Market Microstructure: A Practitioner's Guide

Ananth Madhavan

Knowledge of market microstructure—how investors' latent or hidden demands are ultimately translated into prices and volumes—has grown explosively in recent years. This literature is of special interest to practitioners because of the rapid transformation of the market environment by technology, regulation, and globalization. Yet, for the most part, the major theoretical insights and empirical results from academic research have not been readily accessible to practitioners. I discuss the practical implications of the literature, with a focus on price formation, market structure, transparency, and applications to other areas of finance.

Studies of market microstructure analyze the process by which investors' latent demands are translated into executed trades. Interest in market microstructure is not new. De la Vega's (1688) description of trading practices, market manipulations, futures, and options trading on the Amsterdam stock exchange remains a classic in this field. Interest in markets and trading has increased enormously in recent years, however, because of the rapid structural, technological, and regulatory changes affecting the securities industry worldwide. Beyond immediate concerns about trading and market structure, microstructure has considerable relevance in general for finance practitioners. Specifically, a central concept in microstructure is that asset prices need not equal full-information expectations of value because of a variety of frictions. Thus, the study of market microstructure is closely related to the field of investments, which studies the fundamental values of financial assets, and is of interest to portfolio managers and investment advisors. Moreover, discrepancies between price and value affect the level and choice of corporate financing, providing a link to the field of corporate finance.

Knowledge of microstructure has grown explosively in recent years as complex new models have been developed and rich intraday data from a variety of sources have become available. Despite their practical value, however, many important theoretical insights and empirical results from academic research are not readily accessible to practitioners. A sample of such topics would include models to predict transaction costs for traders or portfolio managers, limit-order models for intraday trading strategies or automated market making, liquidity as a factor in asset returns and in portfolio risk, explanations of return anomalies associated with periodic index reconstitution, the effect displaying the limit-order book may have on liquidity and volatility, the choice between automated or floor trading systems, determinants of international capital market segmentation, and the link between pricing of initial public offerings (IPOs) and the activity of secondary-market dealers.

This article provides a practitioner-oriented review of the literature to complement academic surveys by Biais, Glosten, and Spatt (2002), Lyons (2001), Harris (2001), Madhavan (2000), Keim and Madhavan (1998), and O'Hara (1995). Any survey must be selective, especially for the microstructure literature, which comprises literally thousands of research articles spanning decades. My 2000 paper provides a more complete set of citations than I do here. I should also emphasize that this article is not a survey of topics under current debate (e.g., SuperMontage). Such topics change frequently and receive comprehensive coverage in press and industry publications. Rather, I provide a selective review of the academic literature, with an emphasis on the modern line of thought that focuses on *information*. The objective is to provide the reader with a conceptual framework that will prove valuable in attacking a variety of practical problems, both current and future.

The article is organized around four topics that roughly correspond to the historical evolution of microstructure:

- first, price formation and price discovery—including both static issues (such as the determinants of trading costs) and dynamic issues (such as the process by which prices come to impound information over time). The goal is to

Ananth Madhavan is managing director at ITG Inc., New York.

look inside the "black box" by which latent demands are translated into realized prices and volumes;

- second, market structure and design issues, including the relationship between price formation and trading protocols. The focus is on how various rules affect the black box and, hence, liquidity and market quality;
- third, information, especially market transparency (i.e., the ability of market participants to observe information about the trading process). This topic deals with how revelations of the workings of the black box affect the behavior of traders and their strategies;
- fourth, the interface of market microstructure with corporate finance, asset pricing, and international finance. Models of the black box provide fresh perspectives on such topics as IPO underpricing, portfolio risk, and foreign exchange movements.

Price Formation and Discovery

Price formation, the process by which prices come to impound new information, is a fundamental topic in microstructure.

The Crucial Role of Market Makers. By virtue of their role as price setters, market makers are a logical starting point for an exploration of the black box within which a security market actually works. In the traditional view, market makers passively provide "immediacy," the price of which is the bid–ask spread. (Note that "spread" here refers not solely to *quoted* bid–ask spreads—which have been typically small since decimalization in the U.S. market—but to *effective* spreads—that is, the true cost of a round-trip transaction for an average-sized trade.) Early empirical research confirmed that effective bid–ask spreads are lower in higher-volume securities because dealers can achieve faster turnaround in inventory, which reduces their risk. Spreads are wider for riskier and less liquid securities. Later research provided a deeper understanding of trading costs by explaining variation in bid–ask spreads as part of intraday price dynamics. This research showed that market makers are not simply passive providers of immediacy but must also take an active role in price setting to rapidly turn over inventory without accumulating significant positions on one side of the market. **Exhibit 1** illustrates this literature with a description of "Garman's Logic."

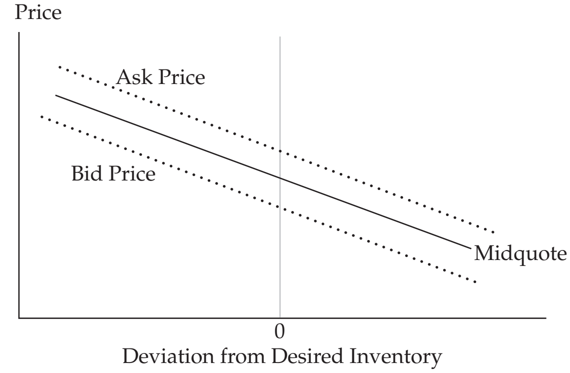
Price may depart from expectations of value if the dealer is long or short relative to desired (target) inventory, giving rise to transitory price movements during the day—and possibly over longer periods. This intuition drives the models of inventory control developed by, among others, Madhavan and Smidt (1993). **Figure 1** illustrates a typical inventory model. As the dealer trades, the actual and desired inventory positions diverge, which forces the dealer to adjust prices, lowering them if the position is long and raising them if it is short relative to desired inventory. Because setting prices away from fundamental value will result in expected losses, inventory control implies the existence of a bid–ask spread even if actual transaction costs (i.e., the physical costs of trading) are negligible. The spread is the narrowest when the dealer is at the desired inventory; it widens as inventory deviations grow larger.

The model has some important practical implications. First, dealers who are already long may be reluctant to take on additional inventory without dramatic temporary price reductions. Thus, price effects become progressively larger following a sequence of trades on one side of the market. This

Exhibit 1. Garman's Logic

Garman (1976) showed that dealer inventory must affect stock prices. The intuition can be easily explained with a simple example. Consider a pure dealer market where a market maker, with finite capital, takes the opposite side of all transactions. Suppose for the sake of argument that the market maker sets price to equate demand and supply, so buys and sells are equally likely. Consequently, inventory is equally likely to go up or down (i.e., it follows a random walk with zero drift). Whereas inventory has zero drift, however, the variance of inventory is proportional to the number of trades. Intuitively, if you flip a coin and win a dollar on heads and lose a dollar on tails, your net expected gain is zero, although your exposure is steadily increasing with the number of coin flips. But if dealer capital is finite, eventual market failure is certain because the dealer's long or short position will eventually exceed capital. Thus, to avoid such "ruin," market makers must actively adjust price levels in relation to inventory.

Figure 1. Price and Deviation from Desired Inventory



consideration is important for institutional traders, who typically break up their block trades over several trading sessions.

The second implication is that inventory considerations are also likely to affect market-impact costs (i.e., price movements associated with trading), which will be greater toward the end of the day because market makers must be compensated for bearing overnight risk. Indeed, Cushing and Madhavan (2001) documented significant increases in market-impact costs at the close of the day. Furthermore, significant return reversals occur overnight and the subsequent day following indications of order imbalances in the closing period. Intuitively, the concessions demanded by dealers are temporary, so large price reversals from the close to the open should occur once market makers have had a chance to lay off excess inventory in other markets or hedge their risk. These transitory inventory effects represent a significant hidden trading cost for traders who use market-on-close orders.

