Mark D. Flood

Mark D. Flood is an economist at the Federal Reserve Bank of St. Louis. David H. Kelly provided research assistance.

Microstructure Theory and the Foreign Exchange Market

A GROWING BODY OF theoretical literature, known as the study of securities market microstructure, deals with the behavior of participants in securities markets and with the effects of information and institutional rules on the economic performance of those markets. These institutional factors may arise from technology, tradition or regulation. Microstructure and its impact are important, because of the vast amounts of wealth which pass through securities markets — including the foreign exchange market — every day.

Microstructure is of interest to students of the foreign exchange market: microstructural analyses of other markets have yielded insight into traders' behavior and the effect of various institutional arrangements. Conversely, the foreign exchange market is also of special interest to students of microstructure, because it combines two very different arrangements for matching buyers and sellers — bank dealers trade with one another both directly and through foreign exchange brokers.[1]

Standard models of exchange-rate determination concentrate on relatively long-run aspects, such as purchasing power parity. While microstructure theory cannot address these issues directly, it can illuminate a more narrowly focused array of institutional concerns, such as price information, the matching of buyers and sellers, and optimal dealer pricing policies. Despite the substantial literature on microstructure, little attention has been paid to the particular microstructure of the foreign exchange market.[2]

[1]Similar arrangements exist for other securities—for example, the federal funds market and the secondary market for Treasury securities—but these too have been relatively neglected in the literature.

[2]The shaded insert on the opposite page provides a context in which the microstructural approach can be compared with more traditional approaches to market efficiency.

Following some early articles by Demsetz (1968), Tinic (1972) and Tinic and West (1972), Garman (1976) performed the crucial task of defining market microstructure as an independent area of the literature, thus focusing the debate. Since then, market microstructure has burgeoned, led by Cohen, Maier, Schwartz and Whitcomb (1978a, 1978b, 1981, 1983), Amihud and Mendelson (1980, 1986, 1988), Stoll (1978, 1985, 1989) and Ho and Stoll (1980, 1981). See also Beja and Hakansson (1977), Cohen, Hawawini, Maier, Schwartz and Whitcomb (1980), Cohen, Maier, Ness, Okuda, Schwartz and Whitcomb (1977), Amihud, Ho, and Schwartz (1985), Schreiber and Schwartz (1986), Schwartz (1988) and Cohen and Schwartz (1989). Cohen, Maier, Schwartz and Whitcomb (1979, 1986) and Stoll (1985) have surveyed the microstructure literature.

In addition to the early note by Allen (1977), very recently there have appeared some microstructural studies of the foreign exchange market: Bossaerts and Hillion (1991), Lyons (1991), Rai (1991) and Flood (1991). There is also an empirical literature measuring the determinants of the bid-ask spread in the foreign exchange market. See Black (1989), Wei (1991) and Glassman (1987) as well as the references therein. Because the focus of this article is on microstructure theory, such empirical studies receive little attention here.

Finally, although a consideration of the results of laboratory experiments would expand the scope of this paper to unwieldy dimensions, their role in establishing the sensitivity of market behavior to institutional factors must at least be acknowledged; see Plott (1982, 1991) for an introduction.

Price Efficiency in a Heterogeneous Marketplace

Implicit in most microstructural models is a presumption that participants in any given market are heterogeneous, that is, that they differ in certain key determinants of economic behavior: information, beliefs, preferences and wealth. Although this assumption consumes little attention in the microstructure literature — it is taken for granted — it is valuable to discuss it in the more familiar theoretical context of market efficiency.

The standard definition of price efficiency is: $f_m(p_t|I_{mt-1}) = f^*(p_t|I_{t-1})$. In other words, the joint distribution over future prices, $f_m(p_t)$, assessed by the monolithic market (or a representative agent in that market) and made conditional on the current information, I_{mt-1}, available to the market is equal to the "true" joint distribution, $f^*(p_t)$, made conditional on all current information, I_{t-1}. Roughly speaking, the market sorts things out as accurately as possible.[1]

This approach breaks down in a microstructural analysis. First, the simplifying assumption of homogeneous participants is abandoned. Although it is widely recognized "that investors do not show the homogeneity of beliefs which characterize our theories," the benefits of realism (i.e., the heterogeneity assumption) are often outweighed by other criteria (e.g., testability, tractability, etc.).[2] Emphasizing testability, Ross offers a standard rejoinder, namely that "since a single ex post distribution of returns is observed by all, over time one would not expect to observe systematic and persistent differences." This is a rational expectations argument, which depends crucially on the stationarity of the returns distribution and which ignores the effect of differences in opinions and beliefs, which go beyond differences in information.

In general then, at the level of detail involved in microstructural studies, the homogeneity assumption is not an excusable flaw; in a homogeneous market why — let alone how — would anyone trade?

More fundamentally, the notion of a "true" price must be questioned. In the context of the literature on price efficiency, the introduction of a "true" distribution as a theoretical conceit leads to joint testing problems, as the "true" distribution is *ipso facto* unobservable. More fundamentally, positing a "true" distribution confuses the chain of causality; it presumes that future prices are drawn from some exogenous probability distribution and that investor behavior is concerned with accurately estimating that distribution.

In fact, investor behavior in the marketplace determines the distribution of future prices, not the other way around. This fact in no way depends on the ultimate basis or motivation for investor behavior. In an explicit model of price discovery, the assertion of an *ex ante* exogenous equilibrium price is meaningless. As Schreiber and Schwartz put it, "the fact that security analysts assess the value of a stock for their own portfolios does not imply that they undertake a treasure hunt to find some golden number which one might call an intrinsic value."[3] In sum, the standard theory of efficient markets is ill-suited to the modeling of price discovery. In comparing observed prices to an imputed "true" distribution, studies of market efficiency ignore more immediate concerns — for example, how well the institutional structure transmits information, whether arbitrage opportunities occur, and how well the market allocates assets among investors. These concerns are the focus of microstructural analysis.

[1] See Fama (1976), chapter 5, for the definitive presentation.

[2] See Ross (1978), pp. 889-90. See Varian (1989) for a more thorough review of the theoretical issues involved in the heterogeneity assumption.

[3] See Schreiber and Schwartz (1985), p. 22.

This paper examines the extant literature on market microstructure to determine how it might be applied to the foreign exchange market.

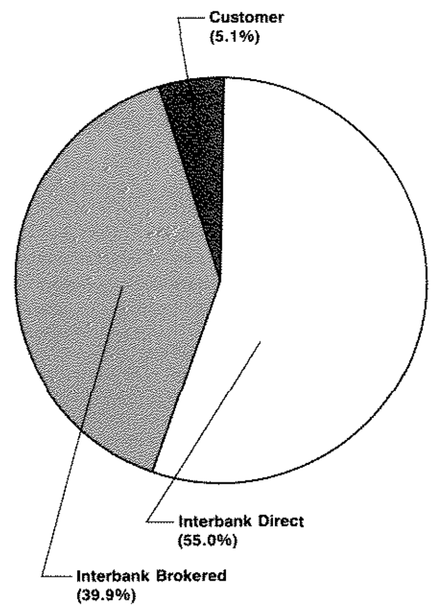
The paper begins with a brief description of the foreign exchange market. Aspects of the literature concerned with institutional details are addressed second, noting how such details can affect the performance of the market. Next, the literature dealing with behavioral details, especially the communication and interpretation of price information, is considered. Finally, the interaction of institutional and behavioral factors, notably the bid-ask spread, is discussed.

INSTITUTIONAL BASICS OF THE FOREIGN EXCHANGE MARKET

The foreign exchange market is the international market in which buyers and sellers of currencies "meet."[3] It is largely decentralized: the participants (classified as market-makers, brokers and customers) are physically separated from one another; they communicate via telephone, telex and computer network. Trading volume is large, estimated at $128.9 billion for the U.S. market in April 1989. Most of this trading was between bank market-makers.[4]

The market is dominated by the market-makers at commercial and investment banks, who trade currencies with each other both directly and through foreign exchange brokers (see figure 1).[5] Market-makers, as the name suggests, "make a market" in one or more currencies by providing bid and ask prices upon demand. A broker arranges trades by keeping a "book" of market-maker's limit orders — that is, orders to buy (alternatively, to sell) a specified quantity of foreign currency at a specified price — from which he quotes the best bid and ask orders upon request. The best bid and ask quotes on a broker's book are together called the broker's "inside spread." The other participants in the market are the customers of the market-making banks, who generally use the market to complete transactions in international trade, and central banks, who may enter the market to move exchange rates or simply to complete their own international transactions. Market-makers may trade for their own account — that is, they may maintain a long or short position in a foreign currency — and require significant capitalization for that purpose. Brokers do not contact customers and do not deal on their own account; instead, they profit by charging a fee for the service of bringing market-makers together.

Figure 1
Spot Market Volume by Transactor (4/89)



The mechanics of trading differ substantially between brokered transactions and direct deals. In the direct market, banks contact each other. The bank receiving a call acts as a market-maker for the currency in question, providing a two-way quote (bid and ask) for the bank placing the call. A direct deal might go as follows:

Mongobank: *"Mongobank with a dollar-mark please?"*

(Mongobank requests a spot market quote for U.S. dollars (USD) against German marks (DEM).)

