



Order flow and the bid-ask spread: An Empirical Probability Model Of Screen-Based Trading

Tim Bollerslev^{a,*}, Ian Domowitz^b, Jianxin Wang^c

^a *Department of Economics, Rouss Hall, University of Virginia, Charlottesville, VA 22901, and NBER, Cambridge, MA 02138, USA*

^b *Department of Economics, Northwestern University, Evanston, IL 60208, USA*

^c *School of Banking and Finance, University of New South Wales, Kensington, NSW 2033 Australia*

Abstract

A probabilistic framework for the analysis of screen-based trading activity is presented. Probability functions are derived for the stationary distributions of the best bid and offer, conditional on the order flows. By identifying the unobservable order and acceptance flows, our estimation method permits the prediction of the stationary distributions of other market statistics. A test is proposed that allows a comparison of predicted and sample bid-ask spread distributions taking parameter estimation error into account. The methodology is applied to the screen-based interbank foreign exchange market, using continuously recorded quotes on the Deutschemark/US dollar exchange rate.

Key words: Screen-based trading; Limit order book; Order flow; Bid-ask spreads; Statistical model evaluation; Foreign exchange quotations

JEL classification: C51; C52; G15; F31

1. Introduction

Screen-based trading markets, characterized by real-time display of quotation information, constitute a rapidly growing form of financial market microstructure; Domowitz (1993) documents 38 such major trading markets operating in this fashion in conjunction with various automated trade execution mechanisms.¹ Modern dealership markets, in which quotes are submitted to a

* Corresponding author.

¹ The CATS and CAC stock trading systems in Toronto and Paris, and the GLOBEX system for the international trading of futures and options are examples.

computer network that immediately displays the bids and offers on the screen, form another important type of market structure. In the dealership markets, final transactions often are consummated over the telephone, and transaction price information is typically not available through the trading screen.² The newest generation of such markets is fully automated with respect to trade execution, but maintains opaqueness with respect to transaction information.³

Despite the growing proliferation of such screen-based market structure, relatively little theoretical or empirical work has yet been devoted to its general study. The purpose of this paper is to help fill part of that gap. A probabilistic framework for the analysis of screen-based trading activity is proposed, together with a methodology for the empirical implementation and testing of the model's predictions with respect to various market statistics.

More specifically, the mathematical model allows for characterization of the stationary conditional probability structure of the best bids and offers in the market, given the distribution of the order flows and the acceptance rates to the market. The order and acceptance flows are only conditioned on the screen information that typically would be available to traders in an electronic dealership market. The probability distributions for the best bids and offers in the market may in turn be used in the derivation of the implied unconditional distributions for the bid–ask spread, the transaction prices, and other market characteristics of interest.

In implementing the model, the conditional probabilities are parameterized in such a way as to allow the separate identification of the unobservable flows into and out of the system. A two-step estimation procedure is developed for estimating the model parameters. The first-stage parameter estimates for the support of the distributions converge at a fast enough rate to allow for conventional inference procedures regarding the second stage estimates for the parameterized conditional order flows. Functional forms for the conditional order flows that automatically satisfy an adding-up constraint are too restrictive to allow identification of all the separate arrival rates, however. We therefore introduce an additional set of moment conditions in the estimation process, which ensure that the adding-up constraints are satisfied asymptotically. An out-of-sample prediction test for the implied distribution of various market statistics, that takes into account the parameter estimation error uncertainty, also is derived.

The methodology is applied to the study of quote and spread distributions in the interbank foreign exchange market. The data set consists of continuously

² The NASDAQ system and London's SEAQ International for the trading of stocks, and the Reuter's network for the interbank foreign exchange market are examples.

³ Recent fully automated dealer markets for interbank foreign exchange trading include the Dealing 2000-2, EBS, and Minex systems.

recorded bid and ask quotations from the Reuter's screen for the Deutsche-mark/US Dollar exchange rate over three noncontiguous weeks in 1989. Between 20,000 and 29,000 quotes are available for each of the three weeks. In order to minimize the degree of data-snooping bias, the first week of the data was used only for preliminary model specification. The model identified over this week was then estimated for the second week of data, with the third week reserved for predictive testing purposes. Interestingly, the structural model estimated over the second week is found to predict the general shape of the bid-ask spread distribution for the third week remarkably well, although the new formal prediction based test with more than 10,000 true out-of-sample spread observations rejects the simple parametric formulations for the order flows employed here.

Our methodology and application are closely related to the literature on discrete pricing in financial markets. The theoretical model and estimation technique both are predicated on the assumption of discrete pricing, thereby avoiding potential biases in inference noted by Ball (1988) and Harris (1991), among others. Other studies that pay close attention to price discreteness include the ordered probit model of Hausman, Lo and MacKinlay (1992) for analyzing discrete stock prices and the conditional Gamma distribution model of Harris (1994), oriented towards a characterization of discrete bid-ask spreads. We combine the two in the context of screen-based trading by explicitly modelling the discrete nature of the screen quotes.

The methods and results here also provide some contribution to the issue of clustering in the distribution of the bid-ask spread, investigated by Goodhart and Curcio (1991) and Harris (1994), for example. In the same spirit as Harris (1994), we find a role for the level of quoted prices on the distribution of the spread. Our route is more indirect, however, since the result follows from the influence of price levels on the stationary distribution of quotations. Despite the differences in methodology and financial market, we nevertheless find the same pattern of predicted spreads in the foreign exchange market as does Harris (1994) in the stock market. Statistical comparisons of predicted spreads from parametric models to nonparametric estimators of the spread distribution also typically are made using root mean-square error criteria on a frequency-by-frequency basis. Our contribution in this regard is to provide a procedure that allows for inference with respect to the entire distribution, accounting for parameter estimation error uncertainty, as well as for uncertainty in the nonparametric estimates of the spread.

The plan of the rest of the paper is as follows. The probabilistic structure of the model, and the distributional results for the market statistics are presented in the next section. Some of the institutional features of the foreign exchange market and the specific data set used in the empirical analysis are discussed in Section 3. Econometric issues related to the practical implementation and estimation of the structural probability model, together with the actual

estimation results are presented in Section 4. The prediction based test of the model implications regarding the stationary distribution of the bid–ask spread is derived and implemented empirically in Section 5. Some concluding remarks are contained in Section 6.

2. Some basic results on probability distributions

Screen-based trading technology constitutes an important component of modern financial markets. In this section, we present a stylized mathematical model of the quasi-book provided by any such screen-based trading system. Specifically, our goal is to characterize the stationary probability structure of the best bids and offers in the system, the bid–ask spread, and the transactions prices produced by the trade execution mechanism. Other aspects of the same theoretical probability model have previously been analyzed by Domowitz and Wang (1994). The focus of the present study centers on the elements of the proposed model that are directly amenable to empirical estimation and testing.

