
Market Structure 2015

MARCH, 2015

Coming in to 2015, a few key interrelated market structure issues have come to the fore: maker-taker, trade-at, and tick size. These elements appear in the SEC's small-cap tick pilot, NYSE's proposed "grand bargain," and BATS' scaled-back answer to that proposal. However, none of these proposals are complete or adequate. Many knowledgeable market structure participants – including exchanges, high frequency traders, and the SEC – have said that market structure reform needs to be considered holistically. This sounds suspiciously like a smooth excuse to do nothing, but it's valid. The market is a complex machine, and attempts to tinker with individual pieces just push the distortions and conflicts from one mechanism to another while doing nothing to improve the overall situation. Meanwhile, the SEC is skeptical of the buy-side's sometimes inchoate and collectively inconsistent complaints and suggested remedies, and are reluctant to make changes that seem likely to result in unintended consequences.

In this note, we review the ways in which broker conflict is produced by maker-taker and proprietary trading opportunities; the connections between these conflicts to dark pools, trade-at, tick size, and high frequency trading; and we present a coherent set of market structure changes that would (without overturning Reg NMS) provide a true improvement to US market structure.

CAPPING REBATES FIXES ONE MAJOR BROKER CONFLICT

Industry opinion seems to have recently coalesced around maker-taker as the first target for reform. The NYSE proposal would reduce the maximum exchange fee (and therefore effectively cap the maximum exchange rebate) to 5 mils, down 85% from today's 30 mils. The BATS proposal would create a cap starting at 5 mils, rising for stocks with bigger spreads.

We've written a series of research notes over the past several years showing how the maker-taker system of fees and rebates creates conflicts of interest for brokers handling client orders. In *To Hop the Queue (Or Not to Hop the Queue)* (March, 2012)¹ we showed how paying to provide liquidity at inverted exchanges provides better execution quality for passive orders, and pointed out that brokers rarely use this capability because getting that execution quality for the client is expensive for the broker. This conflict was also examined in a widely publicized academic study last year.² In *A Conflict Inherent in the Maker-Taker Model: Equities vs. Futures* (July, 2013), we showed that even very simple signals such as quote imbalance have a significant effect on execution quality, but that brokers rarely use such signals when the market is likely to run away because crossing the spread -- and incurring the accompanying take fee -- is, again, costly for the broker. Thus for both passive and aggressive orders, the maker-taker fee structure of the exchanges creates a direct conflict between the institutional client's execution quality and the broker's profitability.

Capping the fees exchanges can charge (and thereby the rebates they can afford to pay) to 5 mils significantly reduces these distortions and conflicts of interest. Totally forbidding rebates (as further proposed by NYSE) would further reduce the conflict, but even without total elimination of rebates, a 5 mil cap would materially mitigate the problem.

However, without a trade-at rule, reducing or eliminating maker-taker will not eliminate broker conflict – it will simply push the business of extracting profit from client order flow further into the dark.

¹ All Pragma research notes and commentaries referenced here are available through our website at <http://www.pragmatrading.com/research/research-notes>

² Battalio et al., *Can Brokers Have It All? On the Relation between Make Take Fees & Limit Order Execution Quality* (March 5, 2014).



TRADE-AT FIXES THE SECOND MAJOR BROKER CONFLICT

In conjunction with capping maker-taker, the NYSE proposal includes a trade-at rule, which would mean trades could only be executed off-exchange if at a significantly better price than that displayed on-exchange. In other words, brokers and dark pools couldn't trade along at the NBBO – they would have to provide a full tick of price improvement (or half-tick in the case of a midpoint order).

This would dramatically reshape the landscape of ATS's and drive much of the current off-exchange volume - about 75-80% of which trades at the bid or offer (or other impermissible prices in a trade-at world) on to the exchanges. The other 20-25% of current off-exchange volume – about 8% of overall volume – could continue to trade at midpoint in dark pools, which can and should continue to provide valuable, specialized functions that can't be achieved on exchanges.