Third, because inventory effects are related to the degree to which dealers are capital constrained, larger inventory effects might be observed for the smaller dealers with less capital.

Finally, inventory models provide an added rationale for reliance on dealers. Specifically, just as physical marketplaces consolidate buyers and sellers in space, the market maker can be viewed as an institution to bring buyers and sellers together in time through the use of inventory. A buyer need not wait for a seller to arrive but may simply buy from the dealer, who depletes his or her inventory.

Inventory is not the only consideration for a dealer. An influential paper by Jack Treynor (published under the pseudonym of Walter Bagehot in 1971) suggested the distinction between liquidity-motivated traders (who possess no special informational advantages) and informed traders (who possess private information about future value). The market maker loses to informed traders, on average, but recoups the losses on trades with liquidity-motivated ("noise") traders. Models of this type (asymmetric-information models) include those of Glosten and Milgrom (1985) and Easley and O'Hara (1987). **Exhibit 2** illustrates the Glosten–Milgrom arguments about asymmetrical information and bid–ask spreads.

Asymmetric-information models have important implications: (1) In addition to inventory and order-processing components, the bid–ask spread contains an informational component because market makers must set a spread to compensate themselves for losses to informed traders. (2) Without noise traders, dealers would not be willing to provide liquidity and markets would fail. (3) Given the practical impossibility of identifying informed traders (who are not necessarily insiders), prices adjust in the direction of money flow.

Exhibit 2. Post-Trade Rationality and Bid–Ask Spreads

That bid–ask spreads contain a component attributable to asymmetrical information can easily be proved. Consider an extreme example with no inventory or transaction costs but in which some traders have information about future asset values. Based on public information, the dealer believes that the stock is worth \$30. This dealer, however, is "post-trade rational"; that is, if a trader buys 100 shares, the dealer knows that the probability that the asset is undervalued is greater than the probability it is overvalued. Why? Because informed traders only participate on one side of the market. Suppose, for the sake of exposition, that (given that the dealer observes a buy of 100 shares) the expected asset value is \$30.15, and assume, symmetrically, that the expected value of a sell of 100 shares is \$29.85. A post-trade-rational dealer will set the bid and ask prices at \$29.85 and \$30.15, good for 100 shares. These prices are free of "regret," in the sense that after the trade, the dealer does not suffer a loss. The result is a nonzero bid–ask spread driven purely by information effects.

Empirical evidence on the extent to which information traders affect the price process is complicated by the difficulties of identifying the effects explicitly because of asymmetrical information.

Both inventory and information models predict that order flow will affect prices—but for different reasons. In the traditional inventory model, order flow affects dealers' positions and dealers adjust prices accordingly. In the information model, order flow acts as a signal about future value and causes a revision in beliefs. Stoll (1989) proposed a method to distinguish the two effects by using transaction data. But without inventory data, the results of such indirect approaches are difficult to verify. Madhavan and Smidt developed a dynamic-programming model that incorporates inventory-control and asymmetric-information effects. The market maker acts as a dealer and as an active investor. As a dealer, the market maker quotes prices that induce mean reversion toward inventory targets; as an active investor, the market maker periodically adjusts the target levels toward which inventories revert. The authors estimated the model with daily specialist inventory data and found evidence of both inventory and information effects.

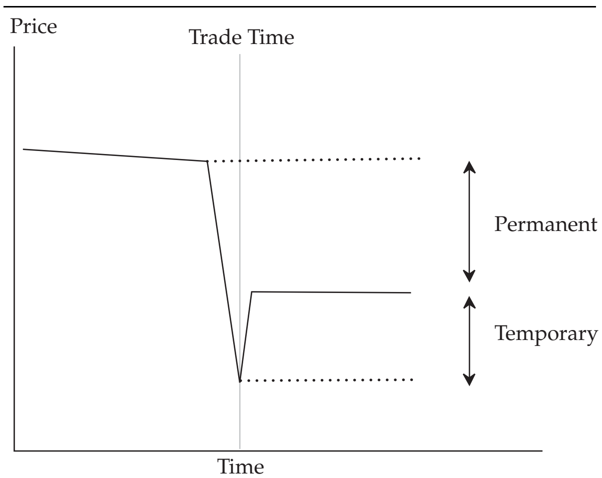
Inventory and information effects also explain why "excess" volatility might be observed, in the sense that market prices appear to move more often than is warranted by "fundamental" news about interest rates, dividends, and so on. An interesting example is provided in **Exhibit 3**.

Exhibit 3. Does Trading Create Volatility?

French and Roll (1986) found that, on an hourly basis, the variance during trading periods is 20 times larger than the variance during nontrading periods. One explanation is that public information arrives more frequently during business hours, when the exchanges are open. An alternative explanation is that order flow is required to move prices to equilibrium levels. To distinguish between these explanations is difficult, but a historical quirk in the form of weekday "exchange holidays" that the NYSE declared at one point (to catch up on a backlog of paperwork) provides a solution. Because other markets and businesses were open, the public information hypothesis predicted that the variance over this two-day period, beginning with the close the day before the exchange holiday, would be roughly double the variance of returns on a normal trading day. In fact, however, the variance for the period of the weekday exchange holiday and the next trading day was only 14 percent higher than the normal one-day return. This evidence suggests that trading itself is an important source of volatility; for markets to be efficient, someone has to make them efficient.

Practical Implications of Information Theories. An important application of the information models concerns the price movements associated with large trades. Given the preceding discussion, the price impact of a block trade can be decomposed into permanent and temporary components, as in Keim and Madhavan (1996). The permanent component is the information effect (i.e., the amount by which traders revise their value estimates based on the trade); the temporary component reflects the transitory discount needed to accommodate the block. **Figure 2** illustrates the effects for a block sale. Let p_{t-h} denote the pretrade benchmark, p_t the trade price, and p_{t+k} the post-trade benchmark price, where h and k are suitably chosen periods, typically half-hour periods. The solid line shows the price path over these time periods; the dotted lines show the continuation of the benchmark price (top of the figure) and the temporary price (bottom of the figure). The price impact of the trade, relative to the pretrade benchmark, is manifested by the drop in price at the trade time. For a sell order, we can represent the impact as $p_t - p_{t-h}$, a negative number usually. The price impact can, in turn, be decomposed into two components: (1) the permanent component, defined as $\pi = p_{t+k} - p_{t-h}$ (the difference between the post-trade price and the price before the trade), and (2) the temporary component, defined as $\tau = p_t - p_{t+k}$, the difference between the trade price and the post-trade price, which is represented in Figure 2 by the "rebound" in the price following the trade.

Figure 2. Price-Impact Components of a Block Sale



Distinguishing the two components in post-trade analysis is valuable for traders and portfolio managers. For example, a passive index fund can avoid significant permanent effects by trading with counterparties who know that it does not trade on information. Consequently, many equity markets have created two economically distinct trading mechanisms for large-block transactions. First, a block can be sent directly to the "downstairs" markets, such as the NYSE or Nasdaq. Second, a block trade can be directed to the "upstairs" market, where a block broker facilitates the trading process by locating counterparties to the trade and then formally crossing the trade in accord with the regulations of the primary market.

The upstairs market operates as a search brokerage where prices are determined through negotiation. Reputation plays a critical role in upstairs markets, where it allows traders who are known not to trade on private information to obtain better prices than in an anonymous market. Liquidity providers, especially institutional traders, are reluctant to submit large limit orders and thus offer free options to traders using market orders. This problem is especially significant in systems with open limit-order books and small minimum price increments. Upstairs markets allow these traders to participate selectively in trades screened by block brokers, who avoid trades that may originate from traders with private information. Indeed, anecdotal evidence from the Toronto Stock Exchange (TSE) suggests a migration of orders upstairs following greater transparency in the primary market. **Exhibit 4** illustrates an astute strategy based on use of the upstairs market.