The quasi-order book on the screen is represented by two queues; one queue for bids and one for offers. Bids and offers arrive to the queue according to Poisson processes, with each arrival being for one unit of the asset. This assumption of Poisson arrival rates is appropriate when there exist a large number of potential traders placing orders at a given time; see, for instance, the discussion of the Poisson distribution in Feller (1971). We implicitly require that such a pool of traders exists for every information signal in the form of a best price on the other side of the market. Of course, the actual number of traders that do place orders, conditional on such information, may be very low.

A trader can either take the best bid or offer, or submit an order at or below the best bid price or above the best asking price. This set of possible limit prices is denoted by $\{p_1, p_2, \dots, p_N\}$, with the minimum tick size normalized to unity for notational simplicity, i.e., $p_{i+1} = p_i + 1$ for $i = 1, \dots, N - 1$. The lowest offer and the highest bid available on the screen are denoted by A^M and B^M , respectively. We adopt the convention that $p_0 = 0$, if there are no buy orders on the screen. Similarly, $p_{N+1} = \infty$ corresponds to there being no sell orders on record. Note, that p_0 and p_{N+1} prices are merely adopted out of notational convenience. The actual observed prices are assumed to be strictly bounded away from zero and some upper finite value. With this notation the discrete bounded supports for the market ask and bid, A^M and B^M , are therefore given by $\{p_1, \dots, p_{N+1}\}$ and $\{p_0, \dots, p_N\}$, respectively.

To simplify the development of the probability model, we shall assume that the arrival rate at each price is conditional on the current best market bid and ask only. That is, the best bid and offer act as sufficient statistics for the state of the market. In particular, the setup is general enough to include optimized strategies that depend on the current market statistics. This simplifying

assumption is also consistent with the limited information structure common to many screen-based trading systems, i.e., at any point in time the screen only shows the top and not the entire book of available orders. The arrival rates of offers are denoted by $\lambda^A(i, j)$, where $i = 1, \dots, N$ and $j = 0, 1, \dots, N$. For $i \geq j$, $\lambda^A(i, j)$ gives the arrival rate of offers at the price p_i conditional on the event $\{B^M = p_j\}$. The probability of an ask arriving at p_i given $\{B^M = p_j\}$ where $i < j$ is zero by definition; i.e., $\lambda^A(i, j) \equiv 0$ for $i < j$. Similarly, we denote the arrival rates of bids by $\lambda^B(j, i)$, where $j = 1, \dots, N$ and $i = 1, \dots, N + 1$. For $j \leq i$, $\lambda^B(j, i)$ gives the arrival rate of bids at price p_j conditional on $\{A^M = p_i\}$. By definition $\lambda^B(j, i) = 0$ for $j > i$.

An order can be taken off the screen in two ways. Firstly, an arrival from the Poisson process with $i = j$ and the screen non-empty clears an order at transaction price p_i . Secondly, a trader may hit the best bid or lift the best offer on the screen at any time. The arrival rates for this activity are denoted by $\gamma^B(i, j)$ and $\gamma^A(j, i)$, respectively. As for the Poisson arrival rates for the bids and offers, the arrival rates for such ‘market orders’ are conditioned on the screen information through the market bid and ask only, i.e., A^M and B^M .

Subject to the requirement that the Poisson arrival distributions are unconditionally stationary, the resulting stationary conditional distributions for the best market ask and bid are summarized in the following theorem.

Theorem 1 (Domowitz and Wang, 1994). Let $\lambda^A(i, j)$, $\gamma^A(i, j)$, $\gamma^B(j, i)$ and $\lambda^B(j, i)$ denote the stationary Poisson offer arrival and acceptance rates at the price points $\{p_1, \dots, p_N\}$. The stationary conditional distributions for the market ask and bid are then given by

$$P(A^M = p_i | B^M = p_j) = \frac{\gamma^A(i, j)}{\lambda^B(i, i) + \gamma^A(j, i)} \equiv \alpha(i, j), \quad 0 \leq j < i \leq N \quad (1)$$

and

$$P(B^M = p_j | A^M = p_i) = \frac{\gamma^B(j, i)}{\lambda^A(j, j) + \gamma^B(i, j)} \equiv \beta(j, i), \quad 1 \leq j < i \leq N + 1, \quad (2)$$

provided that $\alpha(N + 1, j) \equiv 1 - \sum_{i=j+1, N} \alpha(i, j) > 0$ for all $j = 0, \dots, N$, and $\beta(0, i) \equiv 1 - \sum_{j=1, i-1} \beta(j, i) > 0$ for all $i = 1, \dots, N + 1$.

Proof. The theorem is a special case of previous results in Domowitz and Wang (1994), and the proof is therefore omitted. ■

Interestingly, the existence of a stationary conditional distribution for the market ask and bid prices is contingent on a positive probability of no buy or

sell orders at all possible price points, i.e., $\alpha(N + 1, j) > 0$ for all $j = 0, \dots, N$ and $\beta(0, i) > 0$ for all $i = 1, \dots, N + 1$. The inequalities guaranteeing existence represent a direct extension of the stationarity condition in simple single-server queues, depending here on net inflows and outflows to the system, summed over price priority classes.

In order to derive the stationary distribution of other market statistics, it is convenient to first characterize the stationary marginal distributions for A^M and B^M .

Lemma 1. Let the conditions of Theorem 1 be satisfied. The stationary marginal distributions for the market ask and bid prices are then given by

$$P(A^M = p_i) = \frac{\alpha(i, 0)/\beta(0, i)}{\sum_{j=1}^{N+1} \alpha(j, 0)/\beta(0, j)}, \quad 1 \leq i \leq N \tag{3}$$

and

$$P(B^M = p_j) = \frac{\beta(j, N + 1)/\alpha(N + 1, j)}{\sum_{i=0}^N \beta(i, N + 1)/\alpha(N + 1, i)}, \quad 1 \leq j \leq N. \tag{4}$$

The unconditional probabilities for there being no sell or buy orders on record are $P(A^M = p_\infty) = 1/[\sum_{j=0, N} \beta(j, N + 1)/\alpha(N + 1, j)]$ and $P(B^M = p_0) = 1/\sum_{i=1, N+1} \alpha(i, 0)/\beta(0, i)]$, respectively.

