Agenda

Advancing economics in business

Foreign exchange trading: is it time to say goodbye to 'Last Look'?

Since 2013, the foreign exchange (FX) market has been the subject of regulatory investigations resulting in more than US$10bn in fines globally. 'Last Look', a feature of some electronic FX venues, has been under increasing regulatory scrutiny amidst concerns around the impact on end-users of FX trading markets. Oxera Associate, Álvaro Cartea, discusses his recent academic work in this area carried out with colleagues Sebastian Jaimungal and Jamie Walton

The organisation of trading in FX differs from what is observed in other financial markets. Unlike in the equities market, where thousands of individual securities are traded across the world, there are less than 50 actively traded currencies, and more than 87% of FX trading in 2013 involved either US dollars, euros or pounds sterling.[1]

Despite this apparent concentration of trading, there is still a high number of tradable currency pairs, and liquidity in any particular currency pair is generally more fragmented than for individual equities. For example, including both single and multi-dealer platforms, the most common G10 currencies are available to trade on more than 20 electronic trading venues.

One implication of this market structure is that demand for a different currency is often satisfied by a professional liquidity provider, or 'market maker'. Market making—whereby a professional dealer quotes both a buy and a sell price and is willing to agree to one side of a trade without necessarily having an offsetting order on the other side—is therefore important in FX trading.

Risks involved in market making

The fundamental risk in market making arises when a market maker agrees to one side of a trade without having an offsetting order on the other side. In this situation the market maker becomes exposed to the risk that they will incur a loss when offsetting this trade.

The prevalence of this risk, and the costs associated with managing it, have changed alongside technological developments and the increase in the number of FX electronic venues.

First, as technology develops, dealers making markets on multiple venues become increasingly exposed to riskless arbitrage by 'latency arbitrageurs'—professional traders who intentionally take advantage when a market maker's price is slow to update.[2] This is described in more detail in the box overleaf.

Second, as demand becomes fragmented across multiple venues, and market makers compete to increase their market share by offering liquidity across many (or all) of these venues, there is a greater chance that they will commit to delivering higher volumes than they can fulfil. This is because the market maker cannot be sure on which venues their offers will be accepted, and will therefore, in aggregate across all venues, offer higher volumes than they want to (or indeed can) deliver.

The service provided by market makers is important to the real economy. It enables end-users (such as corporates, governments and pension funds) to quickly exchange (large) amounts of value from one currency to another, while minimising the 'effective spread'—the difference between the value of the final currency pre-transaction and the value that the end-user achieves.

Developments in technology and regulation (such as the promotion of electronic trading platforms) create opportunities for the efficiency of the trading system to improve, but also in some instances to worsen. For example, the rise of riskless latency arbitrage opportunities could reduce the efficiency of the system, by increasing the risk faced by market makers—for example, by increasing the potential for quotes offered by the same market maker to differ across the various electronic venues that they support. Reducing this risk (all else being equal) creates the potential to increase the efficiency of the system. One potential way of achieving this risk reduction is explored in the next section.

Riskless arbitrage—taking advantage of stale quotes

Market makers will generally offer to both buy and sell a particular currency (say, US$) in relation to another currency (say, €). For example, at a specific time a market maker may be offering to:

- buy $100.00 for €90.00 (or the equivalent: sell €90.00 for $100.00); and

- buy €90.01 for $100.00 (or the equivalent: sell $100.00 for €90.01).

If the market maker finds equal numbers of counterparties for both sides of these particular offers, they will make €0.01 for every $200.00 of trading that they do ($100.00 in each direction).

However, the prices of currencies do not remain static, and at any one time a market maker is likely to have an outstanding position in one direction or another. As a result, market makers constantly change the prices that they are offering in order to ensure that, over time, they buy each currency at a lower overall price than the price at which they sell it, rather than vice versa.