We view trade-at – with an appropriate Large In Size exemption – as an essential complement to reforming maker-taker because it addresses the second major source of conflict of interest between agency brokers and their clients around order handling: proprietary trading profits. **Brokerage firms acting as agent can make money from client orders that they control by trading against the orders proprietarily, or charging for access to another firm who will trade against it proprietarily.**

Let's walk through a few examples:

The most obvious is explicit payment for order flow as practiced by retail agency brokers selling order flow to "wholesalers." For example, TD Ameritrade reportedly routed essentially all of its aggressive orders in 2013 to proprietary traders such as Citadel, who in turn paid TD Ameritrade \$236mm for doing so. Proprietary trading firms pay for the order flow because they can trade against it profitably, capturing a relatively large portion of the bid-ask spread.

A similar practice is prevalent among institutional agency brokers. Proprietary trading firms like KCG, Citadel, and Two Sigma vie for first pick at aggressive institutional order flow by marketing a free "ping" service to agency brokers. In this model, the agency broker will send an aggressive order (the ping) to the proprietary trader, who then has the option to trade against or reject it. When the proprietary trader believes it's a profitable trading opportunity, he will accept the ping and execute the order. When the proprietary trader believes it won't be profitable – perhaps because the market is likely to tick in the same direction as the order – he simply declines to interact with it, leaving the agency broker to execute the order some other way, for example by taking at an exchange. In this model, the payment to the agency broker is in the form of the free execution, which allows the broker to avoid paying the exchange take fee. Since the value of that aggressive order flow to the proprietary trader wouldn't change even if exchange take fees were capped (say to 5 mils instead of 30), the typical arrangement might well shift from free (avoiding a fee) to an explicit payment to the institutional broker.

Many dark pools are in large part mechanisms to facilitate bringing together proprietary traders and aggressive directional client order flow. In perhaps the most extreme example, Getco's ATS, GETMatched was (at least pre-merger) explicitly nothing but a facility for brokers to wash their order flow against Getco's proprietary trading pods; there was no client-to-client matching. The motivations for a dark pool operator are laid out in an interesting internal Barclays memo quoted in the New York Attorney General's complaint:

“[A]ggregating [order] flow into Barclays LX has strategic and economic value for the entire Equities business,” including the savings Barclays would realize by not having to pay commissions to execute trades on other venues; fees gained from firms paying to trade in the dark pool; and the “internal trading P&L [profit and loss] opportunities” available to internal Barclays trading desks that trade in the dark pool against brokerage client order flow.



But where is the problem in all this? The problem comes when the broker is forced to choose between execution quality for the client and the broker's own profit. The dumber the aggressive flow, the more profitable it is to trade against, and the more valuable the access the broker can sell. Similarly, dark pool features that make it more profitable for proprietary traders allow brokers to charge more for access.

Details of how a dark pool operates – from the speed with which it incorporates prices into the matching engine, to the priority rules of its matching system – can create more or less profitable trading environments for proprietary traders interacting with the pool operator's captive order flow. For example, as disclosed in the SEC's announcement of its recent \$14.4mm fine, UBS broke SEC rules by selectively disclosing and marketing certain advantageous features of its pool to proprietary traders who paid to interact with UBS client order flow through its dark pool, and by not providing most users access to a feature that could prevent buy-side orders from interacting with proprietary traders.

When an algorithmic broker not only operates a dark pool but has full control over a client algo order, it presents yet another level of potential conflict. Perhaps the clearest example is "net pricing," which refers to embedding a fee into the execution price. One prominent algo broker and dark pool operator offers net pricing by executing client orders at a worse price when there's a better price available in its dark pool, and interposing an affiliate who trades as a principal between the two, capturing the difference as a riskless trading profit.³