Proof. From the definition of conditional probabilities,

$$P(B^M = p_j) = \frac{\beta(j, i)}{\alpha(i, j)} P(A^M = p_i), \quad 0 \leq j \leq i \leq N + 1.$$

In particular, in the event of there being no buy orders on record, i.e., $j = 0$,

$$P(A^M = p_i) = \frac{\alpha(i, 0)}{\beta(0, i)} P(B^M = p_0), \quad 1 \leq i \leq N + 1.$$

By the adding up constraint, $\sum_{i=1, N+1} P(A^M = p_i) = 1$, it follows therefore that $P(B^M = p_0) = 1/[\sum_{i=1, N+1} \alpha(i, 0)/\beta(0, i)]$, and the marginal distribution for the market ask price in Eq. (3) is obtained by direct substitution. The derivation of the market bid distribution in Eq. (4) proceeds analogously. ■

The unconditional distribution of the bid–ask spread is easily derived from the results in Theorem 1 and Lemma 1. This distributional result is summarized in the following corollary.

Corollary 1. Let the conditions of Theorem 1 be satisfied. The stationary distribution of the bid–ask spread, $S = A^M - B^M$, is then given by

$$\begin{aligned}
 P(S = s) &= \frac{\sum_{i=g+1}^N \beta(i - s, i) \alpha(i, 0) / \beta(0, i)}{\sum_{j=1}^{N+1} \alpha(j, 0) / \beta(0, j)} \\
 &= \frac{\sum_{j=1}^{N-s} \alpha(j - s, j) \beta(j, N + 1) / \alpha(N + 1, j)}{\sum_{i=0}^N \beta(i, N + 1) / \alpha(N + 1, i)}, \quad 1 \leq s \leq N - 1. \quad (5)
 \end{aligned}$$

Proof. By the definition of conditional probabilities,

$$\begin{aligned}
 P(S = s) &= P(A^M - B^M = s) = \sum_{i=g+1}^N P(A^M = p_i, B^M = p_i - s) \\
 &= \sum_{i=g+1}^N P(B^M = p_i - s | A^M = p_i) P(A^M = p_i) \\
 &= \sum_{i=g+1}^N \beta(i - s, i) P(A^M = p_i).
 \end{aligned}$$

The first expression in Eq. (5) follows by substitution of the marginal distribution for A^M in Lemma 1. The derivation of the second equivalent expression for the unconditional spread distribution in Eq. (5) proceeds analogously by summation over the probabilities that $P(A^M = p_j + s, B^M = p_j)$ for $j = 1, \dots, N - s$. ■

Corollary 1 provides a potentially useful policy tool, for the analysis of transactions costs associated with alternative market structures subject to different order flows. We shall not pursue this aspect of our theoretical results any further here, however. The main focus of the present analysis centers directly on the issues associated with the empirical implementation and testing of the proposed probability model. In that respect, the bid–ask spread distribution in Corollary 1 forms the basis for the out-of-sample prediction test derived in Section 5.

Even though we shall concentrate solely on the implied distributional results for the bid–ask spread distribution in the empirical analysis reported on below, it is worth noting that the implied distributional characteristics of other market statistics may be derived in a similar fashion. To illustrate, Corollary 2 summarizes the unconditional transactions price distribution predicted by the model.

Corollary 2. Let the conditions of Theorem 1 be satisfied. The stationary distribution of the transactions prices, P^T , is then given by

$$P(P^T = p_i) = \rho(i) / \sum_{j=1}^N \rho(j) \quad 1 \leq i \leq N, \quad (6)$$

where

$$\rho(i) \equiv \frac{\alpha(i, 0)(\lambda^B(i, i) + \sum_{j=1}^{i-1} \gamma^A(j, i))}{\beta(0, i) \sum_{j=1}^{N+1} \alpha(j, 0)/\beta(0, j)} + \frac{\beta(i, N + 1)(\lambda^A(i, i) + \sum_{j=i+1}^N \gamma^B(i, j))}{\alpha(N + 1, i) \sum_{j=0}^N \beta(j, N + 1)/\alpha(N + 1, j)} \quad (7)$$

Proof. A transaction occurs at p_i whenever the best buy and sell order both equal p_i , which happens with probability $P(A^M = p_i) \lambda^B(i, i) + P(B^M = p_i) \lambda^A(i, i)$, or alternatively when a trader decides to hit the best market ask or bid at p_i , which happens with probability $P(A^M = p_i) \sum_{j=1, i-1} \gamma^A(j, i) + P(B^M = p_i) \sum_{j=i+1, N} \gamma^B(i, j)$. Substituting for the unconditional probabilities for the market bid and ask from Lemma 1, and rearranging the ratios yields Eq. (7). Eq. (6) follows by normalization. ■

We now turn to the empirical implementation of the stylized probability model outlined above. The estimation results are based on a data set that consists of continuously recorded Deutschemark/US Dollar spot quotes from the screen-based interbank foreign exchange market.

3. Screen trading in the interbank foreign exchange market

Trading in the interbank foreign exchange market is carried out largely through the Reuter's network. The foreign exchange market is in operation 24 h a day, seven days a week, and is the closest analogue to the concept of a continuous time global marketplace. Although the posted quotes on the Reuter's screen are only indicative, with all the transactions done over the telephone, reputation effects rule out the posting of quotes at which a bank would subsequently refuse to deal; for further discussion of the institutional details specific to the foreign exchange market, see, for instance, Bollerslev and Domowitz (1993a), Goodhart (1988), Goodhart and O'Hara (1995), Lyons (1995), and Zhou (1996).

The Deutschemark/US Dollar exchange rate is the most heavily traded spot rate. According to recent estimates by the Federal Reserve Bank of New York, this particular rate accounts for almost one-third of the overall trading volume in the foreign exchange market; see Tygier (1988). The data set analyzed below consists of continuously recorded bid and ask quotations on the Deutschemark/US Dollar rate for the three weeks in 1989 from 9–15 April (week 1), 21–27 May (week 2), and 18–24 June (week 3).⁴ The explicit analysis of three

⁴The data set was generously provided by Charles Goodhart. A detailed description of the method of data capture and the screening for outliers are contained in Goodhart and Demos (1990) and Goodhart and Figliouli (1991).

separate weeks of data was motivated by an attempt to reduce any data-snooping biases in the empirical results; see Lo and MacKinlay (1990). Thus, in the analysis below the data from week 1 was used only in the preliminary model specification searches. The chosen model was then estimated on the data from week 2, with the data originating during week 3 used solely to check the model specification by an out-of-sample prediction test. During each of the three weeks under study 20,408, 29,280, and 29,547 quotes, respectively, appeared on the Reuter's screen.⁵ The original data set consists of more than 300,000 quotes for the full twelve week period from 9th April to 30th June 1989.

To illustrate the actual format of the available screen information, consider the following three pairs of quotes from April 11 in week 1:

093150	COMMERZB	HAM	1.8813/20
093209	ROY SCOT	LDN	1.8815/20
093237	DANSKE	COP	1.8812/22

At 9:31:50 Greenwich Mean Time (GMT), Commerzbank in Hamburg entered quotes on the Reuter's screen to buy US Dollar at 1.8813 Deutschemark per Dollar, and sell Deutschemark at 1.8820 per Dollar. The spread for this quotation pair is therefore 0.0007, or seven ticks. The next two pair of quotes originated from Royal Scottish Bank in London, and Danske Bank in Copenhagen. These two quotation pairs appeared on the screen 19 and 47 s later, respectively.

