

Bid-Ask Spreads: Measuring Trade Execution Costs in Financial Markets

Hendrik Bessembinder *

David Eccles School of Business
University of Utah
Salt Lake City, UT 84112
U.S.A.
Phone: (801) 581 8268
Fax: (801) 581 3956
Email: finhb@business.utah.edu

and

Kumar Venkataraman

Edwin L. Cox School of Business
Southern Methodist University
Dallas, TX 75275
U.S.A.
Phone: (214) 768 7005
Fax: (214) 768 4099
Email: kumar@mail.cox.smu.edu

Final version: March 2009

Forthcoming, Encyclopedia of Quantitative Finance

Keywords: bid-ask spread; trading cost; quoted spread; effective spread; price impact of trade; realized spread; benchmark price; stock exchange; liquidity cost.

Abstract

Bid-Ask spreads, which measure trade execution costs, and reflect the price concessions necessary to complete transactions quickly, are important as indicators of market quality and in determining traders' actual investment results. Execution costs arise because it is costly to provide liquidity, and can be estimated based on comparisons of trade prices to proxies for underlying security value, with the most common proxy being the quote midpoint. Comparisons can be of trade prices to midpoints at or before the time of the trade, as in effective spread measures, or to midpoints after the trade, as in realized spread measures. Recent research indicates that trade execution costs have declined in U.S. markets in recent years, and documents substantial variation in average trading costs across international equity markets.

* Corresponding Author

The literature on asset pricing often assumes an ideal world where security prices are set by market participants in *frictionless* financial markets. However, a variety of market frictions, including trading costs and constraints on short selling, exist in actual markets. It is important for market participants to accurately estimate and incorporate the impact of trading costs. For portfolio managers and investors, implementing investment decisions is costly and will typically lead to a shortfall in investment performance (see [26]) relative to that theoretically attainable in frictionless markets. Decisions need to be conditioned not only on the fundamental soundness of potential investments, but also on the anticipated costs of implementing the required trades. Estimates of trading costs also serve as a measure of market quality, allowing policy makers to assess the impact of regulatory reforms, exchange officials to assess the effects of trading rule and market structure changes, and informing corporate managers' decisions on where to list their shares. This article provides an overview of several issues related to measuring trading costs in financial markets. It focuses on three related measures of trading costs: quoted spreads, effective spreads, and realized spreads.

The Nature of Trading Costs

A fundamental issue in trading is the asynchronous arrival of buyers and sellers (see [9]). This creates uncertainty as to the amount of time that will be required to locate a counterparty, and regarding the market price that will prevail at the time a trading partner is located. This uncertainty can be mitigated by the continual presence of "liquidity suppliers," who stand ready to serve as counterparties, thereby providing immediacy of trade execution, i.e. "liquidity". Liquidity suppliers often take the form of designated market makers or dealers, but liquidity can also be provided by traders in the form of limit orders.

Liquidity providers need to be compensated for the costs involved. Besides order processing costs, dealers incur inventory holding and adverse selection costs. Accommodating investors' order flows generally leaves dealers holding inventory positions that are not optimal in terms of the diversification of risk (see [16] and eqf18/007). Further, dealers need to be compensated for the possibility that some buy or sell orders can originate with traders possessing superior information regarding security value. Dealers on average lose money on transactions with better informed traders (see [22], [13], eqf18/011 and eqf18/012), who sell to dealers ahead of price declines and buy from dealers ahead of price increases.

Dealers recover these costs by purchasing at a lower "bid" price, while selling at a higher "ask" (or "offer") price. The bid and ask prices are the dealer's quoted prices or "quotes", and the difference between the two is the bid-ask or quoted spread, a measure of trading cost. The corresponding quantities offered by the dealers at the quoted prices are referred to as the quoted depth, i.e., the bid depth and the ask depth.

