Journal of Financial Economics 21 (1988) 123–142. North-Holland

ESTIMATING THE COMPONENTS OF THE BID/ASK SPREAD*

Lawrence R. GLOSTEN

Northwestern University, Evanston, IL 60201, USA

Lawrence E. HARRIS

University of Southern California, Los Angeles, CA 90089-1421, USA

Received July 1986, final version received August 1987

This paper develops and implements a technique for estimating a model of the bid/ask spread. The spread is decomposed into two components, one due to asymmetric information and one due to inventory costs, specialist monopoly power, and clearing costs. The model is estimated using NYSE common stock transaction prices in the period 1981–1983. Cross-sectional regression analysis is then used to relate time-series estimated spread components to other stock characteristics. The results cannot reject the hypothesis that significant amounts of NYSE common stock spreads are due to asymmetric information.

1. Introduction

Most economic models of asset pricing assume that the impact of transaction costs on pricing is minor. Although this is arguable and remains, empirically, an open question, most investors consider transaction costs very important in making portfolio management decisions. This may largely explain the substantial interest in 'microstructure' models of the bid/ask spread.

One such model is the asymmetric information model. This model breaks the spread into two components. The first allows market-makers to generate revenue from a seemingly random order flow to cover inventory costs, clearing fees, and/or monopoly profits. This component may be called the *transitory* component, since its effect on stock price time series is unrelated to the underlying value of the securities. The second component arises because market-makers may trade with unidentified investors who have superior information. When such asymmetric information exists, informed traders profit by

*We would like to thank Joel Hasbrouck, Robert Hodrick, Ravi Jagannathan, Tom Lys, Eugene Lerner, Jay Ritter, Mark Weinstein, Hans Stoll (the referee), and seminar participants at New York University, CRSP Autumn 1985 Seminar, and the Institute for Quantitative Research in Finance Spring 1986 Seminar for their many constructive comments on our work. We are also grateful to the Institute for Quantitative Research in Finance for financial support.

submitting orders that will be correlated with future price changes. Rational market-makers in a competitive environment widen the spread beyond what it would otherwise be to recover from uninformed traders what they lose (on average) to the informed traders. The additional widening of the spread is called the *adverse-selection* component because the market-makers face adverse selection in their order flow. This model was first suggested by Bagehot (1971) and was later formally analyzed by Copeland and Galai (1983) and Glosten and Milgrom (1985).

Although the asymmetric information model is important for explaining transactions costs, it is also an important hypothesis about about how private information in the order flow becomes impounded in prices. In the Glosten and Milgrom (1985) model, the adverse-selection spread component is equal to the revision in market-maker expectations of stock resulting from the submission of an order. When someone submits an order to buy (or sell) stock, the uninformed market-maker, knowing that the order might be information-motivated, revises his expectation of the future stock value upward (or downward). Since the revision in expectations, conditional on the type of order received, can be anticipated, the rational market-maker incorporates it into his bid and ask prices. One of these prices will subsequently be observed when an order is filled.

The practical and theoretical interest in the asymmetric information spread model suggests empirical research. In this paper we propose, estimate, and cross-validate a two-component asymmetric information spread model. The results do not reject the asymmetric information theory. Although other models discussed below may also be consistent with the results, we believe there is substantial empirical evidence in favor of adverse-selection spreads. The remainder of this introduction describes how our model and methods differ from and are similar to those in previous studies.

Our estimates are obtained directly from transaction price time series. Recognition that the bid/ask spread is reflected in time-series properties of transaction prices is not new. Several characteristics of the relation between the bid/ask spread and transaction price behavior have been examined by Niederhoffer and Osborne (1966), Cohen, Maier, Schwartz, and Whitcomb (1979), Blume and Stambaugh (1983), Roll (1984), and French and Roll (1986). These papers assume that the entire bid/ask spread is due to factors such as specialist rents, inventory carrying costs arising from risk aversion or other factors, and/or transaction costs that the specialist must pay. These factors explain the transitory spread component, which causes price changes to be negatively serially correlated.

Unlike these other researchers, we also model the adverse-selection component. In contrast to the transitory component, this component, which is due to the revision of market-maker expectations, does not cause serial correlation in our model. It has a permanent effect on all future prices, in the sense that subsequent prices net of the transitory component may go up or down but on average they will stay the same. Glosten (1987a) shows that serial covariance spread estimators like that implemented by Roll (1984) do not estimate the total spread if some part of it is due to adverse selection. Fortunately, the differential time-series properties of the two components allows us to estimate them separately using transaction price series.

Our estimation model allows the adverse-selection spread component to depend on order size. Easley and O'Hara (1987), Kyle (1985), and Glosten (1987b) have theoretical models that suggest this component should increase with the quantity traded (because well informed traders maximize the return to their perishing information). Our empirical results do not reject this prediction. The estimates therefore provide some evidence of the extent to which spreads depend on order size. Although there are other reasons noted below why spreads potentially depend on order size, theoretical predictions and our empirical evidence suggest that at least part of the order size dependency is due to asymmetric information.

Our study is related to the block trading investigation of Holthausen, Leftwich, and Mayers (1987). They measure the temporary and permanent price effects of large-block transactions on the New York Stock Exchange. Interpreted within the asymmetric information model, their estimate of the permanent price change corresponds to an estimate of the adverse-selection spread component for large transactions, while the temporary price change corresponds to the transitory component. Our model and methods allow us to estimate the spread for small as well as large trades.

Our investigation is also related to research reported in Ho and Macris (1984) concerning spread estimation from options market transaction data. Although their model of transaction price changes is similar in spirit to ours, they concentrate on the effect that (risk-aversion-induced) inventory cost has on the location of the spread while ignoring the adverse selection spread. We concentrate on the latter while largely ignoring the former.