Another application of the ideas in this section concerns the determinants of the anomalous returns associated with stocks added to and deleted from widely used stock market indexes. Madhavan (forthcoming) documented economically and statistically significant abnormal returns associated with

Exhibit 4. A Smart Passive Fund and the Upstairs Market

Keim (1999) analyzed the performance of the 9–10 Fund of Dimensional Fund Advisors (DFA). The 9–10 Fund is a passive index fund that attempts to mimic the performance of the bottom two deciles of the NYSE by market capitalization. Keim reported that the 9–10 Fund's mean return from its inception in 1982 exceeded that of its benchmark by an average of almost 250 bps without higher risk. This performance would be envied by many actively managed funds. Keim showed that the outperformance was largely a result of DFA's intelligent use of the upstairs market. Instead of immediately selling or buying shares when a stock moved into or out of the universe, DFA traded in the upstairs market by providing liquidity when approached by block traders who knew DFA's strategy. Thus, DFA earned the liquidity premium in the upstairs market, although it incurred some tracking error because it waited to buy or sell. The rewards to earning the spread through upstairs trading—as opposed to incurring the price-impact costs in illiquid stocks—exceeded 204 bps annually.

the annual reconstitution of the Frank Russell Company indexes from 1996 through 2002. Decomposing returns into permanent and temporary effects provides insights into such return anomalies. Specifically, permanent changes in prices are attributable to shifts in liquidity associated with changes in index membership, whereas temporary effects are related to transitory price pressure. The magnitude of return reversals (temporary effects) following index rebalancing suggests that liquidity pressures help explain the return anomalies associated with the annual reconstitution of the stock indexes of the Frank Russell Company.

Pretrade-Cost Estimation. Intraday models are essential for formulating accurate predictions of trading costs. Traders use pretrade-cost models to evaluate alternative trading strategies and form benchmarks for evaluating the post-trade performance of traders and brokers. Such models can be used as modules in autotrading strategies in which computers automatically generate trades under certain conditions.

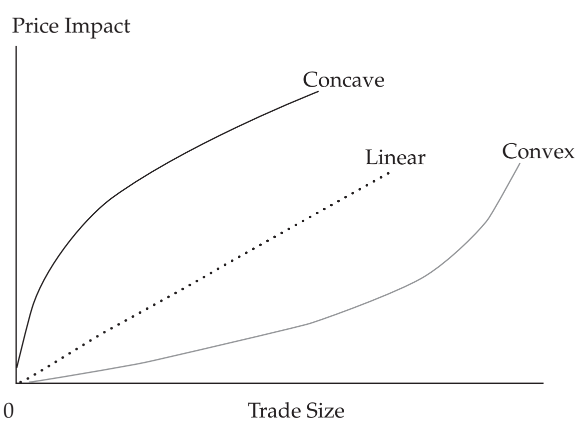
Interest in pretrade-cost models is also motivated by the fact that transaction costs can significantly erode investment performance. Specifically, the net alpha to an investment strategy is the expected alpha less the product of turnover and two-way trading costs. Because alpha is linear in trade size but costs are typically nonlinear, many investment managers use portfolio optimizers that incorporate pretrade-cost models to avoid constructing portfolios that consist of large positions in illiquid small-cap securities.

A successful model has three essential ingredients: (1) Because most investors break their orders into component trades, the model builder must recognize that current trades affect the prices of future transactions by distinguishing between permanent and temporary price impacts. (2) Costs depend on stock-specific attributes (liquidity, volatility, price level, market) and order complexity (order size relative to average daily volume, trading horizon). And (3) because costs are a function of style, no single cost estimate applies to any given order; rather, the model yields cost estimates that vary with the aggressiveness with which the order is presented to the market. In particular, an order traded quickly using market orders will typically incur higher costs than if it were traded passively over a long horizon using limit orders, upstairs markets, or crossing systems. Passive trading, however, involves more risk from adverse price movements on the unexecuted portion of the order.

The first ingredient implies that a realistic model will have to be solved recursively, because the execution price of the last subblock of an order depends on how the last-but-one subblock was traded, and so forth. In technical terms, the optimal trade-break-up strategy and corresponding minimum expected cost is the solution to a stochastic dynamic-programming problem. The problem is *stochastic* because the future prices are uncertain; the analyst has conjectures about their means but recognizes that other factors will affect the actual execution prices. *Dynamic programming*, a mathematical technique, is needed because it provides solutions to multiperiod problems in which actions today affect rewards in the future. Examples of models of this type include those in Almgren and Chriss (1999) and Bertsemas and Lo (1999). In such models, the price-impact function (i.e., the effect of trade on price) is assumed to be linear. This assumption is made partly for analytical tractability but also because theoretical models (such as Kyle's 1985 model) derive linear equilibriums from fundamental principles. A consideration of equal or greater importance is that when the permanent price function is linear, investors cannot manipulate the market by, say, buying small quantities and then liquidating in one go at a future date. This argument does not apply to the temporary price impact, which is likely to be nonlinear (see Keim and Madhavan 1996). In practical terms, the model builder who allows nonlinear price functions often runs into situations in which the recommended trade strategy involves some trades that are in the opposite direction to the desired side. This recommendation is often counterintuitive to traders and risks the possibility of regulatory scrutiny. Thus, a common practice is the imposition of a requirement that all parcels of the order be on the same side as the order itself.

However, although most traders agree that linearity is too simplistic an assumption, they disagree over what form these functions really take. Are they concave (i.e., rising at a decreasing rate in size), convex (rising at an increasing rate), or linear? **Figure 3** shows three possible shapes. Loeb's 1983 study of block quotations suggests a concave shape to the functions. Similarly, Hasbrouck (1991) found that square-root transformations of volume fit well in modeling price impacts. But others argue for convex functions because available liquidity is ultimately limited. Madhavan and Cheng (1997) suggested that the resolution of this problem might be found in the fact that observed prices and volumes reflect the operation of two economic mechanisms. On the one hand, Keim and Madhavan argued that the price functions are concave for upstairs trades, where buyers and sellers are matched. They showed that as block size increases, more counterparties are contacted by the upstairs broker, which cushions the price impact, so the costs of trading are relatively low for large trades. These results are consistent with Loeb's findings from interviews with block traders; he reported a concave price-impact function. Similar logic applies to trades in external crossing systems. Because of commission costs, upstairs trades are relatively uneconomical for small trades.

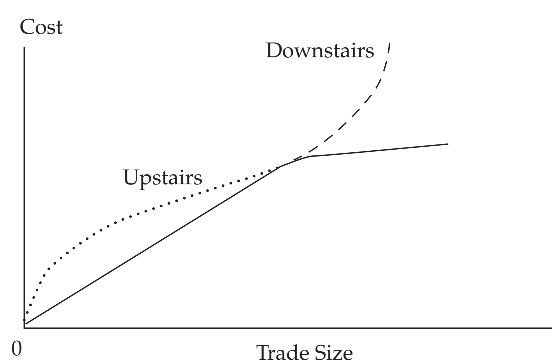
Figure 3. Price-Impact Functions



On the other hand, the price-impact function for market orders sent anonymously to downstairs markets might well be convex. On the NYSE, small orders are executed through the SuperDot system and if they are below the stated depth, have zero price impact; they may even benefit from price improvement. Medium-sized orders may execute against the limit-order book or the specialist, producing some price impact. A very large trade will eat up all the liquidity on the book, and the specialist may demand a large price concession to accommodate the remainder of the trade from inventory, as is sometimes observed at the close if a large order imbalance exists. The result is a convex price function.