In the absence of an automated trade execution algorithm, which is typically the case in a dealership market such as the one analyzed here, identification of the exact state of the quasi-book on the screen and the treatment of stale quotes require some bookkeeping methodology.⁶ The analysis reported below is based on the quote cancellation scheme put forward in Bollerslev and Domowitz (1993a). Firstly, a new quote from a given bank automatically cancels an outstanding old quote from that same bank, regardless of what else is happening in the market. Secondly, the effective life of any bid–ask quotation pair is assumed to depend on the current degree of market activity. In particular, no quote is good for more than 60 min, and in an active market environment with five or more quotation arrivals per minute an incoming quote is only good for 2 min.⁷ The average arrival rate across the trading day is roughly two-and-half

⁵ The original data set consist of more than 300,000 quotes for the full 12 week period from 9 April to 30 June 1989.

⁶ In the case of an automated execution system, the record keeping is done by the computer, and this complication is not relevant; for further discussion of the mechanics of automated trade execution systems, see Domowitz (1990).

⁷ Formally, the life of a quotation pair is modelled as $\tau = 1/(0.0167 + 0.097 \text{ minimum}\{\lambda, .5\})$, where λ denotes the average per minute quote arrival rate over the past hour. The results reported below are fairly robust to the choice of any particular quote cancellation scheme.

quotes per minute. The arrival rate of five quotes per minute corresponds to the average market activity during the peak trading hours from 7 to 9 and 14 to 16 GMT; see Goodhart and Demos (1990) and Bollerslev and Domowitz (1993a).

Based on the quote cancellation scheme outlined above, and the transactions mechanism discussed in Section 2, the number of market bids and asks available for analysis for each of the three weeks under study are 12,909, 20,019, and 20,373, respectively. Thus, a total of 7499, 9261, and 9174 quotation pairs were placed outside the current best market bid and ask, and never made it to the top of the quasi-book before they expired.

4. Estimation of the parameterized probability model

The discussion of the estimation methods will focus on the particular probability structure developed in Section 2. The same ideas may be employed in the implementation of other probabilistic structures in financial market analysis, however. Given the intuitive interpretation in terms of market behavior, it is natural to parameterize the present model by the order arrival and acceptance rates. Subject to the underlying stationarity conditions, the resulting stationary distribution for the market bids and asks are then given in Lemma 1. These unconditional moment conditions form the basis for the estimation procedure.

The first step involves the identification of the positive supports for the bid and offer distributions. Specifically, let n denote the number of observed market quotation pairs.⁸ Let the corresponding order statistics for the bid and ask quotes be denoted by $q_{1,n}^B \leq q_{2,n}^B \leq \dots \leq q_{n,n}^B$, and $q_{1,n}^A \leq q_{2,n}^A \leq \dots \leq q_{n,n}^A$, respectively. It is a well-known result from the theory of order statistics that under general conditions the endpoints of the support of a bounded continuous distribution are consistently estimated by the trimmed minimum and maximum; see, for instance, Lehman (1983, Chapter 5). Thus, in the present context, p_1 and p_N are consistently estimated by $q_{\ell,n}^B$ and $q_{n-\ell,n}^A$, respectively, for any fixed integer value ℓ . Also, the number of price points, N , may be consistently estimated by $q_{n-\ell,n}^A - q_{\ell,n}^B$. Furthermore, with a finite number of price support points, the converge rate of these estimators will be arbitrarily high. In particular, $n(q_{\ell,n}^B - p_1)$, $n(q_{n-\ell,n}^A - p_N)$, and $n(q_{n-\ell,n}^A - p_{\ell,n}^B - N)$, are all $O_p(1)$ for ℓ fixed and n going to infinity.

In the actual implementation we simply took the minimum and maximum quote as our estimates for the two endpoints, i.e., $\ell = 1$. To ease the notational

⁸ For the foreign exchange quotations analyzed here, the number of bids and offers are the same. In other applications of the model, the number of quotations for the two sides of the market may differ. The methodology is equally applicable in that context, with only some extra complication in the notation.

burden, in the discussion below we do not differentiate any further between these estimates and the true lower and upper endpoints of the distributions, p_1 and p_N , or the true number of price points, N . Because of the fast rate of convergence, the second stage inference regarding the parameters for the conditional order flow distributions may be conducted using standard \sqrt{n} asymptotics, without having to adjust for the first stage estimation error uncertainty associated with not knowing the true endpoints of the distributions.⁹

The frequency estimates for the conditional probabilities entering Theorem 1 are computed as follows. Over the set of all possible quotation pairs, we tabulate the number of observations, i.e., the number of market asks and bids over the grid $\{p_1, p_0\}, \{p_2, p_0\}, \dots, \{p_{N+1}, p_N\}$. Note, p_{N+1} and p_0 signify the absence of any asks or bids on record. The observed frequency for the joint event $\{A^M = p_i, B^M = p_j\}$, $0 \leq j < i \leq N + 1$, is then calculated by dividing the number of counts with n . Summing these observed frequencies for $\{A^M = p_i, B^M = p_j\}$ over j and i , yields the marginal observed frequency distributions for $\{A^M = p_i\}$ and $\{B^M = p_j\}$, respectively. Estimates of the conditional probabilities, $P(A^M | B^M)$ and $P(B^M | A^M)$, are then calculated directly from these quantities. Denote these frequency estimates by $\hat{\alpha}_n(i, j)$ and $\hat{\beta}_n(j, i)$, respectively. Then, under general regularity conditions $E[\hat{\alpha}_n(i, j)] = \alpha(i, j)$ and $E[\hat{\beta}_n(j, i)] = \beta(j, i)$, where $0 \leq j < i \leq N + 1$; see Johnson and Kotz (1969) for a discussion of the properties of this maximum likelihood estimator for the parameters in the multinomial model.