The econometric method used to estimate our model is similar to that used by Harris (1986) in his study of discrete prices. This likelihood method permits spread estimation from time-series prices that are unidentified as to bid/ask classification. In this respect, the method is similar to the serial covariance moment method in that both identify the transitory spread component from price reversals. Unfortunately, discreteness-induced errors in the variables can also cause negative serial correlation, and as Harris noted, thereby bias spread estimates. To demonstrate the importance of the problem, we estimate our model taking into account the discreteness problem and also ignoring it. As expected, discreteness has a significant absolute effect on the transitory component estimates. The effect, however, appears to be uniform in cross-section. Accordingly for reasons of cost, we ignore discreteness in our cross-sectional analyses.

The model cross-validation analysis we present at the end of this paper is in the same spirit as the analyses of Benston and Hagerman (1974) (B&H) and Branch and Freed (1977) (B&F). In this cross-sectional regression analysis, we relate the time-series estimated spread components of 250 NYSE common stocks to a number of other stock characteristics. The asymmetric information spread theory provides sign predictions for the regression coefficients. These predictions are compared with the regression results to cross-validate the theory and our time-series estimation methods.

Our cross-sectional analysis differs from those in B&H and B&F in several important respects. First, since we have estimates of both spread components, we can separate the effects of various variables. Second, the transactions data give us access to better independent variables. In particular, while B&H had to use a proxy for a market activity measure and B&F had only the total daily volume, we have two additional variables: average trade frequency and average trade size. Spreads may be different if a given volume is the result of numerous small transactions or a few large ones. These extra variables therefore potentially offer more explanatory power. Third, we use spread estimates obtained from actual transaction data, whereas B&H, B&F, and other authors use quoted spreads. Estimated spreads will differ from quoted spreads when limit orders are crossed with market orders or when floor traders are making market. Finally, we use simultaneous equations methods to estimate our regressions. Many of the right-hand-side variables, such as trade volume, simultaneously depend on the spread components.

This paper is organized as follows. Section 2 introduces our two-component spread model and discusses the estimation technique. The empirical results are presented in section 3. Subsection 3.1 discusses the data, subsection 3.2 discusses the time-series estimation results, and subsection 3.3 analyses the cross-sectional properties of the spread estimates and compares them with results from previous studies. The paper concludes with comments on the limitations of our technology and suggestions for further research.

2. Model and estimation method

This section briefly presents our two-component asymmetric information spread model and describes the estimation method. We omit finer details about the model motivation, derivation, and estimation. These can be found in Glosten and Milgrom (1985), Glosten (1987a), Easley and O'Hara (1987), Kyle (1985), and Harris (1986).

We first present a general two-component asymmetric information spread model in which a number of alternative assumptions about spreads and price evolution are nested. A specification search, described in section 3.2, suggests a parsimonious model that is used in the cross-sectional analysis of section 3.3.

Observed prices in our model are determined from 'true' prices by adjusting for the costs of providing liquidity service and then rounding to the nearest eighth. The following notation is used:

P_t^0 = observed price of transaction t,
V_t = observed number of shares traded in transaction t,
T_t = observed time between transactions $t-1$ and t,
P_t = unobserved price that would have been observed if there were no rounding to discrete one-eighth values,
Q_t = unobserved indicator for the bid/ask classification of P_t^0, $=+1$ if transaction t was initiated by the buyer (ask) and $=-1$ if by the seller (bid),
m_t = unobserved 'true' price, which reflects all publicly available information immediately following transaction t (this price includes any information revealed by that transaction),
e_t = unobserved innovation in 'true' prices between transactions $t-1$ and t due to the arrival of public information,
Z_t = unobserved adverse-selection spread component at transaction t,
C_t = unobserved transitory spread component at transaction t.

Our general two-component asymmetric information spread model is given by

$$m_t = m_{t-1} + e_t + Q_t Z_t \quad \text{('True' price process),} \tag{1a}$$

$$P_t = m_t + Q_t C_t \quad \text{(Unrounded price process),} \tag{1b}$$

$$P_t^0 = \mathrm{Round}(P_t, \tfrac{1}{8}) \quad \text{(Observed price process),} \tag{1c}$$

$$Z_t = z_0 + z_1 V_t \quad \text{(Adverse-selection spread component),} \tag{1d}$$

$$C_t = c_0 + c_1 V_t \quad \text{(Transitory spread component),} \tag{1e}$$

$$e_t \sim \text{iid Normal } (f_1(T_t), f_2(T_t) \mid T_t) \quad \text{(Public information innovation),} \tag{1f}$$

where z_0, z_1, c_0, and c_1 are constants and f_1 and f_2 are currently unspecified functions with $f_2 > 0$.

The 'true' price innovations are of two types. The first, e_t, is due to the arrival of public information, while the second, $Q_t Z_t$, the adverse-selection spread, is due to the revision in expectations conditional on an order arrival. Assuming Z_t is positive, buy orders cause 'true' prices to rise by Z_t while sale orders cause them to fall by $-Z_t$. The adverse-selection spread has a 'permanent' effect on prices since it is due to a change in expectations.

The unrounded price is obtained from the 'true' price by adding or subtracting C_t, the transitory spread component. This component lets market-makers generate revenue by 'buying low and selling high' on average. It causes price changes that reverse on average.

The observed price is obtained by rounding the unrounded price to the nearest one-eighth. The rounding is a purely statistical assumption designed to capture an obvious feature of observed prices.

As we noted in the introduction, the adverse-selection component is expected to be a positive function of order size. To allow for this possibility, we adopt a linear specification for Z_t. For symmetry and to allow for possible economies or diseconomies of scale in the provision of liquidity services, we also adopt a linear specification for C_t, the transitory spread component.