In most financial markets, including the New York Stock Exchange and the Nasdaq Stock Market, liquidity provision by designated dealers is augmented by standing limit orders submitted by public traders. A limit order to buy sets a maximum price to be paid, while a limit order to sell sets a minimum price that will be accepted. A limit order may be viewed as a one-sided quote. Some markets (e.g. The Hong Kong Stock Exchange) operate without designated dealers, in which case the bid-ask spread is determined by the most aggressively priced unexecuted limit orders. In summary, the bid-ask spread, which measures trading costs for investors buying at the ask and selling at the bid, arises to compensate liquidity providers for order processing, inventory and adverse selection costs.

Measures of trading cost

To estimate trading costs, empirical methodologies rely on the simple intuition that transactions would occur at the true underlying security value in the absence of trading costs. Hence, the deviation between transaction price and an estimate of the true underlying security value is an estimate of trading cost. For buyer (seller) initiated trades, the traded price is expected to be higher (lower) than the true security value; the difference being the estimate of trading cost.

The Quoted Spread

The simplest measure of trading cost is the quoted spread (QS), which is defined as the difference between the bid and ask prices. The quoted spread measures the cost of completing a round trip (buy and sell), if trades are executed at the quoted prices. Execution costs for a single trade are often measured as half the spread, described on a percentage basis by equation (1):

$$\text{Quoted half-spread} = \text{QS}_{it} = 100 * (\text{Ask}_{it} - \text{Bid}_{it}) / (2 * \text{M}_{it}) \quad (1)$$

where A_{it} and B_{it} are the posted ask price and bid price for security i at time t , respectively, and M_{it} , the quote midpoint or mean of A_{it} and B_{it} , is a proxy for the true underlying security value.

The Effective Spread

In many dealer markets, including those that trade fixed-income securities and foreign exchange, the quoted prices are simply a starting point for negotiations between customers and dealers, and transactions frequently occur at prices other than the quotes. Also, in some markets, including those relying on trading floors, there may be *latent liquidity* not reflected in the quotes. On the New York Stock Exchange (NYSE), for example, market orders may execute at prices within the quotes when the specialist (the NYSE's designated dealer) or a floor broker elects to improve on the quote (see [28], [30] and eqf18/018). Many electronic exchanges allow traders to

hide some or all of the order size, implying that limit orders offering more attractive prices than the quotes may exist on the book (see [4] and eqf18/019). Further, quoted prices pertain only to the quoted depth; large orders might exhaust the depth at the quote and “walk up the book”, executing against limit orders with less attractive prices and leading to a weighted-average trade price outside the quotes.

When trades occur either within or outside the quotes, a better measure of trading costs is the percentage effective half spread, which is based on the actual trade price, and is computed on a percentage basis as described in equation (2):

$$\text{Effective half-spread} = \text{ES}_{it} = 100 * D_{it} * (P_{it} - V_{it}) / V_{it} \quad (2)$$

where P_{it} is the transaction price for security i at time t , D_{it} is an indicator variable that equals one for customer buy orders and negative one for customer sell orders, and V_{it} is an observable proxy for the true underlying value of security i at time t . The effective spread is based on the deviation between the execution price and the true underlying value of the security, and can be viewed as an estimate of the execution cost actually paid by the trader and the gross revenue earned by the liquidity provider.

It is also possible to distinguish between the non-informational (inventory and order processing) and informational (adverse selection) components of trading costs, based on the behavior of prices *subsequent* to a transaction. The seminar paper using this approach is [21]. The intuition is that non-informational transaction costs should result only in a *temporary* deviation of price from value, evidenced by a price reversal after the trade. That is, while customer purchases (sales) should occur at prices above (below) pre-trade value, we should subsequently observe a partial reversal of the price change. The price reversal is partial rather than full because the informational component of trading costs is on average associated with a

permanent increase (decrease) in security value after buys (sells). The informational component can be measured by the change in the estimate of security value, while the non-informational component can be measured by the reversal from trade price to post-trade value.