In general, traders who have the choice will select the lowest-cost mechanism (Madhavan and Cheng), so the proportion of upstairs trades will rise with size. Unfortunately, publicly available databases (the Trade and Quote database, for example) do not distinguish between upstairs and downstairs trades, so the true cost of large trades directed to the downstairs market is underestimated. The result is, as **Figure 4** illustrates, that a concave price function appears to best fit the data. The dashed line shows the price impact in the downstairs (convex) market, and the dotted line shows the price impact in the upstairs (concave) market; the observed relationship is the lower envelope of the two curves, the solid line. In addition, a concave relationship might emerge if an order that moved prices substantially induced other traders to enter the market as counterparties to supply liquidity.

Figure 4. Upstairs and Downstairs Trading Costs



Market Structure and Design

"Market architecture" designates the set of rules governing the trading process. Many academic studies have shown that market structure matters by affecting the speed and quality of price discovery, liquidity, and the cost of trading. Market architecture is determined by choices regarding a variety of attributes, including the following:

- *Degree of continuity.* Periodic systems allow trading only at specific points in time, whereas continuous systems allow trading at any time while the market is open.

- *Dealer presence.* Auction or order-driven markets feature trade between public investors without dealer intermediation; in a dealer (or quote-driven) market, a market maker takes the opposite side of every transaction.
- *Price discovery.* Some markets provide independent price discovery; others use prices determined in another market as the basis for transactions.
- *Automation.* Markets vary considerably in the extent of automation; floor trading and screen-based electronic systems are at opposite extremes. The technology of order submission is less important, however, than the protocols governing trading.
- *Order forms.* The kinds of order forms permitted include market, limit, stop, upstairs crosses, and hidden.
- *Protocols.* Protocols are the rules regarding program trading, choice of minimum tick, trade-by-trade price continuity, when to halt trading, and circuit breakers and any special rules adopted for opens, reopens, and closes.
- *Pre- and post-trade transparency.* The quantity and quality of information provided to market participants during the trading process constitutes pretrade and post-trade transparency. Nontransparent markets provide little in the way of indicated prices or quotes. Transparent markets often provide a great deal of relevant information before trades (quotes, depths, etc.) and after trades (actual prices, volumes, etc.).
- *Information dissemination.* Markets differ in the extent of information dissemination (to brokers, customers, or the public) and the speed of information dissemination (real time or delayed feed).
- *Anonymity.* A crucial factor, the anonymity afforded by a market takes dimensions ranging from hidden orders to zero disclosure regarding trader identities.
- *Off-market trading.* Some markets permit off-exchange trading and/or trading after hours, and some do not.

Trading systems, a sampling of which is shown in **Table 1**, exhibit considerable heterogeneity in their architecture. For example, automated limit-order-book systems of the type used by the TSE and Paris Bourse offer continuous trading with high degrees of transparency (i.e., public display of current and away limit orders) without reliance on dealers. The foreign exchange market and corporate junk bond market rely heavily on dealers to provide continuity but offer little transparency; other dealer markets (Nasdaq, the London Stock Exchange) offer moderate degrees of transparency. Some exchanges have fairly strict trade-to-trade price continuity requirements; others, such as the Chicago Board of Trade, allow prices to move freely. Most organized markets also have formal procedures to halt trading in the event of large price movements. Crossing systems, such as POSIT, do not currently offer independent price discovery but cross orders at the midpoint of the quotes in the primary market.

How do such differences affect price formation and the costs of trading?

Practical Issues in Market Design. The diversity of systems has spurred considerable theoretical research. Early investigators recognized the presence of strong network externalities. For example, high volume implies a short holding period for market makers and, hence, low inventory-control costs. If volume is initially split equally between two markets and the initial volume allocation is perturbed slightly, the higher-volume market will enjoy reduced costs and volume will migrate to this market. In the long run, volume migration will lead to consolidation of the two markets into a single market. The inclusion of information in this model

Table 1. Variation in Real-World Trading Systems

Architecture Element	Typical ECN	NYSE Open Market	NYSE Intraday Trading	Paris Bourse	POSIT	Chicago Board of Trade	Foreign Exchange Market
Continuous trading	×		×	×		×	×
Dealer presence		×	×			×	
Price discovery	×	×	×	×		×	×
Automation	×			×	×		
Anonymity	×	×		×	×		
Pretrade quotes	×		×	×		×	
Post-trade reports	×	×	×	×	×	×	

only serves to confirm this prediction. With asymmetrical information, rational, informed traders will split their orders between the two markets, providing incentives for liquidity traders to consolidate their trading. Intuitively, if two markets are combined into one, the fraction of informed trading volume will drop, resulting in a narrowing of spreads. Even if the information signals are symmetrical, if they are diverse, then pooling orders will provide prices that are informationally more efficient than decentralized trading among fragmented markets. Indeed, even when multiple markets coexist, the primary market is often the source of all price discovery (as shown by Hasbrouck 1995), with the satellite markets merely matching quotes.

Despite strong arguments for consolidation occurring, however, many markets are fragmented and remain so for long periods of time. This puzzle has two aspects: (1) the failure of a single market to consolidate trading in time and (2) the failure of diverse markets to consolidate in space (or cyberspace) by sharing information on prices, quotes, and order flows. As for the first issue, theory suggests that multilateral trading systems (such as single-price call auctions) are efficient mechanisms to aggregate diverse information. Consequently, researchers are interested in how call auctions operate and whether such systems can be used more widely to trade securities. The *information aggregation* argument suggests that call auctions are especially valuable when uncertainty over fundamentals is large and market failure is a possibility. Indeed, many continuous markets use single-price auction mechanisms when uncertainty is large, such as at the open, close, or reopening after a trading halt. Yet, most trading systems are continuous and bilateral (not periodic and multilateral), which suggests that the benefits of being able to trade immediately at known prices are extremely important.

With regard to the second issue, although consolidated markets pool information, whether they will be more efficient than fragmented markets is not clear; for example, efficiency will diminish if some traders develop reputations based on their trading histories. Indeed, one argument cited for the growth of electronic crossing markets is that traditional markets offer too much information about a trader's identity and motivations for trade. Models emphasizing asymmetrical information provide some rationale for the success of off-market competitors in attracting order flow from primary markets. Technology may eventually resolve the fragmentation debate. Today, a variety of systems are being built to route orders intelligently to the most liquid market centers.

Automated Auctions. Within the class of continuous markets are the designated dealers or limit-order markets without intermediaries. Pure auction markets can be structured as batch (single-price) auctions or, more commonly, as automated limit-order-book markets. Examples of automated auctions are electronic communications networks (such ECNs as Island) and the Paris Bourse. With a limit order, an investor associates a price with every order so that the order will execute only if the investor receives that price or better. Clearly, all orders can be viewed as limit orders; a market buy order is simply a limit buy order for which the limit price is the current ask price or higher. Consequently, recent models of limit-order execution have generated considerable interest. **Exhibit 5** describes two approaches to estimating limit-order models. Such models can help a trader evaluate trading strategies, can be used within an autotrading strategy to automatically submit or cancel limit orders, or can form the basis for automated market making.

Exhibit 5. Estimating Limit-Order Models

Limit-order models provide the probabilities of a limit order being hit as a function of the limit price and other variables. Two types of limit-order models exist—first passage time (FPT) models and econometric models.

FPT models estimate the probability a limit order will be executed on the basis of the properties of the stock-price process; typically, these models assume stock prices follow some type of random walk. To get a feel for these models, consider an extremely simple example. Assume for simplicity that the midquote return over the next 10 minutes is normally distributed and that price changes are independent. If the current price is, say, $50, the probability that the midquote will cross a given limit price in a prespecified period of time is straightforward to compute from the cumulative normal distribution.

But although FPT models are analytically convenient, they are not especially realistic. Econometric models (see Lo, MacKinlay, and Zhang 2001) offer more realism because they can accommodate large numbers of explanatory variables. But estimating econometric limit-order models presents problems. Specifically, unless one knows the investor's strategy for canceling unfilled limit orders, estimating the probability of execution is difficult because only executions for filled orders can be observed. Statistical techniques must be used to handle such "censoring"—for example, survival analysis to model the true time to execution. In particular, if T is the random execution time, one models the probability that $T < t$ as a function of a vector of variables, including where in the current bid–ask spread the limit price is located, current depth, recent price changes, and other such variables.

