interface*, an interface allowing the connectivity between several data feeds or trading tools. In the context of FX trading, API often refers to an interface that enables software used by counterparties to connect in order to obtain real-time pricing data or place trades.

Automated trading system: A computer program that defines decision rules to generate, submit, monitor and revise orders continuously.

Child order: Each of the slots of small FX transactions in which parent orders can be sliced in the execution of a transaction.

Circuit breaker: A type of trading curb where trading is halted for some period of time if the market for an asset moves more than a predefined trigger.

CLOB: *Central limit order book*, a trading protocol in which outstanding offers to buy or sell are stored in a queue and are filled in a priority sequence, usually by price and time of entry. Orders to buy at prices higher than the best selling price and orders to sell at prices lower than the best buying price are executed. CLOBs are common for highly standardised securities and markets in which trade sizes can be small.

Consolidated tape: An electronic system which combines traded volume and price data from various trading venues into a continuous live feed to show a more representative picture of the market.

Dark pool: A private venue that provides for anonymous trading and that does not display the order book to market participants.

ECN: *Electronic communication network*, a system that electronically matches buy and sell orders.

Execution algorithm: An automated trading program designed to buy or sell a predefined amount according to a set of parameters and instructions with the objective of filling the order.

Fat finger: Describes a type of trading error caused by mistyping on a computer keyboard. The term has come to capture more generally any trading error caused by simple human error.

Fill ratio: The traded amount as a share of the submitted amount.

Flash event: Refers to a rapid, deep and volatile move in an asset price usually followed by a quick recovery.

Futures: Standardised, exchange-traded derivative contracts for a pre-agreed quantity and quality of a specified asset for a price agreed today, with delivery and payment occurring at a specified date in the future (delivery date).

Iceberg order: An order where only a fraction of the entire order may be visible to other market participants. As the disclosed portion is filled, the subsequent portions are sent to the market until the order is filled.

Implementation shortfall: The difference between the average price the strategy achieves when completely filling an order compared with a benchmark rate.

Internalisation: A process whereby dealers offset risk (open positions) arising from client transactions against risk (open positions) arising from transactions with other clients.

Kill switch: A functionality in trading software designed to instantly disable all trading activity for a particular participant or group of participants, cancelling all working orders and preventing the ability to enter new orders. Some kill switches may also allow risk-reducing orders while preventing risk-increasing orders.

Last look: A practice whereby a market participant receiving a trade request has a final opportunity to accept or reject the request against its quoted price.

Latent liquidity: Refers to actual liquidity on a trading venue that may not be made visible to market participants by actors who opt to hide the actual size of their trading interest, creating a perception of lower liquidity than that actually available.

Limit order: An order to buy a specified quantity up to a maximum price, or sell subject to a minimum price.

Liquidity: The degree to which an asset or security can be quickly bought or sold in the market at a price reflecting its intrinsic value.

Liquidity pool: Intersections of orders, being the ranges with which it is possible find a lot of long, short, take-profit and stop-loss orders.

Lit venue: A trading venue where the order book is visible to all participants.

Manual trading: A method of trading that involves human decision-making and action to submit orders into the market.

Market depth: Sum of the amount of outstanding orders pending (possibly at different prices) on either side of the order book.

Market impact: The difference between the price observed just before a transaction and the actual execution rate. Also known as *footprint*.

Market-maker: A dealer obliged to quote buy and sell prices in return for certain privileges on a trading platform or an exchange.

Market risk: The risk to a market participant's position emanating from changes in FX prices or rates.

Matching engine: The matching engine refers to the allocation algorithm embedded in an exchange's computers to match marketable buy and sell orders within the central limit order book and convert them into executed trades.

NDF: *Non-deliverable forward*, a forward contract that does not involve an actual exchange of currencies. Instead, it entails a cash settlement of the difference between the actual and a pre-agreed exchange rate in a single payment at maturity.

Order book: A continuously updated list of bid and ask orders.

OTC: *Over-the-counter*, referring to bilateral transactions not conducted on a formal exchange.

Parameterisation: A process of adjusting the configuration of an algorithm to change its behaviour in executing orders.

Parent order: An order which can be sliced by dividing it into smaller lots (known as child orders) in execution of a transaction. A parent order is typically used in many algorithmic trading strategies to structure a trade. Also referred to as *meta-order*.

