Market Regulation Services, Inc. (Canada)

Literature Review

Best Execution and Trade-Through

Carole Comerton-Forde
Bruce Robert Arnold

31 October 2005
### Table of Contents

3.4.3.  Liquidity ........................................................................................................... 25  
3.4.4.  Cost ............................................................................................................... 26  
3.4.5.  Fairness and the Policy Choice ....................................................................... 27  

References ................................................................................................................. 28
About the CMCRC

The Capital Markets Cooperative Research Centre was formed in 2001 to bring together the best in innovative research and technology to the capital markets domain.

Positioning itself as a bridge between the practical needs of capital markets and pioneering academic research, the Capital Markets CRC (CMCRC) was formed under the auspices of the Federal Department of Education Science and Technology (DEST) Cooperative Research Centre (CRC) program. Having achieved the goal of attracting key industry and research participants as partners in the program, the CMCRC was granted funding of $62 Million over seven years to foster research that is both leading edge and industry relevant.

The CMCRC is the initiative of a group of researchers and industry partners who have an impressive track record of commercial success. The CMCRC formula attracts the best of Australia’s researchers to work together with industry partners to develop new and innovative technologies for the capital markets domain.

Headed by Chief Executive Officer, Professor Mike Aitken, the CMCRC harnesses the combined research resources of approximately 40 senior researchers from the fields of finance and technology. A similar quantum of human resources is provided by industry partners who also assist in guiding research resources to problems and issues that will make a significant difference in the future operations of global capital markets.

Contacts

Street Address  
Capital Markets CRC Limited,  
Level 2, 9 Castlereagh Street,  
Sydney NSW 2000,  
Australia

Postal Address  
GPO Box 970,  
Sydney NSW 2001,  
Australia

Phone  
+61 (2) 9233 7999

Fax  
+61 (2) 9236 9177

Email  
cmrc@cmrc.com

ACN  
096 930 406
About the Authors

Bruce Arnold
Bruce Robert Arnold is enrolled in the PhD program in the School of Banking and Finance, in the Faculty of Commerce and Economics at the University of New South Wales, and is a Research Fellow at the CMCRC. Returning to study after a career in investment banking, his principal research interest relates to credit spreads and ratings migration.

Carole Comerton-Forde
Dr Carole Comerton-Forde is a Senior Lecturer in Finance at the University of Sydney. Carole graduated with a Bachelor of Commerce with First Class Honours in Finance and a Doctor of Philosophy (Finance) from the University of Sydney. Carole’s research focuses on security market microstructure and issues related to market efficiency and integrity. Her research has been published in academic and industry journals including the Journal of Financial Markets, Journal of Financial Services Research and the Pacific Basin Finance Journal. Carole has been involved in a large number of research projects for the Australian Stock Exchange and other market organizers and regulators. These projects include an examination of a delayed reporting regime for large block trades, an analysis of the impact of the introduction of a closing call auction on market liquidity and analysis of the impact of the pre-trade display of broker identities on liquidity.
About this Memorandum

We have been asked to undertake a literature review for Market Regulation Services, Inc. (RS) to support RS's strategic review of its Universal Market Integrity Rules, as announced in Market Integrity Notice 2004-026—Strategic Review of the Universal Market Integrity Rules (October 4, 2004). This memorandum is limited to the issues of best execution and trade-through.
Preliminary Comments

For the purpose of this memorandum, the term "investors" refers to participants who generally act through brokers to buy or sell securities for investment purposes, as distinguished from dealers who trade for the purpose of earning profits from intermediation. We do not specifically consider the difference between investors who trade for their own account and those, such as investment managers, who are not the ultimate beneficiary of the monies involved.

With respect to the Canadian market-place, we understand that the Toronto Stock Exchange (TSX), trades only shares that are listed on the exchange, but that the TSX Venture Exchange (TSX-V) and the Canadian Trading and Quotation System (CNQ) now have a slight overlap in trading coverage, but are not linked electronically. As to Electronic Crossing Networks (ECNs) or Alternative Trading Systems (ATSs), we understand that Liquidnet Canada allows Canadian investors to effect crossings, but only in U.S. and foreign shares. We also understand that Markets Securities BlockBook data, which facilitates opaque crossings between institutional investors and other subscribers, will be made available through the TSX’s data distribution system. However, as we understand that BlockBook will quote only TSX-listed shares and clear through the TSX, we consider it to be, in effect, a brokerless upstairs market to the TSX. In sum, there is little competition between exchanges and ECNs or other ATSs at present. That said, we acknowledge that global trends call for competition to increase, both through an increased overlap between the TSX-V and the CNQ, and through the arrival of more ECNs and ATSs.¹

The academic research and industry comment reviewed for this memorandum include many materials originating in the United States. Much of the U.S. debate is focused on the relative merits of an open outcry market—notably the New York Stock Exchange (NYSE) and its specialist system—and the dealer-driven market of the NASDAQ. While we survey the literature relating to competition, we make no reference to the research which compares types of markets (to wit, the NYSE to the NASDAQ). Where competition relates to market quality, we make no unnecessary reference to the types of market structures which create the competition.

¹ We note that Bloomberg’s BondTrader System and ALLQ may allow Canadian institutional investors access to quotes on Canadian bonds and fixed-income hybrids, but this memorandum considers only equity markets.
Executive Summary

To clarify discussion, best execution and trade-through are analysed from the perspective of the interests which are being promoted. Thus we distinguish between investors requiring protection from the acts of their own brokers, and investors seeking to have their limit orders filled. We also consider the general public as a class in its own right, which allows us to examine externalities such as overall market efficiency, liquidity, and other indicators of market quality.

Best Price and Best Execution

We examine whether market integrity rules should impose a best-price obligation in connection with a best-execution obligation.

- From the perspective of investors acting through brokers, imposing a best-price obligation is consistent with best execution in certain circumstances only, but restricts strategic choices and possibly results in poorer execution in other circumstances. As a policy choice, then, imposing a strict best-price obligation does not deliver optimal results.

- From the perspective of the general public and its interest in an efficient capital market, competition between markets is proven to deliver lower overall execution costs, as well as different bundles of benefits and costs which appeal to different investor groups. One aspect of market competition is to offer alternative trading systems, some of which are inconsistent with a best-price obligation. While competing structures deliver cost benefits, there is no conclusive proof that overall liquidity or the ability to form prices efficiently would be compromised. A strict best-price rule effectively limits the allowable markets to those featuring a fully transparent, continuous auction, thereby restricting competition and limiting the efficiency gains that can be derived. At the same time, it does not appear that a best-price rule would deliver compensating benefits in liquidity or price formation.

- The principal benefit of a best-price rule is that it can be easier to enforce than a more general best-execution obligation, the value of which is somewhat doubtful given that the rule has the potential to harm both individual investors and the efficiency of the market as a whole.

Trade-Through

We then look to the market-integrity aspects of a trade-through rule, and further consider the effect of an opt-out provision.

- From the perspective of investors who enter market orders, requiring that the order be traded against the best bid or offer is a best-price rule by another name. As discussed in regard to best price and best execution, a strict best-price rule is not always in the best interests of investors entering market orders, thus any ability whatsoever to opt out of a trade-through rule would be to these investors’ advantage.

- From the perspective of investors entering limit orders, there is little evidence that being traded through results in significant economic harm, although we acknowledge that there may be the perception of unfairness if a traded-through order fails to execute in a timely fashion. To the extent that there is any harm to investors entering limit orders, an opt-out provision would reduce the efficacy of a trade-through rule. However, there are alternative remedies which mitigate any harm suffered, but which would not impair the interests of investors seeking, for strategic reasons, to trade through. The implementation of these alternatives would allow an opt-out provision.
• From the perspective of the general public, the effect of a trade-through rule on market efficiency is the same as that of a strict best-price rule: The rule restricts competition, and thereby increases execution costs to participants. Limit orders by definition increase liquidity, but there appears to be no research indicating that a trade-through rule increases the tendency of investors to enter limit orders—particularly the smaller investors who are usually considered the principal beneficiaries of the rule. Large institutional orders and trades—even on less transparent markets—contribute more to price discovery than their retail counterparts, thus the argument that trade-throughs "free ride" on the bypassed limit orders is not supported. Finally, while a trade-through rule is not supported by efficiency, liquidity, or price-formation rationales, its implementation comes at a cost which is eventually borne by both investors and issuers.
1. Framing the Issues

Public policy expresses itself in preventing a protected class from incurring a specific harm, or in promoting positive externalities and discouraging negative externalities. The benefits of mitigating harm or promoting public goods must be weighed against the costs—both the costs imposed on members of the public, and the cost of enforcement borne by the regulators.

