Journal of Financial Markets 6 (2003) 227–231

Journal of FINANCIAL MARKETS

www.elsevier.com/locate/econbase

Guest editorial

What we measure in execution cost measurement

I am fond of observing that finance and labor economics are the subfields within economics with the highest empirical standards and that market microstructure is the most empirically oriented subfield within finance. I am also fond of quoting (without attribution) a colleague who once observed (disparagingly, I am sad to say) that *real* finance people actually *like* microstructure. The four papers that comprise this special issue of the *Journal of Financial Markets* devoted to execution costs are testaments to the veracity of both propositions.

While these papers provide numerous insights into the magnitude of execution costs across market venues in a number of dimensions, their unifying theme is measurement: what we seek to measure, how we do so, what errors we make, and what, if anything, we can do about them. Bessembinder (2002) measures how trade signs and execution costs change across benchmark prices in Trade and Quote (TAQ) data. Peterson and Sirri (2002) use the NYSE System Order Database (SOD) Daily File to address these same questions; the added benefit of the SOD data is that it identifies buys and sells, permitting direct assessment of the accuracy of the Lee and Ready (1991) (LR) and Ellis, Michaely, and O'Hara (2000) (EMO) trade signing algorithms. Bacidore, Ross, and Sofianos (2002) provide a detailed examination of the effects of benchmark prices and other factors on execution quality measurement using a sample of over 7 million SuperDot market orders from the SOD files, obtaining quotes from the SODQ files that contain the same consolidated quotes that are available in the TAQ database. Finally, Werner (2002) examines NYSE execution quality measurement for all order types, not just system market orders, based on NYSE audit trail data. Hence, these papers reflect our ability to address all four questions, the source of microstructure's high empirical standards. They also provide useful answers that should help guide future research.

In this introduction, I will treat these papers as methodological, that is, as contributions to the study of our methods for execution cost measurement. In thinking about execution costs, I find it useful to consider an accounting framework for measuring execution performance. Such a framework begins with the individual trade in a single security on a single day. Accordingly, let q_s and p_s be the (signed) size and price, respectively, of a trade in this security at time s and let $Q_s = Q_{s-1} + q_s$ be the corresponding position of this trader at that time. We might compare p_s to some benchmark price p_{bs} to assess the quality of the execution, making the cost of trade s:

$$p_s q_s = p_{bs} q_s + [p_s - p_{bs}] q_s = p_{bs} q_s + \kappa_s |q_s|; \kappa_{bs} = \mathrm{Sign}[q_s][p_s - p_{bs}] \tag{1}$$

doi:10.1016/S1386-4181(03)00005-3

and so total expenditure on this security on a given day is given by:

$$\Sigma_s p_s q_s = \Sigma_s p_{bs} q_s + \Sigma_s \kappa_{bs} |q_s|. \tag{2}$$

Finally, the value of the end-of-day position in this security $MV_{c'}$ is

$$MV_{c'} = p_{c'} Q_c = p_{c'}[Q_c + \Sigma_s q_s], \tag{3}$$

where p_c and $p_{c'}$, are the closing price or quote midpoint of the security yesterday and today, respectively, and so marked-to-market daily profits are simply the change in the value of the initial position Q_c less the net costs of any trades during the day:

$$\begin{aligned} \pi &= MV_{c'} - \Sigma_s p_s q_s - MV_c = p_{c'} Q_{c'} - \Sigma_s p_s q_s - p_c Q_c \\ &= p_{c'}[Q_c + \Sigma_s q_s] - \Sigma_s p_s q_s - p_c Q_c \\ &= [p_{c'} - p_c] Q_c + \Sigma_s [p_{c'} - p_{bs}] q_s - \Sigma_s \kappa_{bs} |q_s| \\ &= R_{c'} MV_c + \Sigma_s [p_{c'} - p_{bs}] q_s - \Sigma_s \kappa_{bs} |q_s|, \end{aligned} \tag{4}$$

where $R_{c'} = [> p_{c'}/p_c] - 1$ is the close to close percentage return on the security. Parenthetically, one would subtract a cost of carry $r_{c'}$ from $R_{c'}$ in accounting for proprietary trades.