Pegged order: An order to the bid or ask with, or without, an offset. The display quantity will float with the bid or ask, up to the ultimate limit price of the order.

Post-trade credit control: Limit set by a broker to manage its financial exposure to its customers through the different types of market activity in which they participate.

POV: *Percent-of-volume*, referring to targeting a level of participation in markets on the basis of turnover indicators.

Pre-trade risk control: A controls used to prevent inadvertent market activity due to unauthorised access, system failures and errors. On an exchange, a pre-trade risk control can set limits on the size of an order submitted to the exchange's matching engine.

Price/time priority allocation: An exchange matching engine algorithm that fills buy and sell orders according to price and time priority, also known as "first-in-first-out" (FIFO). An incoming order's quantity immediately matches against each resting order at the same price within the central limit order book queue, decrementing each resting order based on its position within the queue. Resting orders at the same price level are given matching priority based on the time they arrive at the exchange, with the oldest order having the highest priority.

Price tolerance limit: The maximum amount an individual order's limit price may deviate from a reference price such as the product's current market price; is typically applied on orders generated from an automated trading system before the order is sent to the exchange.

Primary venue (PV): A classical exchange for settling trades in a transparent manner. For spot FX, primary venues traditionally include electronic communication networks such as EBS and Refinitiv Spot Matching.

Principal trading firm (PTF): A firm that invests, hedges or speculates for its own account. This category may include specialised high-frequency trading firms as well as electronic non-bank market-making firms. Sometimes also referred to as a *proprietary trading firm*.

Resting order: An order that has been submitted to the exchange but has not yet been executed. Resting orders are placed using a limit price and are said to be passive since they do not trade immediately and will only trade when another participant agrees to trade at their price level.

RFQ: *Request for quote*, a query issued by a trading platform member to another member to request price quotations. Systems for sending RFQs vary according to: whether the sign of the order (buy or sell) is revealed; how many participants and what kind of participants may receive an RFQ; and whether the quotes are executable or indicative. A related trading protocol is *request for market* (RFM). RFM refers to a request for quote where the client does not reveal the sign of the desired trade (buy or sell). An RFM is a request to see a two-sided or "market" quote rather than a one-sided quote.

RFS: *Request for stream*, a query in which market-makers provide continuous streams of firm quotes with available size, and the client receiving the quotes can click to trade.

Risk transfer: A request by the client to trade the full size of a transaction with the liquidity provider, allowing the client to swiftly transfer the market risk of the trade to the liquidity provider.

Smart order routing (SOR): A type of algorithm aimed at rapidly executing smaller orders by simultaneously routing to numerous liquidity venues.

Top-of-book (TOB): The best bid and ask prices in a given security. The difference between the highest bid and the lowest ask is the top-of-book bid-ask spread.

Transaction cost analysis (TCA): A tool that allows end users to monitor the efficiency of their transaction both pre- and post-trade.

TWAP: *Time-weighted average price*, a trading benchmark used by traders that gives the average price of a security over a specified time.

Volatility: A measure of the fluctuation in the market price of a product over time.

VWAP: *Volume-weighted average price*, a trading benchmark used by traders that gives the average price a security has traded at throughout a given period weighted by the volume traded.

References

Bank for International Settlements (BIS) (2019): *Triennial Central Bank Survey of foreign exchange and OTC derivatives markets*, reporting guidelines for turnover in April 2019.

——— (2017): "The sterling flash event of 7 October 2016", Markets Committee, January.

——— (2010): "*Triennial Central Bank Survey of foreign exchange and OTC derivatives market Activity in 2010,* Final Results

Black, F (1971): "Toward a fully automated stock exchange", *Financial Analysts Journal,* vol 27, no 4, pp 28–35 and 44.

Bouchaud, J-P, J Bonart, J Donier and M Gould (2018): *Trades, quotes and prices: financial markets under the microscope*, Cambridge University Press.

Capponi, F, R Cont and A Sani (2019): "Trade duration, volatility and market impact", *Volatility and Market Impact*, 13 March.

Chaboud, A, B Chiquoine, E Hjalmarsson and C Vega (2014): "Rise of the machines: algorithmic trading in the foreign exchange market", *Journal of Finance,* vol 69, pp 2045–84.

Collins, B and F Fabozzi (1991): "A methodology for measuring transaction costs", *Financial Analysts Journal*, vol 47, no 2, March/April, pp 27–36.