'True' price innovations due to the arrival of public information follow the process described in (1f). The assumption that they are serially independent is essentially an assumption about the rationality of market-makers. If there were any serial correlation in the location of the spread, an entering market-maker could profit by incorporating this information into his quotes. We allow the drift term, $f_1(T_t)$, and the variance term, $f_2(T_t)$, to be a function of elapsed time between trades. The conditional normality assumption is suggested by the mixture of distributions hypothesis [see Clark (1973) and Harris (1987)]

It is useful to express eqs. (1a)–(1e) in terms of the observed price change, D_t. Define the round-off error to be $r_t = P_t^0 - P_t = \text{Round}(P_t) - P_t$. Then

$$\begin{aligned} D_t &= P_t^0 - P_{t-1}^0 \\ &= P_t - P_{t-1} + r_t - r_{t-1} \\ &= Q_t C_t - Q_{t-1} C_{t-1} + Q_t Z_t + e_t + r_t - r_{t-1} \qquad (2) \\ &= c_0(Q_t - Q_{t-1}) + c_1(Q_t V_t - Q_{t-1} V_{t-1}) \\ &\quad + z_0 Q_t + z_1 Q_t V_t + e_t + r_t - r_{t-1}. \end{aligned}$$

Evaluating this expression for $Q_{t-1} = 1$ and $Q_t = -1$ gives the round-trip price change for a sale that immediately follows a purchase of equal size. The absolute value of this quantity may be interpreted as a measure of the effective spread. Its average value (assuming that e_t and r_t have zero means) is $2C_t + Z_t$.

The effective spread should be distinguished from the quoted spread, which is the amount paid by a fully uninformed trader. The quoted spread is $2C_t + 2Z_t$. This quantity differs from the first because it is an unconditional measure of the spread. Intuitively, the trader who initiates an immediate buy/sell combination is not fully uninformed at the time of the sell, because he knows he originated the previous buy.

Eq. (2) can also be used to show that even though we allow both the C and Z components to depend on the number of shares traded, all of the parameters in the model are identified. If the Q's were observed, (2) could be inefficiently estimated by ordinary least squares.[1] As long as there is variation in the number of shares traded, all parameters (including the drift in e_t) are identified. To the extent that observable data are sufficient to identify the Q's, the model remains identified. Our likelihood estimation method obtains identifying information about the Q's from time-series context. Since the transitory component causes price changes to be negatively correlated, information about the Q's can be inferred from price reversals. (The adverse-selection component does not cause price change autocorrelation).

Our method of estimating (2) follows that presented in Harris (1986). The likelihood function, conditional on the unobserved round-off errors, $\{r_t\}$, and bid/ask classifications, $\{Q_t\}$, is the product of T normal densities of $\{e_t\}$, where T is the number of time-series observations on D_t. We obtain an average likelihood function by integrating the conditional likelihood over diffuse prior distributions for the unobserved variables. The result is then maximized to obtain point estimates of the parameters.

Uniform distributions defined on $[-\frac{1}{16}, \frac{1}{16}]$ are used to integrate out the round-off errors. The uniform distribution is used because it is a diffuse distribution and because Gottlieb and Kalay (1985) show that the round-off errors are asymptotically uniformly distributed. Although the round-off errors in the theoretical model are not independent, we integrate over independent priors to keep the estimation computationally tractable. Since simulations show that the procedure consistently estimates known population parameters, it is unlikely that the use of independent priors significantly biases the results.[2]

The bid/ask classification variables are integrated out over independent discrete distributions that assign equal probabilities to both outcomes. This diffuse statistical specification is chosen because it gives the data the greatest latitude to imply values for the bid/ask classification variables within the likelihood procedure, and because it is tractable. Its use in the estimation method should not be confused with any theoretical assumption or prediction of our model for the bid/ask order distribution. Although we recognize that the bid/ask quote mechanism and the bid/ask order distribution are jointly dependent, our model provides no theoretical specification for this distribution. Since simulations show that our procedure consistently estimates known

[1] OLS estimation would be inefficient because of the round-off errors and because the variance of e_t might depend on T_t.

[2] Exact computation of the sample probability function is impossible because it involves an $(N+1)$-fold integral over the continuous ranges of the round-off errors. Approximate numeric evaluation is accomplished by assuming that the round-off errors take discrete values within their ranges. We use a lattice of five equally spaced points. Simulations suggest that virtually no additional benefit comes from using a finer lattice.