Hence, daily trading profits on this security have three components: close-to-close returns on yesterday's closing position, market timing profits relative to the closing price benchmark (which is usually thought of as price impact), and execution costs relative to the trade-by-trade benchmark. The trade-specific benchmark price p_{bs} figures in both trade-related components of daily profits, determining the allocation of measured performance between execution costs and intra-day market timing or price impact. This dual role for p_{bs} makes economic sense for it captures both dimensions of execution quality, whether trades were consummated at times when prices were high or low relative to future prices and when trading costs were high or low relative to future trading costs.

It is useful to think of each term as reflecting a distinct decision, even though one person could make all three choices. The first piece is that day's return on the past decisions of a portfolio manager or trading strategist. The second component reflects the decisions of a trader who decides how and when to execute the remaining trades necessary to fill out the position desired by the manager. The final element represents the costs of the trade in the venue selected by that trader. Of course, the latter two choices are often made simultaneously.

The appropriate benchmark depends on the perspective of the evaluator and on the uses to which the evaluation will be put. An internal performance auditor examining, say, proprietary trading within a broker/dealer would likely possess information on both the trades that were consummated and those that were not as well as the decisions that generated overall desired position changes. In contrast, external evaluators like investors going through a broker are not so fortunate, possessing only their own trading intentions and actual trades. Nevertheless, the knowledge of order placement strategies in the first case and investment decisions in the second makes it feasible to use benchmark prices such as the midquote prevailing at the time the order placement or investment decision was made.

The overall implementation shortfall—that is, the sum of the market timing profit (or price impact) and execution cost measures—is identified via this benchmark. In order to further decompose trading costs, the evaluator would need additional information regarding market conditions surrounding the transactions. Said auditor might possess the consolidated tape and, perhaps, quotes and thus is in essentially the same situation as the microstructure researcher in that trade sign must be inferred and reference quotes must be selected. If the goal is to assess broker/dealer performance, the closing price—that is, $p_{\mathrm{bs}} = p_{\mathrm{c}'}$—might be a reasonable choice.

However, what if the goal of the analysis is the evaluation of market venue quality as it is in academic and in many commercial settings? One can still use the closing price benchmark $p_{\mathrm{c}'}$ to estimate the realized spread, thus setting $p_{\mathrm{bs}} = p_{\mathrm{c}'}$ and lumping together price impact and spread-related execution costs. One can also use benchmark prices more proximate to the time of trade; an earlier draft of Werner (2002) showed that replacing $p_{\mathrm{c}'}$ with midquotes five and thirty minutes after the trade seldom changes measured realized spreads much. This benchmark choice leaves trade sign as the only unmeasured variable in k_{bs} in TAQ data and its measurement requires a reference midquote for signing the trades using either the LR or EMO algorithms.

There is good news and bad news in these pages regarding classification error rates. As has been found in prior studies, the LR error rates for system market and limit orders reported in Peterson and Sirri (2002) are modest for most trade and firm size categories and those for the EMO algorithm are typically smaller still, suggesting that the latter should be used in future research. The bad news concerns non-SuperDot market and limit orders and other order types such as percentage and specialist orders as well as block trades, both those consummated on the floor and in the upstairs market. Werner (2002) finds that the LR algorithm misclassifies a remarkably large fraction of these trades in her sample, naturally those trades with the smallest (and often negative) realized and effective spreads.

Realized spreads are comparatively insensitive to trade classification errors. Bessembinder (2002) considers reference quotes from zero to thirty seconds prior to the transaction and finds average realized spreads to be minimally affected by the chosen time lag. Of course, insensitivity does not mean accuracy but Werner (2002) finds that the LR algorithm's classification problems results in only a modest overstatement of realized spreads for most order types, modest compared with effective spreads and, consequently, with price impacts.