But why is there a market for such a service? **The trading profit captured by the dark pool operator allows it to provide free or low-cost algos to its broker-dealer clients, who pass the true cost of the service on to their buy-side customers (and ultimately the investor) through what are effectively marked-up prices.** The broker using the algo service doesn't mark the prices up himself, but the service he uses effectively does it for him. The broker's client, who is already paying an explicit commission, probably doesn't expect that his broker will use an execution tool that intentionally trades at prices that are less favorable than what's available (albeit still inside the NBBO). This example makes the conflict clear because it is a mechanical rather than statistical extraction of trading profit from control of a client order. However, other than statistical uncertainty, it isn't so different from the other conflicts faced by brokers as they handle client orders. Even brokers who are one or two steps removed from actual proprietary trading make decisions every day that put their own P&L in conflict with execution quality for their clients. It is easy for these brokers to put the negative consequences of such decisions out of their mind, because the impact on execution quality is hidden in statistical noise.

Trade-at largely eliminates this conflict by eliminating this source of potential revenue for brokers. Under a trade-at rule, proprietary traders would be forced to compete with everyone else for retail order flow on exchanges or else offer truly significant price improvement. Brokers would only be able to trade client orders at midpoint, which would make them and their dark pools much poorer instruments for proprietary traders seeking to squeeze money out of client order flow. It would eliminate wholesaling, monetizing order flow through internalization pools, and net pricing.

Without a trade-at rule, reducing or eliminating maker-taker will simply push the business of extracting profit from client order flow further into the murkier world of proprietary trading and payment for order flow. Without exchange rebates for price makers, off-exchange proprietary trading would become the main method for agency brokers wishing to squeeze money out of client order flow that they've won control of. The conflicts of interest wouldn't disappear - they would just become more opaque.

Eliminating broker conflict by capping maker-taker and creating a trade-at rule would be a great accomplishment, but is not quite the end of the story.

³ It's worth noting that this provider discloses the existence of its affiliate and the nature of the net pricing service it provides.



CUSTOMIZED TICK SIZES WOULD RESOLVE THE PROBLEMS WITH TRADE-AT

Opponents point out that trade-at would disallow sub-penny price improvement, and that some traders would actually experience wider effective spreads and higher trading costs as a result. This is true only because of one of the puzzlers of current regulatory regime: exchanges are only allowed to trade at standard ticks, while brokers can trade at any price inside the spread.⁴ For stocks where the penny tick size is too big to accommodate the market's need for price formation granularity, this drives volume to hidden midpoint orders, and away from exchanges altogether to broker facilities that can price things with arbitrary price improvement.

There are other negative consequences of the standard tick size, especially for the most liquid stocks where the impact of any market distortions are amplified by the extremely high volume turnover. In *The Difficulty of Trading 'Ultra-Liquid' Stocks* (July, 2012), and *HFT and the Hidden Cost of Deep Liquidity* (July, 2012), we show that low-priced, high-volume stocks are actually more expensive for directional traders to trade, and argue that the most liquid stocks have artificially long queues because the penny tick is too big, creating inflated incentives for market makers and ultimately un-needed and expensive intermediation between directional traders.

The solution is simple: allow smaller ticks for liquid stocks. This would resolve the shortcomings of the trade-at rule. Most of the off-exchange volume would be driven to exchanges, and both retail and institutional traders would benefit from this concentrated competition through narrower spreads.

Bringing it around to the small-cap tick pilot, while the pilot as proposed would not be productive, in general it makes sense that tick sizes should be customized for less liquid stocks as well as for super-liquid stocks. Ideally tick size should be adjusted so that the average bid-ask spread in each stock is around 2 or 3 ticks wide – a big enough tick to eliminate economically meaningless pennying, but enough granularity for true competitive price improvement. For example, if a stock's spread is usually 10 or 15c wide, the tick should be around nickel. Some exchanges globally make attempts in this direction, such as the Tokyo Stock Exchange, which sets tick size based on stock price, though other attributes such as volume and volatility also influence the natural spread for a stock. And the US itself has a 1/10 of a cent tick size for stocks priced below \$1.00, but this threshold doesn't provide help in the high-volume stocks where it would have the most economic impact.

