Market Microstructure: A Practitioner's Guide*

Ananth Madhavan
ITG Inc.
380 Madison Avenue
New York, NY 10017

April 28, 2003

Our knowledge of market microstructure – the process by which investors' latent demands are ultimately translated into prices and volumes – has grown explosively in recent years. This literature is of special interest to practitioners given 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. This paper discusses the practical implications of the literature, focusing on price formation, market structure, transparency and applications to other areas of finance.

---

* 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 and do not necessarily reflect those of officers or directors of ITG Inc.

1 Introduction

Market microstructure is concerned with the process by which investors' latent demands are translated into executed trades. Interest in market microstructure is hardly new[1] but has increased enormously in recent years because of the rapid structural, technological, and regulatory changes affecting the global securities industry. Beyond these immediate concerns, however, there is a broader interest in microstructure. Indeed, a central concept in microstructure is that asset prices need not equal full-information expectations of value because of a variety of frictions. Thus, market microstructure is closely related to the field of investments, which studies the fundamental values of financial assets. But microstructure is also linked to traditional corporate finance because discrepancies between prices and value affect the level and choice of corporate financing.

Our knowledge of microstructure has grown explosively in recent years, fueled by complex new models and rich intraday data from a variety of sources. Yet, despite their practical value, many important theoretical insights and empirical results from academic research are not readily accessible to practitioners. An illustrative sample of such topics include:

- Determinants of transaction costs and models to predict costs before the trade;
- Limit order models for evaluating trading strategies or automated market making;
- Liquidity as a factor in asset returns and in portfolio risk;
- Whether displaying the limit order book affects liquidity and volatility;
- The choice of automated or floor trading systems; and
- The link between IPO pricing and secondary market dealer activity.

This article provides a practitioner-oriented review of the literature.[2] I should emphasize that this article is not a survey of topics under current debate. These topics change frequently and receive comprehensive coverage in press and industry publications. Rather, I highlight the most relevant academic literature, emphasizing 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 present and future. Any survey must be selective and this is especially so for microstructure where the literature comprises thousands of articles. Madhavan (2000) provides a more complete set of citations.

Four categories merit attention:

(1) *Price formation and price discovery*, including both static issues such as the determinants of trading costs and dynamic issues such the process by which prices come to impound information over time. Essentially, the goal is to look inside the "black box" by which latent demands are translated into realized prices and volumes.

(2) *Market structure and Design Issues,* including the relation between price formation and trading protocols. The focus is on how different rules affect the black box and hence liquidity and market quality.

(3) *Information*, especially market transparency, i.e., the ability of market participants to observe information about the trading process. This topic deals with how revealing the workings of the black box affects the behavior of traders and their strategies.

---

[1] A classic description of trading on the Amsterdam Stock Exchange is provided by Joseph de la Vega (1688) who describes insider trading, manipulations, and futures and options trading.

[2] See also Madhavan (2000), Lyons (2000), Harris (2000), Keim and Madhavan (1998), and O'Hara (1995).

(4) *Interface of market microstructure* with other areas including corporate finance, asset pricing, and international finance. Models of the black box provide fresh perspectives on a topics including IPO underpricing, portfolio risk, foreign exchange movements, etc..

These categories roughly correspond to the historical development of research in the informational aspects of microstructure, and form the basis for the organization of this article.

The paper proceeds as follows. Section 2 summarizes the literature on price formation with an emphasis on the role of market makers. Section 3 turns to issues of market structure and design. Section 4 looks at the topic of transparency and Section 5 surveys the interface of microstructure with other areas of finance. Section 6 concludes.

2 Price Formation and Discovery

2.1 The Crucial Role of Market Makers

Price formation, the process by which prices come to impound new information, is the most fundamental topic in microstructure. By virtue of their role as price setters, market makers are a logical starting point for an exploration of the "black box" of a security market actually works. In the simplest model (Demsetz, 1968), market makers play a passive role in supplying "immediacy," the price of which is the bid-ask spread. Empirical research confirms that bid-ask spreads are a function of proxies for the costs of liquidity provision and competition. Spreads are lower in higher volume securities because dealers can achieve faster turnaround in inventory, lowering reducing their risk. Similarly, spreads are wider for riskier and less liquid securities.

A deeper understanding of trading costs came from subsequent studies that explain variation in bid-ask spreads as part of intraday price dynamics. This research shows 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.

Garman's Logic

Garman (1976) shows 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., follows a random walk with zero drift. While inventory has zero drift, the variance of inventory is proportional to the number of trades. Intuitively, if we flip a coin and win a dollar on heads and lose a dollar on tails, our net expected gain is zero, but our 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. It follows that to avoid such "ruin," market makers must actively adjust prices in relation to inventory, altering price levels and not simply spreads.

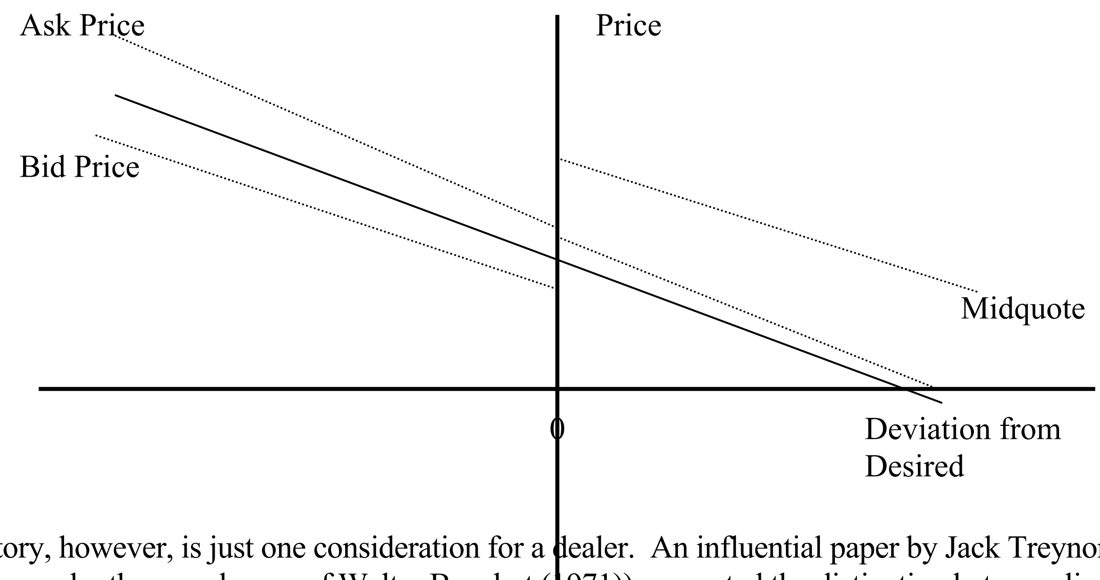
Price may depart from expectations of value if the dealer is long or short relative to desired (target) inventory, giving rise to transitory price movements during the day and possibly over longer periods.

This intuition drives the models of inventory control developed by Madhavan and Smidt (1993), among others.

