The Hidden Cost: Reference Price Latencies

Trading Review and Analysis
Investment Industry Regulatory Organization of Canada

March 9, 2016

Authored by:
Lisa Anderson, Team Lead Analytics, IIROC
Baiju Devani, Director Analytics, IIROC
Yifan Zhang, Data Scientist, IIROC

Acknowledgements:
We thank Corey Garriott and Adrian Walton of the Bank of Canada, and Andreas Park of the University of Toronto for their comments and suggestions. Any errors and omissions are our own.

1. Summary of Findings

This study explores the extent of reference price latency on dark trading venues in Canada and finds that some marketplaces exhibit latency regularly. This results in stale execution prices and latencies that are large enough for fast traders to have a disproportionately large number of trades benefiting from stale prices. While the overall economic impact from such latencies is small, participants may be unaware of the latencies on certain marketplaces which impact their trading costs.

We find that 4% of all dark mid-point trades are executed at stale prices with overall average latencies estimated at 4.8 milliseconds. Observations of latency are not isolated to a single marketplace. However, different marketplaces exhibit variations in both the duration of latency as well as its economic impact. The cost imposed due to this inefficiency is on average $2,969 per day for the TSX60 securities (equivalent to $748,188 per year) in the recent period, and has historically been as high as $5,000 per day. These absolute dollar costs are concentrated in two markets. When measured in relative terms, four marketplaces show high relative costs due to latency.

If these latencies were not large enough to be actionable, then we would expect to see equally distributed costs between active and passive participants. However, we observe that the latency costs are not equally distributed, with 88% of latency-impacted trades favouring the active participant. Further, of all the latency-impacted trades where the active-side benefits, high frequency traders (HFTs) are the dominant beneficiaries and receive 83% of total benefit. When we calculate the latency costs of stale-priced trades as a percentage of the expected half-spread cost of all trading by a participant group, we find that interacting with an active HFT order in the dark with a latency impacted trade will cost 13.1% over the expected half-spread cost (relative latency cost). The majority of dark mid-point trades are between non-HFTs and thus overall relative latency costs represent only 3.1% of the total expected half-spread costs to all participants.

We also shed some light on the potential sources of latency. We find that the number of order messages immediately before a trade is correlated with the duration of latency. In addition, marketplaces' use of co-location services is associated with reduced latency.

2. Introduction

Low latency calculation of the national best bid and offer (NBBO) based on market feeds is crucial in order for participants and marketplaces to react to rapid changes in market conditions. However, not all observers have the most up-to-date view of the NBBO. The time taken for a participant to view the NBBO is affected by their technology for gathering and processing data and their physical proximity to other exchanges.

The latency arbitrage we study is dark execution price latency arbitrage. The price of a dark mid-point trade is computed using the dark marketplace's observation of the NBBO, which is obtained either through direct feeds from lit markets or the IP feed. Latency can result when a dark market has delays in obtaining or processing the feeds or in calculating the execution price. Unlike traditional latency arbitrage where the arbitrage is possible because slow traders directly rely on a delayed IP feed, dark trade execution price latency arbitrage results from inefficiencies at the marketplace level. Interaction between fast and slow traders is now further impacted by a slow marketplace. Fast traders who are able to see the newest NBBO (not yet seen or processed by a dark market) can have their order filled at a stale price before prices move away from them. To take advantage of the latency, fast traders need to both derive the NBBO faster than the marketplace as well as be aware of latencies in marketplaces. These mispriced trades can have a meaningful impact on certain dark market participants. This has recently been discussed in the context of trading by a single buy-side client on US dark marketplaces (Biais, Foucault and Moinas 2014; Easley, Hendershott and Ramadorai 2013; Foucault, Hombert and Rosu Forthcoming) and points to potential systemic latency issues.

To study the phenomenon of dark mid-point trade latency, we:

1. Quantify both the duration of such latencies as well as the scale of its economic impact.
2. Provide some insight into participant level economics and examine the evidence of systematic interaction with marketplace inefficiencies associated with dark mid-point trade latencies.
3. Explore the potential drivers for such latencies, such as the use of co-location services, and how busy the marketplace is prior to a latency-impacted trade.
4. Present a historical view of latency duration and economic impact (over a three year period) and highlight periods where there was significant deterioration.

3. Data and Institutional Details

All of the Canadian equity marketplaces provide trading data in a standardized form, the Financial Information eXchange (FIX) format, to IIROC via a real-time regulatory feed. This data includes all trade, order and quote messages, and contains both publicly available fields and confidential regulatory markers. The regulatory data is used by IIROC's real time surveillance system and is also stored in the Equity Data Warehouse (EDW) database for post trade surveillance and analysis.

All event messages have 1-millisecond timestamp granularity as of October 15 2012. A market-specific sequence number is also provided, which allows for the correct sorting of events for one marketplace within a 1 millisecond interval. However, the order of events from different markets within the same 1 millisecond interval cannot be determined.

The sample period is October 15, 2012 to August 31, 2015. We conduct a historical analysis of dark market latency using the sample period. A subsample from July and August 2015 is used to evaluate economic costs and beneficiaries of this latency, and to shed some light on its potential sources. In-depth analysis is possible in the July-August subsample because mid-point trades can be more reliably identified for most marketplaces from July 2015 onwards due to the inclusion of a new field from the trade message in the EDW.