Debelle, G (2018): "Lessons and questions from the global financial crisis", address to the Business Economist Annual Dinner, 6 December.

Dobrev, D and A Meldrum (2020): "What do quoted spreads tell us about machine trading at times of market stress? Evidence from Treasury and FX markets during the COVID-19-related market turmoil in March 2020", Board of Governors of the Federal Reserve System, *FEDS Notes*, 25 September, https://doi.org/10.17016/2380-7172.2748.

Dobrev, D and E Schaumburg (2015): "The Liquidity Mirage", Federal Reserve Bank of New York, *Liberty Street Economics Blog*, 9 October, https://libertystreeteconomics.newyorkfed.org/2015/10/the-liquidity-mirage-.html.

Edmans, A, I Goldstein and W Jiang (2015): "Feedback effects, asymmetric trading, and the limits to arbitrage", *American Economic Review*, vol 105, no 12, pp 3766–97.

Financial Times (2020): "Fund managers face scrutiny over use of FX benchmarks", 3 July, https://www.ft.com/content/e93798dd-cde5-4f98-9348-545e99b6c984.

Fukuma, N and Y Kadogawa (2020): "An overview of algorithmic trading in foreign exchange markets and its impacts on market liquidity", *Bank of Japan Review*, 2020-E-5.

Global Foreign Exchange Committee (GFXC) (2020): "GFXC 2019 Survey Results", January

Guillaumie, C, G Loiacono, S Kern and C Winkler (2020): "Market impacts of circuit breakers – evidence from EU trading venues", *ESMA Working Paper*, no 1.

Kirilenko, A, A Kyle, M Samadi and T Tuzun (2017): "The flash crash: high-frequency trading in an electronic market", *The Journal of Finance*, vol 72, no 3, pp 967–98.

Kissell, R (2006): "The expanded implementation shortfall: understanding transaction cost components", *The Journal of Trading*, vol 1, no 3, pp 6–16.

——— (2014): "Algorithmic transaction cost analysis", in *The science of algorithmic trading and portfolio management*, Academic Press, pp 87–128.

Kyle, A and J Lee (2017): "Toward a fully continuous exchange", *Oxford Review of Economic Policy*, vol 33, no 4, pp 650–75.

Markets Committee (2011): "High-frequency trading in the foreign exchange market", *Markets Committee Papers*, no 5, September.

——— (2017): "The sterling 'flash event' of 7 October 2016", *Markets Committee Papers*, no 9, January.

——— (2018): "Monitoring of fast-paced electronic markets", *Markets Committee Papers*, no 10, September.

——— (2019): "Large central bank balance sheets and market functioning", *Markets Committee Papers*, no 11, October.

Moore, M, A Schrimpf and V Sushko (2016): "Downsized FX markets: causes and implications", *BIS Quarterly Review*, December.

Oomen, R (2019): "Price signatures", *Quantitative Finance*, vol 19, no 5, pp 733–61.

Perold, A (1988): "The implementation shortfall: paper versus reality", *The Journal of Portfolio Management*, vol 14, no 3, pp 4–9.

Risk.net (2017): "Why Robert Almgren no longer trades using Almgren-Chriss", 3 July, https://www.risk.net/asset-management/5298146/why-robert-almgren-no-longer-trades-using-almgren-chriss.

Schrimpf, A and V Sushko (2019a): "Sizing up global foreign exchange markets", *BIS Quarterly Review*, December.

——— (2019b): "NDF markets thrive on the back of electronification", *BIS Quarterly Review*, December.

Sinclair, J (2018): "Why does fragmentation continue to increase? Increasing entropy in currency markets", *MarketFactory Whitepaper*, February.

US Commodity Futures Trading Commission and US Securities & Exchange Commission (US CFTC-SEC) (2010): "Findings regarding the market events of May 6, 2010", *Joint Staff Report*.

Wagner, W and H Edwards (1993): "Best execution", *Financial Analyst Journal*, vol 49, no 1, pp 65–71.

Annex – Illustrative examples of execution algorithms

This annex describes the typical execution schedules for the six archetypes of execution algorithms discussed in Section 2. All graphs and data presented in this annex are illustrative, and may not correspond with the execution profile for all providers.

Time-sliced/time-weighted average price

Overview. Time-weighted average price (TWAP) algorithms slice up large orders into smaller child orders and space these out evenly over time, with the intention of reducing market impact.