Figure 1 illustrates a typical inventory model. As the dealer trades, the actual and desired inventory positions diverge, forcing the dealer to adjust prices, lowering prices if long and raising them if short relative to target inventory. Since 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 their desired or target inventory; it widens as inventory deviations get 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 impacts get progressively larger following a sequence of trades on one side of the market. This is an important consideration for institutional traders who typically breakup their block trades over several trading sessions. Second, since the concessions demanded by dealers are temporary, we might observe large price reversals from the close to the open, i.e., once market makers have had a chance to layoff excess inventory in other markets or hedge their risk. Third, because inventory effects are related to the degree to which dealers are capital constrained, we might observe larger inventory effects for smaller dealers with less capital. Finally, inventory models provide an added rationale for the reliance on dealers. Specifically, just as physical market places consolidate buyers and sellers in *space*, the market maker can be seen 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 simply buys from the dealer who depletes his or her inventory.

Figure 1: Price and Deviation from Target Inventory



Inventory, however, is just one consideration for a dealer. An influential paper by Jack Treynor (writing under the pseudonym of Walter Bagehot (1971)) suggested the distinction between liquidity motivated traders (who possess no special informational advantages) and informed traders (who possess private information about future value). While the market maker loses to informed traders on average, but recoups these losses on trades with liquidity-motivated (noise) traders. Models of this type include Glosten and Milgrom (1985), Easley and O'Hara (1987), among others.

Post-Trade Rationality

> We can easily prove that bid-ask spreads contain a component attributable to asymmetric information. Consider an extreme example with no inventory or transaction costs. Some traders have information about future asset values, however. Based on public information, the dealer believes that the stock is worth \$30. The dealer, however, is post-trade rational. In particular, given that a trader buys 100 shares, the dealer knows that the probability that the asset is undervalued is greater than the probability is overvalued. Why? Because informed traders only participate on one side of the market. Suppose, for the sake of exposition, that the expected asset value given that the dealer observes a buy of 100 shares is \$30.15, and symmetrically assume the expected value given 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 regret free in the sense that after the trade the dealer does not suffer a loss. There is a non-zero bid-ask spread driven purely by information effects.

Asymmetric 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 will not be willing to provide liquidity and markets will fail, and (3) Given the practical impossibility of identifying informed traders (they are not necessarily insiders), prices adjust in the direction of money flow.

Empirical evidence on the extent to which information traders affect the price process is complicated by the difficulty in identifying explicitly the effects due to asymmetric 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 they adjust prices accordingly. In the information model, order flow acts as a signal about future value and causes a revision in beliefs. Stoll (1989) proposes a method to distinguish the two effects using transaction data, but without inventory data, it is difficult to verify the results of such indirect approaches. Madhavan and Smidt (1993) develop a dynamic programming model that incorporates both 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 towards inventory targets; as an active investor, the market maker periodically adjusts the target inventory levels towards which inventories revert. They estimate the model with daily specialist inventory data and find evidence of both inventory and information effects. Inventory and information effects also explain why we might observe "excess" volatility in the sense that market prices appear to move more often than is warranted than by "fundamental" news about interest rates, dividends, etc. An interesting example is provided below.

> **Does Trading Create Volatility?**

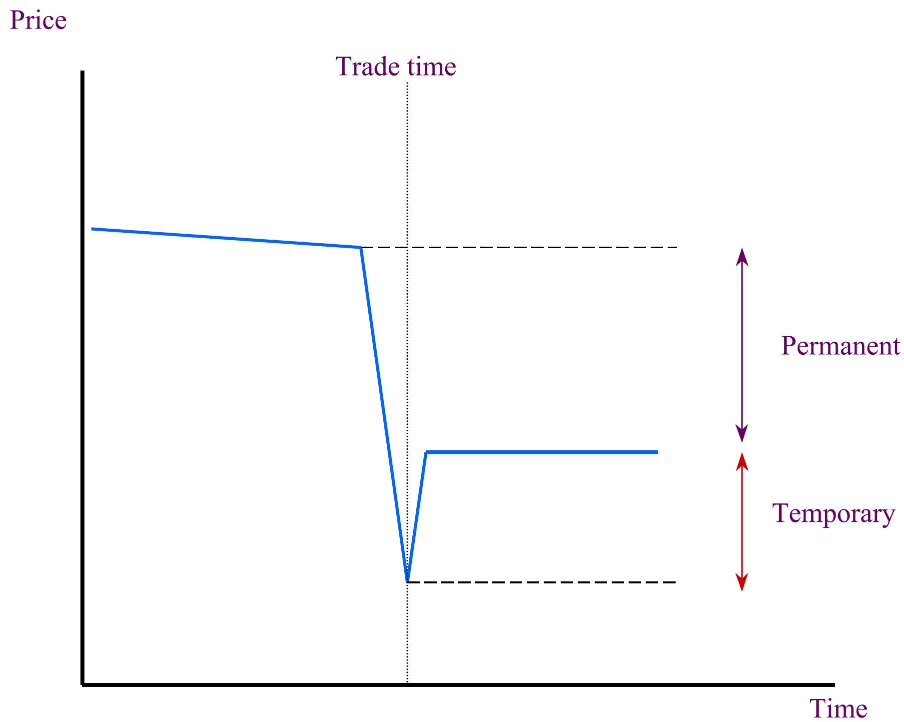
French and Roll (1985) find that on an hourly basis, the variance during trading periods is at least *twenty* times larger than the variance during non-trading periods. One explanation is public information arrives more frequently during business hours, when exchanges are open. Alternatively, order flow may be required to move prices to equilibrium levels. To distinguish between these explanations is difficult. However, a historical quirk in the form of weekday "exchange holidays" that the NYSE declared at one point in time to catch up on a backlog of paper work provides an answer. Since other markets and businesses are open, the public information hypothesis predicts the variance over the two day period beginning with the close the day before the exchange holiday should be roughly double the variance of returns on a normal trading day. In fact, the variance for the period of the weekday exchange holiday and the next trading day is only 14 percent higher than the normal one-day return. This evidence suggests that trading itself is the source of volatility; for markets to be efficient, someone has to make them efficient.

2.2 A Practical Illustration of Information Theories

One important implication of the information models concerns the price movements associated with large trades. In many equity markets, there are two economically distinct trading mechanisms for large-block transactions. First, a block can be sent directly to the "downstairs" or primary markets. These markets in turn comprise the continuous intraday markets, such as the NYSE floor. Second, a block trade may be directed to the "upstairs" market where a block broker facilitates the trading process by locating counter-parties to the trade and then formally crossing the trade in accordance with the regulations of the primary market. The upstairs market operates as a search-brokerage mechanism where prices are determined through negotiation. By contrast, downstairs markets provide immediate execution at quoted prices.