A limit-order provider is offering free options that can be exercised if circumstances change. Consequently, the limit-order trader needs to expend resources to monitor the market—a function that may be costly. Perhaps for this reason, dealers of some form or another who carry out the monitoring for the trader arise so often in auction markets.

Limit-order models provide insights into the consequences of introducing decimalization and changing the minimum tick. Strictly speaking, decimalization refers to the quoting of stock prices in decimals rather than fractions, such as eighths or sixteenths. Proponents of decimalization in the U.S. markets noted that it would allow investors to compare prices more quickly than could be done when fractions were used, thereby facilitating competition, and would also promote the integration of U.S. and foreign markets. They often mistakenly computed large cost savings to investors because quoted spreads were deemed likely to fall dramatically under decimalization. The minimum tick is a separate issue, although it is often associated with decimalization in the literature, that concerns the smallest increment in which stock prices can be quoted. For example, a system could have decimal pricing but a minimum tick of 5 cents or 2 cents. From an economic perspective, what is relevant is the minimum tick, not the units of measurement of stock prices.

If the minimum tick is reduced in a market, the profits from supplying liquidity in the market (assuming a constant book) go down. The result is a reduction in liquidity at prices away from the best bid or offer. The quoted spread itself may fall, however, because of competition. Thus, a reduction in the minimum tick may reduce overall market liquidity (see Harris 1998 for a discussion of this and related points). Recent empirical evidence suggests that quoted spreads were reduced following decimalization in the United States, together with a reduction in quoted depth. In other countries, decimalization and transparency have had similar effects—benefiting small, retail traders but reducing visible liquidity and inducing institutional traders to use alternative venues, such as the upstairs market or low-cost crossing systems.

Intermarket comparisons are difficult because real-world market structures are more complex than simple models suggest. The NYSE, for example, has elements of both auction and dealer markets, whereas SuperMontage moves the Nasdaq closer to an auction model. Furthermore, the definition and measurement of market quality raise serious empirical issues. For example, use of the usual measure of trading costs (or illiquidity), namely, the quoted bid–ask spread, creates problems because quoted spreads capture only a small portion of a trader's actual execution costs.

The early literature argued that competition among market makers on the Nasdaq system would result in lower spreads than in a specialist system of the type used by the NYSE, but the opposite seems to be the case, even after such factors as company age, company size, risk, and price level have been controlled for. One explanation, provided by Christie and Schultz (1994), is that dealers on Nasdaq implicitly collude to set spreads wider than those justified by competition. Institutional practices such as order-flow preferencing (i.e., directing order flow to preferred brokers) and soft dollar payments limit the ability and willingness of dealers to compete with one another on the basis of price and may explain why spreads are large despite easy entry into market making.

Tests of market structure theories face a serious problem—the absence of high-quality data that would allow researchers to pose "what if" questions. However, some interesting natural experiments have been reported. For example, in the late 1990s, the Tel-Aviv Stock Exchange moved some stocks from periodic trading to continuous trading, which allowed researchers to investigate the effects of market structure on asset values by using the stocks that did not move as a control. Amihud, Mendelson, and Lauterbach (1997) documented large increases in asset values for the stocks that moved to continuous trading. But such natural instances for testing are few and far between.

Compounding the problem is the fact that traders adjust their strategies in response to market protocols and information, so assessing the impact of market protocols is difficult. Furthermore, empirical studies are limited by the scarcity of large samples of events to study. An additional obstacle to empirical research is that the changes in structure that markets make are often responses to perceived problems or competitor actions. Such changes are often accompanied by design alterations in other dimensions, such as a switch to automation or greater transparency.

A promising alternative to market-initiated changes is laboratory or experimental studies that allow tests of subtle theoretical predictions about market design. For example, in a laboratory study conducted by Bloomfield and O'Hara (2000), human subjects trading in artificial markets allowed the researchers to examine the effects of various changes in protocols on measures of market quality. A key advantage of such laboratory research is that researchers can accurately measure quality metrics (e.g., deviation of price from intrinsic value, speed of convergence to full-information prices) that cannot ordinarily be observed with real data. Further benefits of laboratory experiments in finance are discussed in **Exhibit 6**.

Practical issues of market design are central to the subject of market microstructure. Although

Exhibit 6. Experimental Finance

The ability to frame controlled experiments in laboratory markets allows researchers to analyze complex information effects. The obvious focus is on metrics, such as the bid–ask spread, market depth or liquidity, and volatility. But an experimental study can also address variables that might not otherwise be possible to observe, including data on traders' estimates of value over time, their beliefs regarding the dispersion of "true" prices, and the trading profits of various classes of traders (e.g., informed versus uninformed, speculators versus hedgers).

Bloomfield and O'Hara used experimental markets to analyze changes in disclosure rules. In their study, lab participants faced various disclosure regimes, and in some experiments, dealers (markets) could decide whether they preferred transparency or not. Bloomfield and O'Hara found that transparency has a large impact on market outcomes.

Several other interesting findings have emerged from lab markets. For example, generating price bubbles is quite easy, even if market participants are aware of bounds on fundamental value. Interestingly, prices in auction markets need not always converge to full-information values; agents may learn incorrectly and settle prices at the "wrong" value.

researchers have learned a great deal, they have come to no uniform view on which structures offer the greatest liquidity and lowest trading costs. Given the considerable complexity of real-world market structures, however, this diversity is hardly surprising. Ultimate decisions on market structure are likely to be decided by the markets on the basis of factors that have less to do with information than most economists believe.

The factor I would single out as most influential is a practical one, namely, the need for automation and electronic trading to handle the increasingly high volumes of trading. Although this factor will favor the increased use of electronic trading systems, it does not imply the demise of traditional floor-based systems. The point to keep in mind is that what ultimately matters is not the medium of communication between the investor and the market but the protocols that translate the order into a realized transaction.

Information

Many of the informational issues about market microstructure concern transparency and disclosure. Market transparency has been defined (see, for example, O'Hara) as the ability of market participants to observe information about the trading process. Information in this context can be knowledge about prices, quotes, volumes, the sources of order flow, or the identities of market participants.

A useful way to think about transparency, which has many aspects, is to divide it into pretrade and post-trade dimensions. Pretrade transparency refers to the wide dissemination of current bid and ask quotations, depths (and possibly also information about limit orders away from the best prices), and other pertinent trade-related information, such as the existence of large order imbalances. Post-trade transparency refers to the public and timely disclosure of information on past trades, including execution time, volume, price, and possibly information about buyer and seller identifications.

Issues of Market Transparency. Both pre- and post-trade transparency issues have been central to some recent policy debates. For example, the delayed reporting of large trades in London has been cited as a factor in intermarket competition and order-flow migration.

In addition, transparency is a major factor in debates about floor systems versus electronic systems. Floor systems, such as the Chicago futures markets, generally do not display customer limit orders unless they represent the best quote. In contrast, electronic limit-order-book systems, such as the TSE's Computer Assisted Trading System (CATS) and the Paris Bourse's Cotation Assistée en Continu (CAC) system typically disseminate not only the current quotes but also information on limit orders away from the best quotes. The trend worldwide has been toward greater transparency. (The NYSE recently created OpenBook, a real-time view of the specialist's limit-order book.)

The practical importance of market transparency has given rise to a great deal of theoretical and empirical research. Several authors have examined the effect of disclosing information about the identity of traders or their motives for trading. These issues arise in many different contexts, including

- post-trade transparency and reporting,
- predisclosure of intentions to trade (known as "sunshine trading") or the revelation of order imbalances at the open or during a trading halt,
- dual-capacity trading, in which brokers can also act as dealers,
- front running, when brokers trade ahead of customer orders,
- upstairs and off-exchange trading,
- hidden limit orders in automated trading systems,
- counterparty trade disclosure, and
- the choice of floor-based or automated trading systems.