The analysis focuses on the TSX60 securities which represent 57.7% of the traded value of all securities in the July-August subsample. The analysis focuses on dark trades which take place between 9:35 am and 3:55pm. Intentional crosses, odd lot trades, and trades during crossed market conditions are excluded. Further, all call-market trades are excluded as they do not have an active side.

Dark order types priced at the mid-point are available on both fully dark and dark plus lit Canadian equity marketplaces (see Table 1). For the purpose of this study, we have anonymized the identity of the six marketplaces which employ mid-point pricing and trade TSX securities (labeled M1 through M6).

4. Identifying latency impacted trades: methodology

i. Identifying dark mid-point trades

We identify dark mid-point trades using the following criteria. In the historical period, we include trades whose prices are an odd multiple of a half-cent[1]. With this method, we eliminate non-mid-point dark trades, though at a cost of omitting some actual mid-point trades with prices which are even multiples of a half-cent. The number of trades identified using this method reflects a lower bound of the true number of mid-point trades. In the July-August subsample, we identify mid-point trades on certain marketplaces using a field available on the trade message (as described in Section 3). This methodology is more precise in identifying mid-point trades as it includes trades with prices which are an even multiple of a half-cent. These trades form the group of dark mid-point trades which we use to quantify latency.

ii. Identifying trades impacted by latency

We then search for latency-impacted trades from the identified dark mid-point trades. The logic used to identify latency-impacted trades is outlined in Figure 1. The steps are as follows:

Step 1. Identify stale-price trades

When we analyze all identified mid-point trades, we would expect both the active and passive sides of such trades to cross half the spread. We identify trades which are executed at prices that are not at the mid-point of the most recent NBBO as stale-price trades. For additional details see Appendix A. We categorize the stale-price trades as follows:

- Active-benefit trade: In this case, the stale execution price favours the active participant as the current mid-point has moved away from the active trader. The active trader receives an execution price better than the current mid-point price (buyer buys lower than the current mid-point, seller sells higher than the current mid-point)
- Passive-benefit trade: Analogous to the previous scenario, the stale execution price favours the passive participant as the current mid-point has moved away from the passive trader.
- Inconclusive: No quote is found to explain the mid-point trade price.
- Timestamp Granularity Limitations (TGL): When events from more than one marketplace are stamped with the same 1 millisecond time interval (the smallest time increment reported to IIROC

[1] The general pattern of an odd multiple of a half-cent is $x.xx5 (for example $1.235).

on the regulatory feed), we define that time interval to be subject to timestamp granularity limitations. Under these conditions, events on different marketplaces cannot be sorted reliably.

Step 2. For the stale-price trades, search for the reference quote

Starting from the trade event at time t_{trade}, search back 0 to 100 milliseconds for a mid-point that equals the trade price. We refer to the most recent quote whose mid-point is equal to the trade price as the reference quote (at time t_{ref}) (for illustration see Figure 2). Then,

- If no quote is found → end the search and label the trade as inconclusive.
- If a reference quote is found and $t_{\text{trade}} - t_{\text{ref}} < 1$ ms → label the trade TGL.
- If a reference quote is found and $t_{\text{trade}} - t_{\text{ref}} \geq 1$ ms → perform Step 3.

Step 3. Calculate the latency

Starting from the reference quote found in the previous step, search forward to the first mid-point (t_{update}) that does not equal the reference quote (for illustration see Figure 2). More specifically:

$$t_{\text{ref}} \leq t_{\text{update}} \leq t_{\text{trade}}$$

Then,

- If $t_{\text{trade}} - t_{\text{update}} < 1$ ms → label trade TGL
- If $t_{\text{trade}} - t_{\text{update}} \geq 1$ ms → label trade as non-TGL stale-price trade and calculate the latency as:

$$Latency = t_{\text{trade}} - t_{\text{update}}{}^{2} \tag{1}$$

We can define the true latency ($Latency_{\text{true}}$) as the time between the establishment of the new NBBO to the dark market's view of it – see Figure 2 for illustration. However, this true latency cannot be observed because we do not know how much longer the dark market continued to use the stale quote after the observed trade. In this study, $Latency$ is a proxy of the lower bound of the true latency. The relation between the two values is $Latency_{\text{true}} \geq Latency$.

iii. Define latency costs

The mid-point is often considered to be the reference for assessing costs. Under this scenario, a mid-point trade price would be considered to be zero cost to both participants. For our analysis, we calculate the cost to each participant relative to the bid (ask) for the buyer (seller). In this case, trading at a mid-point trade price would result in a cost of half the spread for both participants. The total cost for a midpoint trade for the buyer is

$$TotalCost_{it} = (P_{it} - bid_{it}) \times volume_{it} \tag{2}$$

P_{it} is the trade price for the *t*-th trade in stock i (which may be based on a stale mid-point reference price). bid_{it} is the current bid for the *t*-th trade in stock i . $volume_{it}$ is the trade volume for the *t*-th

[2] Multiple quote updates can occur after the reference quote. The first update quote is referred to as the first update quote and the last quote before the trade is referred to as the last update quote. In this case, the term t_{update} corresponds to the time of the first update quote.

trade in stock i . Under normal (non-latency impacted) situations, the total cost for a mid-point trade is equal to half the bid-ask spread. To calculate the total cost for the seller, subtract the price from the current ask.