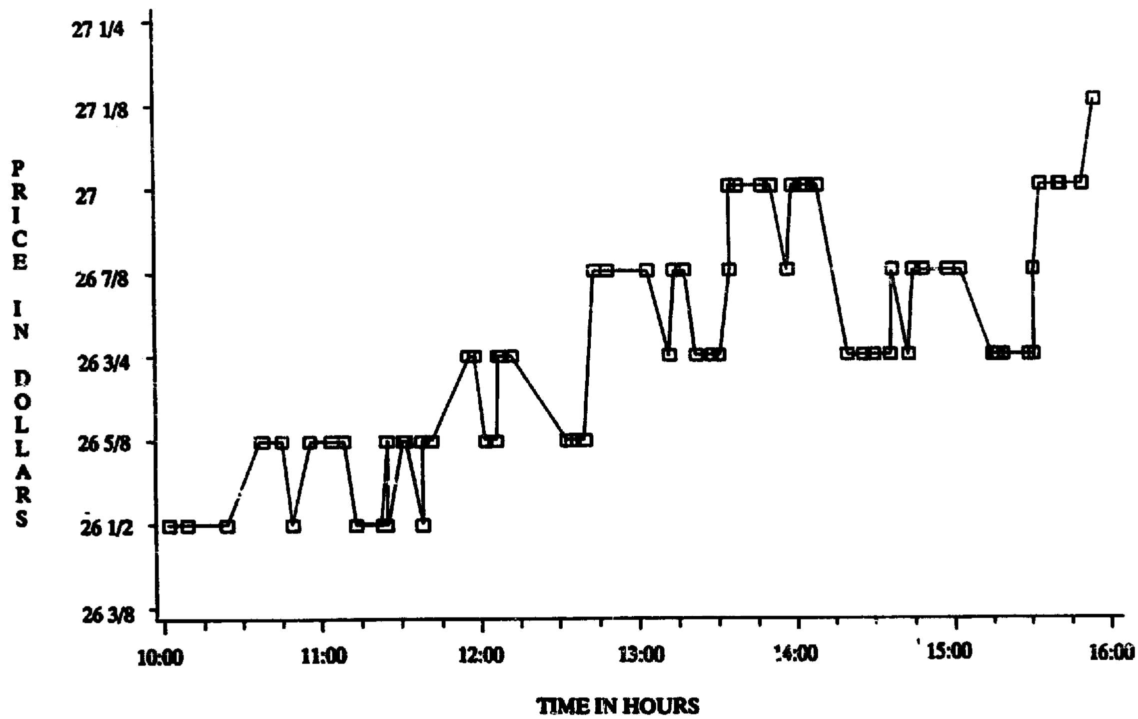

Fig. 1. Complete NYSE transaction price time series for Alcoa Aluminum on December 1, 1981; 78 trades.

population parameters even when the order flow is serially correlated, it is unlikely that the independent priors significantly bias the results.

To give the reader a feel for the data and some intuition as to how our estimation routine works, fig. 1 presents a time-plot of actual transaction prices for Alcoa Aluminum on December 1, 1981. The discreteness of prices and bid/ask bounce are both very apparent in intradaily prices. A cursory examination might suggest that most prices can be readily classified as bid or ask prices. Our estimation procedure obtains information about bid/ask classification by averaging the likelihoods associated with all possible sequences of $\{Q_t\}$, taking into account trading volumes. The sequences that casual guessing would identify as being most probable have likelihood values that are orders of magnitude greater than those of other sequences. They therefore have the most influence on the estimates. The attractive feature of this procedure is that it is able to rigorously organize information about the difficult-to-classify observations, such as those continuations that occurred at about 11:45, 2:15, and 3:45.

Before considering the empirical evidence, it is useful to consider the difference between our model and the Ho and Macris (1984) inventory-theoretic spread model. Ignoring the effects of discreteness, the latter model can be written (in our notation) as

$$D_t = c(Q_t - Q_{t-1}) - b(I_t - I_{t-1}) + e_t, \tag{3}$$

where I_t is market-maker inventory just before trade t and b measures the responsiveness of the spread to inventory changes. Assuming that the specialist takes the other side of every trade gives $I_t - I_{t-1} = -Q_{t-1}V_{t-1}$, so that

$$D_t = c(Q_t - Q_{t-1}) + bQ_{t-1}V_{t-1} + e_t. \tag{4}$$

In contrast, our model with $c_1 = z_0 = 0$ and ignoring discreteness is

$$D_t = c_0(Q_t - Q_{t-1}) + z_1 Q_t V_t + e_t. \tag{5}$$

Although both inventory and adverse-selection considerations lead to changes in bid/ask prices, there are two differences between them. The obvious difference is in timing. In the inventory model, volume has a lagged effect on bid/ask prices, whereas in the asymmetric information model, volume has a contemporaneous effect. The subtle difference lies in the permanence of the volume effect. In the inventory model, bid and ask prices are adjusted by market-makers to maintain their target inventories. After a large buy (sell) order is filled, the bid and/or ask prices are raised (lowered) to increase the probability that the next order will be a sell (buy). The distribution of Q_t therefore depends on lagged Q_t and on lagged V_t. The target inventory adjustment mechanism insures that the cumulative effect of volume on prices is transitory. That is, partial sums of $\{Q_t V_t\}$ regress toward zero. In the asymmetric information model, the adverse-selection component represents a revision in price expectations, conditional on the order. These revisions are permanent in the sense that partial sums of $\{Q_t V_t\}$ do not regress.

Although price-setting mechanisms will in general affect the serial properties of the order distribution, nothing in the asymmetric information model forces this distribution to be serially independent. It is therefore possible that both inventory-theoretic and information-theoretic considerations determine spreads. In particular, inventory-theoretic considerations probably better explain the transitory component, while the information-theoretic considerations explain the contemporaneously correlated permanent component. Our model contains both transitory and permanent components, but we focus primarily on the latter, deferring to additional future work the integration of the two concepts.

3. Empirical results

In this section, we first describe the data. Section 3.2 presents estimates of the spread components under a variety of parametric assumptions. Since estimation is expensive, we examine only 20 common stocks. The most parsimonious model that yields reasonable estimates is then analyzed further.

Table 1

Cross-sectional summary statistics characterizing the specification sample consisting of the first 20 NYSE common stocks chosen in alphabetical order by ticker symbol. Each of the 20-stock time series consists of 800 transactions starting on December 1, 1981, with daily opening transactions deleted.

Cross-sectional summary statistics	Time series attributes				
	Number of price changes	Price level ($)	Average time between trades (minutes)	Market value December 1981 ($millions)	Average daily share volume (thousands)
Mean	696	20	37	444	30
Standard	65	11	18	814	43
Maximum	784	37	61	3308	150
3rd quartile	750	28	52	402	42
Median	703	16	38	60	8
1st quartile	632	10	22	38	4
Minimum	596	7	8	10	2

This cross-sectional analysis examines the estimated spread components of 250 stocks.

3.1. Data

We use transaction by transaction data supplied by Francis Emory Fitch, Inc. The data base consists of a time-ordered record of every common stock transaction on the NYSE for the fourteen months between December 1, 1981 and January 31, 1983. For the model specification search we use the first 20 firms in alphabetical order by ticker symbol, and for the model validation study we use the first 250 firms.