The SEC basically blesses any execution inside the bid-ask spread. For that reason, it's important that the bid-ask spread accurately reflect the market. To the extent it doesn't because of an artificially large tick size, it effectively creates a regulator-sanctioned opportunity to fleece investors, and someone will find a way to exploit that opportunity.

HIGH FREQUENCY TRADING

Finally, if more volume is driven to exchanges, it puts that much more pressure on regulators to ensure that exchanges provide a fair platform, and to eliminate advantages that can be obtained by spending excessive amounts on low-latency technology. High frequency traders pay a lot of money to exchanges for direct, ultra-low-latency market data feeds and proximity colocation, and to other technology providers for specialized low-latency hardware like field-programmable gate arrays and telecommunications like microwave networks.

In our view, high frequency traders are inevitable, and vilifying them is pointless – the SEC sets the rules, and HFTs are playing the game. As long as there is profit to be gained in spending more on technology, it will be spent. Ultimately, this means that money is flowing out of investors' pockets and into technology that provides no real economic benefit to anyone except the technology provider. Regulations that force trading on to public exchanges

⁴ Although the SEC has allowed exchanges to introduce special order types to provide sub-penny price improvement to qualified "retail" orders, because of the lock wholesalers have on retail order flow, this exception remains a footnote.



must, in fairness, be accompanied by regulations to limit the trading advantages gained by such technology spending.

Several ideas have been floated to this end. Our favorite is a randomized delay, anywhere between say 0 and 3 milliseconds, imposed on every message received by an exchange. The basic idea is that this tiny random delay would reduce the value of ultra-low latency technology and thus significantly reduce spending on it. Meanwhile, the delay is so short that it would create no economically meaningful harm to directional traders, to the function of the market, or on price efficiency based cross-market arbitrage.

WINNERS AND LOSERS

Who would be the winners and losers in all this?

Exchanges would be hurt somewhat. Capping take fees wouldn't directly hit exchange profitability on the matching business. Today they only capture about 3 mils per share, the spread between the average fee they charge and the average rebate they pay on each matched share, and there's no reason this spread should go lower. However, to the extent better tick sizes reduced unnecessary intermediation, overall volume would decrease somewhat. In addition, the reduced utility of low-latency market data feeds and proximity colocation – now major sources of revenue for exchanges – might reduce the revenue for these products. Exchanges might have to actually increase their capture from matching to make up for some of that revenue – a good trade for the market as a whole.

Brokers would adjust their commissions and as a group would probably come out even. To the extent the exchanges fee capture stays the same, brokers should also be in roughly the same place on average. The low end of the commissions they charge would have to go up – especially the artificially low commissions charged to mid-tier brokers by the big algo providers. With the mechanisms for monetizing order flow gone, these services would be provided at their true economic cost.

High frequency trading would shrink. In a market where a relatively moderate investment in technology provides all the benefit that can be gotten, where trading opportunities are available only on exchanges (because of trade-at), and where artificial subsidies of large rebates paid by exchanges is gone, proprietary traders would be left functioning their most constructive capacity as market-makers: competing by pricing more intelligently and efficiently than other participants, rather than exploiting technical advantages. A mitigating benefit to HFTs is that their technology expenses would go down significantly.

The big winners would be investors. Spreads would narrow, the hidden fees around trading would be reduced, and the total true cost of trading would go down, benefitting investors. The low end of institutional buy-side commissions, in turn, would go up slightly, but would overall be largely unaffected. And most importantly, the true costs of trading would become more transparent.

CONCLUSION

In summary, our prescription for market structure reform is:

- Cap exchange fees to \$0.0005 (and ideally eliminate rebates completely)
- Impose a trade-at rule
- Set the tick size for each stock so its bid-ask spread is usually 2-3 ticks wide
- Introduce a randomized 0-3 millisecond delay on message processing at all market centers.

Alone, any of these reforms could be useless or even harmful; together they would significantly improve the situation by eliminating many broker conflicts of interest around order handling, and eliminating hidden fees and taxes that make the apparent cost of trading diverge from the true cost.