Some of the most heated public-policy debates relate to a single issue which is concerned with more than one specific harm and/or public good. For example, William Donaldson, then Chairman of the U.S. Securities and Exchange Commission (SEC), testified before a sub-committee of the U.S. House of Representatives, in a hearing relating to the SEC’s Proposed Regulation NMS, that "the trade-through rule is, in the most fundamental sense, a rule that protects investors." (Donaldson, 2005, p. 17.) However, Mr. Donaldson elaborated as follows:

What does this mean? Two things. It means that a broker executing an order will be required to give that order the best price then available in any electronically accessible market, even if the broker internalizes the order or would prefer to trade in another market that may offer the broker itself, if not the customer, an advantage. And second, it means that an investor who is willing to place an aggressively priced limit order on the book will not have his order ignored in favour of a less aggressively priced order. (Donaldson, 2005, p. 16.)

In other words, the rule is intended both to protect principals from the actions of their agents who would trade through a better posted price, and to protect the parties traded through. While both groups are investors, to say simply that trade-through "protects investors" confuses the issue.

To add clarity to the discussion, we have framed the issues in terms of the classes whose interests are affected by the policy issue. We include as a class the general public in its own right, which captures the benefits of public goods, as well as the costs of compliance and enforcement. Note that we do not consider the providers of services (e.g., broker-dealers and exchanges) as a protected class. However, we do consider the burden placed on these parties from the perspective of the general public facing systemic cost increases, which allows us to frame the cost-benefit analysis.
2. Best Price and Best Execution

RS has asked whether market integrity rules should impose a best-price obligation in connection with a best-execution obligation. Following the framing considerations above, we restate this as follows:

From the perspective of investors acting through brokers –

- does an explicit best-price obligation cause the investor to incur lower overall execution costs than a more general best-execution obligation?
- is an explicit best-price obligation more effective than a best-execution obligation in protecting investors from their brokers’ misfeasance or malfeasance?

From a public-goods perspective, does an explicit best-price obligation produce better market-quality measures than a more general best-execution obligation?

2.1. Best Price and Execution Costs

Commentators almost unanimously note that best execution is not necessarily the same as the best price. The consensus view is that best execution is appropriately viewed as a process intended to deliver a desired result, rather than a simple exercise in obtaining the best price. See, e.g., Financial Services Authority (2002) ¶ 3.4 and CFA Institute (2002) p. 4.

Macey and O'Hara (1997), undertake a comprehensive (although slightly dated) review of the legal aspects of best execution, the empirical research measuring execution costs, and the interplay of market practices and best execution. They make the interesting observation that unlike pornography, which is difficult to define but recognisable on sight, best execution is relatively easy to define, but extremely difficult to recognise in the ebb and flow of trading activities. (Ibid., p. 189.)

Factors cited as relevant to the quality of execution include price, commissions and other agency costs, access fees and other transaction-handling costs, speed and certainty of execution, market impact costs, opportunity for price improvement, adverse-selection costs, and trader-specific intangibles such as maintaining anonymity. For example, the European Union's Markets in Financial Instruments Directive (MiFID) states as follows:

Member States shall require that investment firms take all reasonable steps to obtain, when executing orders, the best possible result for their clients taking into account price, costs, speed, likelihood of execution and settlement, size, nature or any other consideration relevant to the execution of the order. (European Parliament and Council, 2004, Article 21, ¶ 1.)

The common understanding is that the determination of best execution is a multi-dimensional exercise, whereby an investor may sacrifice best price in order to optimise in respect of one or more of these other factors. (Lee, 2002, pp. 13-14.) See also Committee of European Securities Regulators (2005) Box 2, which states that a firm should consider the characteristics of the clients, orders, securities, and venues when determining how to prioritise the factors set out in MiFID Article 21(1).

However, we believe that a comment made by Benn Steil in his written Congressional testimony in respect of the initial release of Proposed Regulation NMS makes the analysis more tractable:

“Why should speed be more important than price?”

2 Dr. Steil is the Andre Meyer Senior Fellow in International Economics at the Council on Foreign Relations.
According to this view, the whole debate is about whether traders should be allowed to sacrifice best-price in the pursuit of speed. But the notion that investors would ever sacrifice price for “speed” is nonsensical. In the marketplace, it is always about price. It is about the price for the number of shares the trader wants to trade, not just the 100 shares advertised on the floor of the NYSE, and it is about the price that is really there when the trader wants to trade. Statistics from competing marketplaces about fill rates, response times, and the like make very nice input into a trader’s decision, but they are not substitutes for a decision. (Steil (2004), pp. 2-3, emphasis in the original).

While Dr. Steil made his comment in the context of the trade-through rule, his insight extends to best execution: It is always about the price. Faster execution is sought, price improvement is preferred, and high market impact is to be avoided, but all of these factors are fully reflected in the aggregate price achieved by the investor. If a market order can be executed immediately against the best bid or offer in the limit-order book, price, certainty of execution, market impact costs, opportunity for price improvement, and adverse-selection costs reduce to a single dimension. Under these circumstances, best execution is best price.

2.1.1. Transaction Costs

If execution takes place on the basis of net price, i.e., inclusive of all transaction costs, many other elements relating to best execution—third-party commissions, access fees, and liquidity rebates—can also be incorporated into the one metric. Again, assuming a sufficiently large and accessible limit order book, for a broker to trade an investor's market order against the best net price necessarily constitutes best execution.

As transaction costs are generally less than one cent per share, adjusting quotes to show a net price would be counter to the general regulatory trend to discourage sub-penny pricing. Lawrence Harris proposed an outright prohibition on access fees and liquidity rebates (Harris, 2005, p. 5), and presumably his argument would apply to all other exchange fees affecting net price. Instead, the SEC chose to place a 3-mill cap on access fees, arguably preserving the integrity of the quote. But as long as quotes are not adjusted for access fees and liquidity rebates, quoted prices will be artificially understated.

3 Often brokerage charges levied directly on a principal are excluded from the assessment of best execution, as such costs are fully known and agreed separately by the trader. See, e.g., U.K. Financial Services Authority (2002) ¶ 3.24. Market forces ensure that a principal will pay higher commissions only in exchange for a higher quality of service, thus incremental brokerage costs are discretionary to the principal. In general, we agree. However, we note that compliance costs could force every broker to raise commissions, giving principals a choice between bad and worse. We attempt to pick up this factor below, in the context of a cost-benefit analysis.

4 Sub-penny pricing allows disclosed limit orders to be “picked off” by newly arriving limit orders which only nominally improve the price. Decimation has reduced the cost of “stepping ahead of the book”, and appears to have increased the frequency of dealers or specialists executing orders in front of investors. Harris (Lawrence) (1997) goes so far as to call stepping ahead “parasitic behavior”, and Harris (David) (2004) contends that sub-penny pricing constitutes “economically meaningless price improvement” (ibid., p. 17).

The other side of the argument is put forth by Edwards and Harris (Jeffrey) (2001), who observe that even a nominal improvement of price benefits market orders. Nicoll (2004) places a number on that benefit, noting that sub-penny pricing improvements on a single line of exchange-traded fund saved investors approximately US$150 million in the course of a single year.