In a totally automated trading system, where the components of order flow cannot be distinguished, transparency is not an issue. Most floor-based trading systems, however, offer some degree of transparency regarding the composition of order flow. For example, for trading on the NYSE, the identity of the broker submitting an order may provide valuable information about the source and motivation for a trade.

Theoretical modelers have reached mixed conclusions about the effects of transparency. In some models (e.g., O'Hara), transparency can reduce problems of adverse selection, and thus spreads, by allowing dealers to screen out traders likely to have private information. Other models (e.g., Madhavan 1996), however, indicate that transparency can exacerbate price volatility. Intuitively, disclosing information should allow investors to better estimate the extent of noise trading, thus increasing the market's vulnerability to asymmetric-information effects. Essentially, noise is necessary for markets to operate, and disclosure robs the market of this lubrication. Contrary to popular belief, Madhavan (1996) showed that the potentially adverse effects of transparency are likely to be greatest in thin markets.

These results have important implications for policies on, for example, the choice between floor-based systems and fully automated (typically anonymous) trading systems. Specifically, suppose traders obtain better information on the portion of order flow that is price inelastic on an exchange floor than in an automated trading system. Floor-based systems may be more transparent because traders can observe the identities of the brokers submitting orders and make inferences about the motivations of the initiators of those orders. Unless a system is explicitly designed to function in a nonanonymous fashion, such inferences are extremely difficult in a system with electronic order submission. In this case, traditional exchange floors may be preferred over automated systems for active issues whereas the opposite may be true for inactive issues. Finally, the results of this branch of research provide insights into why some liquidity-based traders avoid sunshine trades even if they could benefit from reputation signaling.

In one model (Madhavan 1996), nondisclosure benefits large institutional traders whose orders are filled in multiple trades by reducing their expected execution costs but imposes costs on short-term noise traders. With the benefit of nondisclosure, the large institutions can break up their trades over time without others front-running them and thereby raising their trading costs. A large trade can be successfully broken up without attracting too much attention and, hence, moving the price in the direction of the trade. Nondisclosure also benefits dealers by reducing price competition. The implication is that, faced with a choice between a high-disclosure market and a low-disclosure market, an uninformed institutional trader will prefer to trade in the more opaque market. This model suggests that one danger of too much transparency is that traders might migrate to other venues, including off-exchange or after-hours trading.

Empirical Research on Transparency and Disclosure. In terms of post-trade transparency, late-trade reporting on the London Stock Exchange and Nasdaq has generated extensive discussion. Research suggests that increased post-trade transparency increases volumes and lowers spreads (see the discussion in Madhavan 2000). This finding makes sense because immediate and accurate post-trade reports on volumes and prices reduce market uncertainty. A validation effect may also be at work; a trader who observes a large trade is more willing to believe that the price of that trade is representative or fair.

In terms of pretrade reporting, few "natural experiments" suggest that too much transparency is detrimental. For example, Madhavan, Porter, and Weaver (2002) found a decrease in liquidity associated with the display of the limit-order book on the TSE after controlling for volume, volatility, and price. Limit-order traders are apparently reluctant to submit orders in a highly transparent system because these orders essentially represent free options to other traders.

Transparency is a complicated subject, but recent research provides several revealing insights. First, there is broad agreement that both pre- and post-trade transparency matter; both affect liquidity and price efficiency (O'Hara; Madhavan 1996). Second, greater transparency, both pre- and post-trade, is generally associated with more informative prices (O'Hara; Bloomfield and O'Hara). Third, complete transparency is not always "beneficial" to the operation of the market. Indeed, many studies demonstrate that too much pretrade transparency can actually reduce liquidity because traders are unwilling to reveal their intentions to trade (Madhavan, Porter, and Weaver). Also, too much post-trade transparency can induce fragmentation as traders seek off-market venues for their trades. Finally, changes in transparency are likely to benefit one group of traders at the expense of others, as discussed in Madhavan (forthcoming). Traders with private information prefer anonymous trading systems, whereas liquidity traders, especially those who can credibly claim their trades are not information motivated (e.g., passive index funds), prefer greater disclosure. Consequently, no market structure will be uniformly preferred by all traders and dealers.

Applications to Other Areas

The recognition that microstructure matters for asset values, liquidity, trading costs, and price efficiency is relatively recent. This section provides examples of some of the applications of microstructure research to other areas of finance—asset pricing, corporate finance, and international finance.

Asset Pricing. Research has generally modeled expected returns as functions of such variables as proxies for size and default risk. Amihud and Mendelson (1986) showed that expected returns are a decreasing function of liquidity because investors must be compensated for the higher transaction costs that they bear in the less liquid markets. The presence of trading costs (asymmetric-information, inventory, and other transaction costs) reduces the equilibrium value of the asset. Indeed, Amihud, Mendelson, and Lauterbach documented large changes in asset values for stocks that moved to more liquid trading systems on the Tel-Aviv Stock Exchange. These and other studies confirm the importance of liquidity as a factor to be considered by portfolio managers in building an alpha-generating model.

From a cross-sectional viewpoint, variation in expected returns among securities arises because of differences in trading costs. In this context, measuring trading costs accurately is critical. As an extreme example, consider a market with risk-neutral investors and two assets—a security that is subject to trading costs and a security that is not. Trading costs in the less liquid security comprise the bid–ask spread, s, and the price impact of the trade (measured by the responsiveness of prices to the average order size), denoted by λ. Keim and Madhavan (1998) showed that the price-impact costs can be substantially larger than the observed spread, but these costs are typically difficult to measure accurately.

In equilibrium, the returns of both assets relative to the price net of trading costs should equal the risk-free rate, r_f. Computation of the return on the security (ignoring transaction costs; that is, using the midquote as the basis for value) produces return premium $r - r_f > 0$, which represents (after other factors affecting returns have been controlled for) the compensation for trading costs $\lambda + s$ across a sample of stocks. This phenomenon may also explain, in part, the observed size effect, because transaction costs are higher in less liquid assets, where the omission of λ in the computation of returns has the strongest effects. Brennan and Subrahmanyam (1996) estimated such a cross-sectional model of returns with some success. **Exhibit 7** provides an example of when portfolio managers should consider liquidity as a factor in their stock selection models.

Exhibit 7. Liquidity and Portfolio Risk

If liquidity is a factor in expected returns, its omission from a risk model can result in substantial understatement of true risk. Suppose, for the sake of simplicity, that the expected return on asset i is determined by a two-factor risk model, $R_i = r_f + \beta_{1i}F_1 + \beta_{2i}F_2$, where the first factor is the market factor and the second is a proxy for illiquidity (i.e., a factor positively related to the implicit costs of trading). Consider a trader who follows a market-neutral strategy but incorrectly ignores the liquidity factor. The trader goes long in securities that have positive alphas relative to the incorrect model (where $\beta_{2i}F_2 > 0$) and short those securities with negative alphas (where $\beta_{2i}F_2 < 0$). If liquidity decreases, the portfolio is exposed to considerable risk, even though it is "market neutral." This liquidity risk can be significant; the failure of the hedge fund Long-Term Capital Management is a good example.

Commonality in liquidity and returns is a logical area for future research. Consider a model in which the price change in each of N stocks is linearly related to order flows in one's own and related stocks, in addition to other factors. In matrix notation,

$$\Delta\mathbf{p} = \mathbf{X\Lambda} + \mathbf{U},$$

where

$\Delta\mathbf{p}$ = an $N \times 1$ vector of price changes
$\mathbf{X}$ = an $N \times k$ matrix of order flows, current and lagged, and other predetermined variables affecting price movements
$\mathbf{\Lambda}$ = a $k \times 1$ vector of coefficients
$\mathbf{U}$ = an $N \times 1$ vector of error terms

Returns to stock i may depend on current and lagged flows in stock j. Commonality in order flows is manifested in that, although $\mathbf{X}$ has full rank, only a few sources of independent variation explain most of the variation in the data.

