



LAST LOOK
MYTHS DEBUNKED

MAY 2021

LAST LOOK: A LONG LOOK

There is now a public commentary period¹ around the topic of Last Look at the three year anniversary of the FX Global Code². There is a general feeling in the industry that the issue of Last Look has been dealt with via disclosures and the issue has been sorted. However, as this paper will argue, there is further improvement that the industry can make around this topic.

Principle 17 of the FX Code states that Last Look can only be used for the purposes of risk control:

*"If utilised, last look should be a risk control mechanism used in order to verify validity and/or price. The validity check should be intended to confirm that the transaction details contained in the request to trade are appropriate from an operational perspective and there is sufficient available credit to enter into the transaction contemplated by the trade request. The price check should be intended to confirm whether the price at which the trade request was made remains consistent with the current price that would be available to the Client."*³

However, because of the lack of a precise definition of Last Look, different industry participants have interpreted the FX Code in different ways and Last Look is still utilised by some Liquidity Providers ("LP's") for the purposes of managing "market impact" and "adverse selection" (i.e. profit optimisation) across many venues and bilateral relationships with their clients. In the view of XTX Markets this is clearly inconsistent with Principle 17 of the FX Code.

Using actual client data, we will show that the use of Last Look for the avoidance of adverse selection (i.e. avoiding the "losing" trades (from the LP's perspective) in a client's portfolio) and for managing market impact (i.e. the post-trade change in the market price after a client's trade has been executed) constitutes an invisible tax on clients and a distortion of their true costs of execution, which we will also quantify.

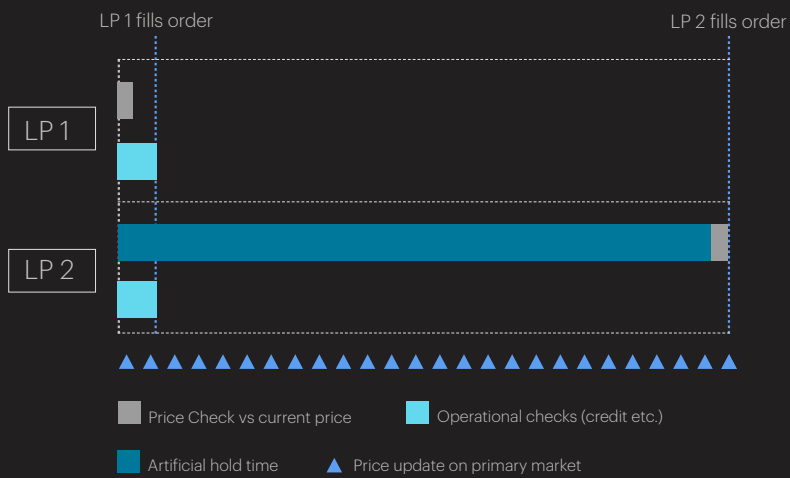
¹ Full consultation paper and details as to how to provide feedback can be found here: https://www.globalfx.org/consultative_process.htm?m=72%7C429

² "FX Global Code: A Set of Global Principles of Good Practice in the Foreign Exchange Market." Globalfx.org, 1 Aug. 2018, www.globalfx.org/docs/fx_global.pdf. Referred to throughout as the "FX Code".

³ Extract from Principle 17 of the FX Global Code

To start with it makes sense to define Last Look. Last Look is actually constructed of **two separate**, but related, practices:

- 1. The right of the LP to accept or reject a trade request (or offer to deal) from a client. This is fundamental to OTC markets and should absolutely be allowed. This control allows LP's to check that the price remains valid in order to protect themselves against latency in their price construction, slow connections to clients etc. It also allows LP's to check client transactions to ensure that they fit within market, credit and operational risk limits. This check against the current price and against limits is vital to the safe provision of liquidity to the market by LP's and is a fundamental tenet of OTC trading. This check can incorporate a small delay in line with the amount of time that it takes an LP to create a new price i.e. the "latency" in its price construction. **This is how Last Look is defined in our view in the FX Code.**
- 2. Additional Hold Time ("AHT"). This is the practice of artificially holding client trade requests (or offers to deal) once they arrive **within an LP's system** for a defined period of time before applying the price check and making the decision to accept/reject the trade request. In many instances the validity/operational checks will have been completed for a considerable period of time before the price check is applied, as the figure below illustrates with LP 2.



Note: This is an illustrative example. LP 1 is not applying an AHT but responds in 10ms as this is how long it takes for it to complete its operational checks. LP 2 responds in 130ms and this is largely due to the AHT of 125ms that it applies before completing its price check, which itself takes 5ms.

The difference between these two is absolutely central to the discussion. No one in the market objects to (1), however, the issues arise with the application of (2): the Additional Hold Time (AHT).

Principle 17 of the FX Code clearly states that Last Look can be used for risk control purposes but not for the purposes of optimising profitability. Using hold times for as long as you need to ensure your price is current is an acceptable practice. However, some LP's are using varying hold times per client and using hold times for longer periods than it takes for them to complete their price check.

How can holding transaction requests and offers-to-deal in a way that varies by client, varies month by month, varies between trade acceptances and trade rejects, and is for a disproportionately long period, be anything other than profitability optimisation - be it profit maximisation or loss mitigation (two sides of the same coin)?

In this paper, we will show clearly why the use of of AHT by some LP's acts against clients and why LP's apply different settings to their clients, not for the purpose of controlling risk but for optimising profitability.

Additionally, we will make a strong argument for why you should all care and why it is critical for an industry that prides itself on its ability to self-regulate, to be applying the FX Code to set the highest standards, rather than setting a low bar or a framework that means that clients need a law degree and a quantitative PhD to understand the behaviours of their LP's.

AT THE OUTSET, WE'D LIKE TO POSE WHAT SHOULD BE A FAIRLY SIMPLE QUESTION.....WHICH OF THESE PRICES GIVES THE LOWEST COST OF EXECUTION

A. 20.1/20.9 with an asymmetric AHT of 20 ms on acceptances and 100ms on rejects (reject rate 5%)

B. 20/21 with an AHT of Zero (reject rate 0.5%)

C. 20.5/20.5 with an AHT of 200ms with an asymmetric price tolerance applied (reject rate 10%)

D. Hard to say for sure but I wish it was just 20/21 and I got done

If you struggled to answer the question easily or answered D , then read on.