5 Dr. Harris is the Fred V. Keenan Chair in Finance at the Marshall School of Business at the University of Southern California. From 2002 to 2004, Professor Harris was Chief Economist of the SEC, and head of the Commission's Office of Economic Analysis.

6 On the Australian Stock Exchange, with a fully automated central limit-order book, traders pay an exchange fee based on the total trade value, and a single clearing fee. However, a per trade settlement fee increases the cost of executing an order against multiple counter-orders.
One consequence of understated spreads is the illusion of a locked or crossed market—illusory because, although the quoted spread is nil or negative, the effective spread net of transaction costs is still positive.7 Goldstein, Shkilko, Van Ness and Van Ness (2005) found crossed or locked spreads occurring frequently—about once every minute—but for periods of time of 10 seconds or less. They find this "consistent with rebates being a cause of non-positive spreads." The SEC is now requiring exchanges to promulgate rules to prevent member brokers from locking or crossing the market, when a locking or crossing limit order could in fact generate a better net price for the client. Thus by failing to adopt a simple net-price rule, the SEC is required both to interject itself into the fee-setting process,8 and then to mandate additional rules to limit investors from optimising their outcomes.9

2.1.2. Insufficient Limit-Order Book

If the limit order book is too thin to absorb the market order, additional liquidity must be found—from new limit orders arriving in the market,10 from dealer quotes, or by way of a block trade. In this situation, best execution is no longer synonymous with price, but is a matter of selecting which of several alternative strategies to pursue. The decision is made under uncertainty, and subject to the risk that some or all of the market order remain unfilled. This introduces a second dimension into the calculus of best execution: There is a trade-off between how much an investor is willing to pay for liquidity and the likelihood of an execution shortfall—a form of substitution function. There have been numerous attempts in the academic literature to impute an appropriate cost of or penalty for execution failures (e.g., Harris & Hasbrouck, 1996), but the substitution function clearly is unique to each investor's own circumstances. Strategy selection therefore constitutes an exercise in risk assessment and risk management, in respect of which investors can be expected to hold strong views.

Note that the issue of how best to obtain additional liquidity must be considered as soon as it becomes apparent that the top of the limit-order book is deficient. Different strategies might take different approaches to existing orders. For example, one might choose not to transact against the limit-order book at all, for fear of spooking the upstairs market. (See Seppi, 1990.) While this would not constitute a best-price strategy—the passed-over limit orders represent money left on the table—it could well embody the optimal trade-off between price and the perceived cost of execution failure.

Macey and O'Hara (1997) are adamant on this point:

[T]o suggest that an individual has not been afforded the benefit of best execution because a trade has not been consummated on an auction market is not appropriate, even though such a market might provide the best structure for the market as a whole. Similarly, it seems unreasonable to argue that best execution requires executing trades at prices drawn from other trading structures, ignoring that each trading structure provides a different vector of execution attributes and services a different clientele. If alternative markets provide other benefits to traders, then focusing narrowly on the

---

7 To illustrate, assume a limit-order offer of $2.00 subject to a 3-cent access fee, of which 2 cents are rebated to the party entering the order. To lift the offer, a buyer would have to pay $2.03, and the seller would net $2.02. Alternatively, the would-be buyer could enter a limit-order bid of $2.01—crossing the market, as the best bid exceeds the best offer. If the bid were hit, the buyer would acquire the stock at a net $1.99. If the access fee and liquidity rebates are less than the minimum price increment on the stock, there is no incentive for a party to cross the market, but locking the market would still ensue.

8 Gerald Putnam, Chairman and CEO of Archipelago Holdings, L.L.C. (which owns the ArcaEx ECN), is one of many commentators who vainly beseeched the SEC to avoid what he called "command economy price-fixing". Putnam (2004) p. 11.

9 The ability of an exchange to claim priority on the basis of a displayed price, but to extract an undisclosed fee constitutes true "regulatory arbitrage", much more than an investor selecting how to execute a trade so as to minimise transaction costs.

10 The hallmark of a resilient market is that trades give rise to new limit orders. However, under a strict best-price regime, a broker would be prohibited from executing part of a market order and waiting for liquidity to restore itself. While the investor could give the market order to its broker in stages, this imposes the cost of order-management time on the investor, when the investor is paying the broker to do precisely that.
trade prices misses the property that trade execution is a multi-dimensional process. (Macey & O'Hara, 1997, p. 220.)

Thus, from the perspective of investors entering market orders through brokers, an explicit best-price obligation is certain to cause investors to incur lower overall execution costs only if the top of the limit-order book were to equal or exceed the size of the market order (and further assuming that transaction costs would not compromise the integrity of the quoted price). However, in all other cases, a strict best-price obligation would not benefit investors, as it would eliminate certain strategies with the potential to generate more favourable outcomes—strategies which certain investors might strongly prefer based on their risk profiles. As a general rule, then, a best-price obligation applied in all circumstances does not deliver the desired policy result, and at the same time interferes with investors' latitude in handling their own affairs.  

2.2. Best Price and Market Quality

Market quality is a public good, resulting in benefits to investors and issuers alike. Indicators of market quality include:

- **Price discovery / price formation**—the extent to which, and the speed with which, information is incorporated into prices.
- **Liquidity**—the extent to which market participants can rapidly execute large-volume transactions with a small impact on prices.
- **Volatility**—the extent to which the prices of the assets traded in a market vary, both individually and collectively.
- **Efficiency**—the extent to which transaction costs in general are minimised.

(Lee, 2002, p. 13.)

2.2.1. Competition Between Markets

Where all trading for a security is conducted on a single market, matching market orders to limit orders is simply a matter of applying priority rules. See Weaver (2004a) for a comprehensive survey of priority rules. When a security is traded on multiple venues, or competing markets, a best-price obligation becomes an order-routing rule, whereby a market order is directed to the venue(s) where the best limit order is displayed. Thus implicit in a best-price obligation is the mandate that orders be sent to that market, among its competitors, which displays the best price.

Competition between markets has been enshrined as public policy by numerous countries, including the United States 12, as it is generally accepted that competition improves market quality, especially as

---

11 Many industry commentators suggest that differential rules should apply to retail and institutional investors, notwithstanding the difficulty in distinguishing between the two categories. We believe that the functional test set out above—whether or not a market order exceeds the depth of the limit-order book—both is relevant to the question of what constitutes best execution, and constitutes an easily enforced standard.

12 Section 11A of the Securities Exchange Act of 1934 provides as follows:

(1) The Congress finds that--

(2) ... (C) It is in the public interest and appropriate for the protection of investors and the maintenance of fair and orderly markets to assure--

(3) ... (ii) fair competition among brokers and dealers, among exchange markets, and between exchange markets and markets other than exchange markets, ...
measured by execution costs. Empirical research supports this premise. Mayhew (2002) found that both the quoted bid-ask spreads and effective half-spreads\textsuperscript{13} for options traded on multiple U.S. exchanges were lower than on those listed on a single exchange, and conversely, spreads widened when a multiple-listed option was delisted by one of the exchanges on which it had been traded. De Fontnouvelle, Fishe, and Harris (2003) found even greater narrowing of spreads—over one-third for effective spreads, and over one-half for quoted spreads—for multiple listings in the wake of a Department of Justice suit and SEC enforcement action. In a different sector, Boehmer and Boehmer (2003) identified double-digit percentage declines in spreads when the New York Stock Exchange began trading exchange-traded funds which had previously been traded only on the American Stock Exchange. Based on a de-composition of spreads, they attribute the tightening to a decline in the rents which market makers were able to extract from investors.

In all of the above studies, the circumstances were ideal for testing the respective hypotheses, as the trading protocols in the competing markets were highly similar. However, "competition between markets" is a more complex issue than merely determining which venue offers a better price under substantially identical parameters.

In its more general sense, competition between markets embraces the idea that different markets can adopt completely different sets of microstructural rules, of which order-routing is but one. Markets also compete on the basis of different levels of pre- and post-trade transparency; different fee structures, and different levels of involvement by brokers, dealers, and specialists (and, per the analysis in Boehmer and Boehmer (2003), different level of rent), creating baskets of benefits and costs which appeal to distinct groups of participants.

