

THE TRANSPARENCY TASK FORCE

Defining What Transparency Means in the Wholesale Foreign Exchange Market

July 2017

Contents:

About the Transparency Task Force	3
Executive Summary	4
Rationale	6
Market Structure Recap	7
Hidden FX Costs	10
The case for Portfolio Compression	17
Electronic Trading: A new timescale	18
How Much Do FX Transactions Cost?	25
Retail FX; the significance for pension savers	28
Recommendations	30

About the Transparency Task Force

The Transparency Task Force (TTF) is a campaigning community, dedicated to driving up levels of transparency in financial services, right around the world. It believes that higher levels of transparency are a pre-requisite for fairer, safer and more efficient markets that will deliver better value for money and better outcomes to the consumer.

Furthermore, because of the correlation between transparency, truthfulness and trustworthiness, the TTF expects its work to improve the reputation of the financial services sector.

The TTF seeks to operate in a collaborative, collegiate and consensus-building way; focusing on solutions, not blame. It has over 160 volunteers organised into 9 teams. Each team is working on separate campaign initiatives. The TTF's Foreign Exchange Team is led by Andrew Woolmer, Co Founder and CEO of New Change FX.

<http://www.transparencytaskforce.org/>

Lead Author: Xavier Porterfield CFA, Head of Research at Newchange FX

Supporting Author: Jessica Bilcock, Government Relations, TransferWise

Executive Summary

The second and final instalment of a new Global FX Code, designed “to promote a robust, fair, liquid, open, and appropriately transparent market” has now been published. MiFID2 and PRIIPs come into force in January 2018, requiring fiduciary managers and service providers to address their FX processes.

The Global Code is a voluntary, principles based set of standards. The purpose of the code is to provide a harmonised, global standard of best practice in the global FX market. Indeed, the first part of the code, released in May 2016 offered useful examples and case studies.

The Global Code came into being as a response to political pressure following various reviews and enquiries into Foreign Exchange practises in numerous jurisdictions around the world. Due to its size, (over one third of the world’s global FX volumes are exchanged in London) the UK government’s Fair and Effective Markets Review played a key role in framing the discussion of the issues surrounding the wholesale FX market.

Achieving agreement on a global set of standards is a laudable achievement. However, the release of the global code does not alter the reality that OTC markets remain fundamentally opaque. A set of standards cannot change that. The areas of the code which caused the most controversy were related to issues related to trading ahead of clients (to source

liquidity for clients) and the practise of “last look”. In the end, no final set of principles that govern these practises was agreed.

To achieve transparency in Foreign Exchange, clients must be able to be able to define who they are trading with, precisely when they are exposing a trade to the market, and at what absolute cost the trade has been done.

This paper seeks to explore where the FX market remains opaque, and how transparency can be achieved in these three critical areas.

The TTF recommends:

- Custodians not transact FX on a principal basis.
- All FX transactions should come with reliable time stamps – this is time immediately prior to the order arriving with the trading desk that executes the order.
- The arrival price for measuring the market impact of point in time FIX transactions should take the market mid-rate available at the moment the fixing order was transmitted to an intermediary.
- Transaction costs should be measured against objective, independent data that reflects the best market price available in the market when the order was submitted to their intermediary

- For retail transactions, these transaction costs should be disclosed in full to the consumer, including the revenue percentage included in the provider's reference rate on the day.

Rationale

The foreign exchange market is a delocalised, global market where foreign exchange transactions can occur around the clock. A client sitting in California can trade with a market maker in London for a fund which is registered in Hong Kong. The perception that the FX market is rigged prompted a co-ordinated response from regulators and market participants, cognisant that unless action was taken in concert, poor behaviour would simply migrate to jurisdictions offering lower regulatory constraints.

Lack of transparency in OTC markets can have a number of undesirable effects. These include a misallocation of capital. If the costs of participating in the FX market are opaque, or if the market is perceived as being rigged, participation will be discouraged. This can lead to a casino image of the FX market where participation is seen as a form of gambling , leaving currency exposures under hedged. This under hedging means that end investors are exposed to unnecessary or sub-optimal risks, all of which increase the cost burden on society as a whole.