[3]For more thorough descriptions of the workings of the foreign exchange market, see Burnham (1991), Chrystal (1984), Kubarych (1983) and Riehl and Rodriguez (1983).

[4]See Federal Reserve Bank of New York (1989a) and Bank for International Settlements (BIS) (1990). Extending this figure over 251 trading days per year, this implies a trading volume of roughly $32 trillion for all of 1989. Volume has roughly doubled every three years for the past decade.

[5]Federal Reserve Bank of New York (1989a) lists 162 market-making institutions (148 are commercial banks) and 14 brokers; an earlier study, Federal Reserve Bank of New York (1980), lists 90 market-making banks and 11 brokers.

Loans 'n Things: *"20-30"*
(Loans n' Things will buy dollars at 2.1020 DEM/USD and sell dollars at 2.1030 DEM/USD —the 2.10 part of the quote is understood.)

Mongobank: *"Two mine."*
(Mongobank buys $2,000,000 for DEM 4,206,000 at 2.1030 DEM/USD, for payment two business days later. The quantity traded is usually one of a handful of "customary amounts.")

Loans 'n Things: *"My marks to Loans 'n Things Frankfurt."*
(Loans n' Things requests that payment of marks be made to their account at their Frankfurt branch. Payment will likely be made via SWIFT.)[6]

Mongobank: *"My dollars to Mongobank New York."*
(Mongobank requests that payment of dollars be made to them in New York. Payment will most likely be made via CHIPS.)[7]

Spot transactions are made for "value date" (payment date) two business days later to allow settlement arrangements to be made with correspondents or branches in other time zones. This period is extended when a holiday intervenes in one of the countries involved. Payment occurs in a currency's home country.

The other method of interbank trading is brokered transactions. Brokers collect limit orders from bank market-makers. A limit order is an offer to buy (alternatively to sell) a specified quantity at a specified price. Limit orders remain with the broker until withdrawn by the market-maker.

The advantages of brokered trading include the rapid dissemination of orders to other market-makers, anonymity in quoting, and the freedom not to quote to other market-makers on a reciprocal basis, which can be required in the direct market. Anonymity allows the quoting bank to conceal its identity and thus its intentions; it also requires that the broker know who is an acceptable counterparty for whom. Limit orders are also provided in part as a courtesy to the brokers as part of an ongoing business relationship that makes the market more liquid. Because his limit order is often a market-maker's first indication of general price shift, Brooks likens the posting of an order with a broker "to sticking out the chin so as to be acquainted with the moment that the fight starts."[8] Schwartz points out that posting a limit order extends a free option to other traders.[9]

A market-maker who calls a broker for a quote gets the broker's inside spread, along with the quantities of the limit orders. A typical call to a broker might proceed as follows:

Mongobank: *"What is sterling, please?"*
(Mongobank requests the spot quote for U.S. dollars against British pounds (GBP).)

Fonmeister: *"I deal 40-42, one by two."*
(Fonmeister Brokerage has quotes to buy £1,000,000 at 1.7440 USD/GBP, and to sell £2,000,000 at 1.7442 USD/GBP)

Mongobank: *"I sell one at 40, to whom?"*
(Mongobank hits the bid for the quantity stated. Mongobank could have requested a different amount, which would have required additional confirmation from the bidding bank.)

Fonmeister: [A pause while the deal is reported to and confirmed by Loans 'n Things] *"Loans 'n Things London."*
(Fonmeister confirms the deal and reports the counterparty to Mongobank. Payment arrangements will be made and confirmed separately by the respective back offices. The broker's back office will also confirm the trade with the banks.)

Value dates and payment arrangements are the same as in the direct dealing case. In addition to the payment to the counterparty bank, the banks involved share the brokerage fee. These fees are negotiable in the United States. They are also quite low: roughly $20 per million dollars transacted.[10]

[6]The Society for Worldwide Interbank Financial Telecommunication (SWIFT) is an electronic message network. In this case, it conveys a standardized payment order to a German branch or correspondent bank, which, in turn, effects the payment as a local interbank transfer in Frankfurt.

[7]The Clearing House for Interbank Payments System (CHIPS) is a private interbank payments system in New York City.

[8]See Brooks (1985), p. 25.

[9]See Schwartz (1988), p. 239.

[10]See Burnham (1991), p. 141, note 16, and Kubarych (1983), p. 14.

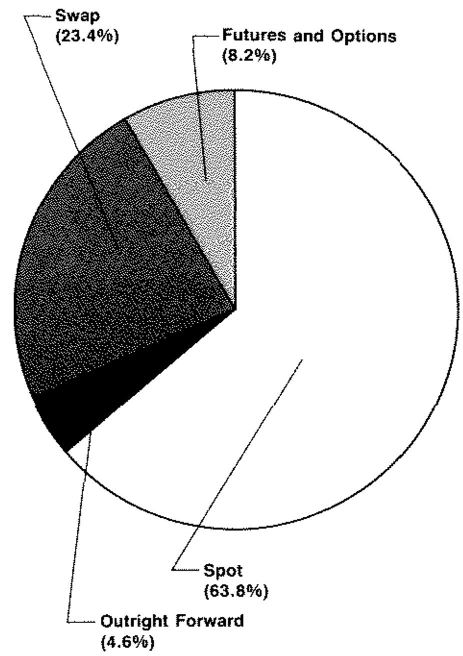
The final category of participants in the foreign exchange market is the corporate customers of the market-making banks. Customers deal only with the market-makers. They never go through brokers, who cannot adequately monitor their creditworthiness. Typically, a customer transacts with a bank with which it already has a well-established relationship, so that corporate creditworthiness is not a concern for the bank's foreign exchange desk, and trustworthiness is not an issue for the customer. The mechanics of customer trading are similar to those of direct dealing between market-makers. A customer requests a quote, and the bank makes a two-way market; the customer then decides to buy, sell or pass. The chief difference between this and an interbank relationship is that the customer is not expected ever to reciprocate by making a market.

Participants in the foreign exchange market also deal for future value dates. Such dealing composes the forward markets. Active forward markets exist for a few heavily traded currencies and for several time intervals corresponding to actively dealt maturities in the money market. Markets can also be requested and made for other maturities, however. Since the foreign exchange market is unregulated, standard contract specifications are matters of tradition and convenience, and they can be modified by the transacting agents.

Forward transactions generally occur in two different ways: outright and swap. An outright forward transaction is what the name implies, a contract for an exchange of currencies at some future value date. "Outrights" generally occur only between market-making banks and their commercial clients. The interbank market for outrights is very small, because outright trading implies an exchange rate risk until maturity of the contract. When outrights are concluded for a commercial client, they are usually hedged immediately by swapping the forward position to spot. This removes the exchange rate risk and leaves only interest rate risk.

A swap is simply a combination of two simultaneous trades: an outright forward contract and an opposing spot deal. For example, a bank might "swap in" six-month yen by simultaneously buying spot yen and selling six-month forward yen. Such a swap might be used to hedge an outright purchase of six-month yen from a bank customer.[11] In effect, the swapping bank is borrowing yen for the six months of the outright deal. The foreign exchange market-maker swaps in yen — rather than simply borrow yen on a time deposit — because banks maintain separate foreign exchange and money market accounts for administrative reasons. Swapping is generally the preferred means of forward dealing (see figures 2 and 3).

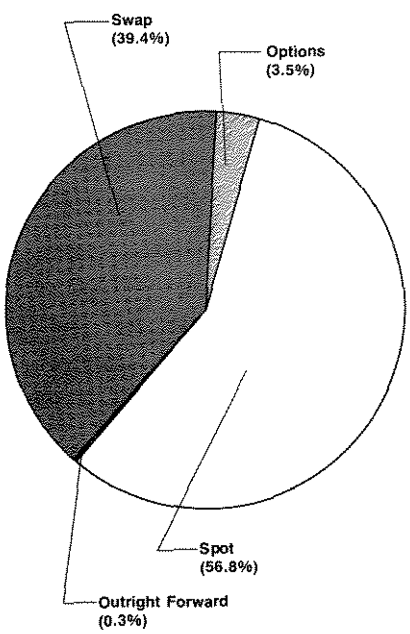
Figure 2
Market-Maker Volume by Type (4/89)



In practice, the vast majority of foreign exchange transactions involve the U.S. dollar and some other currency. The magnitude of U.S. foreign trade and investment flows implies that, for almost any other currency, the bilateral dollar exchange markets will have the largest volume. Consequently, the dollar markets are the most liquid. The possibility of triangular arbitrage enforces the law of one price for the cross rates. The upshot is that liquidity considerations outweigh transaction costs. A German wanting

[11]Hedging an outright purchase of currency with an opposing swap deal still leaves an open spot purchase of the currency. This can be easily covered in the spot market.

Figure 3
Broker Volume by Type (4/89)



pounds, for example, will typically convert marks to dollars and then dollars to pounds, rather than trading marks for pounds directly. Though this is especially true in the American market, it holds for foreign markets as well.