Goldstein, Shkilko, Van Ness, and Van Ness (2005) found that the effective half-spread on NASDAQ-listed equities was 1.10 cents per share on the primary market, but as high as 1.90 cents on the American Stock Exchange, and as low as 0.71 cents on the Island electronic communications network (ECN).\textsuperscript{14} The fact that execution costs on an open-outcry market are different than those on a dealer-based market or an ECN comes as no surprise. Instead, what is interesting is that there is any trading at all on the higher-cost markets. Clearly there is some value to certain investors in executing on the AMEX or the NASDAQ—value which offsets those markets' higher costs.\textsuperscript{15} At the same time, Goldstein \textit{et al}. (2005) report that the lower-cost markets may offer less liquidity, and differ in other key attributes. Investors have therefore selected their venue by assessing the bundle of benefits and costs which have the greatest appeal. This is the principle of competition between markets.

2.2.2. Fragmentation—Execution Costs and Liquidity

A countervailing factor is that competition between markets leads to fragmentation, where fragmentation arises when all orders do not interact with each other via a single order execution mechanism. (Lee, 2002, p. 14.) There is both theoretical Madhavan (1995) is one of several who offer a theoretical basis, and Arnold, Hersch, Mulherin, and Netter (1999) provide empirical evidence, that fragmentation leads to wider spreads. In a more recent survey across 51 stock exchanges around the world, Jain (2001) confirms that a "centralised" market offers lower spreads than a fragmented market.

\textsuperscript{13} The effective half-spread is the difference between a trade price and the average of the best bid and offer. This measure is consistent with the general consensus that execution costs should compare trade prices to an exogenous reference point. That said, there are multiple methodologies for doing so, which makes it difficult to compare execution costs across different research efforts.

\textsuperscript{14} At the time of the study, 2002, Island was the largest ECN on NASDAQ. It first switched reporting of its trades to the Cincinnati Stock Exchange, and then its quotes, because of a more favourable fee structure. Island merged with Instinet, another ECN, later in the year.

\textsuperscript{15} We assume that not all of the volume traded on the higher-cost markets was improperly routed by brokers.
However, there is a difference between a centralised market and a consolidated market. While a centralised market requires that all trading be conducted at a single venue, a consolidated market requires only that all quote and trade information be readily available from a single source, whether an exchange or a private consolidator of information. (Lee, 2002, pp. 20-21.) Harris (2002) observes that, for effective consolidation to take place, investors must adjust their orders to reflect information revealed on other venues, and be able to route their orders to other venues. If arbitrageurs are then able to operate, prices will reach equilibrium across all venues. Galetovic and Zurita (2002) find that a market is effectively integrated as long as every broker meets every other broker on at least one exchange.

Empirical evidence from Europe supports this proposition. Immediately after the Big Bang in 1986, large volumes of continental European securities began to be traded on the London Stock Exchange. However, there appeared to be little decline in domestic volumes, thus overall liquidity seems to have been enhanced. (Lee, 2002, pp. 16-17.) Frequently the quoted bid-ask spreads on the continental exchanges were narrower than those obtaining in London, however, often the amount it was possible to trade at these prices compared to what was possible at the prices available in London was smaller. London offered the possibility of trading immediately against market-maker quotes, while the continental exchanges facilitated a more patient type of trading with smaller execution costs, but greater risk of non-execution.

The evidence in respect of liquidity is less supportive. Mendelson (1987) and Parlour and Seppi (2001) respectively find that fragmentation and competing markets can lead to diminished liquidity. Podpiera (2001), looking at shares cross-listed in Central European exchanges and London also find that volatility increases. However, Menkveld (2001) examined the trading patterns exhibited by stocks dual-listed in Amsterdam and New York, and found evidence that liquidity was rationally exploited across both markets through optimal order-splitting. Even Podpiera (2001) found that prices were co-integrated (albeit with relatively long lags), suggesting that the fragmentation arising from competition is not overwhelmingly negative for market-quality indicators.

2.2.3. "Cream-Skimming"

Another argument offered against an unrestricted competition between different market structures is the objection to "cream-skimming". The concept of cream-skimming dates back to the seminal work by Kyle (1985) suggesting that a dealer faces counterparties who are either uninformed or informed, where informed counterparties cause the dealer to suffer from adverse selection. In such a world, a dealer can afford to make markets only if enough profit can be made from uninformed investors to subsidise the losses at the hands of informed investors.

Many empirical studies have shown that larger orders—orders more likely to originate with informed investors—are more efficiently accommodated at the primary market. If a competing, secondary venue can attract uninformed and discourage informed order flow, its dealers can avoid adverse-selection costs and thereby offer tighter spreads. By skimming this cream, however, the primary market's remaining flow is proportionally more informed, and spreads must widen. (Easley, Kiefer & O’Hara, 1996.)

The response of proponents of the competing-markets ideal is that one of the factors which might influence an investor's choice of venue is whether the investor is informed or not. If not, the investor would naturally gravitate to a venue which doesn't extract disproportionate profits in order to subsidise losses to informed traders. (In fact, Easley et al. note that they cannot reject the possibility that the purportedly cream-skimming markets may well be sharing their profits with end investors.) On the other hand, if the investor is informed, the investor should be paying wider spreads.

2.2.4. Price Formation

As noted above by Harris (2002), consolidation of information, rather than consolidation of trading, is the critical factor to efficient price formation. However, where some markets exhibit less transparency than others (and many ECNs compete on the basis of their lack of transparency), there is arguably the potential
for "free-riding"—misappropriating the information disseminated from the more transparent venue(s). The SEC (2004) itself raised the issue in its re-release of Proposed Regulation NMS, but was not as flowery as Larry Tabb, a columnist in Wall Street & Technology, a journal for financial markets:

Crossing, as it increases, will eventually disenfranchise [proprietary trading organizations]. The transition to crossing networks, which works well when it comprises only 2% to 5% market share and price off a well functioning marketplace, will be exceedingly priced off an increasingly retail price discovery process where the price of matching millions is based on the price of matching hundreds. (Tabb, 2004.)

However, this assertion ignores the fact that already a large volume of institutional trade occurs off the floor, either via ECNs or in the "upstairs" market (which constitutes approximately 30% of volume on even the most transparent venues). Battalio and Jennings (2005) cite a large body of literature which shows that institutional trades are both more informative and have much greater market impact than retail trades. Thus even if a trade takes place on a relatively opaque venue, the information tends to surface in the market, albeit at a slower pace. As noted above, the LSE now sees substantial volume in continental European shares. While domestic spreads appear narrower than those prevailing on the LSE (albeit for smaller parcels), the prices tend to be "closely integrated". (Lee, 2002, p. 17.) Barclay, Hendershot, and McCormick (2003) find that trades executed on ECNs have larger permanent effects on price and thereby evidence more price-forming efficiency. They suggest that the very anonymity which encourages informed traders to patronise ECNs allows them to trade—and thereby disclose their information—with less circumspection. (Ibid., p. 2655.)

Macey and O'Hara (1999) concur:

Security prices do not instantaneously gravitate to their full-information values. Instead, it is through the process of trading that information becomes impounded in security prices. … [I]t does not appear that the underlying price discovery process requires every order to participate, nor does it appear to matter if some traders prefer to trade-off the potential for price-improvement for the promise of lower trading commissions. (Ibid., p. 26.)

As it turns out, there is little evidence that block trades and crossings free-ride off displayed limit orders. Instead, empirical studies suggest that the more usual pattern sees limit orders stepping in front of the book, placed by professionals free-riding off displayed trading interests by institutions. Indeed, retail orders are often internalised or otherwise kept off the floor, and therefore contribute even less to price discovery. (Battalio & Jennings, 2005, p. 30 et seq.)

In light of this evidence, it is somewhat naïve to assume that institutional investors are significantly informed by retail order flow, rather than vice versa, and that the net gains in price formation from increased competition between markets would be reversed if the competition were to continue.