PROFIT IN THE BLINK OF AN EYE

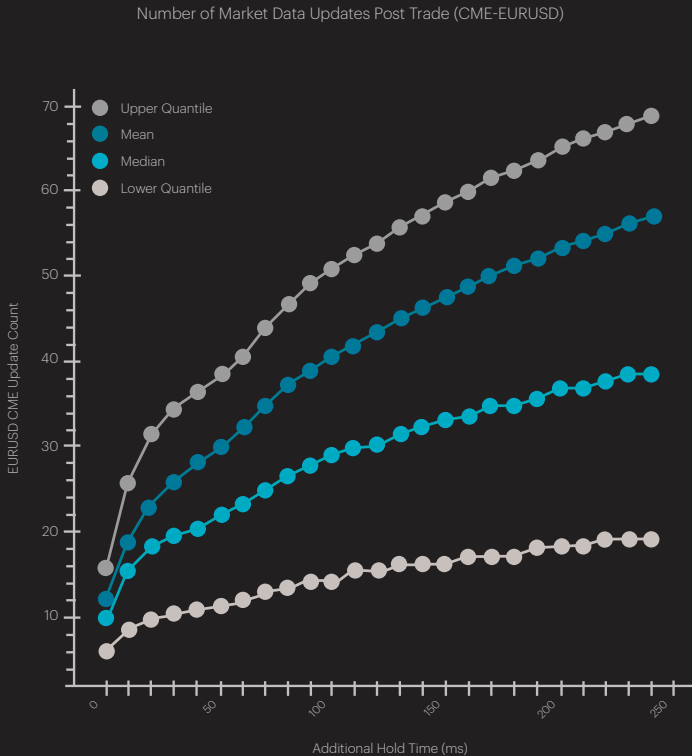
As previously mentioned, although Principle 17 of the FX Code defines Last Look as the price / validity check, looking at the disclosures of certain LP's, many do not explicitly distinguish between Last Look and Additional Hold Time; rather they allow the two to be conflated which, it turns out, is very handy for an LP who wants to use Additional Hold Times to optimise its profitability.

First of all, the price check should be conducted to ensure that the trade request price remains consistent with the current price that would be available to that client. Why would an LP need to wait an additional period before checking that? The FX Code says that the check should be done against the current price, not "at some price in the future at a time determined by the LP". LP's would say that they need to hold the trade request in order to receive the latest price updates from the primary FX Markets to then carry out the current price check.

So how long does it take to receive these updates and check a price? The fastest market data feeds from the primary FX venues in spot FX - which all major LP's subscribe to - update every 5/ms. However, some LP's have Additional Hold Times of 200ms (although many retain the right to extend that period during volatile markets). 200ms - the time it takes for a human eye to blink - doesn't sound a lot of time, does it? What's the big deal?

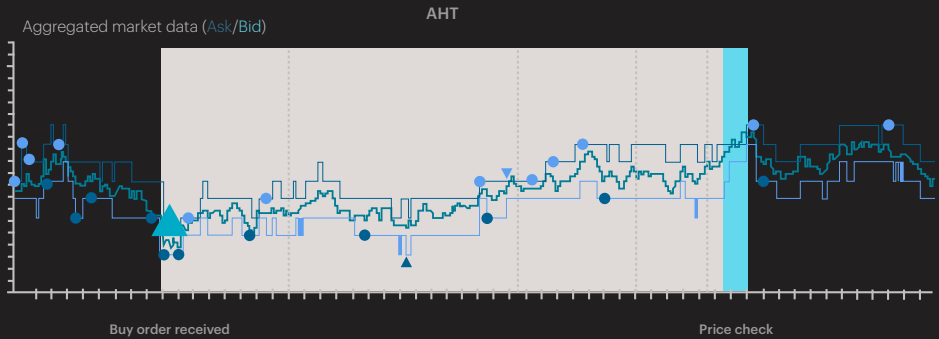
Well, in 200ms, there are an average of 50 price updates for EURUSD on the CME alone (as per the figure below). Now imagine being an LP and also seeing all the price activity on EBS, Reuters, CBOE, LMAX, Currenex, GainGTX, Euronext etc.

Chart produced by XTX using CME data from the period 2020/12/21 - 2021/03/19.



These updates during the AHT are valuable for profit optimisation. Consider the below scenario: because the LP applies an AHT, it can see the market went up after the offer-to-deal was initiated, and after it was received, and then reject the unprofitable buy order. The more updates the LP can see, the easier it is to predict the likely P&L of the trade.

An LP can even set itself fill rate targets (per client) and commercially optimise its logic so that it focuses on rejecting the most unprofitable trades within its reject “quota”. In other words, not all rejects are equal; it is quite possible for the LP to have a relatively low reject rate but absolutely optimise their outcome.



Note: This is an illustrative image only, showing a price stream continuing to tick during an AHT. This gives the LP an option to see the future before deciding whether or not to accept the trade.

Imagine taking this to the extreme and saying that an LP could price choice all day long and then select the trades that they wanted to keep, and those they wish to avoid, by watching the market evolve after the trade request has been submitted by the client.

In extreme, this would be obviously ridiculous - like a bookmaker only deciding to accept a bet after the race had finished. Ah, but what if it wasn't after the race had finished but after say one furlong, or two furlongs, or just before the final furlong? What time would be acceptable? How much of the race should the bookmaker be able to see before deciding whether to accept the bet or not?

Prima facie, this looks bad. Maybe it is because some LP's have poor infrastructure to do their price checks? Let's look at that argument in a little more detail: if AHT is due to the latency of a price check on the LP's own systems, why would any AHT applied in order to meet this current price check be different according to who the client is? And yet LP after LP applies different hold times to different clients. How do we know this? It is in their disclosures....

"After you send an order to us and the [LP systems] receive that order we may place it in a queue for a period of time. That period of time may vary between 0 and 200 milliseconds."

"In addition to applying "last look" trade acceptance parameters to our electronic quotes, [LP] may, for some counterparties, also apply a hold time. Based on observed quote update times at a number of different electronic trading platforms, [LP] has determined that 300 milliseconds is currently the maximum hold time that may be configured for any client."

What possible reason, if not profit optimisation, could exist for holding different clients for different periods? That would mean certain behaviours could be construed as "not complying with the FX Code" or at best "not complying with the spirit of the FX Code". Well, it seems many LP's think otherwise and in reality, of course, this is already happening. How do we know? It is in their disclosures, like these from a major global bank LP.....

"Upon receipt of your electronic offer to execute a transaction at a price and quantity, whether in response to an indication of interest or otherwise, [LP] will determine whether, and at what price, to accept all or any part of your electronic offer, after assessing that electronic offer against a variety of pre-trade factors. These factors may include, among others, available inventory, liquidity, prevailing market prices, anticipated loss or gain based on the [LP]'s analysis of the market and the [LP]'s trading experience with you and/or other market participants, credit and product terms and filters the [LP] may employ for the purpose of implementing credit limits, position limits or other limits on your electronic trading activities. These factors may be changed from time to time without notice to you and may differ from those applied to other counterparties. Due to the speed of the market and execution delays (including time delays that the [LP] may elect to impose on a counterparty-by-counterparty basis, in light of the risks inherent in its electronic market making and risk management activities), the price available for execution of any transaction with you may change between the time of submission of your electronic trade request and the time that you receive a response to your electronic offer, even if the lapse of time is small. This may result in rejection of all or

part of your electronic offer, including in cases where the market has moved favorably to you, even though we may choose to accept electronic offers where the market has moved favorably to us. If we determine to execute, the costs or benefits of any price changes may, in our discretion, be retained by us."