Illustrative execution schedule. To illustrate, a trader may place an order to sell USD 60 million USD/JPY through a TWAP algorithm with a duration set at 60 minutes. Based on this input, the broad schedule of execution would work out to approximately USD1 million every 60 seconds (Graph 4).

- Current versions of TWAP algorithms typically provide for some randomisation in the timing of execution (ie will not attempt to execute exactly every 60 seconds in this case), to reduce the predictability and signalling from orders.
- The algorithms would also typically have some flexibility to opportunistically diverge from the broad execution schedule, to obtain a better transacted price. Users typically can specify how strictly the algorithm needs to keep to this execution schedule. If the user specifies a higher level of flexibility, the profile of execution could deviate more from the linear profile of the schedule, although the execution will complete in any case by the specified end time.

At each time step (eg once per second), the algorithm will compare its progress against the execution schedule:

- If it is ahead of the schedule, the algorithm could decide not to do anything for the current time step. It could transact slightly ahead of schedule if it has been given a higher degree of flexibility.
- If it is in line with the schedule, the algorithm could post small sell orders passively at the best offer price or slightly higher.
- If it is behind schedule, the algorithm could take more aggressive actions, including placing an offer between the current market best bid and offer or, where the execution is far-lagging, aggressively transacting against the current best bid price in the market. The algorithm will shift more quickly to aggressive orders if it has been accorded lower flexibility to deviate from the TWAP schedule.
- To ensure that the full size is completed within the given 60 minutes, the algorithm will also transact against the best bid price as necessary, with little price discretion.

Safety considerations. While TWAP orders are generally meant to transact quite gradually, they can have substantial price impact if too large an order is to be completed in too short a space of time. That being so, some providers may set caps at the point of order entry on the allowable pace of transaction, which may differ by currency pairs.

Variations. Other algorithm providers may allow users to set price or participation limits (explained further in the POV section below) on the execution, but these will no longer strictly be TWAP algorithms, as the limits could halt or pause the transaction mid-way during execution, resulting in a final price that differs substantially from the TWAP.

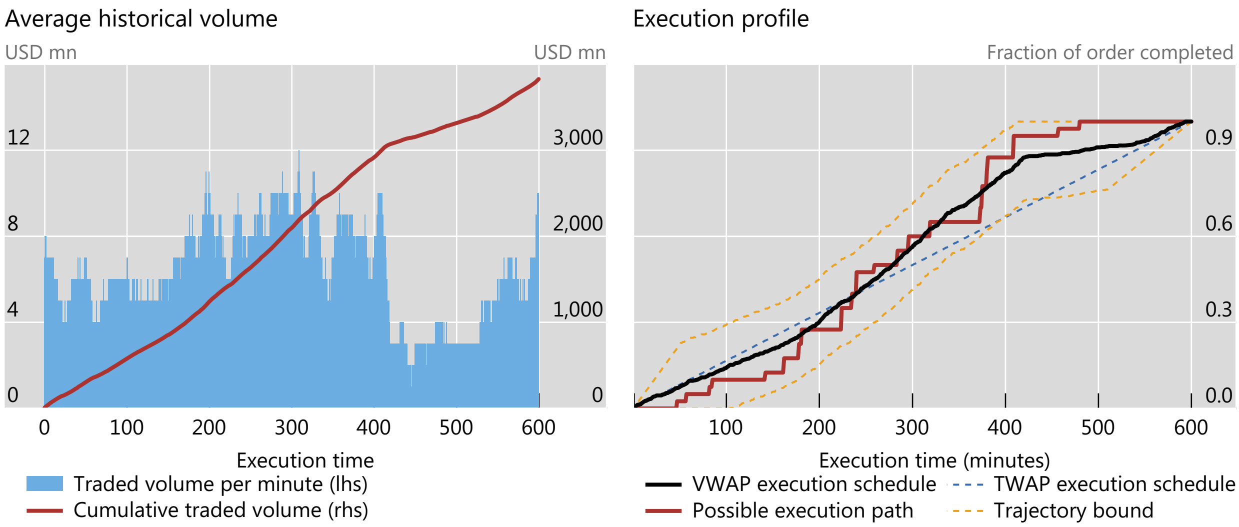
Historical volume-based/volume-weighted average price

Overview. Volume-weighted average price (VWAP) algorithms seek to reduce market impact by tailoring the trading intensity to the expected market turnover based on historical transaction volumes. This is typically useful when large orders need to be executed over a substantial portion of the day such that the intraday profile of market liquidity would matter.