We can decompose the price impact of a block trade into permanent and temporary components. 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. Let p_{t-h} denote the pre-trade benchmark, p_t the trade price, and p_{t+k} the post trade benchmark price, where h and k are suitably chosen periods. The price impact of the trade, relative to the pre-trade benchmark, is just $p_t - p_{t-h}$. In turn, the price impact can be decomposed into two components, a permanent component defined as $\pi = p_{t+k} - p_{t-h}$ and a temporary component, defined as $\tau = p_t - p_{t+k}$. Figure 2 illustrates the impacts for a block sale.

Figure 2: Price Impact Components of a Block Sale



Keim and Madhavan (1996) model the upstairs market as a mechanism to aggregate traders and dampen the price impacts associated with a block trade by risk sharing. They test their model using upstairs market data. Price impacts of block trades are large in small cap stocks; as expected, they rise with trade size and fall with market capitalization. The choice of pre-trade benchmark price makes a large difference in the estimated price impact. For example, using a sample of trades made by an institutional trader, Keim and Madhavan find that the average (one-way) price impact for a seller-initiated transaction is -4.3% when the benchmark ("unperturbed") price is the closing price on the day before the trade. However, when the benchmark is the price three weeks before the trade, the measured price impact is -10.2%, after adjustment for market movements. While part of the difference in price impacts may be explained by the initiating institutions placing the sell orders after large price declines, Keim and Madhavan find little evidence to suggest that institutional traders act in this manner. Rather, they attribute the difference to information "leakage" arising from the process by which large blocks are "shopped" in the upstairs market. If this is the case, previous estimates in the literature of price impacts for block trades are downward biased.

A Passive Fund that Outperforms the Index

Keim (1999) analyzes the performance of the 9-10 fund of Dimensional Fund Advisors (DFA). The 9-10 fund is a passive index fund that attempts to replicate the performance of the bottom two deciles of the NYSE by market capitalization. Keim reports that the 9-10 fund's mean return since its inception in 1982 exceeds that of the benchmark index by an average of almost 250 basis points, without higher risk. This performance would be envied by many actively managed funds but is unheard of in a passive index fund. Keim shows that the outperformance is largely due to DFA's clever use of the upstairs market. Instead of immediately selling or buying shares when a stock moves into or out of the universe, DFA trades in the upstairs market, providing liquidity when approached by block traders who know DFA's strategy. Thus, DFA earns the spread in the upstairs market, although it incurs higher tracking error than many passive index funds are willing to tolerate. The rewards to earning the spread – as opposed to incurring the price impact costs in illiquid stocks – are significant. In recent years, upstairs trading alone generates 204 basis points annually to the 9-10 fund's return.

The upstairs market has been viewed in the literature primarily from the initiator's viewpoint. Upstairs intermediation can reduce trading costs by mitigating adverse selection costs, locating trade counter-parties, and risk sharing. However, every block trade involves willing participants on both sides of the transaction. Thus, one way to interpret the results reported is that the primary benefit offered by the existence of an upstairs market may not be to the initiator but rather to the counter-parties to the transaction. Liquidity providers, especially institutional traders, are reluctant to submit large limit orders and thus offer free options to the market. Upstairs markets allow these traders to selectively participate in trades screened by block brokers who avoid trades that may originate from traders with private information. Thus, the upstairs market's major role may be to enable transactions that would not otherwise occur in the downstairs market. If so, then these traders would perhaps be more willing to trade in downstairs markets if they offered less information about their identities.

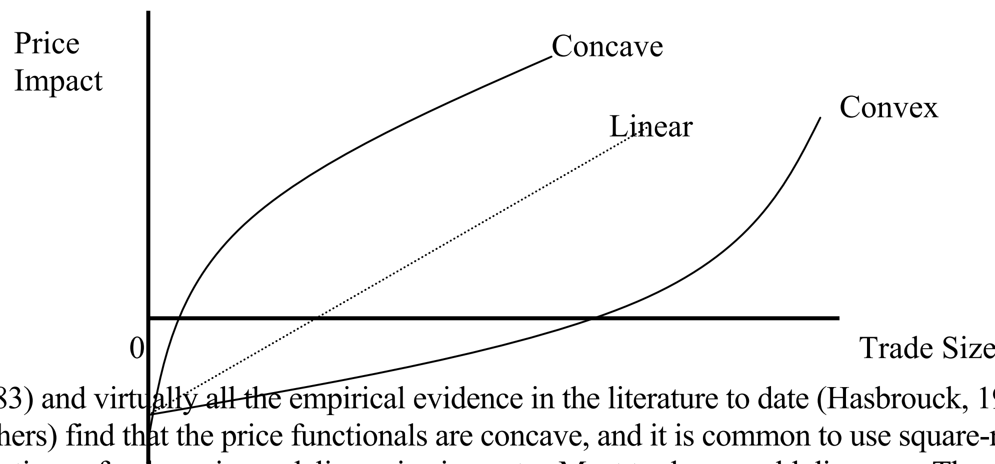
2.3 Pre-Trade Cost Estimation

Intraday models are essential to formulate accurate predictions of trading costs. Pre-trade cost models are increasingly used by large traders who are aware of the impact of trading costs on investment performance. Pre-trade cost estimates are essential to evaluate alternative trading strategies and to form benchmarks to evaluate the performance of individual traders, funds, and brokers. In practice, however, it is difficult to develop models to anticipate the cost of execution. There are several essential ingredients of a successful model: (1) Since most investors break their orders into component trades, the model builder must recognize that current trades affect the prices at which future transactions by distinguishing between permanent and temporary price impacts. (2) Costs depend on stock-specific attributes (liquidity, volatility, price level, and market) and order complexity (order size relative to average daily volume, trading horizon). (3) As costs are a function of style, there will not be a single cost estimate for any given order. Rather, the model should yield cost estimates that vary with the aggressiveness with which the order is presented to the market. In particular, an order traded over a short horizon using market orders will have higher costs than if traded passively over a long horizon using limit orders, upstairs markets, or crossing systems.

The first consideration implies that a realistic model will have to be solved recursively, because the execution price of the last sub-block of an order depends on how the last but one sub-block 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; we have conjectures about their means, but recognize that other factors will affect the actual execution prices. Dynamic programming is a mathematical technique designed to provide solutions to multiperiod problems where actions today affect rewards in the future. Examples of models of this type include Almegren and Chriss (1999) and Bertsemas and Lo (1999). In models of this type, the price impact function, i.e., the effect of trade on price, is linear. This assumption is made partly for analytical tractability but also because theoretical models (such as Kyle (1985)) derive linear equilibria from fundamental principles. A consideration of equal or greater importance is that when the permanent price function is linear, it is not possible to 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 (Keim and Madhavan, 1996) to be non-linear. In practical terms, the model builder who allows non-linear price functions often runs into situations where the recommended trade strategy involves some trades that are the opposite direction of the desired side. This is often counter-intuitive to traders and risks the possibility of regulatory scrutiny. Thus, it is common in more advance models to impose a restriction that all parcels of the order be on the same side as the order itself.

But while most traders agree that linearity is too simplistic an assumption, there is disagreement 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.