But surely LP's have a right to behave differently to handle so-called "toxic" flow? Of course they do, and there is a perfectly good, tried and trusted mechanism for that - it is called **spread**. Tougher flow should attract a wider spread and better flow should attract a tighter spread. Management of "adverse selection" and "market impact" is synonymous with "profit optimisation". Avoiding the worst trades in any client's portfolio is of course vital to the return that the LP can generate from that client's flow. The fair way to do this is by having better, more informed, pricing and pricing that does not have significant latencies built into it.

Some LP's will of course say that removing AHT altogether would create a technology race to zero that would not be beneficial to clients but would increase barriers to entry to market makers. Whilst we believe that AHT should be essentially prohibited, if the market consensus were to allow a short AHT to prevent this theoretical race to zero, what is proportional and would represent a reasonable AHT? Well, one simple method would be for LP's to **explicitly link their AHT to their median "tick to trade" time**, defined as the time it takes an LP to receive a relevant market data update and make the next price. This "median tick to trade" should be a reasonable benchmark for all LP's. We would like to see public disclosures of these times. It is not an inherently difficult calculation for any LP engaged in electronic market making and does not represent a significant overhead.

But do we really need to be so prescriptive? Surely these are issues that can all be covered by disclosures, right?

DISCLOSURES OR DISCLAIMERS?

Many LP's advocate for a broader, less prescriptive, approach arguing that the FX Code should be principles based and that there should be a general assumption of *caveat emptor* (buyer beware). Further, they argue that *caveat emptor* should be applicable because there are clear disclosures as to trading practices of LP's in the public domain.

So, let's look at some sets of disclosures and see how clear they are. These have been anonymised but come from a major global bank LP in the eFX space.

"Along with risk management, regulatory, and staleness checks, LP performs a price check. If the market price for a Transaction is at a level beyond a counterparty-specific threshold applied by LP's trade acceptance logic after receipt of a client request to trade, LP will reject the trade request. LP may take into consideration a counterparty's historical trading characteristics in setting various parameters, including those relating to last look."

"If the above checks have passed and the order is in scope for a price check, the price check will determine whether the order is accepted or rejected. In the price check, we compare the price we are prepared to make available to you in the relevant currency pair at that point in time with your order price request."

"Any actual hold time that is applicable will be set out in our FIX trade reject message. This means that the last look trade acceptance checks described above will not occur until the particular hold time period has elapsed. LP applies a hold time solely to evaluate whether an intervening price movement has occurred."

Reading through these disclosures it is difficult even for relatively sophisticated market participants such as ourselves to understand the practices being deployed and whether or not they are consistent with the FX Code or good market practice. We can only therefore assume it is even harder for the typical client to have a picture of exactly what is going on. At any rate, these hardly form the basis for "informed consent".

**OR CONSIDER THESE EXCERPTS FROM A NON-BANK
(EMPHASIS IN TEXT BELOW IS OUR OWN):**

1. "LP may reject orders asymmetrically. LP believes that a symmetrical execution tolerance will impact the counterparty in an adverse manner."⁴

2. "LP will utilize "Pre-Hedging" as permitted by the FX Global Code. LP may pre-hedge counterparty trade requests where the quote provider is acting as principal, and does so for the benefit of the counterparty. If LP does pre-hedge a counterparty trade request, all of the prehedged quantity will be passed through to the counterparty, unless express counterparty agreements or exchange rules prohibit partial fills.

LP engages in pre-hedging with the intent to advantage the counterparty with optimal fill rates, but also may receive a benefit from pre-hedging the transaction."

3. "At times, LP may deploy orders as pre-hedges in order to facilitate a counterparty trade request. This strategy can possibly result in an "over-hedge" of the amount initially requested by the counterparty."

4. "All of your order data obtained by LP may be used by LP and provided to counterparties on an anonymous and aggregated business for the purpose of data analysis, risk management, compliance and proprietary trading. Order data includes, but is not limited to, the following:

- Orders executed in full or part;
- Cancelled or expired orders;
- Indications of interest;
- Quotes;
- Positions;
- Trade; and
- Other data and analytics utilizing order data."

Unlike the previous examples, these are perfectly clear. Would you as a client agree that pre-hedging, over-hedging, and use of unfilled order information is acceptable market practice so long as it is disclaimed to you in a disclosure document? So if you can't trust the disclaimers....

⁴ See section headed "What about symmetric and asymmetric logic?" below for an explanation of symmetrical acceptance / rejection logic

IN DATA WE TRUST

One client has been kind enough to allow us to use their data to illustrate the behaviour of some of the largest LP's in the market in the period February through April 2020.

This client's liquidity pool consists broadly of 5 of the world's largest FX banks plus XTX. The following tables show the behaviour of all those LP's during that three month period. For the sake of clarity, and without divulging any information about who the client or LP's are, this client executes a significant amount of flow in two books (amongst others) – one of which trades on a Full Amount basis and one of which trades on a Sweep (aggregated) basis.

It is important to note that Round Trip Times are measured at the 95th percentile and can be affected slightly by single day technological anomalies but the data is broadly representative. Note also that all LP's are cross-connected with the client in a single major data centre so there is no difference in latency of connection and that latency could be measured in less than 1-2 milliseconds. Each book trades tens of thousands of trades each month and well in excess of \$1.5bio every day.

Let's take the Full Amount book first. Looking through February, March and April of 2020, we notice how LPD changes Additional Hold Time in March. Acceptances remain as before in February BUT rejects go from 10 to 125 ms. Nonetheless, there is an undeniable shift in Additional Hold Time that clearly has nothing to do with technology or additional latency, as acceptances can clearly be handled at the same speed as in February. In April, as markets start to calm again, the logic changes again and now acceptances are at 75ms and rejects are still at the extended 125ms. However, the reject rate is low so the impact on the client is somewhat minimised, at least during normal market conditions. In other words, the option is there for the LP but it simply doesn't have that much value in relatively benign markets or execution styles.

But of course, options are more valuable when volatility increases as we can see when we look at the behaviour in the sweep book. Now, by definition, it does not take any longer to accept or reject a trade on a sweep book than a full amount book. You may expect the reject rate to be slightly higher as tolerance settings for deal acceptance are likely to be lower as the flow can be expected to be somewhat tougher (as is normal and expected in aggregated trading books). Nonetheless, notice that in general, most LP's employ longer hold times in the sweep book than they do in the Full Amount book. Why could that be - could it be to see more of how the market evolves after the request?