Now let $\lambda^A(i, j; \theta_0)$, $\lambda^B(j, i; \theta_0)$, $\gamma^A(j, i; \theta_0)$, and $\gamma^B(i, j; \theta_0)$ denote the parameterized order arrival and acceptance rates, where $\theta_0 \in \Theta \subseteq \mathbb{R}^k$ refers to the true $k \times 1$ vector of unknown parameters to be estimated. Also define,

$$\alpha(i, j; \theta) \equiv \frac{\lambda^A(i, j; \theta)}{\lambda^B(i, i; \theta) + \gamma^A(j, i; \theta)}, \quad 0 \leq j < i \leq N \tag{8}$$

and

$$\beta(j, i; \theta) \equiv \frac{\lambda^B(j, i; \theta)}{\lambda^A(j, j; \theta) + \gamma^B(i, j; \theta)}, \quad 1 \leq j < i \leq N + 1. \tag{9}$$

In the subsequent analysis, these functions evaluated at the data will be denoted by $\hat{\alpha}_n(i, j; \theta)$ and $\hat{\beta}_n(j, i; \theta)$, respectively. It follows then from Theorem 1 that these parameterized arrival rates must satisfy the following two sets of orthogonality conditions:

$$E[\hat{\alpha}_n(i, j; \theta_0) - \alpha(i, j)] = E[\hat{\alpha}_n(j, i; \theta_0) - \hat{\alpha}_n(i, j)] = 0 \quad 0 \leq j < i \leq N, \tag{10}$$

⁹This is analogous to the results for non-stationary but cointegrated time series, where the OLS estimates of the cointegrating vector converge at rate n , so that any second stage inference may proceed using standard \sqrt{n} asymptotics; see Engle and Granger (1987) and Stock (1987).

and

$$E[\hat{\beta}_n(j, i; \theta_0) - \beta(j, i)] = E[\hat{\beta}_n(j, i; \theta_0) - \hat{\beta}_n(j, i)] = 0, \quad 1 \leq j < i \leq N + 1. \tag{11}$$

The separate identification of the order and acceptance flows makes it difficult to choose functional forms that automatically satisfy the necessary adding-up constraints. The approach taken here is to impose these summability conditions through the use of two additional sets of unconditional moment conditions. In particular, we require that

$$E\left[1 - \sum_{i=j+1}^N \alpha_n(i, j; \theta_0)\right] = 0, \quad 1 \leq j \leq N \tag{12}$$

and

$$E\left[1 - \sum_{j=1}^{i-1} \hat{\beta}_n(j, i; \theta_0)\right] = 0, \quad 1 \leq i \leq N. \tag{13}$$

Provided that the dimensionality of θ is less than $N^2 + N$, the orthogonality conditions in Eqs. (10)–(13) can be combined, and the parameter vector θ estimated by the minimum chi-square methods discussed in Ferguson (1958). In particular, with the addition of the adding up constraints in Eqs. (12) and (13), a version of the Pearson minimum modified chi-square estimator, say $\hat{\theta}_n$, is obtained by minimizing the function

$$Q_n(\theta) = \sum_{j=1}^N \left[\sum_{i=j+1}^N \frac{[\hat{\alpha}_n(i, j) - \hat{\alpha}_n(i, j; \theta)]^2}{\hat{\alpha}_n(i, j)} + \left[1 - \sum_{i=j+1}^N \hat{\alpha}_n(i, j; \theta)\right]^2 \right] + \sum_{i=1}^N \left[\sum_{j=1}^{i-1} \frac{[\hat{\beta}_n(j, i) - \hat{\beta}_n(j, i; \theta)]^2}{\hat{\beta}_n(j, i)} + \left[1 - \sum_{j=1}^{i-1} \hat{\beta}_n(j, i; \theta)\right]^2 \right], \tag{14}$$

with respect to $\theta \in \Theta \subseteq \mathbb{R}^k$. Even though the number of price points, N , will have to be estimated, as discussed above, the fast rate of convergence guarantees that the standard asymptotic distribution of $\hat{\theta}_n$ is unchanged from that delineated by Ferguson (1958). Specifically, for $n \rightarrow \infty$,

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \rightarrow N(0, I(\theta_0)^{-1}), \tag{15}$$

where $n^{-1} I(\theta_0)^{-1}$ denotes the standard variance–covariance matrix of the parameter estimates obtained from the derivatives of the objective function in Eq. (14).

In order to enforce positivity for the order arrival flows, the conditional probability functions entering Eqs. (8) and (9) are all expressed in exponential form. Based on the initial model specification searches for week 1, each order

and acceptance rate is assumed to depend directly on the price at which the order is placed or the acceptance is made, i.e., p_i or p_j . The conditioning of an arrival rate at price p_i on the best price on the other side of the market, p_j , is introduced through the addition of the absolute value of the difference between the two prices, i.e., $|p_i - p_j|$. Thus, the order and acceptance flows at price p_i depend both on that price, and on the relative distance between the price p_i and the best price available on the other side of the market. The preliminary analysis for week 1 also revealed some additional clustering of the quotes for a spread size equal to 0.0005, or five ticks.¹⁰ To account for this clustering phenomenon, which cannot be explained by any observable screen data, we introduced the $I_{i-j=5}$ dummy variable for the case in which $p_i - p_j = 5$.

Turning to the actual estimation results for week 2, the first stage estimation of the support yields $N = 576$. Over the grid of $\frac{1}{2}N(N - 1) = 165,600$ possible market bid-ask pairs, only 2509 have positive sample support, however. In the minimization of the objective function in Eq. (14), only these price pairs were included in the summations. The resulting second stage estimates, with asymptotic standard errors in parentheses, are,

$$\hat{\lambda}_n^A(i, j; \hat{\theta}_n) = \exp(75.645 - 38.276p_i + 9.336|p_i - p_j| + 1.055I_{i-j=5}), \tag{16}$$

(5.778) (3.164) (0.532) (0.100)

$$\hat{\gamma}_n^A(j, i; \hat{\theta}_n) = \exp(89.606 - 44.946p_i + 11.738|p_i - p_j|), \tag{17}$$

(4.113) (2.263) (0.527)

$$\hat{\lambda}_n^B(j, i; \hat{\theta}_n) = \exp(-55.172 + 30.649p_j + 9.001|p_i - p_j| + 0.939I_{i-j=5}), \tag{18}$$

(10.243) (5.400) (3.935) (0.155)

and

$$\hat{\lambda}_n^B(i, j; \hat{\theta}_n)\exp(-98.140 + 53.750p_j + 14.691|p_i - p_j|). \tag{19}$$

(16.777) (8.995) (3.328)

The estimates for the parameterized order flows for week 2 reported above were based on a total of 20,019 market bids and asks. The evidence strongly suggests that the order flows are dependent on the current market conditions. Of course, with this large number of observations, the statistical significance of the parameter estimates should be carefully interpreted. The use of a standard 5% critical value will result in a very unbalanced tradeoff between the type I and type II error rates.

It is certainly possible that an even better characterization of the conditional order flows could be obtained by further expanding on the parameterizations

¹⁰ A similar clustering phenomenon in the quote formation for NYSE listed stocks has been observed by Harris (1991, 1994).

estimated above. Additional explanatory information is limited based on our data set, however, and for illustrative purposes we prefer to keep this first empirical implementation of the probability model relatively simple. The parsimonious representation in Eqs. (16)–(19) captures the interesting features of the sampling distributions very well. For instance, as illustrated below, even though the unconditional bid–ask spread distribution was not used directly in the estimation of the model, predictions of the spread distribution based on the estimated conditional order flows and the theoretical probability structure closely matches the observed sampling distribution. In the next section we derive a rigorous test of model adequacy, predicated on Corollary 1 and the idea that accurate prediction of the spread distribution from the conditional probabilities of quote placement is a natural implication of the model.