Removing opacity could encourage greater market participation and increase the efficiency of FX market participation. The global FX code represents a major achievement in

international co-operation, but unfortunately, a number of the features of the FX market which create opacity have not been dealt with. This paper seeks to remind regulators and stakeholders that the work towards a suitable level of transparency is not complete.

Market Structure Recap

The FX market is decentralised. Liquidity is disaggregated over many different venues and counterparties. Moreover, FX is traded on a bilateral rather than an exchange basis. This means that every FX price is customised for each client. The 'going rate' in the Foreign Exchange market depends entirely on who is asking the price of whom.

Historically, the wholesale FX market was two-tiered. An exclusive 'interdealer' market segment allowed banks to source liquidity from each other and a dealer-customer segment allowed customers to source liquidity from the market-making banks in turn.

The interdealer segment has now lost its exclusivity. In 2005 interdealer prices and data finally became available to the customer segment for the first time. Coupled to changes in the credit structure through the evolution of Prime Brokerage, this gave non-bank market participants the ability to trade on the same prices as banks and, more significantly, to make prices too.

The majority of FX volume is now conducted on market venues where liquidity flows in one direction only: from market maker to customer. This reflects the legacy of a two-tiered system in which customers continue to 'take' prices and pay market makers, despite owning liquidity.

With liquidity disaggregated into numerous pockets or pools, market makers provide the necessary intermediation to source liquidity for clients. Market makers do not typically earn explicit fees, instead, they seek to make profits by earning a bid-ask spread. This spread represents a risk transfer fee. The client closes their risk by transferring it to the market maker – and pays a fee.

When market participants face each other as principal they compete over the terms of the deal. The interests of the principals are diametrically opposed. Market makers seek to maximise the spread they can earn, while customers try to minimize the spread.

In addition to spread, participants compete over the information content of the trade, which can be much more valuable. FX prices are formed *before* a transaction occurs, via expressions of interest. The price shown is just an indication, and the price can be withdrawn if it becomes unfavourable to the market maker showing the price. But in the process of showing a price to a potential buyer, the market maker has gleaned valuable information about market interest at that price. Market makers seek to avoid falling into the trap of the buyer's curse, otherwise known as adverse selection - owning a currency that is worth less than they price they just paid for it.

Changes to the regulatory environment, particularly with respect to the capital charge applied to risk weighted assets, have diminished the appetite of market makers to allocate capital to market making. This has meant that traditional market makers (the banks) have been less willing or less able (or both) to warehouse risk. This has led to much shorter inventory holding periods and a significant increase in 'matched principle' trading.

Matched principle trading entails a market maker effectively white labelling the liquidity solution of another provider. Typically they will add a mark-up to someone else's rates, rather than actually earning the bid ask spread by taking principal risk.

Much of what we call market making actually more closely resembles exploitation of risk free arbitrage. Market making institutions seek to preserve their informational edge, because the source of their profits is not from taking principal risk. Instead, profits come from being better informed than their clients – and charging risk transfer fees for riskless arbitrage.

The new rules on risk weighted assets have encouraged those wanting to take market risk to invest in non-bank market making hedge funds rather than allocate risk directly to in-house market making. Unsurprisingly, the new entrants into the market making space are less constrained in how they use their own capital than the banks. These firms are now stepping into roles that once were the unique preserve of the interdealer banks.

As we discussed above, in principal trading, market participants trade on their own account, putting their own capital at risk. Principal actors seek to buy and sell on terms that are most favourable to themselves. In a matched principle trade, a facilitator brings together a buyer and a seller on an agency basis, without taking the other side of either party's trade. The transaction is completed simultaneously with the facilitator's remuneration remaining independent of the rate at which the transaction occurred.

However, as the recent Fair and Effective Markets review concluded, there are instances when the distinction between agency and principal are blurred.

Hiding FX Costs:

1. Agents trading as Principals

Despite a spate of lawsuits highlighting the problem, a significant number of investment firms have negotiated arrangements with custodians, granting the custodian a monopoly on the FX transactions that originate from the fund. This obliges the underlying fund investors to accept the custodian's FX rates even when they are not competitive.

A Russell Research paper "It's time for more choice in FX" published in 2004 raised awareness of the extent to which FX transactions could be costing investors. Their work, based on thousands of trades, showed that the majority of custodial FX trades were heavily skewed to the worst rates of the day.