Figure 3: Price Impact Functions



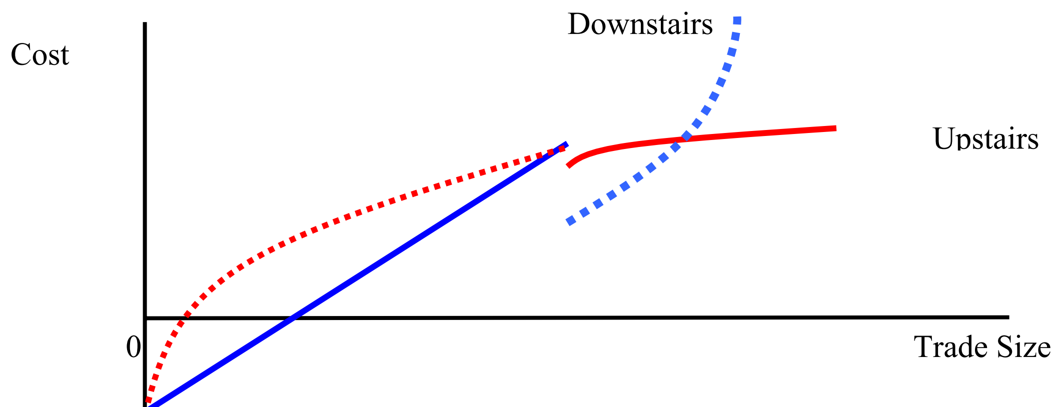
Loeb (1983) and virtually all the empirical evidence in the literature to date (Hasbrouck, 1991, among others) find that the price functionals are concave, and it is common to use square-root transformations of volume in modeling price impacts. Most traders would disagree. They understand that liquidity is limited so that at some stage these functions are convex. How can one resolve this crucial question?

Madhavan and Cheng point out that publicly available databases (TAQ for example) do not distinguish between upstairs and downstairs trades. In their view, this failure can resolve the paradox discussed above. Upstairs trades occur by matching buyers and sellers. Keim and Madhavan (1996) argue that the price functions are concave for upstairs trades. They show that as block size increases, more counterparties are contacted by the upstairs broker, cushioning the price impact, the costs of trading are relatively low for large sizes. These results are consistent with Loeb (1985) who interviews block traders and reports a concave price impact function. Because of commission costs, upstairs trades are relatively uneconomical for small trades.

Simultaneously, the price impact function for large trades in the downstairs market, that is for market orders sent anonymously to various market centers, might well be convex. To see why, consider a market order sent to the NYSE. Small orders are executed through the SuperDot system and if below the stated depth have zero price impact or possibly even receive price improvement. Medium sized orders may execute against the limit order book or the specialist, having 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. The result is a convex price functional.

Overall, traders who have the choice will select the lowest cost mechanism (Madhavan and Cheng, 1997) so that the proportion of upstairs trades rises with size. Without data distinguishing upstairs and downstairs trades, we understates the true cost of large trades that are directed to the downstairs market and makes it appear as if a concave price functional best fits the data. Figure 4 illustrates. The observed relation is the lower of the two curves, in solid line.

Figure 4: Upstairs and Downstairs Trading Costs



3 Market Structure and Design

Market architecture refers to the set of rules governing the trading process. Many academic studies have shown that market structure matters, 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:

- **Degree of Continuity:** Periodic systems allow trading only at specific points in time while continuous systems allow trading at any point in time while the market is open
- **Dealer Presence:** Auction or order-driven markets feature trade between public investors without dealer intermediation while in a dealer (or quote-driven) market, a market maker takes the opposite side of every transaction.
- **Price Discovery:** The extent to which the market provides independent price discovery or uses prices determined in another market as the basis for transactions.
- **Automation**: Markets vary considerably in the extent of automation, with floor trading and screen-based electronic systems at opposite extremes. The technology of order submission is, however, less important than the protocols governing trading.
- **Order Forms** permitted (i.e., market, limit, stop, upstairs crosses, hidden, etc.).
- **Protocols** (i.e., rules regarding program trading, choice of minimum tick, trade-by-trade price continuity requirements, rules to halt trading, circuit breakers, and adoption of special rules for opens, re-opens, and closes);

- **Pre- and Post-Trade Transparency,** i.e., the quantity and quality of information provided to market participants during the trading process. Non-transparent markets provide little in the way of indicated prices or quotes. Transparent markets often provide a great deal of relevant information before (quotes, depths, etc.) and after (actual prices, volumes, etc.) trade occurs.
- **Information Dissemination,** Markets also differ in the *extent* of dissemination (brokers, customers, or public) and the *speed* of dissemination (real time or delayed feed).
- **Anonymity**, is a crucial factor, including hidden orders, counterparty disclosure, etc.
- **Off-Market Trading**, i.e., whether off exchange or after hours trading is permitted.

Trading systems exhibit considerable heterogeneity, as shown in figure 4. For example, automated limit order book systems of the type used by the Toronto Stock Exchange 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. Foreign exchange and corporate junk bond markets rely heavily on dealers to provide continuity but offer very little transparency while other dealer markets (Nasdaq, London Stock Exchange) offer moderate degrees of transparency. Non-continuous markets include the Arizona Stock Exchange and the NYSE open, which differ considerably in transparency and dealer participation. Some exchanges also require fairly strict trade-to-trade price continuity requirements while others, like the Chicago Board of Trade (CBOT), 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 rather cross orders at the midpoint of the quotes in the primary market.

Figure 4: Variation in Real-World Trading Systems

	Island ECN	NYSE Open	NYSE Intraday	Paris Bourse	POSIT	CBOT	FX Market
Continuous	×		×	×		×	×
Dealer Presence		×	×			×	
Price Discovery	×	×	×	×		×	×
Automation	×			×	×		
Anonymity	×	×		×	×		
Pre-trade Quotes	×		×	×		×	
Post-trade Reports	×	×	×	×	×	×	

Do such differences affect price formation and the costs of trading? We turn now to this issue, focusing on some of the key issues in market design.

3.1 Current Issues in Market Design

The diversity of systems above has spurred considerable theoretical research. Early in the literature, the presence of strong network externalities was recognized. Higher volumes imply a shorter holding period for market makers and hence lower inventory control costs. Initially, suppose volumes are split equally between the two markets, but suppose that volume migrates to the market with lower costs. If the initial volume allocation is perturbed slightly, the higher volume market will enjoy reduced costs, attracting further volume, until in the long run there will consolidation into a single market. The inclusion of information into this model only serves to confirm this prediction. With asymmetric 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 narrower spreads. Even if we just assume symmetric, but diverse, information signals, pooling orders will provide informationally more efficient prices than decentralized trading across fragmented markets. Indeed, even when multiple markets coexist, the primary market often is the source of all price discovery (as shown by Hasbrouck, 1995) with the satellite markets merely matching quotes.