CLIENT DATA FULL AMOUNT BOOK

Feb - 20	Round Trip Time		Client Data Full Amount Book	
Full Amount	Accepts	Rejects	Reject Rate	Mkt Share
LPA	20	40	1.0%	20%
LPB	20	20	0.5%	14%
LPC	10	10	4.5%	18%
LPD	10	10	0.3%	16%
LPE	10	10	4.2%	11%
XTX	10	10	0.0%	21%

Mar - 20	Round Trip Time		Client Data Full Amount Book	
Full Amount	Accepts	Rejects	Reject Rate	Mkt Share
LPA	20	30	1.9%	19%
LPB	20	20	0.9%	7%
LPC	10	10	7.1%	7%
LPD	10	125	0.1%	31%
LPE	10	10	17.4%	15%
XTX	10	10	0.2%	20%

Apr - 20	Round Trip Time		Client Data Full Amount Book	
Full Amount	Accepts	Rejects	Reject Rate	Mkt Share
LPA	20	20	1.3%	21%
LPB	20	20	0.8%	9%
LPC	10	10	3.3%	13%
LPD	75	125	0.1%	23%
LPE	10	10	2.0%	14%
XTX	10	10	0.3%	14%

Data provided by a client.

But most significantly, notice the reject rates: LPA - from an already high 12% in February to 57.6% in March and LPD from 3.1% in February to 60.6% in March. Both though obviously continue to price tightly as they win high market shares in both months.

Of course, March 2020 was volatile and market spreads were wider but why would that matter if, as an LP, you could price tightly as you simply don't have to stand behind the trades because you can "select the portfolio". Notice how the LP's that obviously price wider win less share (LPB, LPC and XTX) but have MUCH lower reject rates – they are crowded out by LPA and LPD. For the sake of clarity, the sample set here is significant and 60.6% represents more than 5000 rejected trades in the month. What this data highlights is that some LP's are clearly using the Last Look (and in particular the Additional Hold Time) to "select a portfolio" of trades.

CLIENT DATA SWEEP BOOK

Feb - 20	Round Trip Time		Client Data Sweep Book	
Sweep Book	Accepts	Rejects	Reject Rate	Mkt Share
LPA	30	75	12.0%	14%
LPB	100	100	2.4%	11%
LPC	50	50	2.6%	9%
LPD	125	125	3.1%	27%
LPE	10	20	15.1%	21%
XTX	10	10	3.0%	18%

Mar - 20	Round Trip Time		Client Data Sweep Book	
Sweep Book	Accepts	Rejects	Reject Rate	Mkt Share
LPA	20	10	57.6%	18%
LPB	100	100	3.6%	8%
LPC	10	10	5.7%	9%
LPD	75	75	60.6%	25%
LPE	10	10	15.6%	22%
XTX	10	20	2.3%	16%

Apr - 20	Round Trip Time		Client Data Sweep Book	
Sweep Book	Accepts	Rejects	Reject Rate	Mkt Share
LPA	20	20	8.2%	13%
LPB	100	100	3.7%	6%
LPC	10	10	7.6%	12%
LPD	75	75	2.3%	24%
LPE	10	10	4.0%	29%
XTX	10	10	1.0%	10%

Data provided by a client.

XTX believes this behaviour is inconsistent with the FX Code. The data also shows how difficult it is, even for sophisticated clients, to determine which of these LP's is providing a good service. It is impossible to ascertain the costs associated with each LP, and the total costs incurred, without using sophisticated independent analytical tools.

THE COST OF AHT

The total cost of execution is made up of two parts. The observable or “visible spread” and the invisible cost associated with AHT. But do we know what the total true cost of AHT applied across clients in the market is? No, it is almost impossible to calculate of course because there are so many moving parts - not least the visible cost but especially the invisible cost. *The very fact that LP's try so hard to retain variable AHT as a mechanism, and spend so much time obfuscating clients around the cost suggests it is considerable.* Later in this section, we will explain how we performed a multi-factor analysis of the cost but, for now, let's keep the discussion relatively simplistic. Does it actually cost clients anything or is this just in the way that the yield on a client's flow is apportioned between the LP's?

By definition, these mechanisms distort true price competition. In particular, the cost of the AHT applied and the Last Look mechanisms (rejection/acceptance) applied by some of the LP's in the March 2020 example above imply that the redistribution between LP's is very significant indeed.

Should clients care about this if it is just a redistribution of yield? Yes they should - ultimately real LP's with a genuine alpha based skew, risk capital and real interest to deal, can be crowded out and their value undermined and eroded. That can lead to the best pricing LP's being excluded from the mass of the client's flow and hence affect their ability to price the client in the future or when really needed. *That is neither good nor fair for the client as it will ultimately result in wider spreads.*

But hang on, is it really a zero-sum game? LP's often attempt to frame the discussion as a simple trade-off between spreads and rejections (visible costs and slippage costs). A Head of FX Sales at a tier-one bank explained in FX Markets that, for him, last look is not merely a risk control mechanism as defined in the FX Code but another commercial pricing lever to be used with clients.

“We have a dialogue about [how] this much pause [AHT] is going to relate to this much improvement in spread. Does that work for you? If it doesn't then we need to reapproach it. ... Similarly, who am I to say to a customer who comes to me and says, ‘I don't care if you reject me 80% of the time, I want 0.2 pip spreads?’ Who am I to say that he can't have that? As long as it is a transparent and open conversation then I think it can work because there can be mutual benefits going both ways.”

However, if one looks more deeply at this, we quickly realise that in a best-case scenario for the client, an LP can only price choice (i.e. the value provided to the client is bounded), but the value extracted by the LP from ever longer AHTs and higher rejections can grow in an unbounded manner.

In major instruments such as EURUSD and USDJPY, LP's are no longer able to reduce their 'Visible Costs' (spread) any further. If a client does not forcefully impose limits on AHT/Rejections, the client will over time see their true cost of execution rise as LP's extract additional value through slippage which cannot be passed back to clients through the spread. Slippage in major instruments represents an insidious tax on end client execution costs. In other words, this is not a zero-sum game because the costs imposed on the clients are unbounded, whereas the LP's costs are, by definition, limited.

It's important to remember that some LP's who do not use AHT (XTX is one but not the only one, e.g. JPM, HSBC) have to rely on pricing wider to defend themselves against more directional flow and so they will, by nature, win less trades and have the value of the uniqueness of their price eroded by those LP's using these tactics. Therefore, the outcome becomes self-fulfilling. The best LP's may be crowded out and those that maximise their yield using these mechanisms are able to cherry pick the flow that they want leaving the rest to pick up the trades the AHT wielding LP's do not want. And the client of course has the impression that those LP's are underperforming and their top LP's are those with the high market share using AHT to their advantage.