Estimating the Informational Component: The Price Impact of Trades

The possible presence of informed traders is revealed to liquidity providers in a noisy manner by the order flow imbalance, i.e., the difference between quantities of buy versus sell orders, which will tend to be positive when the security is undervalued and negative when the security is overvalued. Market makers incorporate the information in order flow imbalances by adjusting quotes upward (downward) after buy (sell) orders. These price adjustments reflect both the proportion of the informed traders vs. liquidity traders in the market, and the extent of superior information about security value held by the informed traders. The private information contained in trades, or equivalently the amount of adverse selection cost incurred by the liquidity provider, can be estimated using equation (3):

$$\text{Price impact of trade} = \text{PI}_{it} = 100 * D_{it} * (V_{it+n} - V_{it}) / V_{it} \quad (3)$$

where V_{it+n} denotes the security's true underlying value 'n' periods after the transaction. The price adjustment from V_{it} to V_{it+n} reflects the markets' assessment of the private information conveyed by the trade (see [17], [1] and eqf18/006). Research has documented that markets are particularly sensitive to order flow imbalances ahead of anticipated news disclosures, such as earnings announcements by corporations (see [23]), as price impacts are larger than on non-announcement days. The price impact of trades is an often used empirical proxy in the corporate finance literature for the degree of information asymmetry regarding security value across traders.

Estimating the Non-Informational Component: Realized spreads

The presence of informed traders will cause market prices on average to rise after customer buys and to fall after customer sells. Due to these adverse price movements, market makers earn less than the effective spreads for their services. Market making revenue net of losses to better informed traders can be measured by the *reversal* from the trade price to the post-trade value. The realized spread captures the extent of reversal, computed using equation (4):

$$\begin{aligned} \text{Realized spread} = \text{RS}_{it} &= 100 * D_{it} * (P_{it} - V_{it+n}) / V_{it} \\ &= \text{Effective Spread}_{it} - \text{Price Impact}_{it} \end{aligned} \quad (4)$$

As noted above, some studies refer to price impact and realized spreads as the ‘permanent’ and ‘temporary’ price impacts of a trade (see [21] and [25]). Some authors (see [29]) have argued that trading costs and market quality are better measured by temporary price impacts (realized spreads) than by total price impacts (effective spreads).

Implementation issues

To measure effective or realized spreads, researchers need to identify (1) whether the trade was initiated by a buyer or a seller, and (2) estimates of security value before and after the trade. As documented in recent research, the methodological choices regarding these issues are non-trivial and can significantly affect estimates of trade execution costs (see [27] and [32]).

Algorithms for assigning trade direction

Some publicly-available databases from international markets, e.g., Euronext-Paris and the Toronto Stock Exchange, as well as some proprietary datasets on institutional trading, e.g., those provided by consulting firms Abel/Noser or Plexus, contain information on the buy and sell orders submitted to the markets. In contrast, databases such as Trade and Quote (TAQ)

(released by the NYSE) and Nasdaq (released by Nasdaq), which are widely studied due to their comprehensive inclusion of all trade and quote data for all listed U.S. stocks, do not provide information on underlying buy and sell orders. As a consequence, the direction of the trade (i.e., whether the trade was initiated by a buyer or a seller) must be imperfectly inferred from the available data (see [3] for a detailed discussion).

In datasets where order level data is not available, the most widely used algorithm for assigning trade direction is that recommended by [24]. Their algorithm assigns trades completed at prices above (below) the prevailing quote mid-point as buyer (seller) initiated trades. Trades executed at the quote mid-point are classified based on a “tick test”. The tick test assigns a trade as a buy (sell) if the trade executes at a higher (lower) price as compared to the most recent trade at a different price. An alternative algorithm is proposed by [11], who assign trades executed at the ask (bid) quote as buyer (seller) initiated, while using the tick test for all other trades. Based on proprietary order level data, prior research finds that the algorithm proposed by [24] works fairly well, classifying about 85% of the trades correctly.¹ Error rates are slightly lower for the algorithm proposed by [11] (see [12] and [3]).