Hasbrouck and Seppi (2001) used principal components analysis to characterize the extent to which common factors are present in returns and order flows. Principal components analysis can be viewed as a regression that tries to find a linear combination of the columns of the data matrix $\mathbf{X}$ that best describes the data, subject to normalization restrictions imposed to remove indeterminacy. Hasbrouck and Seppi found that common factors are present in both returns and order flows and that common factors in order flows account for 50 percent of the commonality in returns. Whether such factors can help predict short-run returns, variation in intraday risk premiums, or the observed relationship between price variability and volume is still an open question.

Another interesting application of market microstructure in the asset-pricing area concerns technical analysis, in which past price movements are used to predict future returns. Financial economists are traditionally skeptical of technical analysis; in an efficient market, current prices should impound all available information, so past price patterns should not have predictive power. Yet, microstructure theory suggests several avenues through which technical analysis might have value, at least over short horizons, as described in **Exhibit 8**.

Exhibit 8. The Value of Technical Analysis

Can technical analysis be of value even if markets are efficient? Microstructure theory suggests several possibilities. First, dealer inventories must be mean reverting, as shown in the section "The Crucial Role of Market Makers." So, if inventory affects prices, cyclicality could occur in prices. Second, specialist incentives to smooth prices will induce short-run autocorrelation. Third, if large traders break up their block trades (and if this information leaks slowly to the market), short-run trends will occur. To the extent that order flow contains commonalities, such factors might also aggregate to the overall market level. Finally, technical analysis might help traders discover hidden liquidity. Kavajecz and Odders-White (2002) found that technical support and resistance levels coincide with peaks in depth on the limit-order book. Furthermore, moving-average forecasts reveal information about the relative position of depth on the book. So, technical analysis is valuable, even if markets are efficient, because it reveals patterns in liquidity, which helps traders avoid large trading costs.

Corporate Finance. Close economic ties between corporations and their sources of financing characterize many financial markets. Such arrangements are common in countries where corporations rely primarily on bank financing. Similarly, equity markets for small-cap stocks are characterized by close relationships between new issuers and the underwriters who bring the stock public. In particular, underwriters sponsor new issues by arranging analyst coverage, promote the stock through marketing efforts, and provide liquidity by acting as broker/dealers in subsequent secondary-market trading. Financial economists have only recently recognized the importance of such "relationship markets." Underwriters of small stocks, for example, often dominate trading in the post-IPO market, which gives them considerable ability to affect security prices. The provision of secondary-market liquidity is valuable to the companies, and in a relationship market, it is natural for the broker/dealer that best knows the company (i.e., the underwriter) to also be the primary market maker. Indeed, Ellis, Michaely, and O'Hara (2000) found that for Nasdaq stocks, the lead underwriter is almost always the primary market maker in the aftermarket. The withdrawal of secondary-market liquidity (e.g., by the failure of a broker/dealer) is often associated with significant price drops of the stocks in which that entity made markets.

Another interesting application concerns stock splits. Early analyses of motivation for stock splits focused on traditional corporate finance explanations, such as signaling. From a microstructure viewpoint, however, splits might be a device for a corporation to adjust its stock price relative to tick size to affect its own trading costs and liquidity. For a given tick size, a higher price implies a lower cost of capital and, therefore, higher share values; at the same time, higher prices may discourage liquidity-based trading by small, retail investors. Thus, an optimal price level that maximizes share value may exist. Indeed, average stock prices are relatively constant over long periods of time within countries, despite variations among countries, and the average stock price relative to the minimum tick is even more constant.

International Finance. Microstructure and international finance intersect in several areas, including the microstructure of foreign exchange markets, cross-border competition for order flow, arbitrage in cross-listed securities, capital market segmentation, and international transmission of volatility.

Foreign exchange markets are by far the largest asset markets in terms of volume. Consequently, researchers have shown considerable interest in how they operate and how prices are determined. An unusual feature of the foreign exchange market is the extremely large trading volumes, far larger than one would expect given the level of imports and exports. Lyons provided an elegant explanation for this phenomenon. The intuition behind the model can be explained simply. Suppose an investor initiates a large-block trade with a particular dealer. The trade causes this dealer's inventory to move from the desired level. This change is costly because of the risk of an adverse price movement. In a dealer market, the dealer can offset this added inventory risk by passing a portion of the block trade on to other dealers by hitting their quotes. The block is passed around to successive dealers in a "hot potato" way, so the ultimate trading volume greatly exceeds the size of the initial trade.

What is interesting about this explanation for the volume phenomenon is its reliance on two key assumptions about market microstructure: (1) the dealer structure of the foreign exchange market and (2) a lack of transparency in trade reporting. When dealers trade bilaterally over the telephone, still the most important method of dealing, the trade is informative to them. The advent of electronic trading (e.g., Reuters D2000-2) is changing the structure and availability of information to some extent, however, and could alter the nature of the equilibrium, dramatically reducing volumes.

Another important aspect of the interaction between microstructure and international finance concerns segmentation in internal capital markets. Such barriers to investment are important because they may give rise to various documented "anomalies," such as discounts on international closed-end funds. They also may give rise to arbitrage trading or other cross-border order flows and thus affect market efficiency. Finally, an analysis of segmentation may shed light on the positive abnormal stock returns observed following economic liberalizations. **Exhibit 9** illustrates how microstructure variables can be used to predict exchange rate changes.

Exhibit 9. Exchange Rate Models

Economic theory suggests that exchange rate movements are determined by macroeconomic factors. Yet, macroeconomic exchange rate models, with R^2s below 0.10, do not fit the data well. Evans and Lyons (1999) proposed a microstructure model of exchange rate dynamics based on portfolio shifts that augments the standard macroeconomic variables with signed order flow. The model has the form $\Delta p_t = \beta_1 \Delta(i_t - i_t^*) + \beta_2 x_t + \varepsilon_t$, where Δp_t is the daily change in the (log) spot rate, $\Delta(i_t - i_t^*)$ is the change in the overnight interest rate differential between the two countries, and x_t is the signed order flow. They estimated their model for the mark/dollar and yen/dollar exchange rates. As predicted, both β_1 and β_2 were positive and significant. The estimated R^2 improved substantially when signed order flow was included. More than 50 percent of the daily changes in the mark/dollar rate and 30 percent in the yen/dollar rate were explained by the model. Applications include short-run exchange rate forecasting, targeting of central bank intervention, and prediction of trading costs for large transactions.

A puzzling aspect of international segmentation arises when domestic companies issue different equity tranches aimed at different investors. For example, countries as diverse as Mexico, China, and Thailand have foreign ownership restrictions that mandate different shares for foreign and domestic investors. The objective of such a partition of otherwise identical shares is to ensure that ownership of corporations rests in the hands of domestic nationals. Interestingly, the prices of these two equity tranches vary widely among companies and over time. The share price premiums or discounts can be explained in terms of relative trading costs: If both shares are otherwise equal but one share has higher transaction costs, that share would have to have a lower price if holding-period returns are to be equal. Elimination of market segmentation should reduce costs, thus lowering the cost of capital and boosting share prices in segmented markets.

Similar logic suggests that a stock traded in different markets might trade at different prices (adjusted for exchange rates), even though it essentially represents claims to the same underlying cash flows, because of differences in relative liquidity. In particular, a cross-listed stock that is contained in different stock market indexes in the local and foreign markets need not always trade in the same direction.

Conclusions

Several conclusions from this survey of the literature are especially relevant for practitioners.