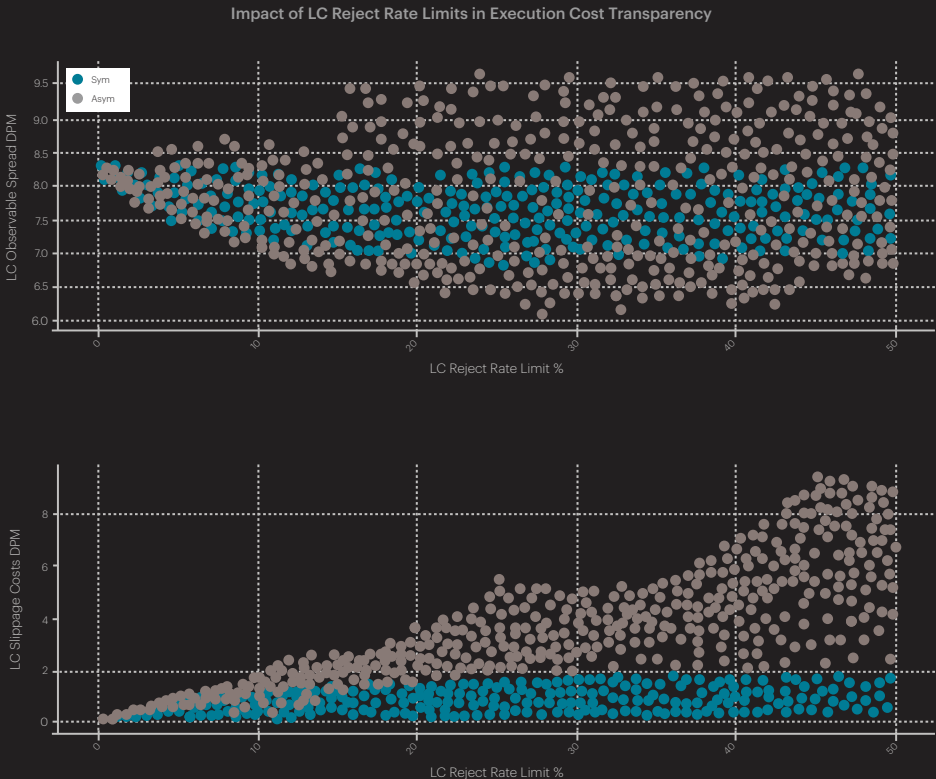
Whilst the above is all true, we did attempt to ascertain the true cost of AHT applied to client portfolios. The results are interesting.

Referencing the work by Oomen (2016)⁵ on pricing within an aggregator, and utilising real market data, we simulate pricing from multiple LP's and generate an aggregate pricing stack equivalent in nature to what a buy side client would observe. We then simulate the activity of a 'noise trader' and demonstrate that **a client's effective spreads are not constant given an allowed limit on rejections; meaning higher tolerable rejections lead to higher effective spreads**. The interested reader is referred to the Appendix for further details of the methodology.

⁵ Oomen, R. (2016). Execution in an aggregator. Quantitative Finance, [online] 17(3), pp.383-404. Available at: http://eprints.lse.ac.uk/67454/1/Oomen_Execution%20aggregator_2016.pdf.

THE RESULTS OF THE MODEL ARE SUMMARISED IN THE BELOW FIGURE WHERE EACH POINT REPRESENTS A UNIQUE SET OF STARTING CONDITIONS AND LIMITS THAT THE CLIENT PLACES ON THEIR LP'S AND DEMONSTRATE:

- 1) The change in visible spread: visible spreads shrink to a limit as rejections increase;
- 2) The change in slippage: slippage costs increase unbounded as rejections increase; and
- 3) The change in true cost: as a client tolerates higher levels of rejection their true costs will grow in a manner that is difficult to predict;



Impact of LC Reject Rate Limits in Execution Cost Transparency

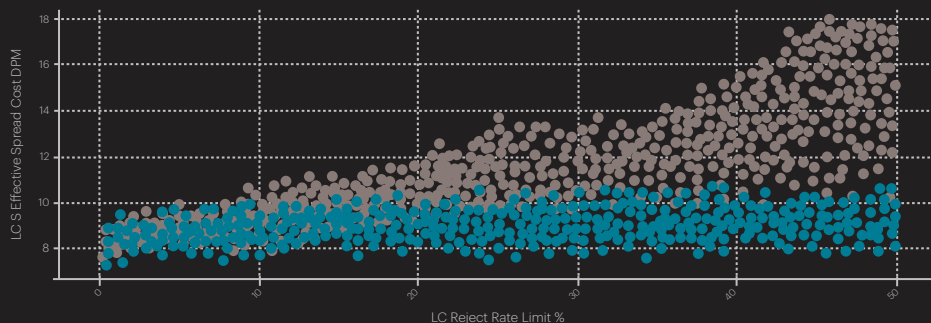


Chart: Simulation data produced by XTX using internal market data.
See Appendix for methodology.

Regardless, it is clear that the client's effective spreads are not constant given an allowed limit on rejections. Higher tolerable rejections lead to higher effective spreads as LP's extract unbounded value from slippage but are bounded in pricing non-negative spreads (especially in the case of asymmetric tolerances).

This demonstrates that the argument that client slippage costs are compensated for through spread competition amongst LP's is not strictly true.

Of course, it may well be that you think performing an analysis such as this ought to be unnecessary in a market which is "*robust, fair, liquid, open, and appropriately transparent*" (FX Code objective statement). Requiring a quantitative modelling PhD to understand the true cost of execution probably is not a feature one would associate with a market defined by those characteristics.

WHAT ABOUT SYMMETRIC AND ASYMMETRIC LOGIC?

The subject of symmetry emerges around treatment of price movements between the submission of the “offer to deal” and the completion of the price validity check (regardless of at what point that check occurs). Most LP’s allow a “price tolerance” i.e., an amount by which the LP’s price may have moved when checked against the price referenced in the “offer to deal”. That tolerance is usually measured in “pips” or fractions thereof, although it can be zero. Symmetric treatment requires that the price tolerance is applied equally regardless of the direction of the price change i.e., regardless of whether the trade has moved in the LP’s favour or against the LP. Such tolerance is often used to ensure that rejection rates are not excessively high and is an entirely reasonable and well intentioned practice. However, in certain circumstances, LP’s apply “asymmetric tolerance” - this means that the LP accepts trades that have moved beyond the tolerance when it is in their favour and they reject when it moves beyond the tolerance against them.