Research based on data from early 1990’s (see [14]) finds that trade report times lagged actual trade times for NYSE stocks. As a consequence, studies such as [24] recommend adjusting the time stamps by five seconds when comparing trades and quotes. However, for recent data, [3] and [11] report that the allowance for reporting lags is not necessary. They recommend that trades are best compared to contemporaneous quotations for both NYSE and Nasdaq stocks when assigning trades as buyer or seller initiated.

¹ However, the accuracy of the algorithms in post-decimalization data, which is characterized by substantial increases in the number of quote revisions and trades, has to our knowledge not been assessed.

Pre-trade benchmark price

Pre-trade price impact refers to implicit trading costs that may be incurred if prices move systematically away from the trader (rising before buys or dropping before sells) between the time of a trade decision and complete execution of the order (see [26]). Pre-trade price impact will tend to arise when larger orders are broken into smaller orders and executed successively. Prices may also move away from a trader because his trading intentions are detected by market participants who then “front run” the order or engage in “predatory trading” (see [7]), or simply infer information from the existence of the trading interest. In addition, prices will tend to move away from traders relying on momentum strategies. Executing a trading program too slowly exposes traders to the risk of larger pre-trade price impacts. On the other hand, executing orders that are larger than quote sizes too quickly may result in larger effective spreads. The skill of a trader handling larger orders lies in balancing these effects. To capture the impact of trader’s “timing” and “liquidity” decisions, Perold [26] recommends that the average of the prevailing bid and ask quote at the time of the trading decision be used as the pre-trade benchmark price.

However, data on trade decision times is often not available, except in specialized proprietary datasets such as that studied by [8]. Studies using publicly available datasets, including [17] and [1] use the quote mid-point at the time of the trade as the pre-trade benchmark price. Although adverse drift in prices ahead of trade executions is most obviously an issue for larger traders, recent evidence (see [3], [27] and [32]) indicates that prices move systematically and adversely in the seconds before even small trades are executed. Bessembinder [3] recommends that researchers use quotation midpoints in effect five seconds prior to the trade report time as a proxy for the true underlying price (V_{it}) when measuring effective spreads and price impact of trades.

Post-trade benchmark price

The post-trade benchmark price (V_{it+n}) should be measured when the market has had sufficient time to incorporate the information contained in the trade (see [17]). If the period after the trade time is too short temporary price effects may still dominate, or alternately, the market may not have had a chance to assess the trade's likely information content. If the period is too long, the measure will become unnecessary noisy due to the arrival of extraneous information. In the absence of theoretical guidance, studies have used different proxies for the post-trade benchmark price. Studies using institutional data have often used the closing price on the day of trade as the post-trade price (see [20]). The practitioner literature commonly relies on the volume weighted average price (VWAP) on the day of the trade. Among datasets with broad coverage, [17] use the first trade price both 5 and 30 minutes after the trade, while [1] use the quote mid-point 30 minutes and 24 hours after the trade. Bessembinder [3] used the mid-quote in effect 30 minutes after the time of the reference quote, or the 4 p.m. quotation for trades completed during the last half hour of trading. Werner [32] reports that realized spread measures obtained in large samples are relatively insensitive to the choice of the post-trade benchmark price.

Since September 2001, the Securities and Exchange Commission (SEC) has required each U.S. stock "market center" to compile and disseminate on a monthly basis various standardized measures of execution quality in nearly all publicly traded securities (see [6]). The intent of SEC Rule 605 (formerly 11Ac1-5) is to provide traders with information on execution quality at different market centers. The execution quality measures that each market now reports include round trip effective spreads, realized spreads, as well as average execution speed. The regulation provides something of an official validation of the effective spread and realized spread

measures described above, and also codifies a specific methodology to estimate trading costs based on the order level data available to each market center. Specifically, the effective spread compares the traded price to the quote midpoint (as a proxy for V_{it}) at order arrival, while the realized spread is based on the quote midpoint (as a proxy for V_{it+n}) five minutes after the trade.