The total cost can be decomposed into two parts, the expected cost, and the latency cost. The relationship is

$$TotalCost_{it} = ExpectedCost_{it} + LatencyCost_{it} \tag{3}$$

The expected cost for a midpoint trade for the buyer is

$$ExpectedCost_{it} = (M_{it} - bid_{it}) \times volume_{it} \tag{4}$$

M_{it} is the current mid-point price for the t-th trade in stock i . This cost is always half the current spread. To calculate the expected cost for the seller, subtract the current mid-point from the current ask.

The latency cost for a buyer is

$$LatencyCost_{it} = (P_{it} - M_{it}) \times volume_{it} \tag{5}$$

To calculate the latency cost for the seller, subtract the current price from the current mid-point. Under normal (non-latency impacted) situations, P_{it} and M_{it} would be the same, resulting in a zero latency cost. A negative latency cost is interpreted as a benefit. When the buyer experiences a cost, the seller receives a benefit, and vice versa.

We are interested in the percentage of latency costs relative to the expected costs for all mid-point trades. The percentage is calculated as

$$RelativeLatencyCost = \frac{\sum LatencyCost_{it}}{\sum ExpectedCost_{it}} \tag{6}$$

The value of $RelativeLatencyCost$ can be interpreted as the extra cost that latency-impacted traders pay as a result of stale mid-point prices, relative to the cost the traders were willing or expected to pay for all of their mid-point trades.

5. Results

In this section, we study the scale, impact, and sources of latency using the July-August subsample. At the end of this section, we provide a history of market latency costs using the entire sample. The trader segmentation used in this study is outlined in previous work by IIROC (2014).

iv. Scale, Impact and Sources of Latency

In the methodology section, we defined the economic value of the latency-impacted trades in both absolute terms (measured in dollar amount) and relative terms (measured as the percentage of the expected trade cost). In the proceeding discussion, we explore the scale, impact and sources of latency in the July-August subsample. This data set provides us with a more accurate identification of mid-point trades as described in Section II.

The scale

Table 2A presents summary statistics which provide context for understanding the scale of the stale price trades over the July-August subsample. Stale-price trades account for 4.0% of dark mid-point trades. Active traders benefit from stale-price trades in 88% of all instances. The summary statistics by average traded value (Table 2B) show similar results.

The observation that active traders benefit more frequently from latency-impacted trades is strongly suggestive that the underlying latencies were observable and predictable enough for latency sensitive participants to take advantage of them. Otherwise we would expect that the latency cost would be evenly distributed between active and passive participants. We explore this idea in more depth below.

The asymmetrical cost/benefit flow

Broken down by trader segment

We have observed above that the latency-impacted trades fall into two buckets: the larger active-benefit group and the smaller passive-benefit group. When looking at active-benefit trades, the passive side pays the cost and the active side receives the benefit of the latency. Similarly, when looking at passive-benefit trades, the benefit is paid to the passive side, and the cost is paid by the active side. As discussed in section III, this benefit/cost is incremental to the half-spread that the participants expect to pay.

Table 3 and Figure 3 shows the benefits earned for the active side of active-benefit trades and the passive side of passive-benefit trades, by trader and counter-party[3]. Although HFT are the active traders for only 32% of all dark mid-point trades they often receive an advantageous stale price. Of the segmented traders who receive a benefit by trading actively at a stale price, HFT take 83% of the benefit, mostly at the expense of non-HFT (97%) rather than other HFT (3%). In contrast, active HFT rarely trade at a disadvantageous stale price, and pay only 11% of all costs when passive traders benefit by trading at a stale price. When non-HFT trade actively, their costs and benefits are of a similar scale.

HFT receive the largest benefit from the relative latency costs observed in dark mid-point trading. When HFT trade actively against non-HFT, the passive non-HFT traders pay an additional 13.1% above the expected cost to trade at mid-point. In contrast, when non-HFT trade with other non-HFT, the relative latency costs amount to an additional 1% above the expected costs.

To be precise, these costs/benefits arise due to the marketplaces' inefficiency in processing and referencing the most up-to-date quote. Our results show that HFT participants can reliably act upon such marketplace pricing inefficiencies. Table 4 shows that when active HFT trade and receive a benefit, they are able to take advantage of the opportunity within 2.9ms on average.

[3] We categorize any trades involving the un-segmented user group as "other". 70% of the monthly traded value during our study period is identified by the HFT and non-HFT user segments.

The overall average daily benefit received by active traders trading at an advantageous stale price is \$2,969 per day in TSX60 securities[4]. Over the recent two month period, this adds up to \$62,349 per month, equivalent to an estimated annual cost of \$748,188. If we were able to expand the identified mid-point trades (for example, by including dark trades with non-fractional price improvement or those trades which cannot be fully attributable to latency due to TGL), we would likely see an increase in the total dollar benefit. The later section on historical latency will show the lower bound of benefits throughout the historical period.

Broken down by marketplace

The benefit flow varies for different marketplaces; see Table 5 and Figure 4 for the results of active-benefit trades by marketplace. Two marketplaces exhibit high total dollar costs to passive traders (and equivalent benefits to active traders). A total of four marketplaces exhibit high relative latency costs (ranging between 4.2% and 2.2% of expected half-spread costs). Such variation in costs across marketplaces may be correlated to multiple factors such as: total trade value, number of trades, volatility, and trader strategies dependent on market conditions.