The first thing to note is that, when it comes to bilateral client flow, there is zero justification for using default asymmetric acceptance logic for what the FX Code defines as a “risk control mechanism”. Asymmetric logic simply means that the LP will fill the client at the original (now off-market) rate if the market moves in the LP’s favour; but reject the client if the market moves in the client’s favour. The reason asymmetric logic exists is that it is more profitable for an LP than symmetric logic on the same flow. This is zero-sum: the client loses precisely whatever the LP gains.

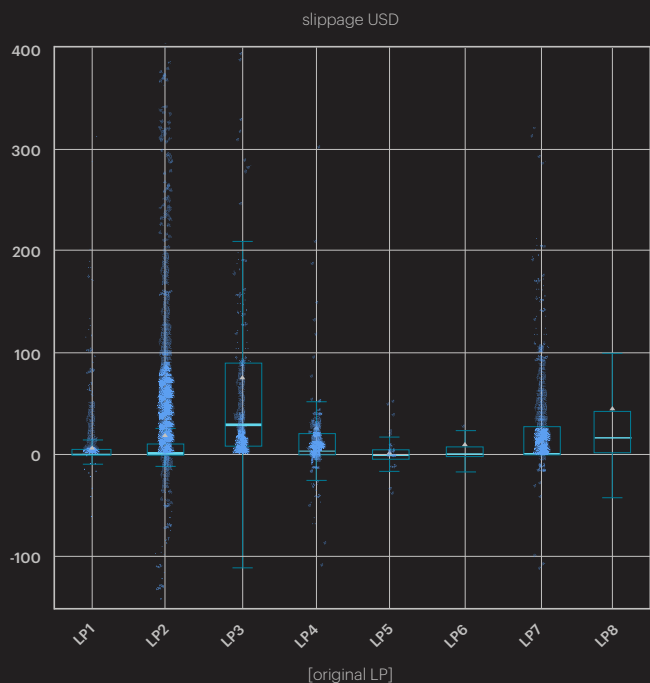
It would be one thing to allow a limited number of sophisticated clients to opt-in to asymmetric logic, following crystal clear disclosures and confirmation the clients have an ability to quantify the cost to them. However, certain top-tier LP’s still use asymmetric

logic as their default setting for clients. Yes, really. A client must specifically opt-out to get treated symmetrically! Do those clients have the analytic transaction cost tools to measure the impact on them, making a fully informed decision, or are they even aware?

Symmetric logic should be preferable to asymmetric logic from a client's point of view. However, even when the logic is applied perfectly symmetrically, it is not the case that the client outcome will be symmetric. The majority of rejects will still occur when the market has moved against the client, rather than in their favour. Symmetric logic or not, this is assuredly not a 50/50 scenario that balances out over time. There is an oft-stated refrain from LP's that "I use a long AHT to maximise the chances that I can fill the client at the end of the last look window". Let's see how likely that is to happen in reality.

Below is an analysis of reject data, provided by a large retail broker. The majority of LP's use symmetric last look logic and yet the outcomes are visibly skewed against the client. This means the slippage cost is far more often positive (if buying, the price moves higher when the client tries again after being rejected) than negative (if buying, the price moves lower when the client tries again after being rejected).

Boxplot grouped by original LP



Analysis produced by XTX, using data provided by a client.

The reason for this is that market movements post receipt of a trade request are not random: for various reasons the price movement upon receipt of a client trade request is likely to be in the same direction of the client trade request, on average. The takeaway is simple: even with symmetric logic, the client will have the wrong kind of slippage on average. The longer the AHT, the more the slippage and thus cost to the client grows and contrary to that oft stated refrain, the reality is that “the longer the AHT, the more likely the LP is to reject you”.

So now that we have ascertained that Additional Hold Times are too long, opaque and costly, and that the trading logic of asymmetric rejects is an arcane tool designed again for profit optimisation, let’s debunk some other myths.

MYTH 1 : REGIONAL BANKS CAN'T PRICE AS QUICKLY AS THE BIG PLAYERS. WON'T THEY BE CROWDED OUT?

Not at all. This is a red herring. One only needs to review the Risk Magazine article of 2019 “How the Top 50 dealers tackle forex last look”⁶ to see that banks such as ING, Santander, TD, Westpac, SEB, Credit Agricole (to name a few) all disclose clearly that they do not apply an AHT. If they can do it, why can’t some of the largest global LP’s?

In actual fact, the reverse is true - regional banks are crowded out by the large global AHT wielding players. Again, think of the logical extreme of the LP who prices choice and then only selects “winning trades”. That LP is distorting real competition, keeping valuable axes and skews offered by regional players away from clients through artificially saying “I have interest to deal on both sides of the market”, when in fact they “only have interest in trading on one side of the market” (they just can’t tell you which side yet.....).

The arguments around regional banks not being able to have such competitive technology and therefore needing some additional protections are also moot points. Regional banks are, in the broadest sense, trading with clients who are less technologically aware and less sophisticated than large institutional clients - i.e. corporates, insurers, standalone pension funds etc. These are not clients with heavily directional flow let alone “toxic” flow and again those that do have such clients should use spread as the means to protect themselves, not AHT.

⁶ Rega-Jones, N. “How the top 50 dealers tackle forex last look.” Risk.net., 08 Aug. 2019.
<https://www.risk.net/derivatives/6888556/how-the-top-50-liquidity-providers-tackle-forex-last-look>.

Take these disclosures from Santander⁷ which are clear and informative.

“Due to the different checks and controls being performed, as described above, an answer to a deal request may not be immediate. However, we wish to make it explicit that, for the purpose of performing price verifications under “Last Look”, Santander does not apply any pre-determined Holding Period, also sometimes referred to as ‘latency buffer’, and trade requests are confirmed as soon as all the relevant checks and controls have been performed. As a general matter, a Holding Period may involve holding a trade request for a prescribed time delay before the price check is performed (the ‘Holding Window’), in order to allow the liquidity provider to see the latest market data updates before applying the price check. Santander does not apply a Holding Window to any trade requests (Zero Hold Time) in respect of our spot FX business. The last look price and/or validity check will be applied as soon as our systems receive the trade request.”

Additionally, our suggestion in the conclusions of this paper that AHT be either removed completely or be capped according to the LP’s median “tick to trade” would level the playing field, allowing LP’s with genuinely more latent pricing-stacks to align their AHT with their own latencies.

MYTH 2 : LIQUIDITY PROVIDERS HAVE GLOBAL CLIENTS TRADING ON GUIs VIA THE INTERNET WITH LOTS OF LATENCY. ISN'T THAT A REASON FOR AN AHT?

This is also a red herring.