CLASSIFYING MARKETS

The microstructure literature is by nature market-specific, and much of it concerns U. S. equity markets. This specificity has the advantage of realism, but it makes the immediate applicability of some microstructural models to the foreign exchange market questionable. The first task is to define some basic microstructural concepts, identifying where the foreign exchange market fits into the context they provide. Such a taxonomy is important, because one of the fundamental lessons of the microstructure literature is that institutional differences can affect the efficiency of pricing and allocation.

As described above, the foreign exchange market combines two disparate auction structures for the same commodity: the interbank direct market and the brokered market. Defying a naive application of institutional Darwinism, whereby only the fitter of the two systems would survive, these trading methods appear to coexist comfortably.[12] The direct market can be classified as a decentralized, continuous, open-bid, double-auction market. The brokered market is a quasi-centralized, continuous, limit-book, single-auction market. The meanings of these classifications are explained below.

Centralization

In a centralized market, "trades are carried out at publicly announced prices and all traders have access to the same trading opportunities." In a decentralized market, in contrast, "prices are quoted and transactions are concluded in private meetings among agents."[13] A New York Stock Exchange's (NYSE) specialist system is a centralized market; the interbank direct market for foreign exchange is a decentralized one.

The distinction between centralized and decentralized markets might seem to provide a neat dichotomy of possible market structures. The multiplicity of brokers in the foreign exchange market violates this simple taxonomy, however. Each foreign exchange broker accumulates a subset of market-makers' limit orders. This network of "brokerage nodes" is as different from a fully centralized system as it is from a fully decentralized one. This arrangement is labeled here as "quasi-centralized."

Most microstructural studies have confined themselves to centralized markets, especially the NYSE's specialist system and the National Association of Securities Dealers Automated Quotation (NASDAQ) System on the over-the-counter (OTC) market.[14] Although there are a number of important decentralized markets, including the interbank direct foreign exchange market, rela-

[12]A similar situation obtains on the New York Stock Exchange, where specialists act as either brokers or market-makers, depending on the level of activity in the market.

[13]See Wolinsky (1990), p. 1. He goes on to analyze theoretically the difference in the price discovery process between centralized and decentralized markets. Schwartz (1988), pp. 426-35, refers to centralization as "spatial consolidation."

[14]For models of specialist systems, see Demsetz (1968), Tinic (1972), Garman (1976), Bradfield (1979), Amihud and Mendelson (1980), Conroy and Winkler (1981), Glosten and Milgrom (1985) and Sirri (1989). For studies of the OTC market, see Tinic and West (1972), Benston and Hagerman (1974), Ho and Macris (1985) and Stoll (1989).

tively few studies have focused on the impact of decentralization.

There is some evidence that differences in the degree of centralization between various markets cause differences in market performance. Garbade, in studying the largely decentralized Treasury securities market, concludes that because brokerage tends to centralize trading and price information, it "uses time more efficiently," "eliminates the most important arbitrages," and benefits dealers by ensuring that orders are executed according to price priority.[15]

The efficiency gains of centralized price information may imply economies of scale and, thus, a natural monopoly for brokers in securities markets. This is entirely consistent with the textbook presentation of the relatively greater operational efficiency of centralized markets.[16] Thus, the fact that a number of brokers service the foreign exchange market seems to represent a discrepancy between theory and reality. Brokers do communicate among themselves, however, to eliminate the possibility of arbitrage between limit order books. While this helps explain the multiplicity of brokers, it does not fully resolve the issue of decentralization in the interbank direct market.

Temporal Consolidation

The distinction between a continuous market and a call market involves what Schwartz refers to as the degree of "temporal consolidation."[17] In a call market, trading occurs at pre-appointed times (the "calls"), with arriving transaction orders detained until the next call for execution. In continuous markets, like the foreign exchange market, trading occurs at its own pace, and transaction orders are processed as they arrive. A range of intermediate arrangements falls between these two extremes.

Most microeconomic models assume call markets. In a Walrasian *tâtonnement* model, for example, an auctioneer calls out a series of prices and receives buy and sell orders at each price. When a price is found for which the quantities supplied and demanded are equal, all transactions are consummated at that price. Interestingly enough, Walras based this price discovery model on the mechanics of the Paris Bourse.

Temporal consolidation can affect the performance of a market. Theoretical work indicates how continuous trading can alter allocations, the process of price discovery and even the ultimate equilibrium price.[18] The basic thrust of these arguments is that, with continuous trading, earlier transactions satisfy some consumers and producers, causing shifts in supply and demand that affect prices for later transactions. As a result, the Pareto-efficiency characteristic of Walrasian equilibria does not necessarily obtain in continuous markets.[19]

On the other hand, the periodic batching of orders that occurs in a call market also has disadvantages. The difference in time between order placement and execution can impose real costs on investors. A recurring argument in the literature is the willingness of investors to pay more — a liquidity premium — for the ability to trade immediately. Similarly, periodic calls delay any information conveyed by prices until the time of the call, introducing price uncertainty in the period between the calls.

In sum, a trade-off exists between the allocational efficiency of the nearly Walrasian call market system and the informational efficiency and immediacy of the continuous market system.[20] It is not clear whether the microstructure of the foreign exchange market represents a globally optimal balance of these relative ad-

[15]See Garbade (1978), p. 497.

[16]The textbook argument counts trips to market. Briefly, if there are N traders, then a total of N trips to a central marketplace are required for each to haggle with everyone else; to pair them bilaterally requires a total of N(N-1)/2 trips. If trips are costly, then centralization is more efficient.

[17]See Schwartz (1988), pp. 435-47. Garman (1976), pp. 257-58, also describes continuous and call markets; he refers to these as asynchronous and synchronous markets, respectively.

[18]See Hahn (1984), Negishi (1962), Beja and Hakansson (1977), as well as the references therein.

[19]A continuous market cannot be viewed as a continuum of infinitesimally lived call markets. Clearing supply and demand in each such call market would require an infinite trading volume over the course of a day. Cohen and Schwartz (1989) recommend an electronic order-routing system for the stock exchanges, to facilitate the placement and revision of orders. This would encourage additional trading volume, making more frequent calls feasible.

[20]See Stoll (1985), p. 72, and especially Schwartz (1988), pp. 442-53, for a more thorough exposition of the pros and cons of temporal consolidation. Intermediate arrangements are also possible. For example, Schwartz argues that many of the problems caused by infrequent batching in a call market might be overcome by expanding access to the market with computer technology, whereby the increased number of traders would allow for more frequent calls.

vantages. A persistent deviation from optimality might be explained, for example, by arguing that the allocational benefits of a call market system are a public good.

Communication of Prices

The terms "open-bid" and "limit-book" refer to ways in which price information is communicated. In an open-bid market — the open outcry system on the futures exchanges, for example — offers to buy or sell at a specified price are announced to all agents in the market. At the opposite extreme, in a sealed-bid market, orders are known only to the entity placing the order and perhaps to a disinterested auctioneer.

Direct trading in foreign exchange approximates the standard open-bid structure. The salient difference between the foreign exchange market and the standard arrangement is the bilateral pairing of participants in the foreign exchange market. In principle, any participant can contact a market-maker at any time for a price quote. The bilateral nature of such contacts and the time consumed by each contact together imply, however, that all participants cannot be simultaneously informed of the current quotes of a market-maker. This practical constraint on the dissemination of price information is significant: it introduces the possibility of genuine arbitrage, that is, of finding two market-makers whose current bid-ask spreads do not overlap.

The limit order book, which is used by both foreign exchange brokers and stock exchange specialists, is another intermediate form of price communication. Although it would be possible in principle for foreign exchange brokerage books to be fully open for public inspection, in practice only certain orders — namely, the best bid and ask on each book — are revealed to market-makers, while the others remain concealed. As in the direct market, market-makers must contact brokers bilaterally to get these "inside spreads." Knowledge of the concealed limit orders would be of speculative value to market-makers, because an imbalanced book suggests that large future price movements are more likely in one direction than the other.

More generally, price communication is intimately related to the role of market-makers as providers of "predictable immediacy."[21] Market participants are willing to pay a liquidity premium, usually embedded in a market-maker's spread, for the reduction in search costs implied by constant access to a counterparty. The costs of "finding" the other side of a transaction can be further broken down into the liquidity concession, the cost of communicating the information and the cost of waiting for potential counterparties to respond.[22] Other things equal, an efficient system of price communication is one that minimizes such transaction costs. While the communication of price information is a central function of securities markets, the fact that the systems of price communication in the foreign exchange market are not fully centralized suggests that these systems do not represent a cost-minimizing arrangement.

Structure of Prices

The terms "double-auction" and "single-auction" refer to the nature of the prices quoted. In a double-auction market, certain participants provide prices on both sides of the market, that is, both bid and ask prices. Participants providing double-auction quotes upon demand are known as market-makers, and they must have sufficient capitalization to back up their quotes. In a single-auction market, prices are specified either to buy or to sell, but not both. In the foreign exchange market, market-makers provide double-auction prices, while brokers try to aggregate single-auction quotes into two-way (inside) spreads. A broker's book may occasionally be empty on one or both sides. Rather than make a market in such cases, the broker provides, respectively, a single-auction quote or none at all.