But despite strong arguments for consolidation, many markets are fragmented and remain so for long periods of time. There are two aspects to this puzzle: (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. In terms of the first issue, theory suggests that multilateral trading systems (such as single-price call auctions) are efficient mechanisms to aggregate diverse information. Consequently, there is interest in how call auctions operate and whether such systems can be used more widely to trade securities. The information aggregation argument suggests call auctions are especially valuable when uncertainty over fundamentals is large and market failure is a possibility. Casual empiricism appears to support this aspect of the argument. Indeed, many *continuous* markets use single-price auction mechanisms when uncertainty is large such as at the open, close, or to re-open following a trading halt. Yet, trading is often organized using continuous, bilateral systems instead of a periodic, multilateral system. For reasons not well understood, there is a surprising demand for continuous trading, even if this necessitates reliance on dealers to provide liquidity. With regard to the second issue, while consolidated markets pool information, it is not necessarily clear that they will be more efficient than fragmented markets if some traders can develop reputations based on their trading histories. One example of such rational fragmentation is off-market trading. Upstairs trading captures the willingness of traders to seek execution outside the primary market, and hence is of interest in debates regarding consolidation and fragmentation. One argument cited for the growth of upstairs markets in the U.S. is that the downstairs markets – in particular the NYSE – offer too much information about a trader's identity and motivations for trade. Models emphasizing asymmetric information provide some rationale for the success of off-market competitors in attracting order flow from primary markets. Established markets could experience competition in the form of cream-skimming of orders likely to originate from uninformed traders and broker-dealers can internalize their order flow, passing on the unmatched orders to the primary market.

3.2 The Automated Auction

Within the class of continuous markets, trading can be accomplished using designated dealers or as a limit order market 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 include ECNs (Island, Archipelago), Paris Bourse, etc. With a limit order, an investor associates a price with every order such 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 where the limit price is the current ask price or higher. So, there is growing interest in developing models of limit order execution.

Estimating Limit Order Models

Limit order models can help a trader evaluate a limit order strategy or may be used as the basis for automated market making. Such models provide 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. First passage time models estimate the probability a limit order will be executed based on the properties of the stock price process typically assuming 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 crosses a given limit price in a pre-specified period of time is straightforward to compute. While analytically convenient, FPT models are not especially realistic. Econometric models (Lo, MacKinlay, and Zhang, 2001) offer more realism because they can accommodate large numbers of explanatory variables. However, estimating an econometric limit order model is problematic. Specifically, unless we know the investor's strategy for canceling unfilled limit orders, estimating the probability of execution is difficult because we only observe executions for filled orders. Statistical techniques are available to handle such "censoring," for example, using survival analysis to model the true time to execution. In particular, if T is the random execution time, we model 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 to the market that can be hit if circumstances change. Consequently, the limit order trader needs to expend resources to monitor the market, a function that may be costly. It is perhaps for this reason that dealers of some form or the other arise so often in auction markets.

Limit order models provide some insights into the consequences of changing the minimum tick or "decimalization." Strictly speaking, decimalization refers to the quoting of stock prices in decimals as opposed to fractions such as eighths or sixteenths. Proponents of decimalization note that it would allow investors to compare prices more quickly, thereby facilitating competition, and would also promote the integration of US and foreign markets. They often mistakenly compute large cost savings to investors because quoted spreads are likely to fall dramatically. By contrast, the minimum tick is a separate issue that concerns the smallest increment for which stock prices can be quoted. For example, one can envisage a system with decimal pricing but with a minimum tick of 5 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, the profits from supplying liquidity (assuming a constant book) go down. It follows that there will be a reduction in liquidity at prices away from the best bid or offer. However, the quoted spread itself may fall through competition. Thus, a reduction in the minimum tick may reduce overall market liquidity. See Harris (1998) for a discussion of this and related points.

The 24-Hour Test

> Differences between periodic and continuous systems might affect returns. Amihud and Mendelson (1987) compare return variances from open-to-open and close-to-close for NYSE stocks. Since both periods span 24 hours, any differences are likely to reflect differences in the trading system, the NYSE opening price being determined in a single-price auction while the closing price is determined in a continuous double-auction. Their evidence seems to support the view that differences between continuous and batch systems are exhibited in observable variables such as price efficiency and return volatility. See also Stoll and Whaley (1990).

Intermarket comparisons are very difficult because real world market structures are more complex than simple models would suggest. The NYSE, for example, has elements of both auction and dealer markets. Further, there are serious empirical issues concerning the definition and measurement of market quality. For example, the usual measure of trading costs (or illiquidity), namely the quoted bid-ask spread is problematic because quoted spreads capture only a small portion of a trader's actual execution costs.

While the early literature argued that competition among market makers on the Nasdaq system would result in lower spreads than a specialist system of the type used by the NYSE, the opposite seems to be the case, even after controlling for such factors as firm age, firm size, risk, and the price level. One explanation is provided by Christie and Schultz (1994) who suggest that dealers on Nasdaq may have implicitly colluded to set spreads wider than those justified by competition. Theoretical studies provide some justification for this view in terms of the institutions of the Nasdaq market. Specifically, institutions 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, resulting in supra-normal spreads despite the ease of entry into market making.

Tests of theories concerning market structure face a serious problem: the absence of high quality data that allows researchers to pose "what if" questions. There are some interesting natural experiments. For example, in the late 1990s, the Tel Aviv Stock Exchange moved some stocks from periodic trading to continuous trading, allowing researchers to investigate the effects on asset values with a control of stocks that did not move. Indeed, Amihud, Mendelson, and Lauterbach (1997) document large increases in asset values for stocks moving to continuous trading on the Tel Aviv stock exchange. But such instances are few and far between. Compounding the problem, traders adjust their strategies in response to market protocols and information. This makes it difficult to assess the impact of market protocols. Further, empirical studies are limited in that there are not large samples of events to study. In addition, changes in structure are often in response to perceived problems. An example is the Toronto Stock Exchange's change in display rules in response to the migration of order flow to U.S. markets. Such changes are often accompanied by design alterations in other dimensions as well, such as a switch to automation or disclosure.

Laboratory or experimental studies offer a very promising way to test subtle theoretical predictions of regarding market design. In a laboratory or experimental study, human subjects trade in artificial markets. Irrespective of method, researchers seek to examine the effects of various changes in protocols (e.g., changes in pre- and post-trade reporting) on measures of market quality.

3.3 Summary

Issues of market structure are central to the subject of market microstructure. While a great deal has been learned, it is fair to say that there is not a uniform view on what structures offer the greatest liquidity and least trading costs. This is hardly surprising given the considerable complexity of real-world market structures. Ultimate decisions on market structure are likely to be decided by the marketplace on the basis of factors that have less to do with information than most economists believe. The factor I would single out is a practical one, namely the need for automation and electronic trading to handle the increasingly high volumes of trading. While this factor will inevitably lead towards the increased use of electronic trading systems, this does not mean that investigations of market structure are irrelevant. The point to keep in mind, however, is that what ultimately matters is not the medium of communication between the investor and the market but the protocols that translate that order into a realized transaction. For instance, it is possible to replace the NYSE floor with a virtual, fully electronic market, while keeping the institutions of the specialist, brokers, etc. Formerly verbal communications between market participants would be replaced with communication by email. Whether this is desirable – or practical – is not the point. Rather, I just want to emphasize the importance of studying protocols rather than focusing on the technological aspects of trading.