It is true that there is natural latency when a client clicks on a GUI in another part of the world. However - even if that trade request takes an hour to arrive - once it has arrived at the LP’s server, it doesn’t take any longer to compare this trade request’s reference price against the current price on the server than it would for any other order. **The AHT clock only starts ticking once the offer to deal has been received by the LP.**

Because this is the most benign flow that exists, most LP’s will simply add tolerance. This means that orders may be filled even if, due to natural latency, they are slightly off market - in either direction - vs the current reference price. This of course does not require an artificial hold time. Truly global banks with large GUI franchises across connectivity challenged regions such as HSBC and JP Morgan both seem to manage this without AHT.

⁷ <https://www.santander.com/en/landing-pages/foreign-exchange-disclosure-notice>

MYTH 3 : LIQUIDITY PROVIDERS NEED AHT TO TRADE ON ANONYMOUS VENUES

What about where the LP does not know the end client or even believes that the person they are trading with is NOT a client. This occurs on ECNs. All ECNs allow the LP's the right to use an AHT and employ last look. Again, there is no issue with the fundamental right of any LP to determine when a trade is executed, that is, as explained above, an expected and necessary behaviour in OTC markets. The real question is the validity of the AHT. In the case of ECN trading, not every price stream is bespoke and multiple "clients" or "tags" are often connected to the same price stream from the LP. This is different to bilateral trading which is done 1-2-1. As a result, there is some justification for the AHT. It helps LP's to price tighter on the stream than the client with the worst flow deserves whilst allowing protection against those worst clients.

Nonetheless, the ultimate protection mechanism does exist on ECNs - that is removing certain tags from an LP's pricing stream(s). This can be done pretty simply and easily. And of course, the old solution of pricing those tags wider, or that whole stream wider, also exists.

Whatever the rights or wrongs of using AHT on an ECN, there is no doubt that all LP's across the ECN could be held to the same AHT (or none) and that pricing by LP's would therefore be like-for-like and truly comparable. Levelling the playing field in this way would, at a minimum, be beneficial for the transparency of pricing. We know that some ECNs, notably CBOE FX, but others also, have been shortening the maximum hold time that LP's can use over the last 1-2 years; this is a good start but those times could be significantly shorter.

Alternatively, the ECN could, as a neutral intermediary, apply the price check on behalf of the LP, using their latest advertised price. This removes one of the biggest conflicts of last look as the LP would only be made aware of a trade request once it had passed this independently applied price check. Should a trade request fail the price check - i.e. the LP's price had moved and was no longer available - the LP would simply not be sent the trade request. LP's would retain the right to perform validity checks and a credit check before confirming execution of the trade request but these are not controversial and account for a tiny proportion of rejects.

MYTH 4: PRICE IMPROVEMENT

Some Liquidity Providers suggest that they use AHT to be able to offer Price Improvement to their clients. However, price improvement is unintuitive. It sounds like a good thing. Imagine a client sends a buy trade request in a 10 / 12 market. Before the request is received by the LP, the market ticks down to 9 / 11. The LP price improves and fills the client at 11 rather than 12. This sounds helpful! However, there is really no guarantee that the same LP who had the best offer in the 10 / 12 market is still the best offer in the 9 / 11 market. The fairest thing to do is reject the trade request and let the client poll all of their providers again and select the best price, rather than fill a captive trade request at what may not be the best price available.

There is also a danger that price improvement as a practice becomes acceptable - this runs altogether new risks similar to those which exist in US retail equities. The practice of price improving the client does not mean that the end client (quite possibly a retail customer) will necessarily benefit (or indeed see the best price that was actually available to them at the time of trade). It also becomes all too easy for large players to use their scale advantage to offer price improvement across a whole portfolio of trades, with a lack of clarity that the improvement ever benefits downstream clients, especially retail customers. This is a slippery slope that we think would be well avoided in the FX market.

If clients wish to receive the market price at the time of LP receipt of the offer to deal, they should send a market order and get filled at the prevailing rate regardless of whether or not it is an improvement versus the market rate at the time of the attempted execution.

CONCLUSIONS

We hope this paper has made you better informed on the topic of Last Look and how LP's are using it in a manner that we interpret is inconsistent with the FX Code. It is worth remembering this statement from Principle 17 of the FX Code:

"A market participant should be transparent regarding its Last Look practices in order for the client to make an informed decision as to the manner in which last look is applied to their trading"

Do you really believe, given the above, that this standard is being satisfied from your perspective?

We would make the following very simple recommendations to improve the situation and which we believe all LP's should feel comfortable applying. This would help make trading practices fair and efficient for all market participants, as well as meet the objectives of the FX Code outlined above. It would also mean that many of the issues with regard to disclosures would simply vanish as they would no longer be applicable.

1. DEFINE Last Look correctly by splitting the definition between Last Look per se (the right of the LP to be the contract determinant) and the AHT applied by LP's.
2. Update the FX Code to specifically PROHIBIT the use of AHT for any reason related to profit optimisation, specifically including for the avoidance of adverse selection and the management of market impact.

Our suggestion would be to include wording along the following lines:

"LP's should remember that the client has an unknown liquidity exposure during the Last Look window and therefore aim to minimise this period of uncertainty. An offer to deal should only be held for such period that it takes for the LP to perform the price check, confirming the price at which the trade request was made against was consistent with the price that would have been available to the client at receipt of offer to deal."

We recognise that this very clean definition, which does not allow for any Additional Hold Time, may not be acceptable to the consensus of market participants, so perhaps we could also suggest a slightly weaker language which aims to allow LP's to see a single price update before they perform the price check. This HAS to be related to

the time that it takes an LP to see one further price update - this price update would not represent a price change, merely the next price update. This means that LP's and clients are protected where the market does not update for considerable periods of time, e.g. Scandinavian currencies outside European hours. As referred to earlier in the document, we call this the "median tick to trade". The suggested wording then becomes:

"LP's should remember that the client has an unknown liquidity exposure during the LL window and therefore aim to minimise this period of uncertainty. A trade request should only be held for such period that it takes for the LP to perform the price check, confirming the price at which the trade request was made remains consistent with the current price that would be available to the client. Where a client's pricing stream updates at a slower rate than the LP receives external reference market price updates, (e.g. throttled streams), it may be reasonable for an LP to wait until it receives the next external reference market update following receipt of the trade request before completing its price check."

Given the problems of creating a truly clean definition that cannot be easily gamed or "mis-interpreted", we also feel it would be helpful for the FX Code to add some examples of what would represent good market practice and what would not. Such clear guidance would enable Compliance departments to more easily ensure that the practices of LP's are in line with the spirit of the Code.

So, an example of good market practice based on the first wording above could read:

"LP takes 5 ms to receive an external market data reference update. Trade request is received at T+0. Price check is begun always at T+0 regardless of market data update speed etc and all checks are undertaken against the current price regardless of whether or not the market data reference update has been received."