First, markets are a great deal more complex than commonly believed. One of the major achievements of the microstructure literature is success in illuminating the black box by which prices and quantities are determined in financial markets. The recognition that order flows can have long-lasting effects on prices has many practical implications. For example, large price impacts may drive institutional traders to lower-cost venues, creating a potential for alternative trading systems. It may also explain the anomalous return behavior associated with periodic index reconstitutions like those of the Frank Russell Company indexes in June each year.

Second, microstructure matters. Under certain protocols, markets may fail and large deviations between "fundamental value" and price may occur. These issues are especially relevant for exchange officials, operators of trading systems, regulators, and traders.

Third, "one size fits all" approaches to regulation and policy making should be avoided. For example, greater transparency does not always enhance liquidity.

Finally, the interface of microstructure with other areas of finance is an exciting new area. A more complete understanding of the time-varying nature of liquidity and its relationship to expected returns is needed; evidence is growing that liquidity is a "factor" in explaining stock returns. Differences in liquidity over time may explain variations in the risk premium and may, therefore, influence stock-price levels.

I thank Ian Domowitz, Margaret Forster, Larry Harris, Don Keim, and Seymour Smidt for many helpful discussions over the years that influenced my thoughts. Of course, any errors are entirely my own. The opinions contained in this article reflect solely the personal views of the author and do not necessarily reflect the views of ITG Inc., its employees, officers, or affiliates.

References

Almgren, R., and N. Chriss. 1999. "Optimal Execution of Portfolio Transactions." Working paper, New York University.

Amihud, Y., and H. Mendelson. 1986. "Asset Pricing and the Bid–Ask Spread." *Journal of Financial Economics*, vol. 17, no. 2 (December):223–249.

Amihud, Y., H. Mendelson, and B. Lauterbach. 1997. "Market Microstructure and Securities Values: Evidence from the Tel Aviv Stock Exchange." *Journal of Financial Economics*, vol. 45, no. 3 (September):365–390.

Bagehot, W. 1971. "The Only Game in Town." *Financial Analysts Journal*, vol. 27, no. 2 (March/April):12–14, 22.

Bertsemas, D., and A. Lo. 1999. "Optimal Control of Liquidation Costs." *Journal of Financial Markets*, vol. 1, no. 1 (February):1–50.

Biais, B., L. Glosten, and C. Spatt. 2002. "The Microstructure of Stock Markets." Working paper, L'Institut d'Economie Industrielle.

Black, F. 1971. "Toward a Fully Automated Stock Exchange: Part I." *Financial Analysts Journal*, vol. 27, no. 4 (July/August):28–35, 44.

Bloomfield, R., and M. O'Hara. 2000. "Can Transparent Markets Survive?" *Journal of Financial Economics*, vol. 55, no. 3 (March):425–459.

Brennan, M., and A. Subrahmanyam. 1996. "Market Microstructure and Asset Pricing: On the Compensation for Illiquidity in Stock Returns." *Journal of Financial Economics*, vol. 41, no. 3 (July):441–464.

Christie, W., and P. Schultz. 1994. "Why Do Nasdaq Market Makers Avoid Odd-Eighth Quotes?" *Journal of Finance*, vol. 49, no. 5 (December):1813–40.

Cushing, D., and A. Madhavan. 2001. "The Hidden Cost of Trading at the Close." In *Transaction Costs: A Cutting-Edge Guide to Best Execution*. Edited by Brian Bruce. New York: Institutional Investor.

De la Vega, J. 1688. *Confusion de Confusiones: Portions Descriptive of the Amsterdam Stock Exchange*. (Translation by H. Kellenbenz, Harvard University, 1957).

Easley, D., and M. O'Hara. 1987. "Price, Trade Size, and Information in Securities Markets." *Journal of Financial Economics*, vol. 19, no. 1 (September):69–90.

Ellis, K., R. Michaely, and M. O'Hara. 2000. "When the Underwriter Is the Market Maker: An Examination of Trading in the IPO After-Market." *Journal of Finance*, vol. 55, no. 3 (June):1039–74.

Evans, M., and R. Lyons. 1999. "Order Flow and Exchange Rate Dynamics." Working paper, University of California at Berkeley.

French, K., and R. Roll. 1986. "Stock Return Variances: The Arrival of Information and the Reaction of Traders." *Journal of Financial Economics*, vol. 17, no. 1 (September):5–26.

Garman, M. 1976. "Market Microstructure." *Journal of Financial Economics,* vol. 3, no. 3 (June):257–275.

Glosten, L., and P. Milgrom. 1985. "Bid, Ask, and Transaction Prices in a Specialist Market with Heterogeneously Informed Agents." *Journal of Financial Economics*, vol. 14, no. 1 (March):71–100.

Harris, L. 1998. "Does a Large Minimum Price Variation Encourage Order Exposure?" Working paper, University of Southern California.

———. 2001. "Trading and Exchanges." Mimeo, University of Southern California.

Hasbrouck, J. 1991. "The Summary Informativeness of Stock Trades: An Econometric Analysis." *Review of Financial Studies*, vol. 4, no. 3 (Fall):571–595.

———. 1995. "One Security, Many Markets: Determining the Contributions to Price Discovery." *Journal of Finance*, vol. 50, no. 4 (September):1175–99.

Hasbrouck, J., and D. Seppi. 2001. "Common Factors in Prices, Order Flows, and Liquidity." *Journal of Financial Economics*, vol. 59, no. 3 (March):383–411.

Kavajecz, K., and E. Odders-White. 2002. "Technical Analysis and Liquidity Provision." Working paper, Wharton School of the University of Pennsylvania.

Keim, D. 1999. "An Analysis of Mutual Fund Design: The Case of Investing in Small-Cap Stocks." *Journal of Financial Economics*, vol. 51, no. 3 (February):173–194.

Keim, D., and A. Madhavan. 1996. "The Upstairs Market for Large-Block Transactions: Analysis and Measurement of Price Effects." *Review of Financial Studies*, vol. 9, no. 1 (Spring):1–36.

———. 1998. "The Costs of Institutional Equity Trades: An Overview." *Financial Analysts Journal*, vol. 54, no. 4 (July/August):50–69.

Kyle, A. 1985. "Continuous Auctions and Insider Trading." *Econometrica*, vol. 53, no. 6 (November):1315–35.

Lo, A., A.C. MacKinlay, and J. Zhang. 2001. "Econometric Models of Limit-Order Executions." Working paper, MIT Sloan School of Management.

Loeb, T. 1983. "Trading Cost: The Critical Link between Investment Information and Results." *Financial Analysts Journal*, vol. 39, no. 3 (May/June):39–43.

Lyons, R. 2001. *The Microstructure Approach to Exchange Rates*. Cambridge, MA: MIT Press.

Madhavan, A., 1996. "Security Prices and Market Transparency." *Journal of Financial Intermediation*, vol. 5, no. 3 (July):255–283.

———. 2000. "Market Microstructure: A Survey." *Journal of Financial Markets*, vol. 3, no. 3 (August):205–258.

———. Forthcoming. "Index Reconstitution and Equity Returns." *Financial Analysts Journal*.

Madhavan, A., and M. Cheng. 1997. "In Search of Liquidity: An Analysis of Upstairs and Downstairs Trades." *Review of Financial Studies*, vol. 10, no. 1 (Spring):175–204.

Madhavan, A., and S. Smidt. 1993. "An Analysis of Changes in Specialist Quotes and Inventories." *Journal of Finance*, vol. 48, no. 5 (December):1595–1628.

Madhavan, A., Porter, D., and D. Weaver. 2002. "Should Securities Markets Be Transparent?" Working paper, Baruch University.

O'Hara, M. 1995. *Market Microstructure Theory*. Cambridge, MA: Blackwell.

Stoll, H. 1989. "Inferring the Components of the Bid–Ask Spread: Theory and Empirical Tests." *Journal of Finance*, vol. 44, no. 1 (March):115–134.