Thus, whether double or single-auction prices are quoted depends largely on whether the agent quoting prices is providing market-making services or simply attempting to acquire (or sell) the commodity. This issue is related to the degree of centralization in the market. The absence of market-makers in a single-auction market, together with the presence of search costs, results in a tendency toward centralization of price information, thus facilitating the search for a counterparty. Inversely, decentralization of price information leads to a tendency

[21] This term is due to Demsetz (1968), p. 35. Tinic (1972), p. 79, calls in "liquidity services."

[22] See Logue (1975), p. 118.

toward double-auction prices, again to facilitate the search for a counterparty.[23]

MODELING TRADERS' BEHAVIOR

The microstructure literature extends well beyond a simple description of market institutions. Modeling the behavior of market participants is central to almost all discussions of microstructure. Although numerous approaches to such modeling have been taken, two common concerns are of special interest. These are the treatment of price information by market participants, and determination of the bid-ask spread. The latter raises the interrelated issues of inventory and quantity transacted.

Price Expectations

Modeling the interpretation of price information is a crucial step in constructing microstructural models of price discovery.[24] Many diverse approaches have been taken in such modeling. An almost universal simplification is to model securities markets in partial equilibrium, so that prices are not determined endogeneously in the traditional general equilibrium sense. This allows the modeler to focus on the microstructure's finer details. Another common simplification is to assume that agents ignore the impact of their own behavior on the market.[25]

Rather than explicitly model such forces as general equilibrium or recursive beliefs, models posit probability distributions that produce the prices of orders in the market. Modelers have included randomness at one or both of two levels, depending on their focus. First, order prices can be generated by objective distributions, that is, by stochastic processes exogenous to the market. For example, there may be a stochastic process that generates the "true" equilibrium price. Second, probability models of participants' subjective beliefs about prices can be used. Conroy and Winkler, for example, attribute subjective normal price distributions to market-makers, who use Bayesian updating to learn about the prices of incoming limit orders.[26] Objective processes can coexist with subjective beliefs about those processes. Harsanyi suggests a consistency requirement for the subjective price distributions of multiple agents; these distributions are each equated with a conditional distribution of a single distribution known to all.[27]

Models can be further classified according to how they relate supply and demand. In particular, there are both models with single price processes and with dual price processes. In dual price models, purchase orders (whether market or limit orders) are generated by one process, while sale orders are generated by another.[28] The salient point here is that purchase and sale orders come from independent distributions. This independence is especially clear in Conroy and Winkler, where the distributional assumptions are explicit; there, independence implies that any sequence of buy orders, regardless of their prices and quantities, has no effect on the subjective probability of a sell order at any price.

Statistical independence implicitly restricts the ways in which orders can be generated. Purchase and sale orders are somehow motivated independently, although the cause of this separation is not always specified. Statistical independence is not a necessary component of a dual price process, however. Cohen, Maier, Schwartz and Whitcomb (1981), for example, assume that actual market bid and ask prices are

[23]Note that the converse does not appear to hold. That is, centralization does not tend to eliminate double-auction quoting. For example, the NASDAQ system on the OTC stock market centralizes price information while still supporting numerous market-makers for every stock.

[24]Notably, the term "price" is generally too inexact in a microstructural context. One must often distinguish at a minimum between quoted prices, transaction prices and equilibrium prices. There are also reservation prices, market-clearing prices and closing prices (see Schwartz (1988), chapter 9, for the distinction between equilibrium and clearing prices). If unspecified here, the intended definition should be clear from the context.

[25]The alternative, which dates at least to Keynes' "beauty contest," is recursive beliefs, in which an agent considers the feedback of her own actions on the beliefs of others, and thence how the behavior on the other agents might affect her own beliefs, etc. See Keynes (1936), p. 156. The limiting case—an infinite recursion of beliefs—presumes extreme informational and computational resources on the part of agents, and models based on it are usually intractable. Intermediate approaches allowing a finite degree of recursion must somehow justify the truncation of recursive beliefs, just as the standard model of atomistic agents allows no beliefs about beliefs and is justified by an assumption on the relative size of individual agents.

[26]See shaded insert on opposite page.

[27]See Harsanyi (1982), especially chapter 9, and the references therein. His consistency requirement identifies a unique equilibrium for the game.

[28]A market order is an order to trade at the best price available; a limit order specifies a price. These models represent a strain of the literature that was pioneered by Demsetz using straightforward supply and demand schedules (see shaded insert on page 63). Similar approaches were later taken by Garman (1976), Amihud and Mendelson (1980) and Conroy and Winkler (1981), among others.

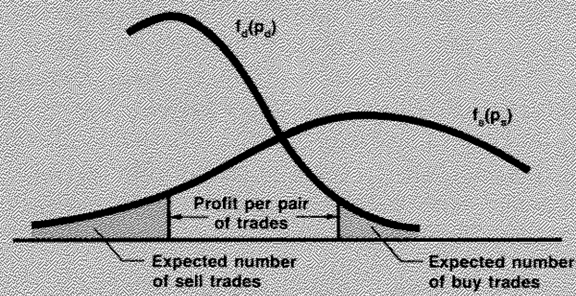
Bayesian Learning of Price Information

Conroy and Winkler (1981) developed a Bayesian model of market-maker price expectations, which is outlined here. Consider an expected-profit-maximizing, monopolistic market-maker who faces streams of buy and sell limit orders from investors.[1] All orders are for a single round lot. Assume that the market-maker believes that reservation prices of buy orders, p_d, are generated by a normal distribution, $F_d(p_d;\mu_d,\sigma_d)$; reservation prices of sell orders, p_s, are generated by a second, independent, normal distribution, $F_s(p_s;\mu_s,\sigma_s)$. That is, the market-maker has two independent, normal, subjective price distributions (with corresponding densities f_d and f_s). Further assume that the market-maker currently holds his desired inventory level. How should he set his spread?

The inventory condition implies that the chosen bid and ask rates, B and A, must satisfy the constraint, $F_s(B) = 1 - F_d(A)$, so that the expected change in inventory is zero.[2] Given this constraint, the expected profit per period is: $E(\pi) = (A-B)\,F_s(B) = (A-B)\,(1-F_d(A))$. Maximizing this over B and A yields: $B^* = M - \sigma_s\Phi(B')/\phi(B')$ and $A^* = M + \sigma_d\Phi(A')/\phi(A')$, where $M = (\sigma_d\mu_s + \sigma_s\mu_d)/(\sigma_d + \sigma_s)$, $B' = (B-\mu_s)/\sigma_s$, $A' = (A-\mu_d)/\sigma_d$, and $\phi(\cdot)$ and $\Phi(\cdot)$ are the standard normal density and distribution functions. It can be shown that this optimal spread shrinks (i.e., $A^* - B^*$ decreases), *ceteris paribus*, as the subjective variances, σ_s and σ_d decrease.

The important aspect of this study is that it provides an explicit mathematical model for a market-maker's interpretation of price information. The market-maker is assumed to behave in a Bayesian fashion, using the observed prices on incoming limit orders to refine the parameters of his subjective distributions. For example, assume that the market-maker views purchase prices as coming from a normal distribution $g_d(p_d;\mu_d,\sigma_d)$, but is unsure about the mean of this distribution. Represent this uncertainty by a normal prior density $h'(\mu_d;m',v')$ over the possible values for the mean, μ_d. Given this, the marginal subjective density over the prices of incoming limit orders, $f_d(p_d) = \int g_d(p_d)\cdot h'(\mu_d)$, is normal with mean m' and variance $(\sigma_d^2 + v'^2)$. Following a sample of n buy orders with mean price m, the market-maker is able to refine his subjective distribution of the mean. The posterior parameters of $h''(\mu_d;m'',v'')$ are $m'' = (m'/v'^2 + nm/\sigma_d^2)/(1/v'^2 + n/\sigma_d^2)$ and $v'' = 1/(1/v'^2 + n/\sigma_d^2)$. The upshot of this refined estimate is that the variance of the marginal subjective price density, $f_d(p)$, is now smaller, and the market-maker's optimal spread, (B^*, A^*), is narrower.

Profit Per Pair of Trades and Expected Number of Trades Per Unit Time



[1]Conroy and Winkler (1981) also consider a risk-averse market-maker and the information conveyed by a market order, which does not specify a price. They do not incorporate the impact of inventory on pricing, nor do they generalize beyond the unrealistic assumption of normally distributed prices.