For each stock, we examine a time series of 800 successive prices beginning on December 1, 1981. Since opening prices are frequently determined by a call auction, we omit them. This breaks the time series into D series of truly successive price changes, where D is the number of days spanned by the 800 prices.[3] The largest and smallest numbers of successive price changes analyzed in the specification sample are 784 and 596 (table 1). These correspond to approximately three weeks of trading for the most actively traded stock and ten months for the least actively traded.

Also reported in table 1 are statistics summarizing the cross-sectional characteristics of the specification sample. There is considerable variation in mean price levels, volumes, trade frequencies, and firm sizes.

[3]The average likelihood for a given stock is computed as the product of the average likelihoods of each of the D days spanned by the data.

Included in our data set is the number of shares traded in each transaction. Many of the larger transactions are arranged off the floor. The prices of these block trades reflect information available at the time of the agreement, and not necessarily all information available at the time the trade was crossed on the floor and recorded by Fitch. To avoid giving too much weight to such nonsynchronous prices, we truncate the number of shares traded at 10,000. That is, if Fitch recorded a trade of 20,000 shares, we use the truncated figure of 10,000 shares for our analysis. The maximum truncation frequency in the specification sample was 3.44%, while the median frequency was only 0.6%.

3.2. *Model specification*

To identify a parsimonious specification that captures the spread effects and leads to estimates that conform to our prior expectations, we estimate the model under a number of varying assumptions. Almost all possible combinations of the following alternatives are examined:

(a) mean and variance of e_t linear in T_t versus constant, and
(b) various zero restrictions in the linear specifications of the two spread components.

In addition, the estimates are computed with and without price discreteness. Several considerations influence our specification decisions.

The mean and variance specification of e_t depends on whether returns are stationary in clock time or transaction time. The latter might be more appropriate for 'microstructure' analysis, since Harris (1987) presents evidence suggesting that the order flow rate is proportional to the number of information generating events.

The asymmetric information theory suggests that in the linear specification of the adverse-selection component, $Z_t = z_0 + z_1 V_t$, the constant should be zero and the slope positive. The latter prediction is discussed in the introduction. The former can be understood by considering the effect of a small trade. Since such a trade is unlikely to have been initiated by an informed trader, it should cause little revision in expectations. This implies that the adverse-selection spread should be insignificant for small trades.

Theoretical considerations concerning the specification of the transitory component are ambiguous. Although cost considerations suggest that the total transitory component should be positive, the sign of the volume coefficient, c_1, depends on whether the per-share cost of supplying liquidity services is increasing, constant, or decreasing in transaction size. If the cost is constant, c_0 will be positive and c_1 will be zero. If it is increasing, as inventory models suggest, c_1 will be positive. If it is decreasing or if there are substantial fixed costs of filling an order, c_1 will be negative. We let the specification search determine the best model.

As noted in the introduction, estimates of the transitory spread component are potentially sensitive to discreteness. Modeling the discreteness should yield more accurate estimates.

Examination of the specification search results suggests that the model with $z_0 = c_1 = 0$ and with constant e_t mean and variance, estimated without accounting for discreteness, is the most useful specification for further analysis. This is the most parsimonious model that captures the essence of the asymmetric information spread theory, and that yields reasonable, economically feasible estimates. Several results from the specification search are worth discussing.

When z_0 is simultaneously estimated with z_1 and c_0, only three of twenty z_0 estimates have asymptotic t-ratios (derived from the Hessian of the maximized average likelihood function) larger than two, and of these, two are negative and one is positive.[4] This evidence and our theoretical prediction that z_0 should be zero support our final specification.

The specification in which c_0, c_1, and z_1 are jointly estimated, while z_0 is constrained to zero is interesting because of its relation to inventory adjustment models. This specification (ignoring discreteness),

$$D_t = c_0(Q_t - Q_{t-1}) + c_1(Q_t V_t - Q_{t-1} V_{t-1}) + z_1 Q_t V_t + e_t, \tag{6}$$

is a linear transform of

$$D_t = c_0(Q_t - Q_{t-1}) + bQ_{t-1} V_{t-1} + zQ_t V_t + e_t, \tag{7}$$

with $b = -c_1$ and $z = z_1 + c_1$. The latter is our adverse-selection specification with an ad hoc inventory adjustment term added in. Only three of twenty of the b estimates in this parameterization have t-ratios greater than two, two of which are negative. Overall, only eleven estimates are negative, as the inventory model predicts. In contrast, fourteen of the z estimates have t-ratios greater than two, all positive as predicted. Only one estimate is negative. Moreover, the z estimates in this model are nearly identical to those obtained when b [or c_1 of eq. (6)] is constrained to be zero. Collectively, these results

[4] In discussing the signs of individual estimates, it is proper to note there is a very limited sense in which the parameters are not fully identified when the $\{Q_t\}$ are not observed. If the vector (c_0, c_1, z_0, z_1) maximizes the likelihood, then so too does $(-c_0, -c_1, -z_0, -z_1)$. This is because the assignment of -1 to bid prices and 1 to ask prices is arbitrary. Our estimation method generally yields estimates with signs that conform to the usual convention (-1 = bid, 1 = ask), given our theory. This is due to the sign of the vector of starting values. When the signs of the maximizing values are negative or are difficult to interpret, we appeal to economic theory to choose the best vector sign consistent with the usual sign convention. For these rare decisions, we take into account estimated t-ratios when making the decision, giving the most weight to the parameters with the greatest significance. In estimating our final specification in the 20-stock sample, we found only one security for which c_0 and z_1 were both significant and opposite in sign, and this was only for the estimates obtained when ignoring discreteness.

suggest that the volume dependency of the spread is mostly due to the adverse-selection component. The transitory component in this sample is nearly constant in volume. We therefore apply the principle of parsimony and restrict c_1 to zero for further analyses.[5]

Panel A of table 2 summarizes the cross-sectional distributions of our final spread component estimates in the specification sample. For reference, results are reported for both discreteness estimation alternatives. The average dollar spread for a round-trip transaction of V shares is $2(c_0 + z_1V)$. In this sample, the average round-trip spread for a trade of 1,000 shares (discreteness modeled) is 2(0.0242 + 0.0133) = \$0.075. For a 10,000-share trade it is 2(0.0242 + 0.0133 * 10) = \$0.31. These results show that in comparison with the transitory spread component, the adverse-selection component is economically significant for large trades but not for small ones.