This approach has the advantage of being very clean and there is no requirement for additional disclosures. Everyone would play by the same rules and there would be no feasible use of Last Look as a commercial tool, as the Code originally foresaw. Example for the second wording could be along these lines:

LP takes 5 ms to receive an external market data reference update. Trade request is received at T0, trade should be checked as soon as practical, so hold time would be limited to the time taken to complete the price check.

In the event that the client's stream updates at a rate slower than every 5ms, any hold time should be limited to a maximum 5ms and the trade request price should be checked against the price which the LP would then make for that client. Any

hold time in excess of 5 ms in this example would be considered to be an example of an unacceptable practice. The price check process should ideally be commenced immediately at T_0 but certainly no later than $T_0+5\text{ms}$.

In the event that the LP updates the client price in a time less than 5 ms, the price check should be commenced immediately as the price is updated. By definition, this time will be before $T_0+5\text{ ms}$.

In the event that the LP updates the client price in a time greater than 5 ms, the price check should be commenced no later than $T_0+5\text{ ms}$. This covers the situation where the client pricing is throttled by the technology provider or the LP themselves.

Additionally, under both wordings, the following could be used as an example of what would be considered unacceptable/acceptable/best market practice:

LP takes 5 ms to receive an external market data reference update. Trade request is made at T_0 . Any check which involves the price check starting at a time greater than $T_0+5\text{ms}$ would be considered UNACCEPTABLE and only immediate price checks would be considered BEST market practice.

For example, trade request received at T_0 , credit and operational check started immediately. Credit and operational checks complete at $T_0+25\text{ms}$. Potential practices would be:

*Price check is commenced at T_0 but completed multiple times up to completion of credit and operational checks. This would equate to an undue hold time.
UNACCEPTABLE*

*Price check is commenced at $T_0+25\text{ms}$. This would equate to an undue hold time.
UNACCEPTABLE*

*Price check is commenced between T_0 and $T_0+5\text{ms}$. NOT BEST BUT
ACCEPTABLE*

Price check is commenced at T_0 . BEST

These examples hold for all clients and all LP's - if LP's have slower market data updates than others, they can utilise a slightly longer hold time so that they are not simply picked off but for all major participants, the additional hold time would be practically removed. This would be a very positive development for the market whilst allowing regional banks and those without the fastest market data to continue to participate effectively.

In addition, we would suggest the following simple but very important wording be added:

There is no justification for commencing price checks at different times based on the client identity or flow type. To do so would be inconsistent with acceptable market practice.

The FX Code should make a clear statement that profit optimisation issues of adverse selection and managing market impact should be handled using spread.

Of course, we accept as would everyone we believe, that there may at times be exceptional circumstances where any LP is unable to guarantee to meet these thresholds, e.g. if they have technology issues etc but that should not preclude an outcome that mirrors these suggestions.

3. DEFAULT deal acceptance logic to be SYMMETRIC.

Thank you for taking the time to read this paper. We would urge you to file a public comment letter (https://www.globalfx.org/docs/Draft_GFXC_Last_Look_Guidance_Paper.pdf) by May 31st 2021. If you don't have the time or resources to do this then we would ask you to complete a simple, confidential survey (answers will only be used in aggregate) – there are only eight questions and it is multiple choice.

Click here to complete: [\[Survey: Last Look and Additional Hold Times\]](#)

[\[Click here for XTX eFX disclosures\]](#)

IMPORTANT NOTICE

This Document is issued by XTX Markets Limited (“XTX”), which is authorised and regulated by the Financial Conduct Authority of the United Kingdom (the “FCA”), with FCA FRN: 711945. XTX is a private limited company incorporated in England & Wales, with company number 09415174. XTX’s registered office and principal place of business is R7, 14-18 Handyside Street, London, N1C 4DN, United Kingdom. This Document may reference products that are not FCA regulated, which include, but are not limited to, Spot FX. XTX endorses and has publicly stated its commitment to conducting its FX market activities in a manner consistent with the FX Global Code.

The distribution of this Document may be further restricted by law. No action has been or will be taken by XTX to permit the possession or distribution of the Document in any jurisdiction where action for that purpose may be required. Accordingly, the Document may not be used in any jurisdiction except under circumstances that will result in compliance with any applicable laws and regulations. Persons to whom the Document is communicated should inform themselves about and observe any such restrictions. This Document is for information purposes only. The Document does not constitute an offer to transact in, or the solicitation of an offer to transact in, securities, derivatives, FX transactions, or other financial instruments, in any jurisdiction.

Although the information in this Document is believed to be materially correct, no representation or warranty is given as to the accuracy of any of the information provided. Certain information included in this Document is based on information obtained from sources considered to be reliable. However, any projections or analysis provided to assist the recipient of this Document in evaluating the matters described herein may be based on subjective assessments and assumptions and may use one among alternative methodologies that produce different results. Accordingly, any projections or analysis should not be viewed as factual and should not be relied upon as an accurate prediction of future results. Furthermore, to the extent permitted by law, neither XTX nor its employees, directors, officers, shareholders or service providers assumes any liability or responsibility nor owes any duty of care for any consequences of any person acting or refraining to act in reliance on the information contained in this Document or for any decision based on it. Past performance cannot be relied on as a guide to future performance.

APPENDIX

METHODOLOGY APPLIED TO ASCERTAIN TRUE COST OF AHT TO CLIENT PORTFOLIOS

In order to determine this, we model a typical last look protocol and allow LP's over time to adjust their settings (spreads, AHT, price check tolerance) to account for gains/losses on the resulting accepted trades. Trade requests generated by the client are routed to the LP with the prevailing best bid/(ask) in their aggregator. The LP then applies AHT to control for adverse selection and/or market impact, and once this period has elapsed the LP performs a price validity check against the most recent price available to the client.

If the client allows for asymmetric acceptance criteria, the trade will be accepted if the price movement (requested price vs current price) is within a tolerance/direction that favours the LP. If the client prohibits asymmetric acceptance, the trade will be accepted if the price movement is within a tolerance regardless of direction.

If the trade is rejected another trade request is generated and routed to the LP with the newly available best bid/(ask) in their aggregator, the price level of the original request is cached to calculate the subsequent slippage/effective spread.

As the model runs, effective spread earned by each LP is monitored, LP's suffering from a negative/(positive) effective spread will act in their own economic interest and will perform one of the three below actions:

- Widen pricing; reduces market share but allows the LP to accommodate greater adverse selection in the aggregator
- Decrease trade acceptance tolerance; reduces the likelihood of an individual trade request being accepted
- Increase AHT; increases the option value available to the LP at the point of price validity check (end of AHT)

This process continues until the simulation reaches an equilibrium state whereby LP's see no marginal benefit from modifying their settings.