The issue came to a head in 2009 when the largest pension fund in the United States, Calpers, sued their custodian. The heart of the problem is that custodians may choose to execute FX trades on a principal basis, putting their own interests in direct competition with the interests of the underlying investors whose assets they hold.

Investors can easily be confused as to who they are dealing with, and on what basis. Are their funds traded with someone looking after their interests, under the protection that agency laws provide, or are they trading with someone on a competitive basis, as a principal, without the protection of agency?

The evidence from years of Transaction Cost Analysis suggests fund investors are much better served when fiduciaries are not given the opportunity to trade against the interest of the customers they serve.

The Fair and Effective Markets Review (2015) calls for the Global FX Code to “set standards for the treatment of clients and counterparties. This section of the code should address issues such as the prevention and management of conflicts of interest, especially concerning mixed principal and agent roles”

And yet, custodial FX continues to be offered to investment fund clients on a principal risk basis, and the Global FX Code did not address this issue.

2. Time Stamps

A study conducted by researchers at Brandeis International Business School and Williams College, (Olser, Savaser, Nguyen 2012) found that asset managers may be choosing to outsource their FX trading in order to “shroud” cost information from underlying investors.

The rationale for this strategy is that underlying fund investors may not notice the efficiency gains from negotiating currency deals individually. Improved performance might be accredited to other factors. On the other hand, negotiating deals individually requires a trader, technology and trade processing staff, all of which are likely to result in higher running costs that are much easier to identify. Transaction efficiency is hard to isolate. Costs are much more visible.

Olser, Savaser and Nguyen studied the complete trading record of a mid-size global custody bank which included 70,000 transactions in 25 currencies. They found that the average cost of non-negotiated FX trades was 19 basis points, well above the 2-3 basis points clients might expect had they executed on a negotiated basis.

When FX orders are “shrouded”, prices are set relative to the high and low for the day. Because a customer’s loss is a principal counter party’s gain, the distribution of customer fills are heavily skewed to the worst prices of the day. As evidenced by research from Russell, (2004 opcit, revisited in 2010) and Record Currency Management (2011).

The movement of asset flows that necessitate FX deals are known long before the FX trade is done. Custodial FX dealers are able to acquire the inventory they require discretely, filling client orders before they make their client a price. They are then well placed to add a mark up to their own acquisition cost to make sure they always make money on the trade. This effectively turns the FX trade into a riskless arbitrage which does not justify the earning of a risk transfer fee.

These effects are particularly amplified in restricted currencies, where the custodian cannot by law do an FX deal without an underlying asset transaction. The custodian will insist on the asset deal settling before the FX hedge is done, which means that investors often wait 4 days before their FX is hedged. During this time the custodian can opportunistically pre-hedge the deal – and guarantee themselves a profit.

Abuse is difficult to spot. Custodial trade reports contain no reference to the time nor the relevant mid-market rate when the trade took place. This makes it extremely difficult for asset managers to identify their realised FX execution costs.

To encourage transparency, the Fair and Effective Markets review called for mandatory time stamping on all FX trades. This simple step has been ignored in the Global Code, allowing custodians to continue their practise of hiding time stamps. Obliging custodians to release time stamps of non-base currency equity and bond deals would of course be simple enough – and enable FX cost measurement.

Even with effective Time-Stamping, automated, non-negotiated FX deals remain vulnerable to other forms of cost shrouding.

3. The 4pm Fix

The WMR Fix is a valuation tool used to value trillions of dollars in common investment benchmarks. Because the underlying equity and bond benchmark providers use the WMR Fix to value the benchmark portfolio, passive fund managers have an incentive to try to peg their FX transactions to the benchmark rate. This serves to minimize the tracking error between the benchmark valuation rate, and the realized execution rates of their portfolios.

Responding to concerns about the integrity of FX Benchmarks, the Financial Stability Board tasked a special working group to review FX Benchmarks. Their work focused on the WMR 4pm fix, and the ECB fixing, the two most widely used benchmarks.

The working group final report (September 2014) concluded, “it is the incentive and opportunity for improper trading behaviour of market participants around the fix, more than the methodology for computing the fix (although the two interact), which could lead to potential adverse outcomes for clients.”