All of the z_1 (discreteness modeled) estimates in the specification sample are positive with 12 of the 20 having t-ratios that are significantly different from zero at the 1% level. Not surprisingly, the cross-sectional sample mean estimate of z_1 is also significantly different from zero. Similar results are obtained when discreteness is ignored. This suggests that adverse selection is important in determining spreads. It does not trouble us that eight z_1 estimates are insignificantly different from zero, because adverse selection is not necessarily a significant problem for all stocks. The cross-sectional analysis in the next subsection shows when the problem is most serious.

As predicted, the transitory component estimates, c_0, are quite sensitive to whether or not discreteness is modeled in the estimation process. The individual estimates are lower in 19 of 20 cases when discreteness is modeled. The average z_1 estimate, however, is relatively insensitive to discreteness.

Unfortunately, the estimation procedure is an order of magnitude more costly when discreteness is modeled. This is an important consideration for our cross-sectional analysis, since we wish to examine 250 stocks. Although the level of the c_0 estimate is very sensitive to whether or not discreteness is modeled, c_0 estimates obtained using the two alternatives are highly correlated in cross-section (0.71), as are the z_1 estimates (0.88). In the interest of

[5]Consider an ad hoc specification that contains a transitory term, an inventory term, and an adverse-selection term, each a function of volume:

$$D_t = c_0(Q_t - Q_{t-1}) + c_1(Q_tV_t - Q_{t-1}V_{t-1}) + bQ_{t-1}V_{t-1} + z_1Q_tV_t + e_t.$$

Since parameters c_1, b, and z_1 in this model are not all jointly identified, additional prior information is necessary for estimation. Ho's and Stoll's 1981 model (which does not consider asymmetric information) suggests that b may be approximately twice c_1 (our notation). Substituting this relation into this ad hoc specification yields

$$D_t = c_0(Q_t - Q_{t-1}) + (b/2)Q_{t-1}V_{t-1} + (z_1 + b/2)Q_tV_t + e_t,$$

which is another reparameterization of (6) and (7). Empirical results in this parameterization are identical to those described for eq. (7). In particular, b (and hence c_1) is near zero, while z_1 is significantly positive for most securities.

Table 2

The cross-sectional distribution of estimated adverse-selection and transitory spread components in the 20-stock specification sample and the 250-stock model validation sample. The two samples consist of the first 20 and 250 NYSE common stocks chosen in alphabetical order by ticker symbol. The stock time series each consist of 800 transactions starting on December 1, 1981, with the daily opening transaction deleted. The model is $D_t = c_0(Q_t - Q_{t-1}) + z_1 Q_t V_t + e_t + r_t - r_{t-1}$, where D_t is the transaction price change, Q_t is an unobserved $(-1,1)$ indicator of bid and ask prices, V_t is trade size, c_0 is the transitory spread component, z_1 is the adverse-selection component, e_t is the unobserved innovation in true prices due to public information, and r_t is unobserved round-off error due to price discreteness. The total spread for a round-trip transaction of V thousand shares is $2(c_0 + z_1 V)$. The model is estimated using likelihood methods described in section 3.1. Estimates obtained ignoring discreteness are computed assuming that all r_t are zero.

	Discreteness ignored		Discreteness considered	
	Transitory component c_0 ($/share)	Adverse-selection component z_1 ($/share/1000 share lots)	Transitory component c_0 ($/share)	Adverse-selection component z_1 ($/share/1000 share lots)
	Panel A. 20-stock specification sample			
Mean	0.0444	0.0113	0.0242	0.0133
Standard	0.0265	0.0073	0.0244	0.0071
t-statistic	7.29	6.74	4.33	8.11
N sig at 1%	15	11	9	12
N positive	19	19	17	20
Maximum	0.0948	0.0280	0.0659	0.0290
3rd quartile	0.0690	0.0138	0.0408	0.0156
Median	0.0422	0.0098	0.0236	0.0119
1st quartile	0.0256	0.0071	0.0063	0.0081
Minimum	−0.0050	−0.0005	−0.0177	0.0027
	Panel B. 250-stock validation sample			
Mean	0.0465	0.0102		
Standard	0.0255	0.0126		
t-statistic	28.87	12.89	Not computed	
N sig at 1%	210	170		
N positive	239	222		
Maximum	0.0984	0.0878		
3rd quartile	0.0637	0.0136		
Median	0.0503	0.0075		
1st quartile	0.0296	0.0028		
Minimum	−0.0377	−0.0071		

economy, we therefore perform the cross-sectional validation tests on estimates we obtain ignoring discreteness. Given the high cross-sectional correlation, we believe that valid inferences can be drawn from the simpler estimates.

To determine whether the specification sample adequately represents the 250-stock validation sample, we collected statistics summarizing the cross-sectional distributions of the spread component estimates in the latter sample (table 2, panel B). Comparison with panel A shows that the two samples are quite similar. The mean c_0 estimate is 0.0444 in the 20-stock sample and

0.0465 in the 250-stock sample. For the z_1 estimate, these means are 0.0113 and 0.0102.