The variation in costs is also impacted by latency duration. When the latency is not long enough for any trader to react to, we would expect to see a symmetrical distribution of active-benefit and passive-benefit trades. When the latency is large enough for faster traders to act upon consistently, asymmetrical benefit flow would be observed (as reported in our study). Therefore, we consider the length of the true latency as the key underlying factor contributing to the scale of latency-impacted trades. Table 6 shows the variation of our observations of latency duration across different marketplaces. Exploring the factors that correlate to latency forms the focus of our next discussion.

Exploring sources of latency in dark mid-point trading

We use a regression analysis to explore the relationship between the observed duration of the latency and a number of factors (using both passive-benefit and active-benefit latency-impacted trades), including:

1) How busy the respective marketplace was in a short window prior to the stale-price trade[5]
2) Marketplaces use of co-location to the TSX[6]

The regression model is as follows:

$$Latency_{it,h,m} = \alpha_i + DCOL_m + \beta N_{it,h} + \sum_{j=1}^{3} \gamma_j \, Control_{j,it} + \varepsilon_{it} \quad (7)$$

Here, $Latency_{it,h,m}$ is the observed latency (as defined in Section III) measured in milliseconds for the trade h of stock i on date t at market m. $N_{it,h}$ is the proxy of how busy the overall market is at the time of the stale-price trade[7]. It measures the number of order messages (for all Canadian equity securities)

[4] We include trades by the unsegmented group to calculate total average daily costs.
[5] Proxied by the number of orders, order amends and order cancels across all Canadian traded equity securities.
[6] The majority of trade execution and quote information originates from the TSX.
[7] For robustness we calculate but do not report $N_{it,h}$ using various time frame durations. Both methods yield similar results.

just prior to the h-th trade of stock i on time t on all marketplaces. This metric is computed for the time frame t_{update} - 20 ms to t_{update}. $DCOL_m$ is the co-location dummy variable that takes value of 0 for marketplaces which are not co-located with the primary exchange and 1 for those which are co-located. α_i is a set of stock fixed effects. The control variables include daily trade value of all marketplaces, security price, and time of day at the time of the trade to account for any trends throughout the day.

The results are shown in Table 7. We find the use of co-location is negatively correlated with the latency. This may indicate that marketplaces without collocation with the primary exchange may take more time in receiving the quote update from the major exchanges. Latency and the level of activity of the marketplace (as measured by the number of orders entered just prior to the latency-impacted trades) are positively correlated with statistical significance. This result indicates that high numbers of order messages result in measurable latencies. We intuit that large delays are more likely to happen when the number of messages arriving per unit time exceeds the technological capacity of the marketplaces.

v. Historical Latency

To provide an estimate of the historical latency, we use the full sample period which includes all latency-impacted trades (both active-benefit and passive-benefit). Figure 5 shows that over this period, the percentage of dark trading has increased steadily over time. Figure 6 illustrates the changes in observed dark mid-point trade latencies and the resulting economic impacts since October 2012. The mean latency over the 35 month period is 4.3 milliseconds, which translates into a cost (or benefit) of approximately $42,500 per month. As discussed in section III, this estimate represents a lower bound of the true cost due to limitations in our observation capabilities (millisecond granularity) and the availability of necessary annotations in the trade data.

We also observe that there is significant variation in latency and its economic impact across the three-year observation window, with some periods showing significant deterioration. The months of December 2012 and January 2013 are notable for exhibiting large average latencies on all marketplaces. This period, highlighted in red in Figure 3, likely corresponds to technology or co-location changes at particular marketplaces. The months between October 2012 and April 2013 exhibit high economic costs relative to most other periods. Total monthly costs in this period peak at about $100,000.

6. Conclusion

Our goal was to investigate reference price latencies in Canadian dark marketplaces. The results show that while latencies do exist, the economic costs relative to overall trading costs are not high. We find the benefit from stale pricing, albeit small, is distributed disproportionally to active HFT participants.

We also show that co-location and ability to deal with large message volumes are correlated with latency. These could be addressed via technology investments by marketplaces to ensure that reference prices remain up-to-date in all market conditions. It is debatable if the resulting benefit to dark marketplace participants would offset the cost of technology investments. We believe this is an important debate and that our work will inform this conversation.

7. Tables

Table 1: Canadian marketplaces and order types

Marketplace	**Lit/Dark**	**Dark Mid-point/Peg**
Aequitas NEO Exchange – Lit Book	Both	
Aequitas NEO Exchange – Neo Book	Both	
Alpha Exchange Inc. (Alpha)	Both	✓
Chi-X Canada ATS Ltd. (Chi-X Canada)	Both	✓
CX2 Canada ATS (CX2)	Both	✓
CNSX Markets Inc. – Canadian Securities Exchange (CSE)	Lit	
Instinet Canada Cross Ltd. (ICX)	Dark	
Liquidnet Canada	Dark	
Lynx ATS	Lit	
MATCH Now	Dark	✓
Omega ATS	Both	✓
Toronto Stock Exchange (TSX)	Both	✓
TSX Venture Exchange (TSXV)	Both	✓
TMX Select	Lit	

Table 2A: Daily average trade count statistics

Trade Category	Passive-benefit trades	Active-benefit trades	Total trades
Stale-price trades	253	1,828	2,081
TGL	173	993	1,166
Inconclusive	417	184	601
Correct trades			48,141
Dark mid-point trades			51,989