The European Central Bank now actively discourages the use of its own ECB reference rate for transaction purposes.

However, the FSB working group did not review the cost to investors of using these benchmarks.

Independent research conducted by numerous firms and institutions (including but not limited to New Change Currency Consultants, Pragma Securities, New City Initiative) reveals that using the WMR Fix is highly inefficient.

As fixing orders must be submitted at least 30 minutes before the fix, market making institutions can (and in many cases must) begin pre-hedging large orders well in advance of the Fix itself. Speculators can of course see this activity, and are able to jump ahead and front run it, which creates skew in pricing, pushing the price against the users of the Fix. This means that the cost to customers is often many times higher than the Fix itself shows – but they cannot see this cost because they have achieved the fixing price, and are only ever shown the cost of execution within the fixing window. The fixing window opens after the skew has been achieved.

Users of the 4pm fix are particularly vulnerable to these skew costs at period ends, when large portfolio rebalances that match popular investment benchmarks occur. The amount of money tracking particular benchmarks ensures that the direction of the market for the fix is predictable. This error is then combined with the popular misconception that FX markets are highly liquid. In fact FX markets are highly illiquid, with a maximum of USD14.4 million dollars a second being traded in EURUSD globally at the busiest time of day.

In lieu of time stamping trades, custodians and managers often pretend to offer transparency for their transactions by declaring that they will convert currency at the WMR fix price.

Because market impact can only be seen in the context of the trading activity immediate before and after the trade, market impact is to all intents and purposes completely invisible to the underlying investors in a fund.

The choice to use the 4 PM fix is simply another shrouding method, chosen to hide transaction costs. What is more, this particular shrouding method is massively expensive.

New Change FX conducted a study comparing execution rates using a time weighted average price and the 4 pm fix. Assuming passive investors will normally be positioned in the direction of the fix, the study found that using the Fix for EURUSD trades resulted in a net cost to NAV of 3.6% over a year compared to transacting at a time weighted average price.

Comparing this result to realised trades in EURUSD of UK and US based asset managers, Newchange FX found that investors were incurring between 60% and 70% of these costs - that is about \$23,000 per million. These are costs that are unreported, as the timestamp used is the open of the fixing window - and will continue to be unreported under MiFID2.

When a fund had no contractual obligation to use the 4pm Fix for some index tracking purpose, cost shrouding may be a factor in the choice to use the Fix.

The Case for Portfolio Compression

Cost shrouding is a special case of opacity, but use of the FIX also presents another problem. It is not efficient. The special working group mandated by the FSB to investigate the use of FX benchmarks “supports the development of industry-led initiatives to create independent netting and execution facilities for transacting fix orders” (Section 7, Point 6 of the Foreign Exchange Benchmarks Final Report)

In their review, the FSB working group found that clients were using the FIX to trade, when netting might be a more cost effective solution. The problem is that appropriate netting facilities do not exist, or are not widely supported.

Market makers live on flow. Netting of trades reduces the volume of flow that is available to market makers. A recent study by Newchange FX based on the FX exposures of just 4 UK pension funds found that netting could reduce their combined FX costs by 60%.

By executing each transaction individually, rather than netting down to produce a smaller market exposure, FX volumes are larger than they need to be. Portfolio compression can offer substantial cost savings. The problem is that market makers have no incentive to support or create external netting facilities. This would seem to offer an opportunity for

disruptive new entrants. But here the FX market presents a structural obstacle. Because the dominant FX market makers are also dominant in clearing and settlement it is difficult for more efficient methods of execution such as external netting facilities to emerge.

Despite the vested interest of incumbents, a number of portfolio compression initiatives have been launched. For the time being these services are only available to banks, but LMRKTS a portfolio compression service provide for non CLS currencies does plan to open up to some buy-side firms. OTC markets remain dependent on credit, which creates a tiered hierarchy of who has access to these services. In effect, unless a universal access approach can be found, portfolio compression can become another differentiator, re-introducing a softer form of two tiered market where the best names can benefit from compression, while others may not.

Electronic Trading- A new time scale

Electronic trading has become the dominant medium for conducting Foreign Exchange, accounting for up to 65% of average daily volumes. The development of electronic trading has introduced a new dimension to how market makers can gain an advantages over price takers. Time is money!