4 Information

Many informational issues regarding market microstructure concern information and disclosure. Market transparency is defined (See, e.g., O'Hara, 1995) as the ability of market participants to observe information about the trading process. Information, in this context, can refer to knowledge about prices, quotes, or volumes, the sources of order flow, and the identities of market participants. It is useful to think of dividing transparency into pre- and post-trade dimensions. Pre-trade 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, as well as other pertinent trade related information such as the existence of large order imbalances. Post-trade transparency refers to the *public* and *timely* transmission of information on past trades, including execution time, volume, price, and possibly information about buyer and seller identifications. Consequently, transparency has many dimensions.

4.1 Current Issues Concerning Market Transparency

Issues of transparency have been central to some recent policy debates. For example, the issue of delayed reporting of large trades has been highly controversial and continues to be an issue as stock exchanges with different reporting rules form trading linkages. A closely related issue concerns the effects of differences in trade disclosure across markets. These differences, some argue, may induce order flow migration, thereby affecting liquidity and the cost of trading. Transparency is a major factor in debates over floor vs. electronic systems. Floor systems such as the New York Stock Exchange (NYSE) generally do not display customer limit orders unless they represent the best quote. By contrast, electronic limit order book systems such as the Toronto Stock Exchange Computer Assisted Trading System (CATS) and the Paris Bourse *Cotation Assistee en Continu* (CAC) system disseminate not only the current quotes but also information on limit orders away from the best quotes. In general, the trend around the world has been towards greater transparency.

The practical importance of market transparency has given rise to a large theoretical and empirical literature. Specifically, 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;
- Pre-disclosure of intentions to trade such as sunshine trading or the revelation of order imbalances at the open or during a trading halt;
- Dual-capacity trading, where brokers can also act as dealers;
- Front-running, where brokers trade ahead of customer orders;
- Upstairs and off-exchange trading;
- The role of 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. However, most floor-based trading systems offer some degree of transparency regarding the composition of order flow. For example, on the New York Stock Exchange (NYSE), the identity of the broker submitting an order may provide valuable information about the source and motivation for the trade.

Theoretical models reach mixed conclusions. In some models, transparency can reduce adverse selection problems, and hence spreads, by allowing dealers to screen out traders likely to have private information. However, other models show transparency can exacerbate the price volatility. The rationale is that disclosing information about "noise" in the market system increases the effects of asymmetric information, thereby reducing liquidity. Essentially, noise is necessary for markets to operate, and disclosure robs the market of this lubrication. Contrary to popular belief, the potentially adverse effects of transparency are likely to be greatest in thin markets.

These results have important policy implications concerning, 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 the 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 regarding the motivations of the initiators of those orders. Unless it is explicitly designed to function in a non-anonymous fashion, such inferences are extremely difficult in a system with electronic order submission. If this is the case, traditional exchange floors may be preferred over automated systems for the active issues while the opposite may be true for inactive issues. Finally, the results provide insights into why some liquidity-based traders may avoid sunshine trades, even if they can benefit from reputation signaling.

Non-disclosure benefits large institutional traders whose orders are filled with multiple trades by reducing their expected execution costs, but imposes costs on short-term noise traders. The rationale is that these traders can breakup their trades over time without others front-running them and hence raising their trading costs. However, non-disclosure benefits dealers by reducing price competition. The implication of this analysis is that faced with a choice between a high disclosure market and a low disclosure market, an uninformed institutional trader will prefer to direct trades towards the more opaque market. Why? Essentially, a large trade can be successfully broken up without attracting too much attention and hence moving the price in the direction of the trade. 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.

4.2 Empirical Research on Transparency and Disclosure

Porter and Weaver (1998a) find a decrease in liquidity associated with the display of the limit order book on the Toronto Stock Exchange even after controlling for other factors that may affect spreads in this period, including volume, volatility, and price. Limit order traders are less willing to submit orders in a highly transparent system because these orders essentially represent free options to other traders. In terms of post-trade transparency, Porter and Weaver (1998b) study the effects of late trade reporting on Nasdaq. They find that large numbers of trades are reported out-of-sequence relative to centralized exchanges such as the NYSE and AMEX. Porter and Weaver (1998b) conclude that there is little support for the hypothesis that late-trade reporting is random or is the result of factors (such as "fast" markets, lost tickets, and computer problems) outside Nasdaq's control. Indeed, the trades most likely to be reported late are large block trades, especially those at away prices. This suggests that late-trade reporting is beneficial to Nasdaq dealers. This view is consistent with the arguments put forward by dealers on the London Stock Exchange against post-trade reporting.

Experimental Finance

The ability to frame controlled experiments in laboratory markets allows researchers to analyze difficult issues relating to information. The obvious focus is on metrics such as the bid-ask spread, market depth or liquidity, and volatility. But an experimental study also study quality variables that might not otherwise be possible to observe. These include data on traders' estimates of value over time, their beliefs regarding the dispersion of "true" prices, and the trading profits of various classes (informed or uninformed) of traders. Bloomfield and O'Hara (1999) use experimental markets to analyze changes in disclosure rules. In their study, lab participants face different disclosure regimes and in some experiments, dealers (markets) can decide whether they prefer transparency or not. Bloomfield and O'Hara find that transparency has a large impact on market outcomes. More generally, several interesting findings emerge from lab markets. It turns out to be quite easy to generate price bubbles, 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 price settle at the "wrong" value.

4.3 Summary

Transparency is a complicated subject, but recent research provides several revealing insights. First, there is broad agreement that both pre- and post-trade transparency matters; affecting liquidity and price efficiency. Second, greater transparency, both pre- and post-trade, is generally associated with more informative prices. Third, complete transparency is not always "beneficial" to the operation of the market. Indeed, many studies demonstrate that too much pre-trade transparency can actually reduce liquidity because traders are unwilling to reveal their intentions to trade. 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. Traders with private information prefer anonymous trading systems while liquidity traders, especially those who can credibly claim their trades are not information-motivated (e.g., passive index funds), prefer greater disclosure. Consequently, we will never find a market structure that all traders and dealers uniformly prefer.

5 Applications

The recognition that microstructure does matter – affecting asset values, liquidity, trading costs, and price efficiency – is relatively recent. This section provides examples of some of the applications of this research to other areas of finance, including (1) Asset pricing, (2) Corporate finance, and (3) International finance.

5.1 Asset Pricing

Previous research has modeled expected returns as functions of variables including proxies for size and default risk. Amihud and Mendelson (1986) show that expected returns are an decreasing function of liquidity because investors must be compensated for the higher transaction costs that they bear in less liquid markets. The presence of trading costs (asymmetric information, inventory costs, and other transaction costs) reduces the equilibrium *value* of the asset. Indeed, Amihud, Mendelson, and Lauterbach (1997) document large changes in asset values for stocks moving to more liquid trading systems on the Tel Aviv Stock Exchange. These and other studies confirm the role of liquidity in asset pricing.