Illustrative execution schedule. To illustrate, a trader may place an order to sell USD 600 million USD/JPY through a VWAP algorithm, with the duration set at 10 hours (600 minutes):

- The algorithm will determine its broad execution schedule at the start of the transaction, based on the specified duration and the profile of historical transaction volumes (Annex Graph 1) at a similar period of the day.
- As opposed to the linear execution schedule of the TWAP algorithm, it will seek to transact more during periods where high trading volume is expected, and transact less or not at all when trading activity tends to be low. For instance, as the major currency pairs tend to be deeper at the overlap of Asia and Europe trading hours and at the overlap of Europe and US trading hours, a VWAP algorithm would tend to transact a larger portion of its order at these times compared with less liquid hours.
- Depending on the level of sophistication, algorithms may also be able to take into account various factors that would affect trading volumes and liquidity patterns (eg seasonality across days of the week and months of the year, and releases of key economic data such as US non-farm payroll).
- As with the TWAPs, users of VWAP algorithms typically have the option to specify how strictly the algorithm needs to keep to this execution schedule. The VWAP algorithm is also a scheduler and will aim to complete the order by the specified end time.

Illustrative example of a VWAP algorithm — Annex Graph 1



Source: Markets Committee study group calculations.

Illustrative example of a POV algorithm

Annex Graph 2

Source: Markets Committee study group calculations.

Safety considerations. Similar to TWAP algorithms, VWAP algorithms can have substantial price impact if too large an order is to be completed in too short a space of time. That being so, some algorithm providers may set caps at the point of order entry on the allowable pace of transaction.

Variations. Some VWAP algorithms may seek to improve the execution by taking into account real-time estimates of actual trading volumes. For instance, the algorithm could accelerate its pace of transaction if actual volumes were deeper than expected. Price or participation limits (explained further in the POV section below) may also be set but will mean that the algorithm is no longer strictly executed like a time-scheduled VWAP when these limits are implemented.

Percent-of-volume

Overview. Similar to VWAP algorithms, percent-of-volume (POV) algorithms aim to achieve a level of participation of the market (eg 20% of the traded volume). However, instead of historical traded volume like VWAP, POV uses real-time indicators of liquidity (eg recent transaction volumes or estimates of, and depth of order books). POV algorithms typically are not constrained by the need to complete the order by a predefined time.

Illustrative execution schedule. To illustrate, when a trader places an order to sell USD 600 million USD/JPY through a POV algorithm:

- Users typically specify the level of urgency or aggression, and this is translated by the algorithm into a level of participation in the market (eg 10%, 15% or 20%), taking into account other factors such as the size of the order and market volatility.
- POV algorithms do not follow a predefined schedule, but will adjust its level of activity in line with real-time estimates of available liquidity. This assessment and determination will repeat itself until the order is completed. In the example shown in Annex Graph 2, the POV algorithm transacts at a much slower pace than the VWAP algorithm, as the actual liquidity is much worse than what the historical volume would predict.

- As the POV algorithm is also more opportunistic than the historical VWAP algorithm, the pace of execution varies significantly across the lifespan of the execution.

Safety considerations. POV algorithms are typically used for passive executions to maximise spread capture, and will typically post interest on the passive side of the market. However, their intrinsic execution logic is to adjust order throughput according to market turnover. This has the potential, especially with growing use of FX EAs, to self-reinforce feedback loops (eg 15 concurrently running POV algorithms each targeting 10% market participation). In fact, POV EAs with a target participation of only 9% were a key driver of the 4 May 2010 equity flash crash.[31]