[2]This is depicted in the figure above, where price is on the horizontal axis, and the relative frequency of orders is on the vertical axis. The market-maker's spread is the interval from B to A, and the inventory constraint is satisfied when the two shaded tails have equal area. Their optimization problem is similar in spirit to that of Allen (1977), although the latter does not consider learning.

independent Poisson processes and give investors joint subjective distributions over those prices. For the latter distributions, probabilistic independence of bid and ask prices is not explicitly required. Black (1989) models quantities (independent of prices) of market orders. Quantities supplied and demanded are drawn from different distributions, but the distributions are constrained to have the same mean. Garbade (1978), on the other hand, assumes a single, unknown and fixed equilibrium price, around which market-makers set their spreads. Incoming buy and sell orders arrive via random processes whose mean arrival rates depend on the difference between the quoted bid (or ask) price and the exogenous equilibrium price and, thus, are not independent.

The most common alternative to separate purchase and sale processes is to model prices as some function of a single scalar process. This approach is in the spirit of the efficient markets literature, which posits a unique value for a security conditional on the available information. Ross (1987) points out that this approach can be regarded conceptually as a special case of the dual price process, with supply and demand infinitely elastic at a common price. Many authors reveal their theoretical roots by using terminology drawn from the literature on efficient markets. Thus, for example, Barnea describes a stock's "intrinsic value," which follows a random walk.[29] Similarly, Copeland and Galai posit a "'true' underlying asset value ... known (ex ante) to all market participants."[30] In contrast, Garbade's (1978) exogenous equilibrium price is unknown.

It is possible to extend the single price approach beyond the efficient markets tradition by modeling the value of a security subjectively rather than as an objective fact. Glosten and Milgrom (1985), for example, begin with an exogenous random value representing the consensus value of a stock given all public information. Investors do not act on this exogenous value directly; instead, they act on their expectation of it, conditional on their information set. Ho and Stoll personalize price expectations in a similar fashion:[31]

> We take the dealer's opinion of the "true" price of the stock to be exogenously determined by his information set and ask how the dealer prices relative to his "true" price...

This subjectivization of the pricing process is significant, because it allows for heterogeneous expectations and thus for more realistic modeling of price discovery.

Research into the microstructure of the foreign exchange market should presume such heterogeneity among market-makers. There are numerous market-makers for foreign exchange: The Federal Reserve Bank of New York (FRB-NY) (1989a) lists 162 dealing institutions in the U.S. interbank market. There would be little point in such superfluity if all market-makers were identical. Furthermore, it is well known that "taking a view," that is, speculating on future prices, is routine for many participants.[32] To omit this heterogeneity from a model is to ignore an important characteristic of the market.

The large proportion of market-makers in the foreign exchange market has another important modeling implication. It implies that a single-price process is more appropriate as a theoretical representation of agents' expectations. Market-makers consistently face other market-makers, who can hold positive or negative inventories of foreign currency with equal ease. A quote that is "off the market" on the high side will be hit (i.e., traded upon) just as surely as a quote that is off on the low side. This is also true of customers, who normally enter the market with a predilection to either buy or sell. As Burnham notes:[33]

> The customer knows that if the first marketmaker is too far off the market price, he can unexpectedly take the other side of the quote and resell the position to a second marketmaker.

The point is that the market-maker must expect to be penalized for underestimating as well as overestimating his counterparty's valuation of the currency. From the perspective of the market-maker, who quotes a spread and observes a response, the forces determining short-run effective demand and supply are not merely related, but indistinguishable.

[29]See Barnea (1974), pp. 512-14.

[30]See Copeland and Galai (1983), p. 1458.

[31]See Ho and Stoll (1981), p. 48. For a similar example, see Stoll (1978), especially p. 1136.

[32]See, for example, Kubarych (1983), p. 29, or Burnham (1991), p. 139.

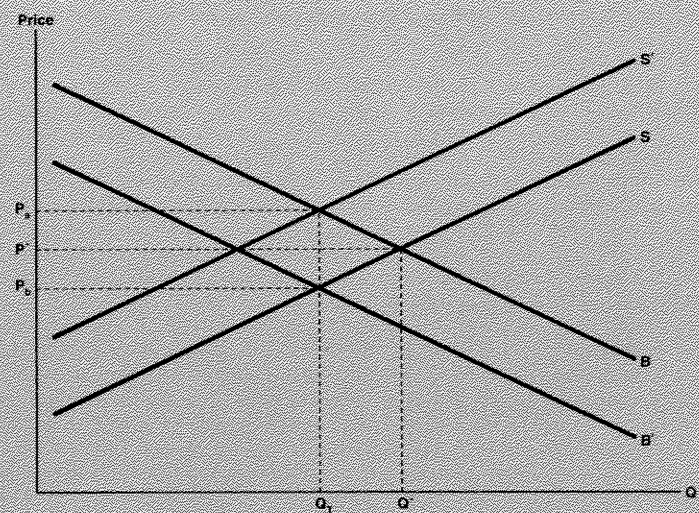
[33]See Burnham (1991), p. 136.

Dealer Services and the Bid-Ask Spread

Traditional wisdom refers to the bid-ask spread as the "jobber's turn," suggesting that it provides compensation to the dealer for the provision of services.[1] Demsetz (1968) formalized this rationale for the spread, defining the particular service provided as "predictable immediacy" and offering a simple model to describe the spread.

Consider a continuous market with aggregate supply (sell) and demand (buy) schedules, S and B, for a security (see figure at right). In an idealized world, investors would come together simultaneously, and the market would clear at price P^* and quantity Q^*. In this marketplace, however, such coordination of trading is impossible. By assumption, the market is continuous, and there is no mechanism (e.g., a limit order book) for holding orders over time. Thus, S and B do not represent standard static supply and demand schedules, but rather time rates of supply and demand. At any given instant, there may be no orders on either side.

Market Buy and Sell Curves with and without the Provision of Immediacy



Instead, we introduce a monopolistic market-maker who allows the trading to occur by standing in as a counterparty to all trades. In the process, he provides a service to investors that Demsetz labels "predictable immediacy." The market-maker knows the aggregate supply and demand propensities. The supply and demand curves that he presents to the public, S′ and B′, however, are both shifted leftward.

Investor purchases clear for price P_a, at the intersection of the market demand schedule, B, with the market-maker's supply schedule, S′. Similarly, investor sales clear at the intersection of S and B′, for price P_b. The differences: $P_a - P^*$ and $P^* - P_b$ are liquidity premia. In the figure, the quantities, Q_T, purchased and sold by the market-maker happen to be equal, so that no market-maker inventory is accumulated. His profit thus equals $Q_T(P_a - P_b)$; this is the "jobber's turn."

[1]See, for example, Keynes (1936), p. 158, or Stigler (1964), p. 129.

A market-maker's constant contact with other well-capitalized market-makers implies that this is not a theoretical fine point. In the words of one market-maker:[34]

> Ninety percent of what we do is based on perception. It doesn't matter if that perception is right or wrong or real. It only matters that other people in the market believe it. I may know it's crazy. I may think it's wrong. But I lose my shirt by ignoring it.

In other words, as a direct implication of their readiness to buy or sell, market-makers must strive first to achieve a price consensus. The im-

[34]James Hohorst, as quoted by Mossberg (1988), p. 29R. Mr. Hohorst directed foreign exchange trading in North America for Manufacturers Hanover.

perative of arbitrage avoidance must be regarded as the first priority in individual market-maker pricing, to which all other factors (e.g., purchasing power parity) must be subordinated.

Market-makers' Bid-Ask Spreads

The bid-ask spread has attracted considerable interest in the literature on market microstructure. The complexity of modeling the spread is largely because it requires incorporating a substantial amount of institutional detail. At a facile theoretical level, a market-maker's spread appears to be a direct violation of the law of one price, since it assigns two prices to the same commodity. Several explanations have been offered to resolve this seeming inconsistency. They can be roughly categorized as involving the cost of dealer services, the cost of adverse selection and the cost of holding inventory.[35]

The dealer services argument can be traced back at least as far as Stigler (1964), who argues that stock exchange specialists charge a "jobber's turn" as compensation for the costs of acting as a specialist. The analysis of dealer services was formalized by Demsetz (1968), who identified "predictable immediacy" as the particular service for which investors are willing to pay. This identification hints at the complex question of what liquidity is and where it comes from. In a busy market, liquidity is a public good: a continuous stream of buyers and sellers generates predictable immediacy as a by-product of their trading.

The determinants of the level of compensation are themselves a topic of debate. Stigler argues that, because centralization of exchange limits fixed costs and aggregates separate transaction orders into less risky actuarial order flows, it implies economies of scale and thus a natural monopoly for market-making.[36] Smidt (1971) counters that barriers to entry among NYSE specialists allow them to exact monopoly rents from other investors. In his view, the natural monopoly argument, while used as an apology for barriers to entry, remains unsupported empirically: "There is no empirical evidence to support the proposition that [market-making] is, in fact, a natural monopoly."[37] Indeed, if market-making is a natural monopoly, barriers to entry should be unnecessary.

The foreign exchange market has no apparent barriers to entry other than the need for sufficient capitalization. It also has no apparent barriers to exit. The market supports a large and increasing number of competing market-makers. Unless it can be shown that there is some subtle restriction in the foreign exchange market that prevents consolidation of the market-making function, one must conclude that market-making per se is not a natural monopoly.[38] The multitude of market-makers also implies that they cannot earn monopoly rents by embedding a premium for predictable immediacy in the spread, although the spread may still cover the costs of processing orders.