From a cross-sectional viewpoint, variation in expected returns across securities arises because of differences in trading costs. Of course, it matters how we compute returns. In the simple model above, suppose there are two assets: a security that is subject to trading costs and one that is not. If we correctly measure the return to the illiquid security based on the actual purchase price, it will equal the risk-free rate, which is the return provided by the fully liquid asset.

Liquidity and Portfolio Risk

If liquidity is a factor in expected returns, its omission from a risk model can substantially understate 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. In particular, the trader will go long assets that have positive alphas relative to the incorrect model (where $\beta_{2i}F_2 > 0$) and short those with negative alphas (where $\beta_{2i}F_2 < 0$). If illiquidity increases, 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. Portfolio managers today increasingly use trading cost estimates when using optimization engines to form mean-variance efficient portfolios.

As noted above, measuring the price impact of the trade (i.e., λ) is difficult, especially without transaction level data. Keim and Madhavan (1998) show that these costs can be substantially larger than the observed spread s. If we compute the return on the security ignoring transaction costs (i.e., using the midquote as the basis for value) we obtain return premium $r - r_f > 0$ that represents the compensation for illiquidity, λ. Thus, even if we correctly account for the spread, we are likely to observe that expected excess returns are positively related to the trading costs ($\lambda_i + s_i$) across a sample of stocks after controlling for other factors affecting returns. 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) estimate such a cross-sectional model of returns with some success.

A promising area for research in this area is the subject of commonality in liquidity and returns. So far, our analysis, like much of the microstructure literature has focused on a single stock. Consider a model where the price change in each of N stocks is linearly related to order flows in own and related stocks, and other factors. In matrix notation, we write $\Delta\mathbf{p} = \mathbf{X}\Lambda + \mathbf{U}$, where $\Delta\mathbf{p}$ is a $N\times1$ vector of price changes, $\mathbf{X}$ is a $N\times k$ matrix of order flows, current and lagged, as well as other predetermined variables affecting price movements, Λ is a $k\times1$ vector of coefficients, and $\mathbf{U}$ is an $N\times 1$ vector of error terms. Returns in stock i may depend on current and lagged flows in stock j. Commonality in order flows is manifested in the fact 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 (1999) use principal components analysis and canonical correlation 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 the normalization restrictions imposed to remove indeterminacy. Hasbrouck and Seppi (1999) find that common factors are present in both returns and order flows. Common factors in order flows account for 50% of the commonality in returns. Whether such factors can help predict short-run returns, variation in intraday risk premia, or the observed relation between price variability and volume is still an open question.

Technical Analysis

Financial economists are traditionally skeptical of technical analysis where past price movements are used to predict future returns. In an efficient market, current prices should impound all available information, so that past price patterns should not have predictive power. Yet, modern microstructure theory suggests several avenues through which technical analysis might have value, at least over very short horizons. First, dealer inventories must be mean reverting from above. If so, and if there are inventory effects on prices, there should be cyclicality in prices. Specialist incentives to smooth prices may also lead to short run autocorrelation. Finally, if large traders break up their block trades (and if this information leaks slowly to the market), there will be short-run trends. To the extent that there are commonalities in order flow, such factors may also aggregate to the overall market level.

5.2 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 smaller capitalization 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. Yet, despite their prevalence, many basic questions concerning the operation of relationship markets remain unanswered.

An important issue is the role of underwriters in linking the primary and secondary stock markets for the firms they bring public. Underwriters of smaller stocks often dominate trading in the post-IPO market, giving them considerable ability to affect security prices. Ellis, Michaely, and O'Hara (1999) examine the role of the underwriter in after-market trading. They find that for Nasdaq stocks, the lead underwriter is almost always the primary market maker in the after-market. Why is this arrangement so common? Is there a link between the degree of underpricing and the secondary market? How does this affect the IPO decision?

To answer these questions, consider a model where an entrepreneur wishes to sell shares of a privately owned corporation to the public. Without market making by the lead underwriter, outside dealers provide liquidity. Since these dealers do not observe firm quality they set a spread that will compensate themselves for the risk of insider trading should they underestimate risk. In other words, the spread is determined by the worst possible risk on cash flows. By contrast, if the underwriter also acts as the dealer in the after market, he can set lower spreads because of the information advantage acquired during the IPO process.

The underwriter selects the spread with the objective of maximizing total revenues from commissions at the IPO stage plus future trading revenues. If the commission rate is sufficiently high, the underwriter undercuts the competitors and sets a lower spread. In turn, this increases the IPO price. Thus, the features of the relationship market are shown to have crucial effect on the ability of small firms to raise capital in the primary market. This model's predictions are consistent with stylized facts concerning both primary and secondary markets. Investment bankers who subsequently function as broker-dealers will have higher trading volume and will also offer to buy at higher prices, on average, than other competitive market makers. Several studies have documented that among firms qualified to list on both Nasdaq and the NYSE, smaller firms tend to list on Nasdaq. This result is puzzling because recent evidence suggests that both issue costs and bid-ask spreads tend to be higher on Nasdaq. This model, although simple, resolves this puzzle by showing that smaller firms might prefer a relationship market to a centralized market.

Stock Splits

Previous analyses of stock splits focus on corporate finance explanations such as signaling. Average stock prices are relatively constant over long periods of time within countries, despite variation across countries. However, the average stock price, relative to the minimum tick, is more constant. This suggests a possible microstructure based explanation for stock splits because a corporation can adjust its stock price relative to tick size through splits. Higher prices imply lower costs of capital and hence higher share values but at the same time, may discourage liquidity based trading by smaller retail investors. Thus, there might be an optimal price level that maximizes share value.

5.3 International Finance

In the international area, an important aspect of the interaction with microstructure concerns internal capital market segmentation. 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 hence affect market efficiency. Finally, an analysis of segmentation may shed light on the positive abnormal stock returns (Karolyi, 1996) observed following liberalizations.

One interesting and puzzling aspect of international segmentation arises when domestic firms issue different equity tranches aimed at different investors. For example, countries as diverse as Mexico 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 across firms and over time. Again, if both shares are otherwise equal but one share has higher transaction costs, that share will have a lower price if holding period returns are to be equal. Thus, share price premia or discounts can be explained in terms of relative trading costs. Elimination of market segmentation should reduce costs, lowering the cost of capital and boosting share prices in segmented markets. This model can explain the large jumps in share prices in emerging markets following economic liberalizations.

5.3.1 The Microstructure of Foreign Exchange Markets

Foreign exchange markets are by far the largest asset markets in terms of volume, and consequently there is considerable interest in how they operate and how prices are determined. An unusual feature of the foreign exchange market are the extremely large trading volumes, far larger than one would expect given the level of imports and exports. Lyons (1997) provides an elegant explanation for this phenomenon. The intuition of 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 depart from the desired level. This 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 through a "hot potato" effect, so that 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 aspects of market microstructure: (1) the dealer structure of the FX market, and (2) a lack of transparency in trade reporting. This is so particularly when dealers trade bilaterally over the telephone, still the most important method of dealing. The trade is then informative to them. The advent of electronic trading, e.g., EBS and Reuters D2000-2 systems, is changing the structure and availability of information to some extent. A violation of either of these assumptions would alter the nature of the equilibrium, dramatically reducing volumes.