Pegged/tracker

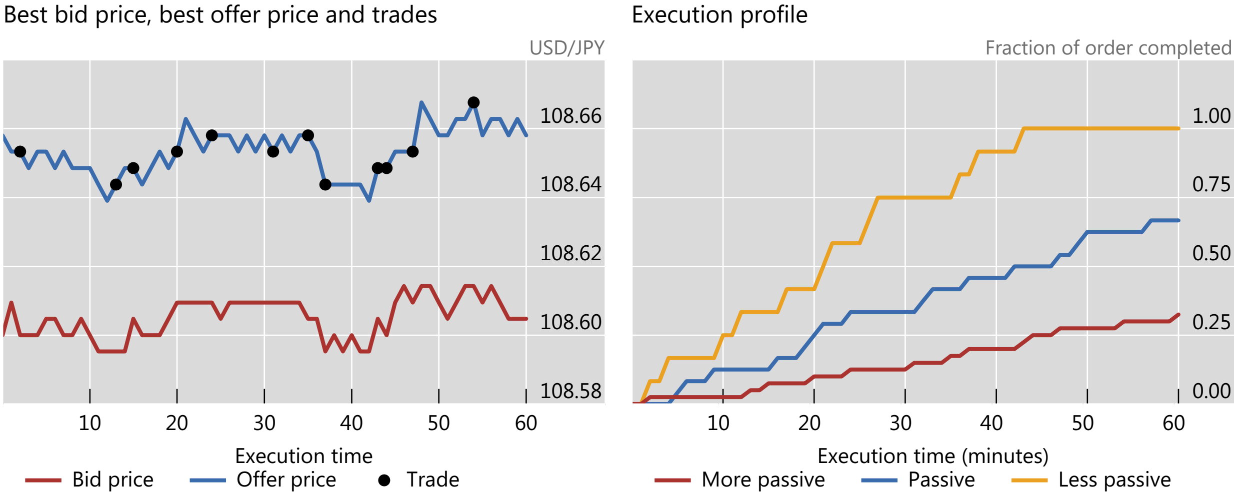
Overview. Pegged or tracking algorithms employ passive execution strategies that are aimed mainly at executing on the passive side of the bid-ask spread, with the intent of minimising spread cost. Such algorithms generally cannot guarantee the completion of an order within the time allocated.

Illustrative execution schedule. To illustrate, when a trader places an order to sell USD 60 million USD/JPY through a pegged algorithm:

- The algorithm will passively place limit orders of small amounts (eg USD 2 million) at or near the best offer price (Annex Graph 3).
- The algorithm will replenish the limit orders when initial orders are transacted, until the entire order (USD 60 million) is completed.
- The algorithm will cancel and resubmit its orders at or near the best offer price if the market price moves away (ie lower in this example).

Illustrative example of a pegged/tracker algorithm

Annex Graph 3



Source: Markets Committee study group calculations.

[31] US CFTC-SEC (2010).

Users can typically specify how passively the algorithm should transact:

- A less passive execution could allow for the replenishment of sell orders slightly inside of (ie sell slightly lower than) the best offer price, and more frequent resubmission of orders when market prices adjust.
- A more passive execution could involve replenishment of orders strictly behind (ie sell slightly higher than) the best offer price, and possibly with restrictions on the amount of time the resting order can remain on the top of the book after prices shift towards it.
- The user's settings will also help determine the size of child orders, taking into account the overall transaction size and depth of market – from historical transaction volumes (like volume-weighted algorithms) or real-time transaction volumes (like POV) in the currency pair.

Safety considerations. Pegged algorithms are exposed to substantial market risk. Pegged algorithms can run into adverse selection problems, which over time can lead to unfavourable trading patterns. For instance, the orders would tend to be filled when least desired (eg sell orders filled just as prices start to rise). Tracking algorithms are also exposed to substantial market risk.

Implementation shortfall

Overview. Implementation shortfall (IS) algorithms seek to minimise slippage relative to arrival price, typically measured as the primary market mid-price at the time of receipt of the order. Execution cost is determined ultimately by a trade-off between market impact (from a faster pace of execution) and market risk (from a slower pace of execution). IS algorithms often rely on proprietary statistical models, including market impact and market pricing models, to determine the optimal execution schedule.

Illustrative execution schedule. To illustrate, when a trader places an order to sell USD 60 million USD/JPY through an IS algorithm:

- The execution schedule is determined by the IS algorithm's statistical models, taking into account the order details (currency/size/direction) and current market conditions (eg available liquidity, price volatility and trend). This could be dynamically reviewed over the lifespan of the transaction, by taking stock of order completion and changes to market conditions (Annex Graph 4).
- IS algorithms do not follow a predetermined execution schedule, and thereafter do not have a fixed time to complete the orders. However, in practice, most IS algorithms would tend to execute more quickly during the early stage of the transaction before prices move significantly, then slow down towards the tail end of the transaction.
- Most IS algorithms allow the user to specify a level of urgency/aggression for the transaction. While it may be argued that transaction costs are optimised to a neutral setting, a user may want to specify a higher level of urgency if there is a sense that the market is trending or may move unfavourably away from the current price. The user may also want to increase the level of urgency to increase the probability of completing the order within a set amount of time.