Electronic trading and voice trading do not occur in the same time scale. Reported response times from a number of ECN platforms are in the order of 300 *micro*seconds.

However, it is extremely unlikely a customer would be able to complete a transaction with a market maker within this time frame. The response time from a market maker might be anywhere between 10 and 1000 times slower.

Price slippage occurs when the actual transaction price differs from the price when the decision to trade was made. Firm pricing in Foreign Exchange is still rare. The majority of prices are displayed as an indication of interest, which can be withdrawn at a moment's notice. If the price can be withdrawn, customers can wonder whether the price was really there at all. This gives rise to a phenomenon known as phantom liquidity.

One of the areas that has caused controversy in the FX market is how banks have been applying **last look**. Last look ensures the client is always wrong and that the price-maker secures a risk-free profit. It is an embedded component of trading, and thanks to very weak handling of the issue by the sponsors of the Global Code, it remains so. The issues are complicated and the Global Code simply ensures that last look can be given more respectable attire than used to be the case.

The issues and definitions are as follows:

Latency slippage is caused by the time it takes for a signal to travel over a network.

Latency slippage should be symmetrical. After all it is simply the result of a signal being out of date by the time the recipient of the signal receives the message. A market maker has updated a price, but the updated price is not visible to the customer yet, who is attempting to trade on a price that is stale.

Last Look slippage occurs when a market maker has left prices in the market that have been superseded by events. This sounds like latency, but it is not about the time it takes for a signal to propagate through a system. It is about whether the market has moved and the market maker has been slow to adjust their prices.

Last look addresses a problem where the market maker has been slow to respond to a change in market conditions, as opposed to a latency issue which describes the time it takes signals to travel over a network.

The problem arises, and the opportunity for abuse occurs because latency is reported in different ways. It may be used to refer to the time it takes a venue to show a price update from liquidity providers, or it may refer to the time it takes for a client to trade on price that has been displayed.

Last Look – Adverse Selection

Last look was introduced by market makers as a means to protect themselves from latency arbitrage. Prices change in response to new information. Market makers will lose if they make prices to a better informed trader (a trader who can respond to new information quicker than the market maker) , and win when they trade with customers who are less well informed.

A price that updates too slowly is vulnerable to being picked off by a trading algorithm system that is able to react faster. This creates adverse selection risk for the market maker- where the market maker might end up long (owning) a currency that has fallen in value.

Last look has created controversy because some market makers have applied last look asymmetrically, which is to say unfairly.

Consider the following process of how an order is transacted on a price stream:

- Client opens a live stream. The live stream is a series of FiX Quote Messages which are continually cancelled and updated.
- Client hits a price and in doing so instantly sends an order message to the market maker.
- The market maker opens a hold window for the client.
- At the end of the hold window, the market maker runs a tolerance check.

- If the deal is within the tolerance check, a confirmation (AC) message is sent, if not then a rejection message is sent (NAC).

This process applies to all clients. The tolerance check is either defined in basis points or USD terms. By using USD terms, the client will not be rejected on small deals, which can happen if basis point variances are used. The hold window is continually updated as the machine learns how a client behaves. Both variables are set on a bespoke basis per client. The tolerance level can be 'reversed' whereby the bank actually accepts a certain level of loss per transaction before it is rejected.

As we noted earlier, response times are based on the time it takes a signal to travel across a system. This signal itself has variability - it follows a sine wave. It is very difficult to achieve full transparency on signal lengths because of this variability, but clearly, moving response time arbitrarily, such as explaining to a client the reason for a reject was because the system was running very slowly today are clearly areas that are vulnerable to abuse.

The Request for Quote (RFQ) protocol is typically handled much the same way, but the opening procedure starts with the client pulling in a price, rather than receiving prices that are "pushed" via a stream. Last look can be employed on both streaming and request for quote prices.

Whereas applying last look on streaming prices might be justified to handle latency issues, applying last look on RFQ is simply banks protecting themselves from the possibility of being

picked off in a 'drive by shooting' as liquidity gets consumed in a market sweep. The last to price in the sweep is likely to be disadvantaged in much the same way that a game of musical chairs always has one chair less than the number of players.