3.3. Cross-sectional analysis

The asymmetric information spread theory provides a number of cross-sectional predictions relating the two spread components to other stock characteristics. We examine these predictions using estimated spread components for 250 stocks.

The analysis has two interpretations. If we accept the asymmetric information spread theory, these cross-sectional investigations provide evidence of whether the estimates we obtain from our time-series model actually contain information on the concepts we claim to be estimating. Alternatively, if we accept that the time-series estimates are estimates of contemporaneously correlated transitory and 'permanent' components in the stock price innovation process, these cross-sectional analyses provide evidence of whether these components can be interpreted as spread components within the asymmetric information context. Of course, since neither conditioning argument is known, tests in this cross-sectional analysis are joint tests of the time-series estimates and of the asymmetric information spread theory.

Our cross-sectional model consists of four simultaneous equations. The first two explain the two spread components in terms of a number of variables, one of which is trade frequency. Since trade frequency is probably itself a function of the spread, it is modeled in the third equation. The fourth equation, trade size, is included for interest. The entire system is jointly estimated using appropriate simultaneous methods.

Rather than modeling the absolute spread components, we examine them as a percentage of price. This specification, which Branch and Freed (1977) also use, focuses attention on the economic significance of the spread to a trader.

We begin by discussing determinants of the transitory spread component. Ho and Stoll (1981) consider an inventory-theoretic model in which security risk and transaction frequency determine risk-aversion-induced inventory costs. In a competitive market, these costs are recovered through the transitory spread component (Ho and Stoll do not consider an asymmetric information environment). The higher the security risk and the more time between trades, the higher the transitory spread should be. We adopt these predictions. As a proxy for security risk, we use the weekly return standard deviation calculated over the prior eleven months (*WKSD*).[6] As a proxy for trade frequency, we use the inverse of the average number of trades per day (*INVNT*). Adding an

[6]Branch and Freed (1977) argue that firm-specific risk is the appropriate risk measure. The analysis of Ho and Stoll (1981), however, suggests that total risk is appropriate, and we adopt this formulation. A weekly rather than daily or intradaily measure is used because the transitory component is a source of total price variation. Using the weekly standard deviation minimizes the fraction of the risk measure that can be explained by c_0/P.

error term yields the first equation in our model:

$$c_0/P = a_0 + a_1 INVNT + a_2 WKSD + e_1. \tag{8}$$

The dependent variable in the adverse-selection component equation is the average adverse-selection spread paid on a typical trade: z_1 times the average number of shares traded per transaction, divided by the price level ($AVGZ/P$). This should be a function of the informed trade frequency, the liquidity trade frequency, and [as shown in Glosten (1987a)] the transitory spread component.

As a proxy for informed activity, we use insider ownership concentration (IC), defined as the proportion of shares owned by legally defined insiders (top management and 5% reporters) and persons with an obvious relationship to top management. This information is collected from the firms' proxy reports for the previous year. We expect that the larger this variable is, the more likely it is that a trade is initiated by someone with information, and hence the larger the adverse-selection spread.

If there are many shareholders, however, the probability that any trade is information related could be small even if insider ownership concentration is high. We use the number of noninsider shareholders (NSH) as a proxy for the frequency of liquidity motivated trade. We expect that the larger the number of noninsider shareholders, the smaller should be the adverse-selection spread.

Finally, the adverse-selection spread component should be positively related to the transitory spread component. The adverse-selection component is essentially the revision in expectations resulting from a trade. The wider the transitory spread, the less likely is a trade of any type, but especially a liquidity motivated trade. When the transitory spread is small, the relative frequency of informed trade should increase, and so should the adverse-selection spread. Moreover, in the presence of a large transitory component, profitable informed trade can take place only if informed signals are very large. This also implies a large adverse-selection spread. Our second equation to be estimated is thus:

$$AVGZ/P = b_0 + b_1 C/P + b_2 IC + b_3 NSH + e_2. \tag{9}$$

Although the return standard deviation, insider concentration, and number of shareholders can reasonably be assumed to be exogenous, the same cannot be said for the inverse average number of trades. We expect average number of trades per day to be negatively related to the total spread, since a large spread reduces the attractiveness of all types of trade. Rather than modeling the inverse of this average, we model the average itself as a function of the total proportional average spread, $AVGSP/P = 2(c_0/P + AVGZ/P)$, and the number of noninsider shareholders. The more shareholders there are, the more

Table 3

Estimates obtained from cross-sectional regressions of the 4-equation model (described in section 3.3). The first two equations of the model relate time-series estimates of the transitory and adverse-selection spread components to a set of predictors which include proxies for security risk, adverse-selection risk, and trading activity. The transitory component is expected to increase with security risk (represented by the weekly stock return standard deviation) and with thin trading (represented by the inverse average number of trades per day). The adverse-selection component is expected to increase with the risk of informed trade (represented by insider concentration), decrease with the extent of liquidity trade (represented by the number of shareholders), and increase with the size of the transitory component. Two of these predictors, the average number of trades per day and the average volume per trade depend on the spread components. The third and fourth equations model the joint dependency. The average number of trades per day is expected to decrease with the total size of the spread and increase with the number of shareholders. The average volume per trade is expected to decrease with the adverse-selection component of the spread and increase with the average shareholdings by outsiders. The system is estimated using three-stage nonlinear least squares. The sample consists of the first 250 NYSE common stocks chosen in alphabetical order by ticker symbol. Spread components for each stock are obtained from time series estimations of (2) with $c_1 = z_0 = 0$ and ignoring discreteness. The stock time series each consist of 800 transactions starting on December 1, 1981, with the daily opening transaction deleted. Asymptotic t-statistics are in parentheses.