Table 2B: Daily average trade value statistics (million $)

Trade Category	Passive-benefit trade volume	Active-benefit trade volume	Total trade volume
Stale-price trades	2.5	14.7	17.2
TGL	1.6	7.6	9.2
Inconclusive	6.5	1.9	8.4
Correct trades			314.1
Dark mid-point trades			348.9

Table 3: Daily average latency benefit broken down by trade counterparty

active	passive	active-benefit trades		passive-benefit trades		totals	
		benefit to active trader ($)	trade value ($1 million)	benefit to passive trader ($)	trade value ($1 million)	expected cost ($)	trade value ($1 million)
HFT	non-HFT	1,353 (13.1%)	6.6	17 (0.2%)	0.1	10,264	50.1
non-HFT	non-HFT	261 (1.0%)	1.4	117 (0.5%)	0.4	25,268	85.9
HFT	HFT	39 (5.3%)	0.19	5.0 (0.6%)	0.03	763	3.2
non-HFT	HFT	30 (0.4%)	0.15	59.7 (0.7%)	0.3	8,974	26.8
Other		1,286 (2.6%)	6.4	332 (0.01%)	1.7	49,293	182.8
All		2,969 (3.1%)	14.7	531 (0.6%)	2.5	94,562	348.8

Values in brackets show the relative latencycosts in percent.

Table 4: Duration of observed latency (by trade) broken down by trade counter-party for active-benefit trades (ms)

active side	value weighted average latency	mean	standard deviation	median	3^{rd} quantile	95^{th} percentile
HFT	3.65	2.93	7.00	1	2	10
non-HFT	6.90	6.59	12.59	1	5	34
other	6.59	4.73	10.26	1	3	23
All	4.75	3.72	8.65	1	2	16

Table 5: Daily average benefit for active-benefit trades broken down by marketplace

	active-benefit trade		passive-benefit trade	
	benefit to active trader ($)	relative latency cost (%)	benefit to passive trader ($)	relative latency cost (%)
M1	1,214	4.2%	397	1.4%
M2	1,420	3.9%	15	0.04%
M3	179	3.5%	97	1.9%
M4	14	2.2%	13	1.0%
M5	150	0.8%	13	0.1%
M6	4	0.1%	7	0.2%
All	2,969	3.1%	531	0.6%

Table 6: Duration of observed latency (by trade) broken down by marketplace (ms)

	value weighted average latency	Mean	standard deviation	Median	3rd quantile	95th percentile
M1	8.56	7.28	13.27	2	6	35
M2	2.50	2.16	5.20	1	2	5
M3 M4 M5 M6	7.76	6.46	13.08	1	5	33
All	5.98	4.80	10.61	1	3	24

Table 7: Regression analysis of latency

latency (ms)	
Co-location dummy	-2.34
	(-15.86***)
N_{order}	0.00949
	(18.04***)
daily trade value (in 1 m)	0.115
	(3.75***)
price	-0.0261
	(-1.43)
time of day	0.0629
	(3.62***)
Observations	66445
Adjusted R^2	0.4067
F-statistic	701.8
Fixed effect	stock

Standard errors are double-clustered by stock and date. Data used for the regression includes both active and passive side benefit trades.

At market-latency level, outliers for very large $N_{it,h,m}$ are removed at 99% level.

The calculation of the control variables are: 1) price uses time-weighted average spread; 2) time of day uses fraction of hours of the day e.g. 09:30:00 is represented by 9.5.