The new global code reaffirms market makers retain sole discretion on whether a trade is accepted or not. This means that the Global Code has retreated from encouraging symmetric last look (rejecting both favourable and unfavourable trades equally). If the goal of the Global Code is to promote appropriate transparency, failing to insist on symmetric last look can only be a failure.

Symmetric last look allows the bank the opportunity to re-price to the customer, based on the information the client has transferred to the bank. It will always create slippage, but whereas asymmetric Last Look invariably creates negative slippage, applying last look symmetrically theoretically creates a situation where the client might experience positive slippage. Given a scenario where the customer's intent has already been transferred to the market maker, in practice the opportunity for positive price slippage is slim.

On the other hand, asymmetric last look simply provides a market maker with the option, but not the obligation to trade with a customer, after the customer has fully disclosed their trading intentions. How is this fair?

The Global Code's approach legitimises Last Look, when a more transparent solution would be fairer. Just as bilateral markets are based on market makers being able to quote different

prices to different people, Last Look is designed so that response times are tailored individually. When things are different, it is difficult to compare them.

First Look

Electronic trading has opened up a new avenue for market intermediaries to gain an unfair advantage over their clients. By definition, when custodians have a monopoly on the FX transactions of a fund they benefit from inside knowledge.

In the last few years there have been at least two instances of financial intermediaries offering agency trading or direct market access (DMA), whilst simultaneously operating an undisclosed principal market making desk.

Spotting this type of abuse can be very challenging. In a recent case the CFTC withdrew the retail fx broking licence from a leading FX platform, FXCM. The firm did not disclose their interest in a market making firm that clients were trading with via their platform, misrepresenting to clients that it's "no dealing desk" platform did not have a conflict of interest with its clients. Customers continued to benefit from very tight spreads, so that it might appear that clients were not disadvantaged by FXCM's behaviour.

Earning the spread was not the goal of the hustle. The value of first look is the benefit of being able to selectively choose which trades to price, and which to let flow on to outside

market makers. The market making desk, by having first look, was able to choose trades that were favourable to their trading book. Last Look can be applied retrospectively, increasing adverse selection. First look forces clients into adverse selection prospectively.

Until senior management spotted the behaviour, employees of ITG, the broker dealer, ran a similar first look operation in their US equities business. (SEC press release 12th August 2015) ITG and its affiliate Alternet Securities agreed to pay a \$20.4 million fine to settle charges that they operated a secret trading desk and misused confidential trade information of users of their dark pool.

In order to spot first look or last look abuses, forensic transaction cost analysis is required, to quantify fill ratios and post trade decay with individual counter parties.

How much do FX transactions cost?

The early pioneers of FX Transaction Cost Analysis, firms such as Record Currency Management or Russell Investments faced particular challenges when they attempted to uncover the hidden costs that pension funds were paying for their FX transactions.

As noted above, custodians do not typically provide time stamps, nor do they provide a reference to the actual market mid-rate prevailing at the time the transaction took place.

Time stamping is becoming more prevalent and custodians have made improvements to their operations. However, the question remains, if everybody receives a slightly different price, what should we measure against, to enable comparability across fund managers, market makers and venues?

Customers have typically addressed the question of slippage by comparing their execution to the prices that were made available to them. This method seemed to be supported by MiFID I which produced a requirement to put liquidity providers into competition.

The trouble with this approach is that determining execution costs in this way is entirely circular.

For instance, a large UK asset manager created a mid-rate database from Bids and Offers collected from 10 liquidity providers. This ignored whether there was not a better quote elsewhere, offered by a counter party that they did not deal with. Choosing to measure their costs in this way caused the asset manager to understate their FX transaction costs by approximately £10 million a year.

In method, this is only a little better than comparing execution to the prices that were made available to the fund by the custodian's FX desk, dealing with a monopoly on the fund's FX deals.

From an investor's perspective, we need to know how a transaction occurred relative to the best available price in the market as a whole. This can only be determined by aggregating bids and offers across a large cross section of venues that is representative of the wider market.

Because the FX market is disaggregated, the prices from single aggregated venues such as EBS or Reuters do not necessarily indicate the best price available in the market. Using single venue data introduces sampling error.