Exchange Rate Modeling

Economic theory suggests that exchange rates movements are determined by macroeconomic factors. Yet, macroeconomic exchange rate models do not fit the data well, with R-square below 0.10. Evans and Lyons (1999) propose a microstructure model of exchange rate dynamics based on portfolio shifts that augments the standard macroeconomic variables with signed order flow. They estimate their model for the Deutsche Mark /Dollar and Yen/Dollar exchange rates. The model takes 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. As predicted, both β_1 and β_2 are positive and significant. The estimated R-square improves substantially when signed order flow is included. Over 50% of the daily changes in the DM/$ rate and 30% Yen/$ rate are explained by the model. Applications include short-run exchange rate forecasting, targeting of central bank intervention, and prediction of trading costs for large transactions.

6 Conclusions

Several lessons emerge from a careful reading of the literature. First, markets are a great deal more complex than commonly believed. One of the major achievements of the microstructure literature has succeeded illuminating the "black box" by prices and quantities are determined in financial markets. We now understand the role of inventory and asymmetric information in determining the responsiveness of prices to order flows. 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. Large price reactions to flows might also explain why proxies for liquidity appear to do so well in explaining the cross-sectional variation in returns. Second, microstructure does matter. Specifically, markets may fail under certain protocols, and there may be large deviations between "fundamental value" and price. Third, and a consequence of the discussion above, we must guard against "one size fits all" approaches to regulation and policy making. Greater transparency, for example, need 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 relation to expected returns is needed; there is growing evidence that liquidity is a "factor" in explaining stock returns. Differences in liquidity over time may explain variation in the risk premium and hence may influence stock price levels.

References

Almgren, R., Chriss, N., 1999. Optimal Execution of Portfolio Transactions. Working paper, NYU.

Amihud, Y., Mendelson, H., 1986. Asset Pricing and the Bid-Ask Spread. Journal of Financial Economics 17, 223-249.

Amihud, Y., Mendelson, H., 1987. Trading Mechanisms and Stock Returns: An Empirical Investigation. Journal of Finance 42, 533-553.

Amihud, Y., Mendelson, H., Lauterbach, B., 1997. Market microstructure and Securities Values: Evidence from the Tel Aviv Stock Exchange. Journal of Financial Economics 45, 365-390.

Bagehot, W., 1971. The Only Game in Town. Financial Analysts Journal 27, 12-14, 22.

Bertsemas, D., Lo, A., 1999. Optimal Control of Liquidation Costs. Journal of Financial Markets, 1, 1-50.

Black, F., 1971. Toward a Fully Automated Stock Exchange: Part I. Financial Analysts Journal 27, 28-35, 44.

Bloomfield, R., O'Hara, M., 1999. Can Transparent Markets Survive? Journal of Finance (forthcoming).

Brennan, M., J., Subrahmanyam, A., 1996. Market Microstructure and Asset Pricing: On the Compensation for Illiquidity in Stock Returns. Journal of Financial Economics 41, 441-464.

Christie, W., Schultz, P., 1994. Why do NASDAQ market makers avoid odd-eighth quotes? Journal of Finance 49, 1813-1840.

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

Demsetz, H., 1968. The Cost of Transacting. Quarterly Journal of Economics 82, 33-53.

Easley, D., O'Hara, M., 1987. Price, Trade Size, and Information in Securities Markets. Journal of Financial Economics 19, 69-90.

Ellis, K., Michaely, R., O'Hara, M., 1999. When the Underwriter is the Market Maker: An Examination of Trading in the IPO After-Market. Journal of Finance (forthcoming).

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

French, K., Roll, R., 1986. Stock Return Variances: The Arrival of Information and the Reaction of Traders. Journal of Financial Economics 17, 5-26.

Garman, M., 1976. Market microstructure. Journal of Financial Economics 3, 257-275.

Glosten, L., R., Milgrom, P., 1985. Bid, ask, And Transaction Prices in a Specialist Market With Heterogeneously Informed Agents. Journal of Financial Economics 14, 71-100.

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

Harris, L., 2000. Trading and Exchanges. Mimeo, University of Southern California.

Hasbrouck, J., 1991. The Summary Informativeness of Stock Trades: An Econometric Analysis. Review of Financial Studies 4, 571-595.

Hasbrouck, J., 1995. One Security, Many Markets: Determining The Contributions To Price Discovery. Journal of Finance 50, 1175-1199.

Hasbrouck, J., Seppi, D., 1999. Common Factors in Prices, Order Flows, and Liquidity. Working paper, New York University.

Karolyi, A., 1996. What Happens to Stocks that List Shares Abroad? A Survey of the Evidence and its Managerial Implications. Financial Markets, Institutions, and Instruments 7, 1-60.

Keim, D., 1999. An Analysis of Mutual Fund Design: The Case of Investing in Small-Cap Stocks, Journal of Financial Economics 51.

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.

Keim, D., Madhavan, A., 1998. The Costs of Institutional Equity Trades: An Overview. Financial Analysts Journal 54, 50-69.

Kyle, A., 1985. Continuous Auctions and Insider Trading. Econometrica 53, 1315-1335.

Lo, A., MacKinlay, A. C., Zhang, J., 2001. Econometric models of limit-order executions. Journal of Financial Economics, Forthcoming.

Loeb, T., F., 1983. Trading Cost: The Critical Link between Investment Information and Results. Financial Analysts Journal, 39, 39-43.

Lyons, R., 2000. The Microstructure Approach to Exchange Rates. (MIT Press, forthcoming).

Madhavan, A., 2000, Market Microstructure: A Survey. Journal of Financial Markets 3, 205-258.

Madhavan, A., Cheng, M., 1997. In Search of Liquidity: An Analysis of Upstairs and Downstairs Trades. Review of Financial Studies10, 175-204.

Madhavan, A., Smidt, S., 1993. An Analysis of Changes in Specialist Quotes and Inventories. Journal of Finance 48, 1595-1628.

O'Hara, M., 1995. Market Microstructure Theory (Basil Blackwell, Cambridge, Mass.).

Porter, D., Weaver, D., 1998a. Transparency and Liquidity: Should U.S. Markets be More Transparent? Working paper, Marquette University.

Porter, D., Weaver, D., 1998b. Post-Trade Transparency on Nasdaq's national market system. Journal of Financial Economics 50, 231-252.

Stoll, H., 1989. Inferring the Components of the Bid-Ask Spread: Theory and Empirical Tests. Journal of Finance 44, 115-134.

Stoll, H., Whaley, R., 1990. Stock Market Structure and Volatility. Review of Financial Studies 3, 37-71.