2.2.5. Competition Between Orders

The debate has been complicated by a new phrase to have recently entered the lexicon: "competition between orders". SEC Chairman Donaldson testified as follows:

The fundamental challenge of the national market system can be expressed in a nutshell as promoting and balancing two essential, yet distinct, types of competition: competition among markets for trading services and competition among the orders of buyers and sellers in individual...
stocks. Each of these forms of competition is essential to the well-being of investors and listed companies. The importance of competition among markets for trading services is self-evident, yet competition among orders is at least as important, for this competition to be the best price produces narrow spreads and deep liquidity. (Donaldson, 2005, p. 10.)

The provenance of this form of competition is somewhat obscure. Jain (2001), attributes fragmentation with "split[ting] the trading volume across trading venues and decreas[ing] price competition between orders thus decreasing liquidity". (Ibid., p. 9). However, the authority cited, Madhavan (1995), doesn't consider at all the effect of fragmentation on liquidity. Instead, the "article analyzes the impact on price formation and market fragmentation of rules requiring timely disclosure of trading information …." (Ibid., p. 579.)

Regardless of provenance, "competition between orders" has become shorthand for the premise that every limit order should be displayed to every participant in the market, and every market order should have access to every limit order, regardless of the venue(s) chosen by the two parties. In the context of a best-price obligation, "competition between orders" means that each market order should be executed against the best-priced limit order, regardless of venue.17

Dr. Steil suggests that this standard is somewhat narrow-minded:

There can be no doubt that thinking about “the right” market structure at the SEC and the European Commission has been profoundly shaped by the proliferation of continuous electronic auction market systems, operated both on and off exchanges, over the course of the last decade. The NASDAQ and London dealer market structures were buried once and for all in 1997, and floor auctions in the derivatives markets have been driven to near extinction by Eurex and others. It should not be surprising, therefore, that regulators have adapted their conception of what “good trading” looks like accordingly.

The problem is that continuous electronic auction markets, as useful as they are, have flaws that are apparent to any institutional trader. They require institutional-sized orders to be chopped up into small bits, each often as little as 1% of actual order size, and executed over days or weeks in order to avoid huge market impact costs. That’s why in every major US or European marketplace—New York, NASDAQ, London, Frankfurt, Paris—about 30% of trading volume is executed in blocks, “upstairs”, away from these systems. (Steil, 2005.)18

In other words, using "competition between orders" to determine what constitutes best execution—meaning that every order should be admitted to the same execution queue—pre-empts an investor from choosing an alternative market structure which may better suit one's immediate needs.

17 More cynically, "competition between orders" has become shorthand for preservation of the U.S. Intermarket Trading System (ITS). See, e.g., McCooey, Jr. (2004): "The competition between orders represented by brokers at the point-of-sale on the Floor of the NYSE helps to ensure fair, orderly and liquid markets." (Ibid., p. 4.)

The ITS was adopted in 1975, and links nine exchanges in the so-called National Market System. The ITS requires exchanges to route orders to competing exchanges if the originating exchange is unable to post the national best bid or offer. In practice, the specialists on the NYSE almost invariably post the best bid or offer, but only for a minimum parcel. This, however, is sufficient for the entire order to be directed to the NYSE. As a result, the NYSE's share of trading in NYSE-listed stocks has declined from 82% to 80%, over the same period that NASDAQ saw more than half of the trading in NASDAQ-listed stocks migrate to other venues, including the NYSE.

18 Macey and O'Hara (1997) are also cynical:

The NYSE has further argued that agency-auction market trading is the most advantageous market mechanism in terms of achieving best execution. This conclusion has prompted calls for regulatory reforms such as the repeal of SEC Rule 19c-3, which allows off-exchange trading of listed stocks. Essentially, the NYSE view of best execution is that it is impossible for rival markets to achieve best execution prices because the other markets do not utilize the agency-auction market. Consequently, the duty of best execution can only be met by sending orders to the best market, which by this definition is the NYSE. (Ibid., p. 217, footnote omitted.)
Indeed, SEC Commissioners Cynthia A. Glassman and Paul S. Atkins reject the very legitimacy, legislative and regulatory, of Chairman Donaldson’s adherence to the principle of competition between orders: "We find no mention of different types of competition in the language of Section 11A, the source of the Commission’s authority in this area", and arriving at the same conclusion at Dr. Steil, "and we believe the rule is anti-competitive." (Glassman & Atkins, 2005, p. 29.)

**Best Price and Better Market Quality**

From a public-goods perspective, an explicit best-price obligation requires investors to subscribe to a fully transparent continuous auction market structure, regardless of investor preference. Best price is thereby anti-competitive in its bias against alternative market structures. Macey and O'Hara (1997) go further, and conclude that any best-price standard would be "at best futile and at worst detrimental to the overall efficiency of markets". (Ibid., p. 219.)

This raises the more pertinent question, whether better market-quality measures are produced by competition between markets or by competition between orders. Competitive markets not only match the preferences of many investors, but the evidence definitively supports the proposition that competitive markets improve execution costs, not least by reducing the rents which intermediaries are able to extract from the trading process. Furthermore, there is little evidence that the routing of business to multiple venues decreases liquidity, particularly when the venues are populated by many of the same brokers and dealers.

As to price formation, we note that the process is not undermined by a substantial portion of trades taking place on less-transparent venues, provided there is a sufficient level of effective consolidation through informational linkages. From a regulatory perspective, then, it should be a priority to determine precisely what linkage requirements RS may wish to impose on competing markets.

**2.3. Best Price and Enforcement**

If best execution is defined any way other than best price, the determination of broker mis- or malfeasance requires a case-by-case assessment of the facts—a costly and time-consuming process that can only reduce the prophylactic nature of the rule. On the other hand, establishing best price as a bright-line standard for best execution makes enforcement a matter simply of comparing times and prices—a process amenable to cheap and efficient automation.

That said, even enforcing best price is not trivial if multiple venues are involved. As discussed in detail infra, section 3.2.1, orders on electronic markets are entered, traded upon, or withdrawn in fractions of seconds. Only if clocks are synchronised to even smaller tolerances, and every broker and dealer in the market enters its time-stamps within a similarly small window, is the matching of trades against prevailing quotes possible. Synchronisation should be an easier issue in Canada than in the United States. Most of any future overlap in trading will likely be attributable to ATSS which are not yet established in the Canadian marketplace. Requiring new entrants to synchronise to existing entrants can simply be imposed as a condition of entry. This presumes, however, that the existing exchanges have sufficiently critical-tolerance technology and procedures.

In the final analysis, even if the benefits of easier enforcement were able to be realised, the unsuitability of best price as the sole arbiter of best execution renders the discussion moot.
3. Trade-Through

RS has asked of the impact of trade-throughs on market integrity, and the implications of an opt-out provision. Following the framing considerations first set out above, we restate this as follows:

From the perspective of investors who enter a market order –

- does a trade-through rule entail lower overall execution costs?
- does an opt-out provision confer a benefit?

From the perspective of investors who have entered limit orders –

- how effectively would a trade-through rule mitigate harm arising from being traded-through?
- how would an opt-out provision reduce the effectiveness of a trade-through rule?

From a public-goods perspective –

- how do trade-throughs affect market quality?
- does a trade-through rule promote liquidity by encouraging the placement of limit orders?

3.1. The Form of the Trade-Through Rule

As noted in our preliminary comments supra, while the U.S. market provides much of the data on which the relevant academic research is based, we have tried to conduct our analysis so as to be context-neutral. In respect of trade throughs, this is complicated by the current version of the U.S. rule as embodied in the Intermarket Trading System (ITS). The ITS requires that a market order be routed to the market where the best limit order is shown, regardless of the volume represented by the limit order. To the extent that the volume of the market order exceeds that of the limit order, it is executed at the same venue, regardless of any other limit order shown on any other venue.