Other research suggests that a market-maker's job is more complex than the mere sale of counterparty services. A second explanation for the bid-ask spread — adverse selection — can be traced to Bagehot (1971). He starts with "liquidity-motivated transactors" who pay the market-maker the price of the spread in exchange for the service of predictable immediacy. The market-maker also confronts traders who have inside information, however, and who can therefore speculate profitably at the expense of the market-maker.[39] The market-maker must charge everyone a wider spread to compensate for losses to the information-motivated traders.

Because of the relatively abstract nature of currencies as commodities, it is difficult to construct examples of "inside" information on foreign exchange rates. One exception is money supply announcements, which, if known before

[35]This is essentially the same taxonomy as provided by Barnea and Logue (1975), although they use the terms "liquidity theory," "adversary theory," and "dynamic price/inventory adjustment theory," respectively.

[36]See Stigler (1964), p. 129.

[37]See Smidt (1971), p. 64.

[38]For example, in the context of the OTC stock market, Benston and Hagerman (1974), p. 362, conjecture that, "dealers may face positively sloped marginal cost curves which shift down as industry output increases." The idea is that market-making per se is not a natural monopoly, even though the industry as a whole experiences economies of scale. Hamilton (1976) also addresses the natural monopoly question; Reinganum (1990) provides evidence on liquidity premia for NYSE vs. NASDAQ stocks.

[39]This situation is called adverse selection, because, in a market with competing market-makers, the one who gets the insider's business is a loser rather than a winner. Bagehot also posits a third class of investors, who only think they have inside information; they speculate, but lose on average, and are indistinguishable to the market-maker from the liquidity-motivated traders.

publicly distributed, might provide a basis for profitable speculation. Another form of information that can be construed as inside information is knowledge of an arbitrage opportunity. Consider a hypothetical market in which there are numerous decentralized market-makers who do not quote spreads, but single prices at which they are willing both to buy and sell. Unless there were a perfect consensus among the market-makers on the value of the foreign currency, all of them would be vulnerable to arbitrage. A decentralized market makes a perfect consensus difficult to achieve. Without centralizing price information, it is impossible to know if no arbitrage opportunities exist. A bid-ask spread, in contrast, allows a market-maker to include an error tolerance in her prices, thus facilitating a price consensus: it is easier to get bid-ask spreads to overlap than to get scalar prices to coincide. The spread also provides the market-maker with some degree of protection from adverse selection in the form of arbitrage.

The bid-ask spread is also affected by inventory considerations. This idea dates back at least as far as Barnea and Logue (1975).[40] The notion of a desired inventory level for the market-maker underlies all of these models. In the simplest case, the desired level is set at zero, and a constant spread is shifted up and down on a price scale to equalize the probability of receiving a purchase order with that of receiving a sale order. The result is that the expected change in inventory is always equal to zero, and (with all trades for one round lot) the inventory level follows a simple random walk.

An undesirable implication of random-walk models of inventory is the inevitable bankruptcy of the market-maker. Finite capitalization levels for market-makers impose upper and lower bounds on allowable inventories. Because inventory follows a random walk, with probability one it will reach either its upper or lower bound in a finite number of trades.[41] The dynamic optimization models of Bradfield (1979), Amihud and Mendelson (1980) and Ho and Stoll (1981) resolve this problem. They conclude that a market-maker, optimizing his bid and ask prices over time in the face of a stochastic order flow, will shift both bid and ask rates downward (upward) and increase the width of the spread when a positive (negative) inventory has accumulated.[42]

We should expect two of these three rationales for the spread to apply to market-makers' bid-ask spreads in the foreign exchange market. Because there are numerous market-makers, competition should eliminate their ability to earn monopoly rents by charging a premium for predictable immediacy per se. The adverse selection argument does apply in the foreign exchange market, however, since the spread allows market-makers some protection against arbitrage opportunities. Arbitrage opportunities can be construed as a form of inside information in a market where price information is not centralized. In accordance with the dynamic optimization models, a market-maker's inventory level should affect the spread, widening and shifting it as inventories accumulate.

Brokers' Spreads

So far, the discussion of the bid-ask spread has focused on models in which bid and ask prices are set by individual market-makers. The dual role of the stock exchange specialist suggests that this is only part of the story. Spreads are produced in two fundamentally different ways. It is only when limit orders are sparse that a NYSE specialist must step in as a market-maker to provide an "orderly market."[43] When limit order volume is sufficient, the specialist acts as a broker, accounting for incoming limit orders on the limit order book, and pairing market orders against them. Cohen, Maier, Schwartz and Whitcomb (1979) note that inadequate attention has been given to the fact that not all prices are market-maker spreads. The market often makes itself without specialist assistance, through the aggregation of limit orders on the book.

The foreign exchange market differs from the NYSE in that the market-making and brokerage roles are separated: market-makers do not act as brokers, and brokers do not make markets.

[40]Barnea and Logue attribute it to Smidt (1971), although Smidt's paper does not explicitly develop the connection between the market-maker's inventory and his spread. Formal models of the relationship between inventories and spreads can be found in Stoll (1978), Amihud and Mendelson (1980), Ho and Stoll (1981) and Sirri (1989), among others.

[41]See, for example, Ross (1983), pp. 106-07.

[42]See shaded insert on page 66.

[43]The NYSE defines this role in rule 104: "the specialist should maintain a continuous market with price continuity and close bid and asked prices, and minimize the effect of temporary disparity between public supply and demand." See Leffler and Farwell (1963), pp. 211-12.

Dynamic Price-Inventory Adjustment Models

Amihud and Mendelson (1980, 1982) provide a model of market-maker spread-setting that takes inventory into account. Assume that a market-maker faces order flows of buy and sell market orders that arrive according to independent Poisson processes. The buy and sell arrival rates (i.e., process intensities), d and s, respectively, depend on the ask and bid prices, P_a and P_b, that the market-maker quotes: $d = D(P_a)$ and $s = S(P_b)$. Denote the inventory level by $k \in \{-\Sigma,...,\Lambda\}$, where Σ and Λ are the largest allowable short and long positions, respectively. Let d_k and s_k denote the order arrival rates when prices are set as functions of the inventory level: $d_k = D(P_a(k))$ and $s_k = S(P_b(k))$.

The expected sojourn at k (i.e., the time until an order arrives) is given by the Poisson processes known as $1/(d_k+s_k)$. The probability that the next order will be a buy order is $d_k/(d_k+s_k)$, and the probability that it will be a sell order is $s_k/(d_k+s_k)$. Thus, the expected cash flow per unit time at position k is given by:

$$Q(k) = \left[\frac{d_k}{d_k+s_k}\cdot P_a(k) - \frac{s_k}{d_k+s_k}\cdot P_b(k)\right]\cdot(d_k+s_k)$$

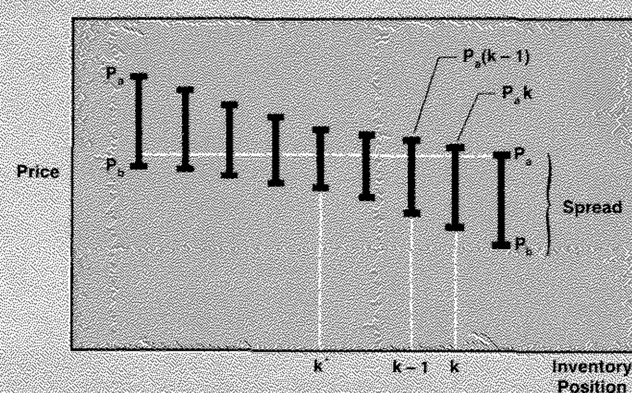
$$= d_k \cdot P_a(k) - s_k \cdot P_b(k).$$

The market-maker's objective is to maximize the expected profit per unit time, given by:

$$\pi = \sum_{k=-\Sigma}^{\Lambda} \Phi_k Q(k).$$

Where Φ is the probability of being at inventory level k. The solution to this optimization problem gives the values for $P_a(k)$ and $P_b(k)$, which are depicted in the figure below. The market-maker controls inventory by adjusting prices up (down) to make an investor sale (purchase) more likely when inventory is low (high). The spread must widen as the inventory nears its bounds, in order to avoid the problem of a random walk for inventory; at the extremes, $d_{-\Sigma} = s_\Lambda = 0$.

Optimal Prices and Bid-Ask Spread as a Function of Inventory Position



Therefore, it is even more appropriate to model brokered spreads as determined in a fundamentally different way from market-maker spreads. The separation of roles also has other implications for modeling foreign exchange brokerage.

A brokered spread is the combination of the best bid and best ask, received by the broker as separate limit orders. This arrangement might be modeled as a pair of extreme order statistics from independent distributions of purchase and sale limit orders. The distribution of these statistics would have to be conditional on limit order volume and on the fact that the best ask must always exceed the best bid, since crossing orders transact immediately and are removed from the book.[44] Perhaps because of its complexity, such a derivation has not been attempted.