8. Figures

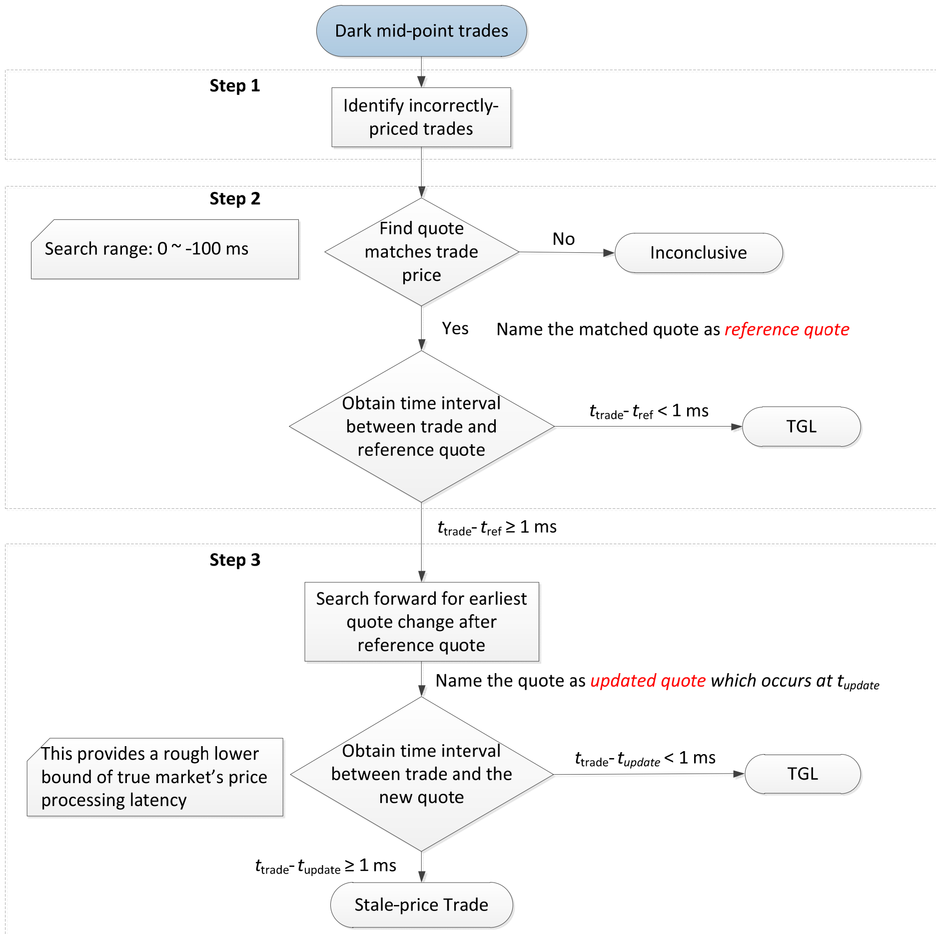

Figure 1. Logic flow for determining stale-price trades.

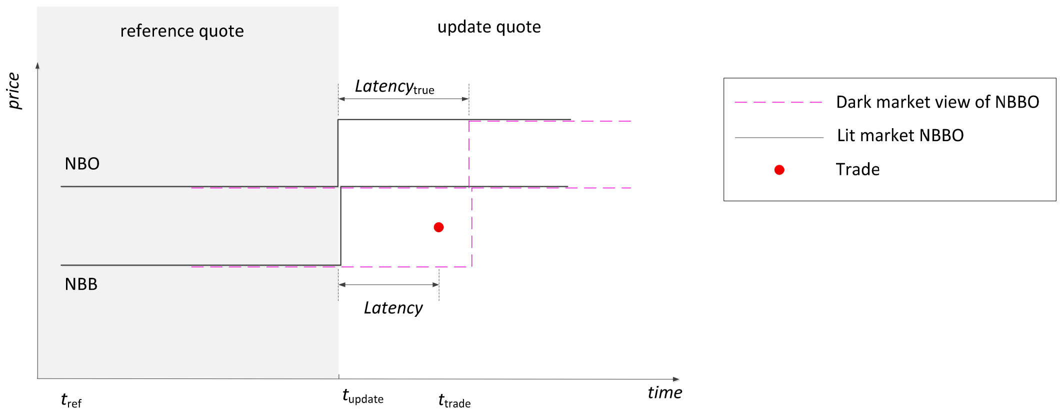

Figure 2. Illustration of the relation between the observed latency and the true latency.

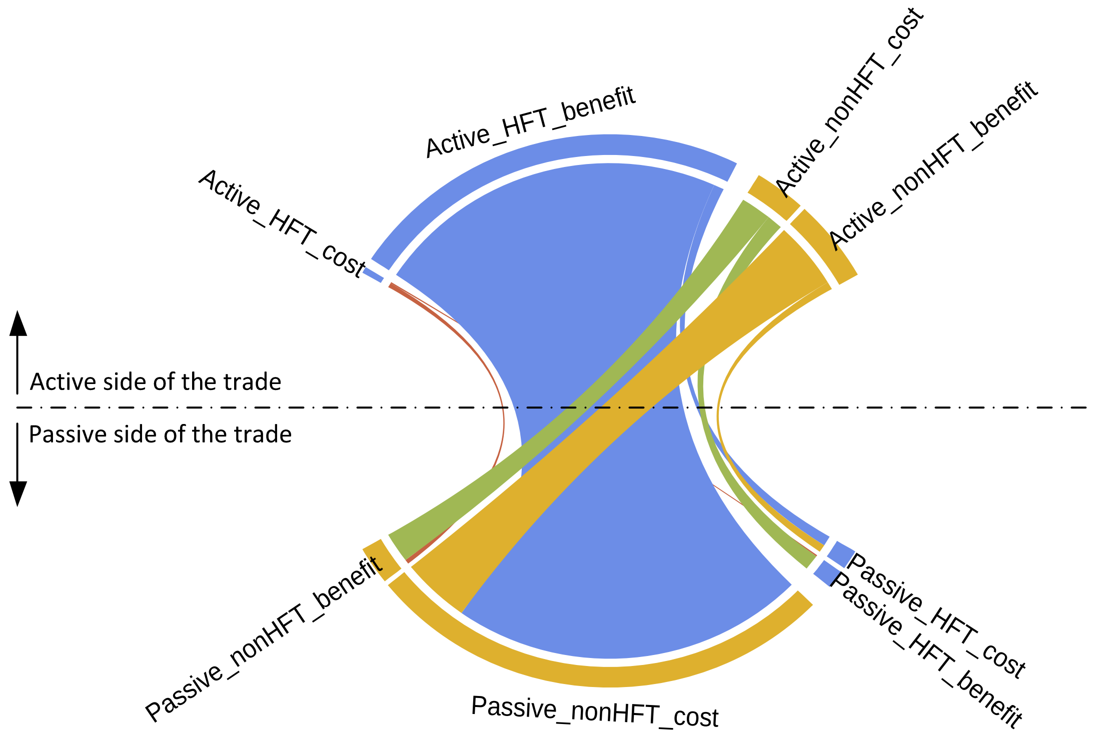

Figure 3. Benefit/cost for active-benefit and passive-benefit latency impacted trades by trader segment and counter-party.