To be polite, the rule reflects its 1975 vintage, when the technology simply didn't support complex order-splitting and -routing. From a policy perspective, however, the rule is deficient in every respect:

- If the purpose of the trade-through rule is to protect market orders from inferior execution, there can be no justification for forcing a 1,000-share market order to a venue showing the best price for 100 shares, but a markedly worse price for the balance. The investor placing the market order is demonstrably better off if the order is directed to the venue offering the best aggregate price. Better still if the investor's order can be broken into parcels that take advantage of the best prices at multiple venues. In any case, though, if a market order is deemed worthy of protection, not to protect the entire order is nonsensical.

- An alternative policy goal of the trade-through rule is, by protecting limit orders from being bypassed, to encourage investors to place limit orders and reward them for the liquidity and better price formation which the orders lend to the market. An ITS-style protects one order only—even parasitical trades (see note 4 supra)—and patently fails to reward the informational value of large limit orders placed just off the best bid or offer. It is one thing to promote the interests of investors placing limit orders vis à vis the interests of investors placing market orders (see discussion infra, section 3.4.5), but it is another matter to give a small subset of limit orders a high level of protection, and offer no protection whatsoever to the majority of limit orders. Instead of rationally advancing the policy objective, this type of trade-through protection merely invites "gaming". (Glassman & Atkins, 2005, p. 31.)
That said, whether a trade-through rule protects only the top of the order book or its entire depth is not critical to the analysis below. The focus of the discussion is whether any trade-through rule, regardless of particulars, does indeed protect market orders and encourage and properly reward limit orders.

3.2. Protecting Investors Trading Through

3.2.1. Quantifying the Losses

The SEC's Office of Economic Analysis performed an analysis of trade-throughs in over 5,000 U.S. stocks on three different days in 2003. The OEA estimated that from 2% to 10% of trades and from 2% to 13% of share volume "traded through", i.e., was executed at a price worse than the best-priced limit order (the national best bid or offer, or NBBO) on other exchanges and linked ECNs. Extrapolating to the full year, the OEA estimated that the aggregate difference between the NBBO and the price at which trade-throughs were executed ranged from US$252 million to US$321 million. This range, according to the OEA, represents the amount by which investors were hurt by their brokers not executing at the NBBO. (Office of Economic Analysis, 2004.)

The reaction of the research community to the OEA's methodology was scathing. James J. Angel, commenting to the SEC, noted that clock synchronisation rules require that time-clocks be only within 3 seconds of a standard time source. Thus a trade which is stamped at 12:00:03 by a clock three seconds fast appears to have followed a quote which is stamped at 12:00:02 by a clock three seconds slow, when in fact the trade preceded the quote by five seconds. Furthermore, delays of up to 90 seconds are permissible, and multiple-leg block trades often are consolidated into a single report. (Angel, 2005, pp. 3-4.)

These errors are compounded by widespread compliance failure in trade reporting. Ellis, Michaely, and O’Hara (2000) found that 64% of sampled NASDAQ trades were reported with delays of more than 14 seconds (ibid., p. 545), and Battalio Jennings (2005) found "[i]n unpublished work, … that only 56% of a sample of large NYSE trades in an audit-trail database can be matched to prints in the trade data used by the OEA within a 10-second window" (ibid., p. 15, emphasis in the original). Against this pattern of late trade reporting, one must also consider patterns in the limit-order book. Hasbrouck and Saar (2004), studying the Island ECN, found that 25% of all limit orders were cancelled within two seconds, and more than 40% were cancelled within 10 seconds. (Ibid., p. 15.) Given the clock synchronisation problem, permissible reporting delays, compliance failures, and the transience of the limit-order book, the OEA's estimates of trade-through rates simply are not dependable.

Even if the estimated trade-through rates were accurate, the estimate of economic loss was overstated. The OEA acknowledged that no adjustment was made for the depth of the limit-order book.

---

19 The OEA used the trade and quote logs of the primary exchanges (NYSE and NASDAQ), but the log itself was consolidated from data provided by numerous contributors, including competing exchanges. The OEA assigned quotes and trades into three-, five- and eight-second "windows", and then conservatively used the least aggressive NBBO during the window to identify trade-throughs. Most of the headline numbers in the OEA report used the three-second window.

20 Dr. Angel is Associate Professor of Finance at Georgetown University's McDonough School of Business.

21 During the time of the study, Island was the largest ecn on the NASDAQ. Island later merged with Instinet.

22 As further evidence, the OEA reported that 1.6% of all trades on ArcaEx were trade-throughs. However, ArcaEx is a fully automated system with strict price-time priority—its systems simply do not allow trade-throughs. As Professor Angel comments, "this demonstrates the weakness of the staff’s methodology more than a weakness with ArcaEx". (Angel, 2005, p. 3.)

23 Assume the limit-order book contains two buy orders, 100 shares at $10.00, and another 100 shares at $9.99. If a market order to sell 200 shares was executed at $9.99—i.e., it traded through the $10.00 order—the OEA calculates the loss at $0.01 on 200 shares. However, had the market order traded against both orders in the book, the difference in proceeds would have been only $0.01 on 100 shares, or only half the amount calculated by the OEA. This aberration is magnified when the trade-through is significantly larger than the size of the order at the NBBO, and exacerbated by the fact that the OEA allows for multiple trade-
were the US$252 million to US$321 million range accurate, on total annual volume of US$16.8 trillion, the losses to investors amount of "mere rounding error" of between 15 and 19 hundredths of a basis point. (Glassman & Atkins, 2005, p. 12.) It is easy to agree with Commissioners Glassman and Atkins that "[t]hese percentages do not indicate a significant problem with trade-throughs or best execution". (Ibid.)

Finally, many (if not most) of these apparent trade-throughs relate to orders executed on ECNs and other ATS, i.e., away from an NBBO posted on a conventional exchange. It is inappropriate to look solely at any putative losses suffered on ATS-transacted market orders. One must also consider the "drastic reduction in execution costs" which ATS have collectively delivered, which constitutes evidence of "diminished dealers’ and specialists’ rents". (Goldstein et al., 2005, p. 13.) In other words, a trade-through rule may protect market-order investors from the misdeeds of their own brokers, but ATS definitively protect investors from "losses" suffered at the hands of other market insiders. Any trade-through rule which impairs the effectiveness of ATS may just deliver a small benefit while destroying a larger one.

3.2.2. Opting Out

We refer to the discussion supra, section 2.1, relating to execution costs in a best-price regime. As noted, strategies other than blindly trading against the limit-order book offer the potential for better execution. Accordingly, giving investors the opportunity to opt out of a trade-through rule—best-price by another name—confers a benefit.25

Larger investors are better positioned to monitor the actions of their agents and would therefore be less likely to suffer agency losses from trading through. (See Harris (1996), who notes the difficulties, especially facing retail investors, of auditing best execution. (Ibid., pp. 3-4.) Smaller investors thus seem likely to bear a disproportionately large share of any agency losses which would be avoided by a strict trade-through rule. This would suggest that opting out might be limited to investors with larger portfolios or transacting larger trades. However, this too is an issue of fairness: The policymaker must balance the (miniscule) losses which might be prevented by a trade-through rule against the perception of giving smaller investors fewer rights than larger investors to optimise execution.

RS has proposed to extend existing trade-through prohibitions to Access Persons, i.e., investors accessing alternative trading systems directly, without the services of a broker.26 We note that, from the perspective of these investors, a trade-through rule confers no benefits at all, as there is no agent in a position to extract improper emoluments.

3.3. Protecting Investors Traded Through

3.3.1. Quantifying the Losses

Despite the inconclusiveness of the OEA's trade-through data, most market observers agree that trading through is often observed and rarely enforced. Dr. Steil testified that "the [trade-through] rule is not even enforced at present against its leading supporter and only systematic violator, the New York Stock Exchange, which trades through other markets hundreds, even thousands of times a day.” (Steil, 2004,
p. 4.) Similarly, ArcaEx’s CEO testified as follows: "Industry insiders have known for years that the trade-through rule is the least enforced rule this side of the double nickel speed limit on America’s highways. For example, despite the fact that there is a trade-through rule for NYSE-listed securities, ArcaEx quotes are traded through on average of over 2,000 times per day." (Putnam, 2004, p. 9, footnote omitted.)