Safety considerations. IS algorithms tend to be more aggressive, usually executing with a higher rate of transaction at the start. IS algorithms would also tend to increase their pace of transaction in response to an exogenous spike in price volatility, or when prices start to shift away from it. However, IS algorithms also rely on internally tuned market impact models, which by design are aimed at mitigating against taking action that moves prices significantly.

Limit-based/sweeping algorithm

Overview. Limit-based or sweeping algorithms will try to rapidly fill an order with available liquidity from across multiple trading venues at prices better than or equal to the user-specified limit price.

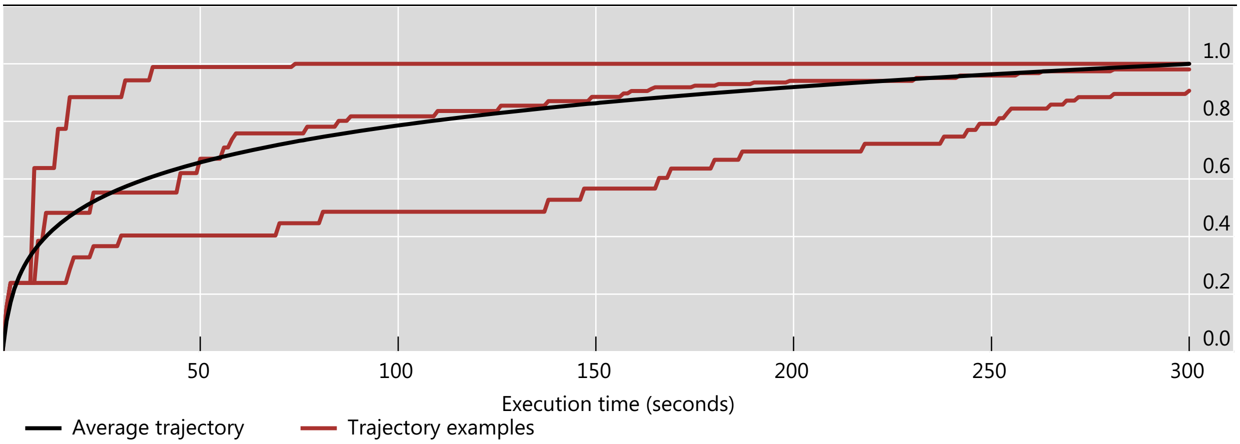
Illustrative execution schedule. To illustrate, when a trader places an order to sell USD 60 million USD/JPY through a limit-based algorithm:

- Such algorithms will generally first take stock of available liquidity in the various trading venues (Graph 5). Some algorithms could make adjustments to the visible amount of liquidity by accounting for historical fill ratios and order rejection rates. Should there be sufficient liquidity, the algorithm will instantaneously consume (ie sweep) all available liquidity up to the order amount (USD 60 million in this example).
- If the liquidity is insufficient, the algorithm would have several options to complete the remaining part of the order. Depending on user inputs on the urgency of execution, the algorithm could proactively consume whatever liquidity is available to minimise exposure to market volatility risk, or wait till there is sufficient visible liquidity before filling the total order amount to minimise market impact.

Safety considerations. Limit-based/sweeping algorithms are the most aggressive and usually the fastest among execution algorithms, and hence are typically used for smaller-sized orders. User-specified price limits are usually required for such transactions, to prevent a substantial overshooting of market prices.