5. The distribution of the bid–ask spread as a test of the model

Early studies on the distribution of the bid–ask spread in the foreign exchange market by Boothe (1988) and Glassman (1987) analyzed the own temporal dependencies in the spread process. More recent studies by Bollerslev and Domowitz (1993a), Bollerslev and Melvin (1994), Bossaerts and Hillion (1991), and Goodhart (1989), provide further evidence on the determinants of the temporal dependencies in foreign exchange market spreads.¹¹ The results reported in this section differ from these earlier studies by our explicit focus on the unconditional spread distribution. We also offer no new evidence on the adverse information based explanations of the size of the bid–ask spread, such as Glosten and Milgrom (1985), or the inventory cost models of the spread, such as Stoll (1978). While it would be interesting to further explore the linkage between these more traditional models of the spread and the new theoretical probability model developed in this paper, our data do not currently allow it. The analysis here simply utilizes the empirical bid–ask spread distribution in providing a direct testable implication of the probabilistic structure and the parameterized order arrival rates. It is the analysis of this specific model implication to which we now turn.

To set up the notation, let $S_m(s)$, $s = 1, \dots, N - 1$, denote the sample distribution for the spread, where $m = 20,373$ refers to the number of quotations in week 3. The unconditional bid–ask spread distribution implied by the structural probability model is summarized in Corollary 1. The model predicted bid–ask spread distribution may be calculated directly from this formula, and the parameterized conditional probability functions in conjunction with the actual order flows for week 3. We denote this predicted spread distribution by $S_m(s; \theta_0)$,

¹¹ In a related context, Black (1991) examines the use of vehicle currency and trading volume as determinants for the magnitude of the bid–ask spread.

Table 1 Bid-ask spread distributions

Spread	Sample week 3	Predicted week 3	Sample week 1
1	0.043	0.049	0.109
2	0.179	0.175	0.229
3	0.150	0.140	0.228
4	0.060	0.103	0.137
5	0.378	0.370	0.231
6	0.009	0.025	0.007
7	0.046	0.059	0.023
8	0.022	0.036	0.006
9	0.002	0.004	0.001
10 ⁺	0.111	0.041	0.030

The columns denoted sample week 3 and sample week 1 give the unconditional sample distributions for the 20,373 and 12,909 market bid-ask spreads for each of the two weeks, i.e., $S_m(s)$, $s = 1, \dots, 10^+$. The column labeled predicted week 3 gives the predicted spread distribution for week 3 based on that week's order flow and the parameterized probability model estimated on week 2 data, i.e., $S_m(s; \hat{\theta}_n)$, $s = 1, \dots, 10^+$.

$s = 1, \dots, N - 1$. Of course, in practice the true parameter vector is unknown. The corresponding feasible predicted spread distribution for week 3 based on the estimates for θ from week 2, are denoted by $S_m(s; \hat{\theta}_n)$, $s = 1, \dots, N - 1$.

The sample spread distribution, $S_m(s)$, and the model predicted spread distribution, $S_m(s; \hat{\theta}_n)$, for week 3 are tabulated in the first two columns of Table 1. Since the market spread very rarely exceeds 10 ticks, we combined the spreads for 10 or more ticks into the 10⁺ category. These same two spread distributions are also graphed in Fig. 1. It is immediately apparent from the figure that the probability model does a remarkably good job of predicting the overall shape of the spread distribution.¹²

It is worth noting that the sample unconditional distributions of the bid-ask spread do vary somewhat across the different weeks. The last column of Table 1 reports the unconditional sample distribution for the 12,909 market spreads for week 1, which provided the quotes used in the original formulation of the estimated probability model. In contrast to the very close correspondence between the actual and the predicted spread in columns one and two, the numbers in the third column are clearly different.

Even though the estimated model is able to predict the general shape of the spread distribution very well, it is important to note that the structural model and the particular parameterization for the conditional order flows only provides an approximation to all the complicated dependencies in the market. In

¹² It is interesting to note that the spread distributions for the foreign exchange quotations analyzed here have the same general shape as the bid-ask spread for equities studied in Harris (1991), especially in the spread analysis for higher price stocks.

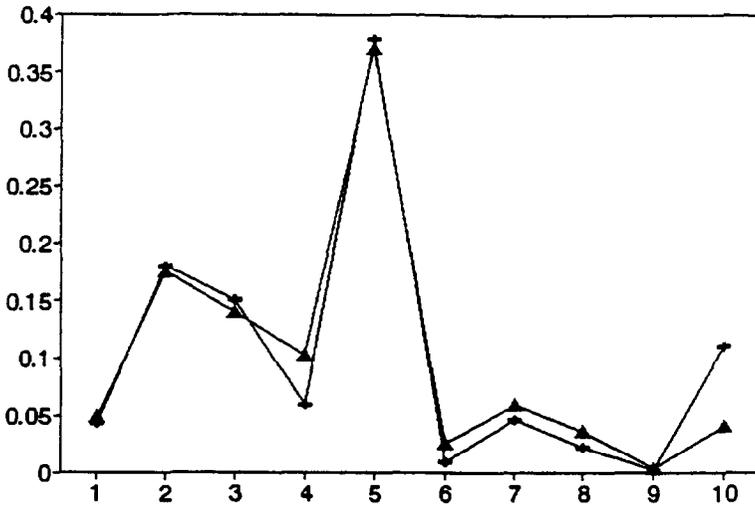


Fig. 1. Bid-Ask Spread Distributions. Note: The points denoted by a + give the unconditional sample distribution for week 3; i.e., $S_m(s)$, $s = 1, \dots, 10^+$. The points denoted by a Δ give the predicted spread distribution for week 3 based on that week's order flow and the parameterized probability model estimated on week 2 data, i.e., $S_m(s; \hat{\theta}_n)$, $s = 1, \dots, 10^+$.

order to more forcefully stress that point, we next develop a formal test for analyzing the out-of-sample predictive performance of the model. Our discussion will focus directly on the particular probability model put forward in Section 2 and its implications for the bid-ask spread. However, as was the case for the estimation methods outlined in Section 4, the testing methodology proposed here is general and should be useful in a number of other situations also. For example, in markets for which transaction price data is available, the methodology can be used to test the model and the parameterized order flows against a predicted price distribution, based on Corollary 2.