We have found no research into quantifying the cost of these trade-throughs to investors whose limit orders have been bypassed. The same OEA analysis cited above, which estimated agency losses from trade-throughs, stated only that "[i]t is difficult to assign a cost to the bypassed orders, as the orders may or may not subsequently be filled ...."  If indeed the bypassed order is later filled, the limit-order investor suffers no harm in terms of the price paid/received, as the very nature of a limit order predetermines its value. While not conclusively causing harm, a trade-through in fact creates the opportunity for a benefit: When traded through, an alert investor can cancel the bypassed order and re-enter it at a different price, perhaps obtaining a better result than if the order were filled at the original price. (Cf. supra, section 3.2.1, noting the high cancellation rates of limit orders entered on the Island ECN.) Thus there is no quantifiable evidence that a trade-through harms the bypassed investor, while there is the possibility that some traded-through investors actually improve their outcomes. However, being traded through delays execution for the investor, therefore increasing execution risk.

It would be possible to design a research project to measure the exact harm, if any, suffered by investors traded through. The Australian Stock Exchange is an example of a market with a properly time-synchronised central limit-order book, but which also allows trade-throughs in the form of both block trades and on-market crossings. The analysis would require that each limit order traded-through be followed until such time as the order is either filled or cancelled—information which is available to researchers. If the order were filled, then the investor would have suffered no economic effect. If the order were cancelled, a comparison of the market mid-point immediately prior to the trade-through, to the market mid-point at the time of cancellation would indicate any benefit or harm from the trade-through. The CMCRC would be pleased to undertake this project should RS so require.

### 3.3.2. Alternative Remedies

The trade-through rule puts the interests of investors entering market orders in direct conflict with the interests of investors entering limit orders, and unequivocally favours the latter. By examining alternative remedies—ways in which investors placing limit orders can be mollified other than be forcing execution on investors placing market orders—it may be possible to diffuse the conflict.

Under the current version of the U.S. trade-through rule, investors traded through can require their limit order to be executed. (Angel, 2005, p. 4.) Alternatively, a traded-through investor could be given cash compensation. In either case, the net effect is to monetise the externalities of trading through. To the extent that policymakers determine that being traded through gives grounds for recompense, the cost would be borne by the party causing offence.

An alternative remedy might complicate the calculus of determining the optimal execution strategy, but at least investors would still be able to exercise their preferences. These investors would choose not only their strategy, but also the market structure through which to implement that choice, and thus the benefits of competition would be unabated.

### 3.3.3. Opting Out

We note that, if one accepts that investors traded through actually do suffer economic harm, any type of opting-out provision reduces the efficacy of a trade-through rule. However, by adopting an alternative remedy, opting out becomes a simple matter of assessing the costs and proceeding if it still remains attractive.
3.4. Trade-Throughs and Market Quality

3.4.1. Execution Costs

Effective 4 September 2002, the SEC promulgated a three-cent de minimis exemption to the trade-through rule as it applied to the three heavily traded ETFs. The OEA (2004) compared quote and trade data from the months before and after the rule change for the ETF representing the NASDAQ-100 index. It found a slight improvement in quoted spreads, and a slight decline in the number of trades filled at worse prices than the NBBO. "Based on this analysis, we do not find evidence of a deterioration in execution quality …." (Ibid., p. 1.)

Hendershott and Jones (2005) examined at the same event, expanding the scope to all three exempted ETFs and performing a more detailed analysis. As to execution costs, they found that share-weighted effective half-spreads declined in all three lines, and in two, by a statistically significant amount. The investigators then sought whether the de minimis exemption was being used disproportionately to the disadvantage of smaller investors. Looking on a venue-by-venue basis at trades of 1,000 shares or less, there were statistically significant reductions in execution costs in several venues for two of the ETFs; for the third ETF, spreads actually fell in every venue, but not by a statistically significant amount. (Ibid., pp. 8-9.)

In theory, smaller effective spreads indicate either that liquidity providers are taking less out of each trade, or that the trades themselves are associated with smaller price impacts. To decompose the savings in execution costs, the investigators looked to the 5-minute realised spread. This compares the price as at the time of the trade to its counterpart five minutes later. The extent to which the price moves against the liquidity provider constitutes an estimate of the cost of the limit order. Again looking only to smaller trades, the investigators determined that realised spreads fell modestly, but significantly. From this, they were able to conclude that "the market for supplying liquidity to smaller orders actually becomes slightly more competitive with the de minimis exemption …." (Realised spreads were also compared across venues, to identify any systematic attempts by any one venue to use the de minimis exemption to improperly route orders. There was no evidence of any such behaviour.) (Ibid., pp. 11-12.)

The investigators also looked at the overall prevalence of trade-throughs before and after the rule change, applying more rigorous statistical tests. Like the OEA, they found an overall slight decline. They also found that in every ETF there were statistically significant declines in at least one category of trade-throughs. (A venue-by-venue analysis did show that the AMEX and Boston Stock Exchange—the exchanges with the two highest trade-through rates—saw modest, but statistically significant, increases in trade-throughs post-rule change. The investigators limited their comment to a "very interesting"). (Ibid., p. 19.)

Professors Hendershott and Jones conclude that "[t]here is no evidence that relaxing the trade-through rule worsens market quality." However, they also made the telling point that these particular results are not conclusive as to the effect of trade-throughs on market quality. As the OEA also found, the rule change did not increase the number of trade-throughs. Accordingly, one can only conclude that the trade-throughs made under a de minimis exemption do not adversely impact market quality, compared to trade-throughs made in violation of a strict no trade-through regime. They continue: "[T]his highlights the importance of enforcement. … Thus, we would advise the SEC and other regulators to consider not just the trade-through rules themselves but also their enforcement implications." (Ibid., p. 22.) (Cf. supra, section 2.3, relating to the advantages of enforcing a strict best-price rule.)
3.4.2. Price Formation

The investigators also used the _de minimis_ exemption to test the impact of trade-throughs on price discovery. They measured autocorrelation of returns, both before and after the rule change, to determine how quickly information was incorporated into price. For one ETF, the rate at which innovations were fully incorporated into price speeded up; for another, it declined; and for the third, it was unchanged. From this, the investigators found it easy to conclude that the rule change had neither a significantly positive nor significantly negative impact on price formation.

There is an extremely active market in derivative contracts (futures and futures options) which use these ETFs as reference securities. Theory predicts, and past studies indicate, that price discovery takes place first on the derivative markets, due to their lower relative cost. Professors Hendershott and Jones compared the interaction of the trading in the ETFs with the trading in the futures market, to see if the pattern had been changed by the rule change. Most measures of cross-autocorrelation showed that the ETF market continued to lag slightly behind the futures market. There were a few cohorts, though, where the lag shortened by a statistically significant amount, indicating a marginal pick-up in the ability of the ETF market to process new information into price. (_Ibid._, p. 16.)

The investigators concluded as follows: "When the trade-through exemption goes into effect, some tests indicate that prices become slightly more efficient, and some tests indicate that prices become slightly less efficient. Some tests find no evidence of a change in efficiency, and that is probably the best way to interpret the overall evidence. (_Ibid._, p. 16.)

Professors Hendershott and Jones’s focus on smaller limit orders reinforces the popular intuition that trade-throughs primarily bypass retail investors. Given this underlying premise, it is informative to revisit the findings of Macey and O’Hara (1997) discussed _supra_, section 2.2.4. (Large trades, even when conducted in less-transparent markets, contribute proportionally more to the price formation process than smaller trades.) Professors Macey and O’Hara note that "all trades are not created informationally equal; the orders of small, retail traders tend to be less informative than those of market professionals." (_Ibid._, p. 16.) Thus even if frequent trade-throughs discourage the placement of limit orders, the overall loss to price discovery would seem to be tolerable. However, this argument is weakened by the fact that institutional investors are known to break-up their orders into smaller parcels in order to disguise their trades and reduce their market impact (Chakravarty, 2001). Therefore it is possible that small institutional orders also get traded through.