For each trade, the active side either recieves a benefit, or pays a cost. The passive counter-party of the trade pays a cost when the active side has received a benefit, and recieves a benefit when the active side has paid a cost. Figure 3 is a visual representation of the data presented in the first 4 lines of Table 3, and illustrates the proportion of costs and benefits borne by HFT and non-HFT trader segments, on the active and passive side of the trade, showing the counterparty associated with each.

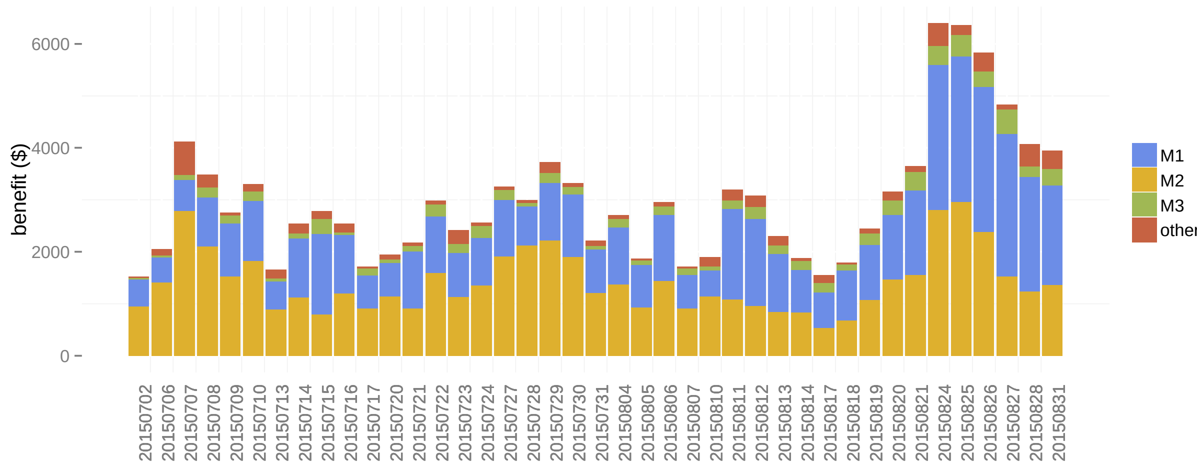

Figure 4. Benefit of active trader broken down by marketplace.

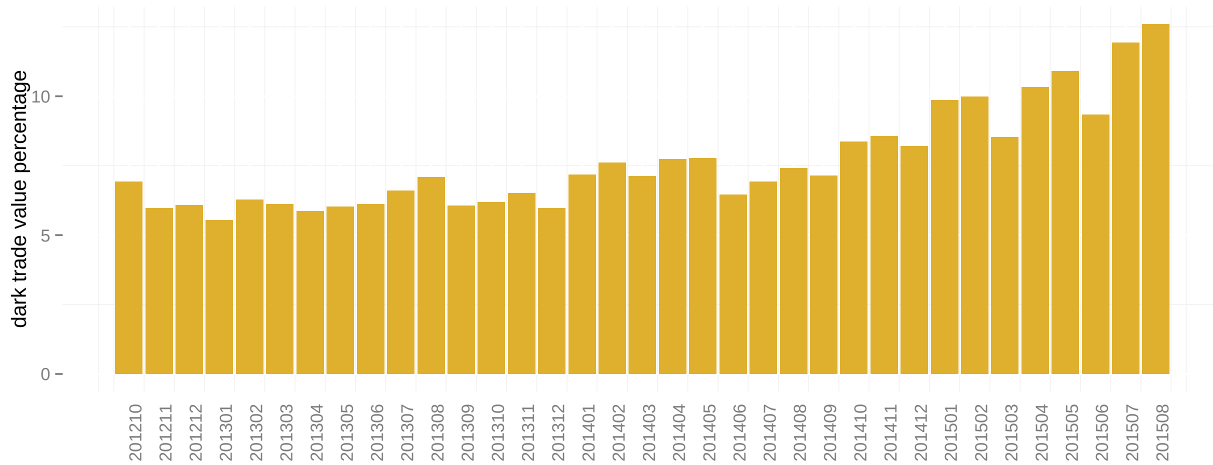

Figure 5. Percentage of dark trade value.