Illustrative example of an implementation shortfall algorithm

Execution profile | Annex Graph 4

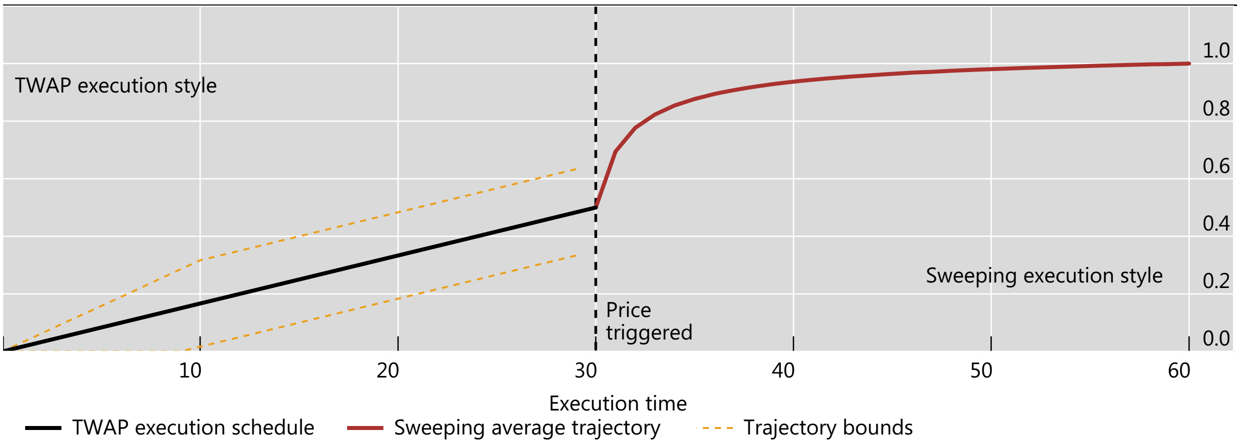


Source: Markets Committee study group calculations.

Illustrative example of a hybrid algorithm (TWAP and sweeping)

Execution profile

Annex Graph 5



Source: Markets Committee study group calculations.

Hybrid

Overview. There are many types of hybrid algorithms for execution. One type would be an algorithm which can switch its execution style (eg from TWAP to sweeping) based on certain trigger criteria (eg price, time). Annex Graph 5 provides an illustration.

Illustrative execution schedule. To illustrate, when a trader places an order to sell USD 60 million USD/JPY through a hybrid algorithm to be completed in 60 minutes, where the algorithm will start with a TWAP execution style and switch to a sweeping style if bid prices for USD/JPY rise above 108.60:

- As per the TWAP algorithm, the hybrid algorithm will determine the broad schedule of execution based on the user input duration. In this case, it decides on a pace of USD 1 million (ie the minimum size at primary venues) every 60 seconds, and will transact at this pace for the transaction if bid prices for USD/JPY stay below 108.60.
- However, if bid prices rise above 108.60, the algorithm will collate available liquidity over all possible trading venues; and should there be sufficient liquidity, the algorithm will rapidly consume all available liquidity for the remaining amount.
- Similar to both TWAP and sweeping algorithms, the hybrid algorithm would typically allow for either some flexibility over its predefined execution schedule (eg TWAP), or vary its execution speed based on user input on urgency of the order (eg sweeping).

Members of the study group

Chair, Swiss National Bank	Andréa M Maechler (Ms)
Reserve Bank of Australia	Jason Griffin
Central Bank of Brazil	Marcelo de Almeida Oliveira
Bank of Canada	Zahir Antia Nicholas Broz
People's Bank of China	Wang Haina (Ms)
European Central Bank	Tobias Helmersson Guy-Charles Marhic Toma Tomov
Bank of France	Alexis Laming
Deutsche Bundesbank	Rafael Zajonz
Hong Kong Monetary Authority	Jing Liu (Ms) Brian Ng
Reserve Bank of India	R S Ratho
Bank of Italy	Alfonso Puorro Alessio Ruggieri
Bank of Japan	Yoichi Kadogawa Takuya Tsuda
Bank of Korea	Dongkyu Choi Jisung Ko
Bank of Mexico	Juan Rafael García Padilla Mayte Rico Fernández (Ms)
Netherlands Bank	Pieter Moore
Monetary Authority of Singapore	Alvin Teo Lim Hui Wen Chai Han Xie
Bank of Spain	Carlos González Pedraz Jesús Ibáñez
Sveriges Riksbank	Jens Vahlquist
Swiss National Bank	Benjamin Anderegg Lukas Frei Carolin Reiss (Ms)
Bank of England	Pelagia Neocleous (Ms) James O'Connor
Board of Governors of the Federal Reserve System	Dobrislav Dobrev
Federal Reserve Bank of New York	Matthew Raskin Pertshuhi Torosyan (Ms)
Bank for International Settlements	Nikhil Patel Andreas Schrimpf (Secretary)