In deriving the test it is convenient to express the discrete distributions under study in vector form. Thus, let S_m and $S_m(\hat{\theta}_n)$ denote the $(N - 1) \times 1$ vectors of the sample and predicted spread distributions, respectively. A joint test for structural stability and model specification may then be based on the observed difference,

$$S_m - S_m(\hat{\theta}_n) = [S_m - S_m(\theta_0)] + [S_m(\theta_0) - S_m(\hat{\theta}_n)]. \tag{20}$$

If the parameterized conditional probability model is correctly specified, it follows by standard asymptotic theory that the term in the first bracket on the right-hand side of Eq. (20) will be asymptotically normally distributed. Specifically,

$$\sqrt{m} (S_m - S_m(\theta_0)) \rightarrow N(0, \Omega_m(\theta_0)), \tag{21}$$

as $m \rightarrow \infty$, where the $(N - 1) \times (N - 1)$ covariance matrix of the corresponding multinomial distribution is given by

$$\Omega_m(\theta_0) = \text{diag}(S_m(\theta_0)) - S_m(\theta_0) S_m(\theta_0)', \tag{22}$$

and $\text{diag}(S_m(\theta_0))$ denotes the diagonal matrix with $S_m(\theta_0)$ along the diagonal. Let $\Omega_m(\theta_0)^+$ denote the Moore–Penrose generalized inverse of the reduced rank covariance matrix in Eq. (22). Then, by Eq. (21) $m(S_m - S_m(\theta_0))' \Omega_m(\theta_0)^+ (S_m - S_m(\theta_0))$ is asymptotically χ^2_{N-2} distributed for $m \rightarrow \infty$. This result underlies the standard chi-square goodness of fit test for a multinomial distribution with known categories; see, for instance, Cramer (1946).

In most applications, however, θ_0 will have to be estimated, and the parameter estimation error uncertainty arising from the second term on the right-hand side of equation (20) should also be taken into account. By a standard first-order Taylor series expansion of the parameterized probability function, it follows from Eq. (15), that for $n \rightarrow \infty$,

$$\sqrt{n} (S_m(\theta_0) - S_m(\hat{\theta}_n)) \rightarrow N(0, \Xi_m(\theta_0)), \tag{23}$$

where

$$\Xi_m(\theta_0) = \frac{\delta S_m(\theta_0)}{\delta \theta'} I(\theta_0)^{-1} \frac{\delta S_m(\theta_0)}{\delta \theta}. \tag{24}$$

The two covariance matrices in Eqs. (22) and (24) are consistently estimated by $\Omega_m(\hat{\theta}_n)$ and $\Xi_m(\hat{\theta}_n)$, respectively. Under the assumption that the two samples used for estimation and prediction purposes are asymptotically independent, and both $m \rightarrow \infty$ and $n \rightarrow \infty$, a model specification test for the estimated parameterized probability model may therefore be calculated using the statistic,

$$m(S_m - S_m(\hat{\theta}_n))' [\Omega_m(\hat{\theta}_n) + \frac{m}{n} \Xi_m(\hat{\theta}_n)]^+ (S_m - S_m(\hat{\theta}_n)) \rightarrow \chi^2_{N-2}. \tag{25}$$

This test statistic is an extension of earlier results in Heckman (1984), in which the goodness of fit test and the actual model estimation are done via maximum likelihood over the same in-sample period, i.e., the test statistic is based on the vector difference $S_n - S_n(\hat{\theta}_n)$.¹³ The statistic in Eq. (25) defines an out-of-sample structural stability test, which is likely to have superior power properties in many situations. It also allows for more general method of moments estimation techniques in the analysis of auxiliary distributional implications from an

¹³ Along these lines, it would also be possible to overidentify the estimated model by combining the objective function in Eq. (14) with the orthogonality conditions arising from the probability function for the spread given in Corollary 1. A standard in-sample test of the overidentifying restrictions would then provide a model diagnostic; see Newey and West (1987).

estimated model. With respect to the latter, the testing methodology is in the same spirit as Christiano and Eichenbaum (1992), in which individual predicted moments provide a basis for specification analysis.

Not surprisingly, on evaluating the formal prediction test derived above based on more than twenty-thousand observations for weeks 2 and 3, the simple parameterizations for the order flows are easily rejected. The calculated asymptotic p -value is equal to zero for all practical purposes. Thus, even though the particular parametric model estimated in Section 4 goes a long way in terms of capturing the important features of the data, there is still room for an improved specification of the order arrival flows. The test statistic in (25) provides a useful benchmark against which to judge any such further model developments. We do caution, however, that with the current sample sizes any consistent test is almost certain to detect only minor, and economically inconsequential, misspecifications of the parametric order flows. Contrasting the highly significant value of the test statistic in Eq. (25) for the present parameterizations with the graphs in Fig. 1 illustrates that point.

6. Conclusion

The study of screen-based trading is still in its infancy, although this form of market structure is now seeing widespread application. The methodological contribution of this paper in this regard is in two parts. First, a model is provided that permits the characterization of the stationary conditional probability structure of quotations in a screen-based trading market. Conditional distributions for the bid–ask spread and transaction prices follow as implications of the quotation model. Second, econometric methodology for estimation and inference in such a discrete pricing model is suggested. Unconditional moment restrictions that allow the separate identification of quotation and acceptance flows in a probabilistic framework are introduced. A testing procedure is derived that examines the performance of a model via predicted distributions that themselves do not enter into estimation of the underlying structural model. The test provides a rigorous basis for examining predicted spreads in other settings, such as those of Harris (1994) and Lyons (1995).

The empirical results are of independent interest. Standard theories of financial market structure do not suggest that price levels should affect bid–ask spreads. Modelling the stationary joint distribution of the best bid and ask in terms of such levels yields a statistically significant impact on the distribution of the spread in our framework. This is in line with the results of Harris (1994) on stock market spreads and those of Lyons (1995) using data from a single foreign exchange market maker. Perhaps more surprising, however, is the fact that the predicted spread distribution in the foreign exchange market has the same basic shape as the predicted distribution of Harris (1994) for higher priced stocks on

the New York Stock Exchange. This result is obtained despite radical differences in modeling techniques, market structure, financial instruments, and information used in conditioning the distribution of the spread. This may suggest that theories of price resolution and the so-called "attraction hypothesis" may be more important than market structure in explaining clustering effects in the spread distribution.