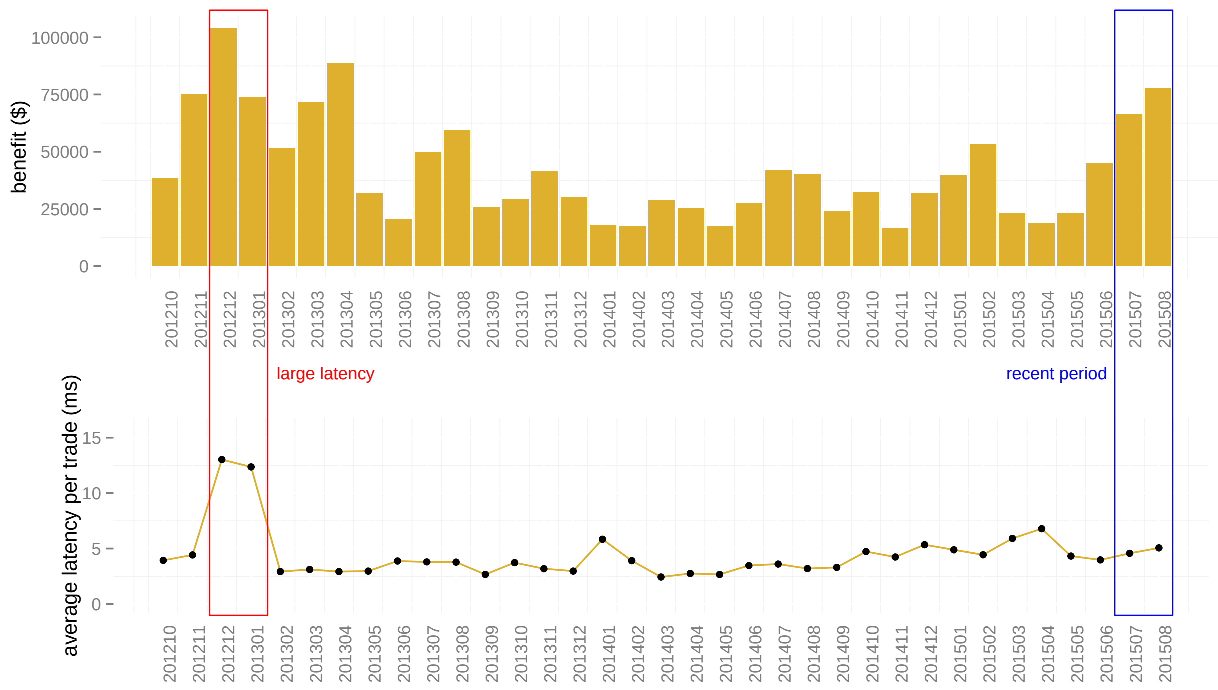

Figure 6. Historical latency – duration and economic benefit.

References

Alexander, Jeff, Linda Giordano, and David Brooks. *Dark Pool Execution Quality: A Quantitative View.* Tabb Forum, 2015.

Bessembinder, Hendrik, and Kumar Venkataraman. "Bid-Ask Spreads." In *Encyclopedia of Quantitative Finance*. John Wiley & Sons, Ltd, 2010.

Biais, Bruno, Thierry Foucault, and Sophie Moinas. "Equilibrium Fast Trading." *AFA 2013 San Diego Meetings Paper.* 2014.

Ding, Shengwei, John Hanna, and Terrence Hendershott. "How Slow is the NBBO? A Comparison with Direct Exchange Feeds." *Financial Review* 49, no. 2 (May 2014): 313-332.

Easley, David, Terrence Hendershott, and Tarun Ramadorai. "Levelling the Trading Field." *Journal of Financial Markets, Forthcoming*, March 2013.

Foucault, Thierry, Johan Hombert, and Ioanid Rosu. "News Trading and Speed." *Journal of Finance*, Forthcoming.

Hoffmann, Peter. "A Dynamic Limit Order Market with Fast and Slow Traders." *Journal of Financial Economics*, 2014: 156-169.

IIROC. "Identifying Trading Groups, Methodology and Results." 2014.

Appendix A

Effective Spread as an Indicator of Potential Stale-Price Trades

The effective spread (ES) is an estimate of execution cost paid by the liquidity taker and the revenue earned by the liquidity provider. The half effective spread is defined as (Bessembinder *et al*, 2010):

$$ES = D \times \frac{P_{it} - M_{it}}{M_{it}}$$

where $D = 1$ for buyer initiated orders and $D = -1$ for seller initiated orders; P_{it} is the trade price for security i at time t; M_{it} is the observable proxy for the true underlying value of security i at time t. Here, we choose M_{it} to be the latest quote before the trade.

For mid-point trades, regardless of which side (buyer/seller) initiates the trade, theoretically we always have $P_{it} = M_{it}$ and

$$ES = D \times \frac{M_{it} - M_{it}}{M_{it}} = 0$$

However, negative ES can occur when there is a quote change in favour of the passive trader but the marketplace still sees an "old" inferior quote (i.e. the reference quote we define in Step 2 of Section 4.iv) and execute the trade based on the old mid-point. Analogously, positive ES can occur when the quote change is in favour of the active trader but the marketplace still sees and executes the trade referencing the "old" quote.

For example, at time 11:30:00.000 the NBBO is 10.01 x 10.05. At time 11:30:00.003 market A sends a new quote feed which changes the NBBO to be 10.03 x 10.05. However, at time 11:30:00.004 the dark market B may still see the quote as 10.01 x 10.05 due to latency in receiving and/or processing the other markets' feed. If an active buy order comes in at a time before the market B sees the new quote, the orders would be executed at the mid-point of 10.01 x 10.05. The active buyer should buy at (10.03+10.05)/2 = 10.04, but instead they buy at (10.01+10.05)/2 = 10.03.

In this example, the ES for the active buyer is negative and it is calculated as

$$ES = D \times \frac{P_{it} - M_{it}}{M_{it}} = 1 \times \frac{(10.01 + 10.05)/2 - (10.03 + 10.05)/2}{(10.03 + 10.05)/2} < 0$$

In our study, we identify trades with non-zero ES as stale-price trades.