Cohen, Maier, Schwartz and Whitcomb (1979) model limit orders as generated by "yawl" distributions. These distributions satisfy three heuristics for the incentives of investors placing limit orders.[45] The heuristics are motivated by a notion of the centralized exchange as a market for immediacy; placers of limit orders produce immediacy, and placers of market orders consume it. This relationship between limit and market orders is formalized in Cohen, Maier, Schwartz and Whitcomb (1981), where each half of the brokered spread is assumed to be generated by a compound Poisson process. A minimum brokered spread results: if the limit order's bid (ask) price is sufficiently close to the specialist's ask (bid), the benefit to the investor of being able to specify the price of a limit order is overwhelmed by the cost of foregone immediacy.

Because models of the informational content of brokered spreads are few, the literature offers little guidance in modeling brokered quotes in the foreign exchange market. The yawl distribution is the only explicit distributional form for brokered spreads in the literature. Unfortunately, its heuristic basis cannot be transferred directly to the foreign exchange market, because market-makers there differ from stock market investors. Indeed, this may be an instance in which the foreign exchange market informs microstructure theory rather than the other way around. The extant approaches to brokerage treat it as a service facilitating predictable immediacy. This aspect of brokerage is redundant in the foreign exchange market, because of the multitude of market-makers, each providing immediacy. This redundancy suggests instead that foreign exchange brokerage serves some other function.

One motive for trading through a foreign exchange broker is to maintain anonymity — the name of the bank placing a limit order is not revealed unless a deal is consummated and then only to the counterparty.[46] Anonymity is valuable, because revealing a need to buy or sell a currency puts a market-maker at a bargaining disadvantage. In addition, anonymity can help pair market-makers who ordinarily would not contact each other directly. These issues have not been explored at a theoretical level. Until an adequate microstructural model of the strategic benefits of anonymity is developed, the theoretical understanding of foreign exchange brokerage will be limited.

CONCLUSIONS

Students of the foreign exchange market can draw several lessons from the literature on market microstructure. The most fundamental of these is that the institutional details of exchange in a market can affect all aspects — price, allocational, informational and operational — of the market's efficiency. A multitude of market-makers who can provide liquidity, or predictable immediacy, arises in response to the decentralization of the market. As a result, search costs are reduced relative to a world without market-makers, because finding one of many market-makers amounts to finding a counterparty. Brokerage also reduces search costs by achieving a degree of centralization in price information.

An unanswered question is why the specific combination of trading structures characteristic of the foreign exchange market — a decentralized, open-book, direct arrangement and a quasi-centralized, limit-book, brokered arrangement — should coexist. Apparently, each structure has relative advantages, but a full analysis of these advantages is lacking. Is there a single microstructure that would combine the relative advantages of the direct and brokered arrangements? Put another way, why does the microstructure of the foreign exchange market differ from that of the stock exchanges, the futures pits and the OTC stock market? Answering these questions will require a fuller specification of the objectives of a trading system and a better understanding of the impact of microstructural arrangements on those goals.

These issues provide a motive for deeper investigation of the behavior of the foreign ex-

[44] An order statistic is defined as follows: the sample realizations of a finite number of independent random variables are ranked in increasing order, and the kth order statistic is the kth number in that list. For the foreign exchange market, the modeling is still more complex, since brokers compare books amongst themselves in the sense that incoming orders can cross against any book.

[45] The yawl distribution, named for its resemblance to a sailboat, is a probability distribution contrived for modeling the generation of buy (or sell) limit orders. See Cohen, Maier, Schwartz and Whitcomb (1979, 1983, 1986) for details.

[46] See Kubarych (1983), p. 16, Burnham (1991), p. 141, and Federal Reserve Bank of New York (1989b), p. 41-3.

change market and its participants. Market-makers are the crucial element: they provide all transaction prices in the market and are involved in at least one side of every deal. The microstructure literature has developed numerous models of the interpretation and setting of prices by traders. The diversity of expectations models used in the literature illustrates the importance of tailoring such models to the specific environment confronted by market participants. Given that a foreign exchange market-maker's double-auction quote can be hit on either side (bid or ask) with equal ease, he must try to maneuver his spread to bracket the market's consensus valuation of the foreign currency. In other words, suppliers and demanders of currency are indistinguishable to the market-maker ex ante. The inability to separate the forces determining effective demand from those determining effective supply in the very short run imply that a single-price expectations process (rather than a dual-price process) is appropriate in modeling market-makers in the foreign exchange market.

A market-maker's bid-ask spread serves several purposes. Competition among market-makers in the foreign exchange market implies that they should be unable to charge a monopoly premium for the service of predictable immediacy. Instead, the spread obviates the need for perfect price consensus by giving the market-maker some protection from arbitrageurs with superior price information. While arbitrage avoidance must be considered a primary goal in setting a market-maker's bid and ask quotes, the spread provides flexibility elsewhere. Just as arbitrage avoidance is concerned with accurately estimating current prices, speculation is concerned with estimating future prices. By changing in size and shifting up or down, the spread can control stochastically the market-maker's foreign currency inventory in the face of random order flows. Systematic empirical study of the effect of inventories on market-makers' spreads is still needed, however.

The brokered spread is less well understood than the market-maker's spread, and certain areas are ripe for further research. Theoretical models of brokered spreads are few. The existing rationales for brokerage maintain that it provides liquidity services. In the foreign exchange market, however, numerous market-makers make the liquidity services provided by brokerage superfluous. Descriptions of the foreign exchange market suggest instead that anonymity is an important motive for trading in the brokered market. Yet the strategic value of anonymity in foreign exchange quoting is not well understood at a theoretical level. In addition, there is not a clear understanding of the differences in price information between a market-maker's spread and a broker's spread; this too remains a topic for future research.

From a broader perspective, a better understanding of institutional choice and change as regards securities market microstructure is necessary. Most microstructural research has been devoted to analyzing the impact of microstructural factors on important economic variables, such as price and allocation. Relatively little attention has been paid to the effect of economic factors on the choice of an institutional microstructure.

REFERENCES

Allen, Wiliam A. "A note on uncertainty, transaction costs and interest parity," *Journal of Monetary Economics* (July 1977), pp. 367-73.

Amihud, Yakov, Thomas S. Y. Ho, and Robert A. Schwartz, eds. *Market Making and the Changing Structure of the Securities Industry*, (Lexington Books, 1985).

Amihud, Yakov, and Haim Mendelson. "Dealership Market: Market-Making with Inventory," *Journal of Financial Economics* (March 1980), pp. 31-53.

________ . "Asset Price Behavior in a Dealership Market," *Financial Analysts Journal* (May/June 1982), pp. 50-59.

________ . "Asset Pricing and the Bid-Ask Spread," *Journal of Financial Economics* (December 1986), pp. 223-49.

________ . "Liquidity and Asset Prices: Financial Management Implications," *Financial Management* (Spring 1988), pp. 5-15.

Bagehot, Walter. "The Only Game in Town," *Financial Analysts Journal* (March/April 1971), pp. 12-14, 22.

________. *Bank for International Settlements*. "Survey of Foreign Exchange Market Activity," (February 1990).

Barnea, Amir. "Performance Evaluation of New York Stock Exchange Specialists," *Journal of Financial and Quantitative Analysis* (September 1974), pp. 511-35.

Barnea, Amir, and Dennis E. Logue. "The Effect of Risk on the Market Maker's Spread," *Financial Analysts Journal* (November/December 1975), pp. 45-49.

Beja, Avraham, and Nils H. Hakansson. "Dynamic Market Processes and the Rewards to Up-to-date Information," *Journal of Finance* (May 1977), pp. 291-304.

Benston, George J., and Robert L. Hagerman. "Determinants of Bid-Asked Spreads in the Over-the-Counter Market," *Journal of Financial Economics* (1974), pp. 353-64.

Black, Stanley W. "Transaction Costs and Vehicle Currencies," *IMF Working Paper* WP/89/96, International Monetary Fund (November 21, 1989).

Bossaerts, Peter, and Pierre Hillion. "Market Microstructure Effects of Government Intervention in the Foreign Exchange Market," *Review of Financial Studies* (forthcoming).

Bradfield, James. "A Formal Dynamic Model of Market Making," *Journal of Financial and Quantitative Analysis* (June 1979), pp. 275-91.

Brooks, J. A. "The dealing room and the dealer," in Rudi Weisweiller, ed., *Managing a Foreign Exchange Department* (Woodhead-Faulkner, Cambridge, 1985), pp. 18-33.

Burnham, James B. "Current Structure and Recent Developments in Foreign Exchange Markets," in Sarkis J. Khoury, ed., *Recent Developments in International Banking and Finance, Volume IV* (Elsevier Science Publishers, Amsterdam, 1991), pp. 123-53.

Chrystal, K. Alec. "A Guide to Foreign Exchange Markets," this *Review* (March 1984), pp. 5-18.

Cohen, Kalman J., Gabriel A. Hawawini, Steven F. Maier, Robert A. Schwartz, and David K. Whitcomb. "Implications of Microstructure Theory for Empirical Research on Stock Price Behavior," *Journal of Finance* (May 1980), pp. 249-57.