Evidence on trading costs

Jones [19] provides a detailed perspective on trading cost in U.S. markets over the last century. He estimates that quoted spreads on Dow Jones stocks were in the range of 0.60% for sustained periods until the beginning of 1980s, with spikes in spreads observed during market downturns, such as Great Depression. During the last two decades, he documents that trading costs have fallen dramatically to around 0.20% for Dow stocks, partly facilitated by regulation but mainly by increased competition among market centers with the onset of electronic trading systems (see [15] for recent evidence).

Several recent studies have examined execution quality for comparable firms on the two major U.S. market centers – the NYSE and the Nasdaq. In 2001, subsequent to regulatory changes, [2] reports that effective spreads are similar for comparable firms in the two markets. The most recent evidence available appears to be that reported by [6], using Rule 605 data over November 2001 to December 2003, who reports an average effective spread for NYSE stocks of 6.2 cents per share, versus 8.8 cents for comparable Nasdaq stocks.

Several studies have examined execution quality in market outside the U.S. For example, [31] reports that effective spread for large firms traded on the Paris Bourse were 0.25% in 1997, relative to 0.21% for comparable firms on NYSE. Jain [18] reports on trading costs during the year 2000 for liquid firms on 51 stock exchanges around the world. The average effective

(realized) spread across exchanges is 2.13% (2.15%). However, there is considerable variation in effective (realized) spreads across markets, ranging from 0.10% (0.25%) at the NYSE (Luxembourg) to 14.47% (14.6%) in Ukraine. Research indicates that differences in execution quality across markets are related to both exchange-design features, such as tick size and order handling rules, and the regulatory environment, such as the enforcement of insider trading laws and the protection of shareholder rights (see [5] and [10]).

Conclusions

Trade execution costs, which reflect the price concessions necessary to complete transactions quickly, are important indicators of market quality and important determinants of traders' actual investment results. Execution costs arise because it is costly to provide liquidity, including order processing costs, inventory holding costs, and losses suffered to better-informed traders. Trading costs can be estimated based on comparisons of trade prices to proxies for underlying security value, with the most common proxy being quote midpoints. Comparisons can be of trade prices to midpoints at or before the time of the trade, as in effective spread measures, or to midpoints after the trade, as in realized spread measures. Further, the amount of asymmetric information present in a market can be estimated by assessing trades' price impact, measured as the difference between post and pre-trade estimates of security value. Recent research indicates that trade execution costs have declined steadily in U.S. markets in recent years, particularly subsequent to decimalization in 2001, and documents substantial variation in average trading costs across international equity markets.

Acknowledgements

We would like thank Charles Jones (the editor) for his comments.

Related articles

Eqf18/006

Eqf18/011

Eqf18/012

Eqf18/018

Eqf18/019

References

- [1] Bessembinder, H, Kaufman, HM (1997) A comparison of trade execution costs for NYSE and NASDAQ-listed stocks, *Journal of Financial and Quantitative Analysis* 32, 287-310.
- [2] Bessembinder, H. (2003a) Trade execution costs and market quality after decimalization. *Journal of Financial and Quantitative Analysis* 38 (4), 747-778.
- [3] Bessembinder, H. (2003b) Issues in Assessing Trade Execution Costs, *Journal of Financial Markets* 3, 233-257.
- [4] Bessembinder, H, Panayides, M, Venkataraman, K. (2009) Hidden Liquidity: An analysis of order exposure strategies in electronic stock markets, *Journal of Financial Economics*, forthcoming.
- [5] Bhattacharya, U, Daouk, H. (2002) The World Price of Insider Trading, *Journal of Finance* 57, 75-108
- [6] Boehmer, E. (2005) Dimensions of execution quality: Recent evidence for U.S. equity markets. *Journal of Financial Economics* 78, 463-704.
- [7] Brunnermeier, M, Pedersen, L. (2005) Predatory Trading, *Journal of Finance* 60 (4), 1825-1863.
- [8] Conrad, J, Johnson, K, Wahal, S. (2003) Institutional Trading and Alternative Trading Systems, *Journal of Financial Economics* 70, 99-134.