This updating process is not instantaneous. The response of the market maker's customers is also not instantaneous, and different customers can (and do) respond in different timeframes. This creates the possibility that, by the time a transaction is actually agreed, the market price may have moved

and the market maker will be exposed to making a loss. For example, if the price movement has been large enough, at the point when the market maker has finalised a deal to buy $100 for €90.015 they may only be able to sell $100 at €90.01. This raises the possibility that the same counterparty can execute a trade that is, for them, 'riskless'—as they can, in practice, 'simultaneously' buy $100 for €90.01 (the market maker's stale quote) and sell the $100 for €90.015 (the market maker's new quote). On a $1m trade (in each direction), this nets the riskless arbitrageur $50. This is illustrated in the figure.

By being extremely quick—quicker than the market maker can adjust their prices—a latency arbitrageur can exploit any instances of where a 'negative spread' becomes available between 'stale quotes' and 'new quotes'. The latency arbitrageur's profit is the market maker's loss, so to remain in business the market maker must recover any such losses from other traders.
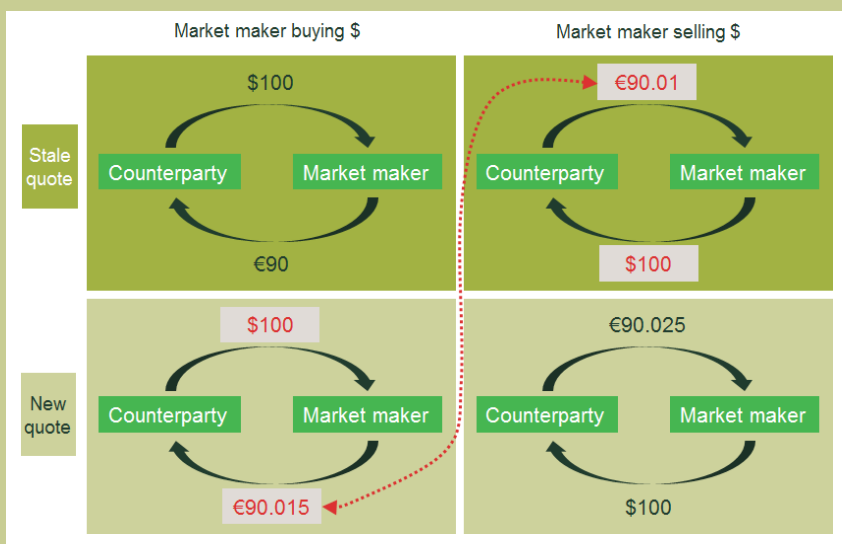
Last Look

Market makers (and the electronic venues that they operate on) can take certain measures to limit their exposure to latency arbitrageurs. For example, some electronic venues allow liquidity providers a 'Last Look' before they agree to undertake a trade—as defined in the box below.

A market maker will generally reject a trade if, within a pre-agreed period of time, the market price has moved

Principle of Last Look: after a market maker has made an offer to trade, and a potential counterparty has responded that they will take up that offer, the market maker has a fixed period of time within which they can decline to actually undertake the transaction (in other words, to renege on their original offer to transact).

against the market maker by some pre-set amount. In this scenario, the market maker is inferring that the counterparty may be a trader who wishes to take advantage of the liquidity being offered at a (now) stale price, in order to exploit a (potentially riskless) trade opportunity at the expense of the market maker. In this way, Last Look can (help to) neutralise the effect of latency arbitrageurs (at least over the interval of time during which the Last Look option is valid—typically measured in milliseconds).

However, Last Look is controversial in the FX marketplace, particularly in the context of the recent fine imposed on Barclays for misuse of Last Look (see the box overleaf). Although Last Look is intended to provide a degree of protection against latency arbitrage, it may also result in market makers rejecting counterparties that genuinely want to convert currencies, and thus effectively undermining the veracity of the offer that they had made a few milliseconds earlier.[3] This raises questions about how fair these markets

really are, when viewed from the perspective of a corporate in the real economy that simply wishes to convert one currency to another. Indeed, some electronic venues actively advertise that they do not allow liquidity providers to operate a Last Look policy on their platforms.

Despite this, Last Look can support more efficient markets. By protecting market makers from more aggressive latency arbitrage behaviour (which is a cost to the market maker), allowing Last Look can result in tighter spreads.