Cognizant of the increasingly fragmented nature of the market, European regulators proposed a transparent framework for reporting transaction costs. The Packaged Retail Investment and Insurance based Products (PRIIPS) requires firms to capture the market mid-rate prevailing at the moment the order was communicated to a third party for execution. Moreover, the market mid-rate (called the arrival price) must be a consolidated price, and not a price from a single counter party or platform.

Applying and recording a consolidated mid-price (arrival price) is a major breakthrough in cost transparency.

Conflicted Data

As investors learnt to their cost in 2013, when liquidity providers can directly influence the price that they are measured against, they will have both opportunity and incentive to manipulate the rate.

Single source data that an investor has traded on provides a circular measurement. When liquidity is fragmented, measuring execution from a single source means measuring against a much smaller liquidity pool. For instance, if the average volume of a trading venue is 10 billion a day across all currencies, normalising volumes for the trading pair will reduce the volume to about 2.5 billion in EURUSD, which, spread out over the trading day, might mean that the actual liquidity available *in all currencies* during a single minute period might be as little as USD 4 million, or even less.

Measuring execution against a rate from the same platform means that it is very likely the investor is measuring the quality of their trade from their own trade, drinking their own bathwater so to speak.

Retail FX; the significance for pension savers

For consumers, it's even harder to understand how much they are being charged. As noted above, custodians do not typically give a timestamp, or reference to an actual mid market rate, when conducting transactions on behalf of pension funds - making it difficult to ascertain the exact amount pension funds are paying for their transactions.

On a smaller scale, there is a similar issue at a retail level. Brokers are not obliged to provide an independent, non tradeable rate on an ex-ante basis as a reference to the customer attempting to transact in a different currency.

Right now, banks and bureaux de change are only obliged to provide a “reference rate” which, under the definition provided in the Payment Services Directive and in the second Directive, can either be an independent publicly available mid rate or the provider's rate offered on the day - which includes their revenue percentage and as such constitutes a charge to the consumer.

The provider’s rate on the day is an arbitrary construct, set entirely at the discretion of the PSP, with no regulatory oversight. It is far from an independent yardstick by which a consumer can understand the total amount charged for a transaction - including the profit percentage imbedded in the rate offered by the provider on the day.

This setup requires an unreasonably high level of financial literacy on the part of the consumer. Research conducted by YouGov shows that only 10% can understand how to calculate the charge when presented with a typical banking structure¹. On average, high street providers add £29.54 on a typical £1,000 GBP > EUR transaction² in an exchange rate mark-up in addition to transaction fees. Over the course of 2015, UK consumers and small

¹ Research conducted by YouGov surveyed 19,277 European adults between the 8th-22nd February

² Average representative of a typical £1,000 GBP > EUR transfer, using traditional high street providers. Research conducted by Consumer Intelligence in February 2017

businesses paid £5.6bn in charges imbedded in the exchange rate, despite just 20% of UK consumers understand that their provider imbeds a charge in the exchange rate³.

Recommendations

The goal of transparency is not to make the provision of liquidity unprofitable. Rather, it is to remove the opacity surrounding liquidity provision so that customers can be sure they are being treated fairly.

As we have seen, the rise of electronic trading provides new ways that information asymmetries and plain old fashioned cheating can occur. The push of regulation and new rules on the capital charge to support market making has made liquidity provision less about market makers taking principal risk and more about dealers earning a riskless spread. In a riskless trade, the economic interest of principles go head to head over information. The spread is just the cherry that sits on top.

Identifying and managing conflicts of interest is vital. However, because cost shrouding strategies are so difficult to spot, and potentially so costly to investors, certain practices need to be challenged.

- Custodians should not transact FX on a principal basis.

³ Research conducted by Capital Economics, August 2016

- All FX transactions should come with reliable time stamps – or the time stamp should be taken from the asset sale or purchase and not the FX deal.
- The arrival price for measuring the market impact of point in time FIX transactions should take the market mid-rate available at the moment the fixing order was transmitted to an intermediary.
- Transaction costs should be measured against objective, independent data that reflects the best market price available in the market when the order was submitted to their intermediary.
- For retail transactions, these transaction costs should be disclosed in full to the consumer, including the revenue percentage included in the provider's reference rate on the day.

These proposals might be considered something of a big bang, but greater transparency will result in better risk management, encouraging market participants to trade in the best interests of their customers.