3.4.3. Liquidity

Limit orders provide liquidity to the market, thus any policy which encourages the entering of limit orders necessarily improves market quality. At first it may seem self-evident that a trade-through rule increases the propensity of investors to provide liquidity. Nazareth28 (2005) offers:

> [T]rade-throughs also can undermine incentives to display limit orders, which are the building blocks of price formation. In addition, many of the investors whose market orders are executed at inferior prices may not even be aware they received an inferior price. A uniform rule establishing strong price protection on an order-by-order basis would both protect the interests of investors by reinforcing the duty of best execution and promote the display of limit orders, and thereby improve the fairness and efficiency of the marketplace as a whole.

---

27 In a perfectly price-efficient market, the return earned by a stock in any given period is independent of previous returns. In an inefficient market, it may take two or more periods for a given piece of information to be fully absorbed into price. If so, there is a tendency for returns to be autocorrelated: a positive return is more likely to be followed by another positive return (and likewise for negative returns), as the market takes time to react.

28 Annette L. Nazareth is Director, Division of Market Regulation, U.S. Securities and Exchange Commission.
However, the academic literature is not so conclusive. We have been unable to identify any research, either theoretical or empirical, which addresses whether strong trade-through protection increases the volume of orders in the limit-order book. Atkins (2005) accurately asserts:

The Commission has not concluded through any study that investor confidence has been affected by the minimal number of trade-throughs in the current market, yet we conclude that investor confidence will improve even though trade-throughs will theoretically decrease, not stop. … The failure to cite an example of a market that has implemented the trade-through rule and increased quoted depth makes me believe that this is a theoretical experiment with an unpredictable outcome.

Weaver (2004b) offers another seemingly plausible explanation:

There is an old adage that “liquidity begets liquidity.” In other words, limit order traders will submit limit orders where market orders are. It is similar to the fact that the more traffic exists on a highway, the more gas stations will exist. If the traffic goes away—so will the gas stations. Similarly, if market orders get routed away from the venue with the best price, limit orders will leave that venue as well. There will be less price competition and as a result spreads will widen. (Weaver, 2004a, p. 17.)

Yet here too, what appears to be common sense is not supported by research findings. In fact, Biais, Hillion, and Spatt (1995) find that, at least at the Paris Bourse in the early 1990s, limit orders are more likely to be submitted when the limit-order book contains relatively few orders, or when the bid-ask spread widens. Limit orders do not beget limit orders.

In short, there is no proven causal link between trade-through rules and a sustained or enhanced level of limit order submission. Anecdotal evidence suggests otherwise, as the OEA found that the NASDAQ, which is not subject to a trade-through rule, attracts more limit orders than the NYSE with such a rule. Conversely, there is no research to the effect that a weakening of trade-through protection results in a decline in limit orders.

The studies conducted by the OEA and Professors Hendershott and Jones into the de minimis trade-through exception for ETFs constitutes a missed opportunity. It would have been rather straightforward to have examined the flow of limit orders before and after the rule change. Absent a statistically significant decrease in limit orders after the loosening of trade-through "protection", one cannot reject the proposition that the trade-through rule is irrelevant to market liquidity.

3.4.4. Cost

The potential savings in enforcement costs resulting from a bright-line best-price test was discussed supra, section 2.3. That discussion ignored the cost of compliance. Clock-synchronisation issues aside, however, meeting a best-price obligation to an investor would merely require a broker to be aware of all prices in the market. One would expect that the broker would be so aware as a matter of course, thus the ongoing cost of compliance would be minimal.

Dr. Larry Harris, the former SEC Chief Economist, believes that the same can be said of linking exchanges to the level of interoperability required to implement a comprehensive trade-through rule: "[I]t merely requires that traders do what they should be doing anyway. Accordingly, the unanticipated costs of the rule should be small." (Harris, 2005, pp. 8-9.)

Professor Harris's view is not shared by the SEC. The Commission estimated a US$144 million start-up cost (including US$5 million in systems modifications for each of nine self-regulatory organisations and US$3

---

29 Paul S. Atkins is a Commissioner, U.S. Securities and Exchange Commission. The speech cited here was made on the same day as Director Nazareth's.
million for each of twenty trading centres) and annual costs of US$22 million.  (SEC Proposing Release No. 34-50870, pp. 106 et seq.; SEC Release No. 34-50870, pp. 329 et seq.)  (Nor by the industry.  In a Securities Industry News Editors’ Poll, 52% of respondents strongly agreed, and 28% agreed, with the following statement: "The SEC has badly underestimated the costs of compliance with Regulation NMS.")

Amihud and Mendelson (2005) approve of a trade-through rule, but one which only protects the top of the limit-order book.  They view such a "loose coupling" as nothing more than the promulgation of standards, similar to protocols for accessing the internet:

> Keeping the standards requirements to a minimum while preserving the ability to innovate on the network nodes quickly, at low cost and with minimal coordination has had very substantial benefits ….  Creating tighter coupling in an attempt to optimize a short-term objective is short-sighted, as it ignores the costs of complexity and interdependence ….  A tightly-coupled system creates substantial interdependencies that impede innovation and change.  (Ibid., pp. 3-5.)

We suggest that the extent of the coupling required to implement a fully integrated trade-through rule is yet to be determined.  In any event, the incremental expense of compliance will eventually be borne by market participants, and increase the general level of transaction costs.

3.4.5. Fairness and the Policy Choice

A trade-through rule benefits all investors—both those placing market orders and limit orders—only in those cases when the displayed limit-order book is sufficient to absorb a market order with little or no price impact.  Beyond those restricted circumstances, a trade-through rule specifically favours the interests of the investors placing limit orders over the interests of investors placing market orders.

It is not clear that preferring limit-order investors is generally accepted as equitable.  Dr. Benn Steil testified as follows:

> [I]t takes two to trade.  The fellow who puts down a limit order in market X has no moral standing over the gal who sees a better package deal in market Y. Appeals to 'fairness' favor neither one over the other.  (Steil, 2004, p. 3, emphasis in the original.)

While others may choose to disagree, there is by no means a consensus that, on the individual investor level, limit orders should be preferred as a matter of public policy.

Ironically, it is fairly simple to identify plausible circumstances in which market-order investors who are forced to trade against limit orders experience significantly increased execution costs.  At the same time, there is no evidence that limit-order investors suffer economic losses from trade-throughs.  Even the most generous estimates of losses are tiny relative to the total transactional flow, thus the public-good value of eliminating these losses is equally tiny.

From a public-goods perspective, trade-through rules seem to have only minimal effects on execution efficiency and price formation.  Most importantly, there is no evidence of a causal link between the strength of a trade-through rule and liquidity in the form of an increased supply of limit orders.  The only impact on externalities that a trade-through rule seems certain to attain is to restrict the competition between markets—a competition which has repeatedly delivered reduced execution costs—and a formidable compliance challenge.

From a policy perspective, then, the only proven benefit—eliminating whatever harm is suffered by investors whose limit orders have been traded through—seems to be significantly outweighed by the private and public costs.  However, if the principles of market integrity seem compelling—that limit-order investors must be protected—the adoption of alternative remedies seems very attractive.  The monetisation of externalities is increasingly accepted as the best way to let market forces assist in the attainment of policy goals, and in the case here, would significantly decrease the unintended negative externalities of an unyielding trade-through rule.
References


Financial Services Authority (United Kingdom), Consultation Paper 154, Best Execution, October, 2002.


Harris, Lawrence, 1997, Order exposure and parasitic traders, Deutsche Börse AG Symposium on equity market structure for large- and mid-cap stocks, 12 December 1997, Frankfurt.