Extensions of the theoretical arguments in Domowitz and Wang (1994), combined with the empirical methodology developed here, can be used to produce additional new insights with respect to screen-based trading. In particular, the stationary distribution of the structure of the order book in terms of the numbers of buy and sell orders on the book and the expected waiting time to trade execution can be calculated. The length of time until an order is executed may be considered a measure of market quality for continuous trading markets. A trader seeking to transact by means of a limit order instead of the hit/take option faces a tradeoff between the possibility of receiving a better price and waiting longer for execution. Longer waiting times increase liquidity risk, defined by Garbade and Silber (1979) in terms of the difference between the 'equilibrium' value of an asset at the time of order placement and the actual execution price. The length of the order book also is of interest. In their investigation of floor trading, Cohen et al. (1985) argue for the reduced importance of the book for facilitating information flow as the book length diminishes. In contrast, the work of Bollerslev and Domowitz (1993b) on automated auctions suggests that concentration of the distribution of book length around shorter numbers of orders results in large changes in measures of market efficiency as the number of bids and offers on the book increases on the margin. The results of this paper lay the groundwork for carrying out future empirical tests of these different hypotheses.

Finally, it should be noted that the conditioning information used here is based solely on screen data that would typically be directly available to traders in an electronic environment. The theoretical model can accommodate virtually any stationary form of dependence on order book information, however. From an empirical point of view, this information need not be so limited as for the example analyzed here. For instance, if trader support system data is available, such as inventories in Lyons (1995), this information can be incorporated easily into the probabilistic structure suggested in this paper. We leave further extensions along these lines for future research.

References

- Ball, C.A., 1988, Estimation bias induced by discrete security prices, *Journal of Finance* 43, 841–865.
- Black, S.W., 1991, Transactions costs and vehicle currencies, *Journal of International Money and Finance* 10, 512–526.

- Bollerslev, T. and I. Domowitz, 1993a, Trading patterns and the behavior of prices in the interbank foreign exchange market, *Journal of Finance* 48, 1421–1443.
- Bollerslev, T. and I. Domowitz, 1993b, Some effects of restricting the electronic order book in an automated trade execution system, in: D. Friedman and J. Rust, eds, *The double auction market: Institutions, theories, and evidence* (Addison-Wesley, Reading, MA).
- Bollerslev, T. and M. Melvin, 1994 Bid–ask spreads and volatility in the foreign exchange market: An empirical analysis, *Journal of International Economics* 36, 355–372.
- Boothe, P., 1988, Exchange rate risk and the bid–ask spread: A seven country comparison, *Economic Inquiry* 26, 485–492.
- Bossaerts, P. and P. Hillion, 1991, Market microstructure effects of government intervention in the foreign exchange market, *Review of Financial Studies* 4, 513–541.
- Christiano, L.J. and M. Eichenbaum, 1992, Current real business cycle theories and aggregate labor market fluctuations, *American Economic Review* 82, 430–450.
- Cohen, K.J., R.M. Conroy and S.F. Maier, 1985, Order flow and the quality of the market, in: Y. Amihud, T. Ho, and R. Schwartz, eds., *Market making and the changing structure of the securities industry* (Lexington Books, Lexington, MA).
- Cramer, H., 1946. *Mathematical methods of statistics*. (Princeton University Press, Princeton, NJ).
- Domowitz, I., 1990, The mechanics of automated trade execution systems, *Journal of Financial Intermediation* 1, 167–194.
- Domowitz, I., 1993, A taxonomy of automated trade execution systems, *Journal of International Money and Finance* 12, 607–631.
- Domowitz, I. and J. Wang, 1994, Auctions as algorithms: Computerized trade execution and price discovery, *Journal of Economic Dynamics and Control* 18, 29–60.
- Engle, R.F. and C.W.J. Granger, 1987, Cointegration and error correction: Representation, estimation, and testing, *Econometrica* 55, 251–276.
- Feller, W., 1971. *An introduction to probability theory and its applications*. (Wiley, New York).
- Ferguson, T.S., 1958, A method of generating best asymptotically normal estimates with application to the estimation of bacterial densities, *Annals of Mathematical Statistics* 29, 1046–1062.
- Garbade, K.D. and W.L. Silber, 1979, Structural organization of secondary markets: Clearing frequency, dealer activity and liquidity risk, *Journal of Finance* 34, 577–593.
- Glassman, D. 1987, Exchange rate risk and transactions costs: Evidence from bid–ask spreads, *Journal of International Money and Finance* 6, 479–790.
- Glosten, L. and P. Milgrom, 1985, Bid, ask, and transaction prices in a specialist market with heterogeneously informed traders, *Journal of Financial Economics* 13, 71–100.
- Goodhart, C.A.E., 1988, The foreign exchange market: A random walk with a dragging anchor, *economica* 55, 437–460.
- Goodhart, C.A.E., 1989, News and the foreign exchange market, *Proceedings of the Manchester Statistical Society*, 1–79.
- Goodhart, C.A.E. and R. Curcio, 1991, The clustering of bid/ask prices and the spread in the Foreign exchange market, Manuscript, (London School of Economics).
- Goodhart, C.A.E. and A. Demos, 1990, Reuter screen images of the foreign exchange market: The Deutschmark/Dollar spot rate, *Journal of International Securities Markets*, 333–348.
- Goodhart, C.A.E. and L. Figliuoli, 1991, Every minute counts in financial markets, *Journal of International Money and Finance* 10, 23–52.
- Goodhart, C.A.E. and M. O'Hara, 1995, High frequency data in financial markets: Issues and applications, Manuscript, (London School of Economics).
- Harris, L., 1991, Stock price clustering and discreteness, *Review of Financial Studies* 4, 389–415.
- Harris, L., 1994, Minimum price variations, discrete bid–ask spreads, and quotation size, *Review of Financial Studies* 7, 149–178.
- Hausman, J., A.W. Lo, and A.C. MacKinlay, 1992, An ordered probit analysis of transaction stock prices, *Journal of Financial Economics* 31, 319–379.

- Heckman, J.J., 1984, The χ^2 goodness of fit statistic for models with parameters estimated from microdata, *Econometrica* 52, 1543–1547.
- Johnson N.L., and S. Kotz, 1969. *Discrete distributions*. (Wiley, New York).
- Lehman, E.L., 1983, *Theory of point estimation*. (John Wiley, New York).
- Lo, A.W. and A.C. MacKinlay, 1990, Data-snooping biases in tests of financial asset pricing models, *Review of Financial Studies* 3, 431–467.
- Lyons, R.K., 1995, Tests of microstructural hypotheses in the foreign exchange market, *Journal of Financial Economics*, 39, 321–351.
- Newey, W.K. and K.D. West, 1987, Hypothesis testing with efficient method of moments estimation, *International Economic Review* 28, 777–787.
- Stock, J., 1987, Asymptotic properties of least squares estimators of cointegrating vectors, *Econometrica* 55, 1035–1056.
- Stoll, H., 1978, The supply of dealer services in securities markets, *Journal of Finance* 33, 1133–1152.
- Tygiel, C., 1988, *Basic handbook of foreign exchange*. (Euromoney Publications Limited, London, UK).
- Zhou, B., 1996, High frequency data and volatility in foreign exchange rates, *Journal of Business and Economic Statistics* 14, 45–52.