Cohen, Kalman J., Steven F. Maier, Walter L. Ness, Jr., Hitoshi Okuda, Robert A. Schwartz, and David K. Whitcomb. "The impact of designated market makers on security prices," *Journal of Banking and Finance* (November 1977), pp. 219-47.

Cohen, Kalman J., Steven F. Maier, Robert A. Schwartz, and David K. Whitcomb. "The Returns Generation Process, Returns Variance, and the Effect of Thinness in Securities Markets," *Journal of Finance* (March 1978a), pp. 149-67.

________. "Limit Orders, Market Structure, and the Returns Generation Process," *Journal of Finance* (June 1978b), pp. 723-36.

________. "Market Makers and the Market Spread: A Review of Recent Literature," *Journal of Financial and Quantitative Analysis* (November 1979), pp. 813-35.

________. "Transaction Costs, Order Placement Strategy, and Existence of the Bid-Ask Spread," *Journal of Political Economy* (April 1981), pp. 287-305.

________. "A simulation model of stock exchange trading," *Simulation* (1983), pp. 181-91.

________. *The Microstructure of Securities Markets* (Prentice-Hall, Englewood Cliffs, 1986).

Cohen, Kalman J., and Robert A. Schwartz. "An Electronic Call Market: Its Design and Desirability," in Henry C. Lucas, Jr. and Robert A. Schwartz, eds., *The Challenge of Information Technology for the Securities Markets: Liquidity, Volatility, and Global Trading* (Dow Jones-Irwin, Homewood, 1989) pp. 20-58.

Conroy, Robert M., and Robert L. Winkler. "Informational Differences between Limit and Market Orders for a Market Maker," *Journal of Financial and Quantitative Analysis* (December 1981), pp. 703-24.

Copeland, Thomas E. "A Model of Asset Trading under the Assumption of Sequential Information Arrival," *Journal of Finance* (September 1976), pp. 1149-68.

Copeland, Thomas E., and Dan Galai. "Information Effects on the Bid-Ask Spread," *Journal of Finance* (December 1983), pp. 1457-69.

Demsetz, Harold. "The Cost of Transacting," *Quarterly Journal of Economics* (February 1968), pp. 33-53.

Fama, Eugene F. *Foundations of Finance* (Basic Books, 1976).

Federal Reserve Bank of New York. "Summary of Results of U.S. Foreign Exchange Market Turnover Survey Conducted in April 1989," (September 13, 1989a).

________. "The Foreign Exchange Committee: Annual Report 1988," (June 1989b).

________. "Gross Foreign Exchange Transactions by Banks in U. S. Market Exceeds $23 Billion Daily: Broker Turnover Separately Amounts to $8.5 Billion Daily, New York Fed Survey Shows," (June 23, 1980).

Flood, Mark D. "Market Structure and Inefficiency in the Foreign Exchange Market," Working Paper 91-001A, Federal Reserve Bank of St. Louis (July 1991).

Garbade, Kenneth D. "The Effect of Interdealer Brokerage on the Transactional Characteristics of Dealer Markets," *Journal of Business* (July 1978), pp. 477-98.

Garman, Mark B. "Market Microstructure," *Journal of Financial Economics* (June 1976), pp. 257-75.

Glassman, Debra. "Exchange Rate Risk and Transaction Costs: Evidence from Bid — Ask Spreads," *Jornal of International Money and Finance* (December 1987), pp. 479-490.

Glosten, Lawrence R., and Paul R. Milgrom. "Bid, Ask and Transaction Prices in a Specialist Market with Heterogeneously Informed Traders," *Journal of Financial Economics* (March 1985), pp. 71-100.

Hahn, Frank H. *Equilibrium and Macroeconomics,* (Basil Blackwell, Oxford, 1984).

Hamilton, James L. "Competition, Scale Economies, and Transaction Cost in the Stock Market," *Journal of Financial and Quantitative Analysis* (December 1976), pp. 779-802.

Harsanyi, John C. *Papers in Game Theory,* (D. Reidel, Dordrecht, 1982).

Ho, Thomas S. Y., and Richard G. Macris. "Dealer Market Structure and Performance," in Yakov Amihud, Thomas S. Y. Ho, and Robert A. Schwartz, eds., *Market Making and the Changing Structure of the Securities Industry* (Lexington Books, Lexington, 1985), pp. 41-66.

Ho, Thomas S. Y., and Hans R. Stoll. "On Dealer Markets under Competition," *Journal of Finance* (May 1980), pp. 259-68.

________. "Optimal Dealer Pricing under Transactions and Return Uncertainty," *Journal of Financial Economics* (March 1981), pp. 47-73.

Keynes, John M. *The General Theory of Employment, Interest and Money* (MacMillan, London, 1936).

Kubarych, Roger M. *Foreign Exchange Markets in the United States* (Federal Reserve Bank of New York, 1983).

Leffler, George L., and Loring C. Farwell. *The Stock Market*, (Ronald Press, New York, 1963).

Logue, Dennis E. "Market-Making and the Assessment of Market Efficiency," *Journal of Finance* (March 1975), pp. 115-23.

Lyons, Richard K. "Private Beliefs and Information Externalities in the Foreign Exchange Market," Working Paper, Columbia University (November 1991).

Mossberg, Walter S. "Making Book on the Buck," *Wall Street Journal*, September 23, 1988.

Negishi, Takashi. "The Stability of a Competitive Economy: A Survey Article," *Econometrica* (October 1962), pp. 635-69.

Plott, Charles R. "Industrial Organization Theory And Experimental Economics," *Journal of Economic Literature* (December 1982), pp. 1485-527.

________. "Will Economics Become an Experimental Science?" *Southern Economic Journal* (April 1991), pp. 901-19.

Rai, Anoop K. "A Market Micro-Structure Investigation of Dealer Quotes and Risk Premium in the Foreign Exchange Markets," Working Paper, (1991).

Reinganum, Marc R. "Market Microstructure and Asset Pricing: An empirical investigation of NYSE and NASDAQ securities," *Journal of Financial Economics* (November/December 1990), pp. 127-47.

Riehl, Heinz, and Rita M. Rodriguez. *Foreign Exchange and Money Markets: Managing Foreign and Domestic Currency Operations* (McGraw-Hill, 1983).

Ross, Sheldon M. *Stochastic Processes*, (John Wiley & Sons, 1983).

Ross, Stephen A. "The Current Status of the Capital Asset Pricing Model (CAPM)," *Journal of Finance* (June 1978), pp. 885-901.

________. "The Interrelations of Finance and Economics: Theoretical Perspectives," *American Economic Review* (May 1987), pp. 29-34.

Schreiber, Paul S., and Robert A. Schwartz. "Efficient Price Discovery in a Securities Market: The Objective of a Trading System," in Yakov Amihud, Thomas S. Y. Ho, and Robert A. Schwartz, eds., *Market Making and the Changing Structure of the Securities Industry* (Lexington Books, Lexington, 1985), pp. 19-39.

________. "Price Discovery in Securities Markets," *Journal of Portfolio Management* (Summer 1986), pp. 43-48.

Schwartz, Robert A. *Equity Markets: Structure, Trading, and Performance* (Harper & Row, New York, 1988).

Sirri, Erik R. "Bid/Ask Spread, Price, and Volume in a Specialist Market," Working Paper, University of California, Los Angeles (February 1989).

Smidt, Seymour. "Which Road to an Efficient Stock Market," *Financial Analysts Journal* (September/October 1971), pp. 18-20, 64-69.

Stigler, George J. "Public Regulation of the Securities Markets," *Journal of Business* (April 1964), pp. 117-42.

Stoll, Hans R. "The Supply of Dealer Services in Securities Markets," *Journal of Finance* (September 1978), pp. 1133-51.

________. "Alternative Views of Market Making," in Yakov Amihud, Thomas S. Y. Ho, and Robert A. Schwartz, eds., *Market Making and the Changing Structure of the Securities Industry* (Lexington Books, Lexington, 1985), pp. 67-91.

________. "Inferring the Components of the Bid-Ask Spread: Theory and Empirical Tests," *Journal of Finance* (March 1989), pp. 115-34.

Tinic, Seha M. "The Economics of Liquidity Services," *Quarterly Journal of Economics* (February 1972), pp. 79-93.

Tinic, Seha M., and Richard R. West. "Competition and the Pricing of Dealer Service in the Over-the-Counter Stock Market," *Journal of Financial and Quantitative Analysis* (June 1972), pp. 1707-27.

Varian, Hal R. "Differences of Opinion in Financial Markets," in Courtenay C. Stone, ed., *Financial Risk: Theory, Evidence and Implications* (Kluwer Academic Publishers, Boston, 1989), pp. 3-37.

Wei, Shang-Jin. "Anticipations of Foreign Exchange Volatility and Bid — Ask Spreads," International Finance Discussion Paper No. 409, Board of Governors of The Federal Reserve System (August 1991).

Wolinsky, Asher. "Information Revelation in a Market with Pairwise Meetings," *Econometrica* (January 1990), pp. 1-23.