The disadvantage of Last Look is that traders (both end-users, such as corporates and pension funds, and speculators) no longer have a guarantee that, when they respond to an offer by a market maker, that offer will actually be fulfilled. If it is not fulfilled, there is then no guarantee that the next offer that they can access will not be worse for them. End-customers will also be better off only if the reduction in the market maker's costs translates into lower prices (i.e. effective spreads) rather than merely increased profits for the market maker.

Theoretical modelling of market dynamics with and without a Last Look capability can shed light on whether such a

capability will result in benefits for end-users, or whether it will merely enable the intermediaries to exploit those end-users for their own benefit.

Findings from theoretical research

Ongoing research[4] that I am undertaking with Sebastian Jaimungal and Jamie Walton explores the effects of Last Look on quoted spreads in the context of competitive electronic trading venues, each supported by a single market maker, with two types of trader:

- 'slow traders'—those converting currency to meet a fundamental need (e.g. to invest in foreign stock);

- 'latency arbitrageurs'—those trading only when the difference between a dealer's stale and new quotes provides an opportunity for riskless arbitrage.

Competition is assumed to ensure that the market maker supporting each trading venue can only break even. This means that any losses that a market maker incurs to latency arbitrageurs must be recovered from slow traders by quoting wider spreads. Thus, the higher the value of latency arbitrage transactions, the wider the quoted spread that is needed for the dealer to break even.

Do slow traders benefit from Last Look?

This is not a straightforward question. When a venue has a Last Look policy, slow traders benefit from tighter spreads, but some of their trades will be rejected. Therefore, whether a slow trader prefers Last Look depends on the value that they place on certainty and immediacy relative to having narrower spreads. Generally speaking, the greater the ratio of latency arbitrageurs to slow traders, the more likely it is that slow traders will prefer Last Look.

In the framework we have developed, the impact on latency arbitrageurs is more definitive than the impact on slow traders. In the presence of a Last Look policy, some of the arbitrageurs' trades will be rejected, and those traders will therefore unambiguously be worse off. Furthermore, because any profits achieved by latency arbitrageurs lead to costs that are borne by slow traders, this in turn suggests that, except when slow traders place particular emphasis on the certainty of immediate execution, they are likely to be better off when Last Look is adopted.

In terms of the outcome of competition, we find that multiple equilibria can exist. When the costs to traders from switching between FX venues are low (and slow traders care less about the possibility of their trades being rejected), the whole market typically migrates to a venue with Last Look, and venues without Last Look are abandoned. However, as switching costs increase, traders become less inclined to leave the venue at which they started trading, and therefore the starting positions (the initial proportions of slow traders

and latency arbitrageurs at each venue) are more important in determining the equilibrium market outcome.

Conclusion

At first sight, Last Look may appear to be a market practice that is designed mainly to benefit market makers at the detriment of their clients—and the recent misuse of the practice by one particular firm does little to change that perception. However, new academic research demonstrates that Last Look can benefit end-users, particularly those that are unable (or unwilling) to participate in a technology arms race to minimise trading latency. This suggests that the effects of Last Look need careful analysis before any conclusions can be drawn. The potential for Last Look to improve market efficiency also raises interesting questions about the optimal design of financial markets more generally. Rather than saying goodbye, is it time for other markets to say hello to Last Look?

Álvaro Cartea

---

[1] Bank for International Settlements (2013), 'Triennial Central Bank Survey. Foreign exchange turnover in April 2013: preliminary global results. Monetary and Economic Department', September.

[2] In practice, the distinction between a 'trader' and a 'market maker' used in this example is unlikely to be so clear-cut. Any particular market participant may be simultaneously looking for both market making and riskless trading opportunities.

[3] There are also 'grey' areas where it is not clear whether the counterparty's market conduct should result in a rejected trade. For example, this is the case when traders trade at a higher frequency than 'real money' and/or fragment their trades across many venues simultaneously—which is often referred to as 'spraying the market'.

[4] Cartea, Á., Jaimungal, S. and Walton, J. (forthcoming), 'Foreign Exchange Markets with Last Look', Working paper, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2630662.