- [9] Demsetz, H. (1968) The cost of transacting, *Quarterly Journal of Economics*, 33-53.
- [10] Eleswarapu, V, Venkataraman, K. (2006) The impact of legal and political institutions on equity trading costs: A cross-country analysis, *Review of Financial Studies* 19 (3), 1081-1111.
- [11] Ellis, K., Michaely, R, O'Hara, M. (2000) The Accuracy of Trade Classification Rules: Evidence from Nasdaq, *Journal of Financial and Quantitative Analysis* 35, 529-551.
- [12] Finucane, T. (2000) A Direct Test of Methods for Inferring Trade Direction from Intra-Day Data, *Journal of Financial and Quantitative Analysis* 35, 553-576
- [13] Glosten, L, Milgrom, P. (1985) Bid, ask and transaction prices in a specialist market with heterogeneously informed traders, *Journal of Financial Economics* 14, 71-100.
- [14] Hasbrouck, J, Sofianos, G, Sosebee, D. (1993) New York Stock Exchange systems and trading procedures, NYSE working paper 93-01.
- [15] Hendershott, T, Jones, C, Menkveld A. (2007) Does algorithmic trading improve liquidity? working paper, University of California, Berkeley.
- [16] Ho, T, Stoll, H. (1983) On Dealer Markets Under Competition, *Journal of Finance* 35, 259-267.
- [17] Huang, R, Stoll, H. (1996) Dealer versus auction markets: A paired comparison of execution costs on NASDAQ and NYSE, *Journal of Financial Economics* 41, 313-357
- [18] Jain, P. (2002) Institutional design and liquidity at stock exchanges around the world, working paper, University of Memphis.
- [19] Jones, C. (2002) A century of stock market liquidity and trading costs, working paper, Columbia University.
- [20] Keim, D, Madhavan, A. (1996) The upstairs market for large block transactions: analysis and measurement of price effects, *Review of Financial Studies* 9, 1-36.
- [21] Kraus, A, Stoll, H. (1972) Price impacts of block trading on the New York Stock Exchange, *Journal of Finance* 27, 569-588.
- [22] Kyle, A,. (1985) Continuous auctions and insider trading, *Econometrica* 53, 13-32.
- [23] Lee, C, Mucklow, B, Ready, MJ. (1993) Spreads, depths, and the impact of earnings information: an intraday analysis, *Review of Financial Studies* 6, 345-374.
- [24] Lee, C, Ready, MJ. (1991) Inferring trade directions from intraday data, *Journal of Finance* 46, 733-746.

- [25] Madhavan, A, Cheng, M. (1997) In search of liquidity: block trades in the upstairs and downstairs market, *Review of Financial Studies* 10, 175-203.
- [26] Perold, AF. (1988) The implementation shortfall: paper vs. reality, *Journal of Portfolio Management* 14, 4–9.
- [27] Peterson, M, Sirri, E. (2003) Evaluation of the biases in execution cost estimation using trade and quote data, *Journal of Financial Markets* 6 (3), 259-280.
- [28] Ready, MJ. (1999) The specialist's discretion: stopped orders and price improvement, *Review of Financial Studies* 12, 1075-1112.
- [29] Seppi, D. (1997) Liquidity provision with limit orders and a strategic specialist, *Review of Financial Studies* 10, 103-150.
- [30] Sofianos, G, Werner, I. (2003) The Trades of NYSE Floor Brokers, *Journal of Financial Markets*, 139-176
- [31] Venkataraman, K. (2001) Automated versus floor trading: an analysis of execution costs on the Paris and New York exchanges, *Journal of Finance* 56, 1445-1885.
- [32] Werner, I. (2003) NYSE Order Flow, Spreads, and Information, *Journal of Financial Markets* 6, 309-335.