Effective spreads are quite a different story because effective spread biases can be and are large even when classification error rates are small. Bessembinder (2002) shows that average effective spreads increase monotonically with the time between the transaction and the reference midquote by economically and, at conventional levels, statistically significant amounts and, accordingly, he recommends the use of contemporaneous quotes. For SuperDot market and limit orders, Peterson and Sirri (2002) show that this bias is lowest when using contemporaneous quotes and the EMO trade signing algorithm. Once again, Werner (2002) tells a much more cautionary tale; the misclassification of low and negative cost trades results in substantial overstatement of effective spreads, changing even their signs for limit, percentage, and specialist orders. Given the short lag between order submission and execution on SuperDot, it is not surprising that Peterson and Sirri (2002) and Bacidore et al. (2002) find small differences in effective spreads based on either order submission or execution times.

The papers provide additional insight into many other issues in execution cost measurement. For example, Bessembinder (2002) compares transactions-based spread measures with the Roll (1984) estimator while Peterson and Sirri (2002) analyze midquote errors. Peterson and Sirri (2002) and Bacidore et al. (2002) study the distribution of the time between order arrival and execution. Bacidore et al. (2002) describe in detail how seemingly inconsequential differences in how effective spreads are averaged can lead to substantial differences in mean effective spread estimates. Both Bacidore et al. (2002) and Werner (2002) examine other aspects of execution quality including depth improvement and executions inside and outside the spread. Werner (2002) analyzes numerous liquidity measures including volume-weighted average price (VWAP), all sorted by trade type.

There is a natural tendency to think of benchmark-based execution cost measurement as a natural experiment within which orders routed to different venues are like different (perhaps randomized) treatments within some experimental design. As noted earlier, there is a shift in perspective in the passage from the performance accounting framework with which I began and the examination of the realized and effective spread and price impact measures studied in these papers. The portfolio manager or trading strategist knows the link between current and *expected* future market conditions on the one hand and their order placement strategies on the other, making their order decisions correlated with measures of both. Nowhere does this show up more clearly than in the analysis of single price executions in Bacidore et al. (2002) 97 per cent of their (exclusively SuperDot market) orders execute at a single price, sometimes receiving depth improvement. That is, the system market order placement strategies followed by traders are such that they almost never exceed quoted or *expected* depth and so it is depth that causes order size rather than the other way around. Similarly, the differences in execution costs across order types documented by Werner (2002) clearly show that differentially informed investors choose to use different types of orders in the execution of their trading strategies.

Endogeneity matters. Differences in measured execution performance across trading venues might reflect either the quality of the venue or the difficulty of the orders that are routed to it. Such ambiguities need not arise when looking across securities or order types and sizes within a venue or when trader incentives such as payment for order flow make the appropriate interpretation obvious but there are many situations that are much murkier. Accordingly, consideration of both polar interpretations (as is often done informally in the literature) provides needed discipline regarding the potential effects of endogeneity.

Perhaps more importantly, the endogeneity of the demand for and supply of immediacy places limits on what we can measure. *Observed* spreads, depths, and executions can only show displayed demand and supply conditions. One can think of any order (including a quote) as having state contingent visibility in terms of its price, size, time, identity of submitter, and location at different points in the execution process. The mandated increase in the transparency of *submitted* orders of all sorts in American stock markets has made the decision of traders as to what to display and what to keep in their pockets an even more important one. Not surprisingly, the state contingent visibility of orders is changing throughout the whole submission and execution process at this time. Understanding the economics of order visibility remains an important theoretical and empirical research topic.

References

Ellis, K., Michaely, R., O'Hara, M., 2000. The accuracy of trade classification rules: evidence from Nasdaq. Journal of Financial and Quantitative Analysis 35, 529–552.

Lee, C.M.C., Ready, M.J., 1991. Inferring trade direction from intraday data. Journal of Finance 46, 733–746.

Roll, R., 1984. A simple implicit measure of the effective bid-ask spread in an efficient market. Journal of Finance 39, 1127–1139.

Bruce N. Lehmann
Graduate School of International Relations and Pacific Studies,
University of California at San Diego,
9500 Gilmandrive, La Jolla,
CA 92093-0519, USA
E-mail address: blehmann@ucsd.edu