Dependent variables

c_0/P = estimated transitory spread component as a percent of average price,
$AVGZ/P$ = estimated adverse-selection spread component for a typical trade, computed as z_1 times the average volume per trade (measured in thousands of shares) divided by price,
NT = average number of trades per day,
$AVGVOL$ = average volume per trade (in 1,000s of shares).

Endogenous variables

$INVNT$ = inverse of the average number of trades per day,
$AVGSP/P$ = the total average spread as a percentage of price, computed as twice the sum of c_0/P and $AVGZ/P$,
z_1/P = estimated adverse-selection spread component per 1,000 shares transacted, divided by price.

Exogenous variables

$WKSD$ = weekly return standard deviation in percent for the eleven months prior to December 1, 1981,
IC = insider concentration ratio, defined as the percentage of shares held by officers, directors, and 5% reporters with obvious relation to officers or directors, from the 1982 proxy reports,
NSH = number of shareholders (thousands) not including those counted in IC, from 1982 proxy reports,
AH = average share holdings (thousands) of noninsiders.

Dependent variable	Constant	Endogenous variable				Exogenous variable			
		$INVNT$	c_0/P	$AVGSP/P$	z_1/P	$WKSD$	IC	NSH	AH
C_0/P	−3.34	1.24				0.115			
	(−2.84)[a]	(2.73)[a]				(5.36)[a]			
$AVGZ/P$	0.0172		0.0215				0.000218	−0.000105	
	(4.48)[a]		(3.71)[a]				(1.70)	(−2.00)[b]	
NT	15.40			4.66				0.398	
	(3.85)[a]			(1.29)				(6.38)[a]	
$AVGVOL$	0.848				−4.49				0.813
	(21.74)[a]				(−9.41)[a]				(5.45)[a]

[a]Significant at the 1% level.
[b]Significant at the 5% level.

trades per day there should be. Thus, the third equation in our model is

$$NT = c_0 + c_1 AVGSP/P + c_2 NSH + e_3. \tag{10}$$

The last equation in our system models the average volume per trade, *AVGVOL*. It is not essential to the objectives of this subsection. Rather, it is included to demonstrate how trade size might depend on the spread. We model average volume per trade as a function of the relative adverse selection slope coefficient, z_1/P, and the average holdings by outsiders, *AH*. The larger the relative adverse-selection slope coefficient, the more costly are large trades in relation to small ones. We therefore expect average volume per trade and the relative adverse-selection slope coefficient to be negatively related. The larger are average outsider holdings, the more likely is it that liquidity-motivated trades will be large. We therefore expect average volume per trade and average outsider holdings to be positively related. The last equation in our model is

$$AVGVOL = d_0 + d_1 Z/P + d_2 AH + e_4. \tag{11}$$

Table 3 reports the results of using three-stage nonlinear least squares to estimate the model. The signs of the estimated coefficients agree with the above discussion in every case but one – the coefficient of the total proportional spread is positive in the number-of-transactions equation. This estimate, however, is not statistically different from zero. Of the other estimates, all but one are significantly different from zero at the 5% level. The insignificant estimate is the insider concentration coefficient in the adverse-selection spread component equation. Perhaps information from which market-makers must protect themselves is related to superior analytical ability among some investors rather than information obtained by legally defined insiders.

Overall, we find these results encouraging. The data are unable to reject this specification of the asymmetric information spread model. Although other models might be consistent with these results, we believe the evidence suggests that the adverse-selection component is at least one determinant of the total spread.

4. Conclusion

We have presented a simple asymmetric information model in which the bid/ask spread is broken into a transitory component and an adverse-selection component. The model was estimated using transaction price time series and the estimates were analyzed in cross-sectional regressions. The results from the time-series analysis are unable to reject the hypothesis that the adverse-selection component is positive. The cross-sectional analysis is unable to reject related predictions of the asymmetric information theory. Spreads appear to be determined to some extent by the exposure of market-makers to trader who are better informed than themselves.

We should mention some of the limitations placed on us by the data. Although we implicitly treat every trade recorded by Fitch as independent, this occasionally is not so. A large trade may include executions of several separate limit orders at different prices. They will be recorded as separate trades but this fact is not included on the Fitch tape. Sensitive to this problem, Hasbrouck and Ho (1986) ignore trades that occurred close in time. We do not, because not all close trades result from this process. In using all trades, we may bias upward our estimates of the adverse-selection slope coefficient, since there will be times when a seemingly small trade is associated with a large 'permanent' price change. Some evidence gathered in the specification search, however, suggests that this may not be a serious problem. When we estimated a specification of the adverse-selection component that included a constant term, the constant was near zero and the slope estimate was not significantly smaller than that estimated for the restricted model. If there were many small transactions caused by the breakup of large orders, the constant would have been positive and the slope smaller.

The inventory cost model of Ho and Macris (1984) and the asymmetric information spread model are similar but not identical. As discussed above, spreads probably are determined both by asymmetric information and by inventory considerations. Further research should combine these two effects in a more rigorous model than that postulated in eq. (7) as an adverse-selection specification with an ad hoc inventory adjustment term. Doing so will require much additional work, since the transitory and adverse-selection components of the spread interact. If inventory considerations cause bid or ask prices to change, the inference problem faced by market-makers changes causing the adverse-selection part of the spread to change.

The model and estimation procedures presented in this paper assume that neither spread component changes through time. In reality, this is unlikely, especially near events that generate new information. Further research should estimate and examine spread components around such significant events as earning announcements, dividend announcements, and takeover attempts. If spreads widen, as seems likely, it would be interesting to see whether the widening is due to the adverse-selection component, as the information asymmetry model would predict.

Finally, our results showing that the spread is a function of trade size have implications for additional studies into the relation between transaction costs and expected returns. Recent work by Constantinides (1986) and Amihud and Mendelson (1986) derive relations between expected returns and liquidity measures. Since an important aspect of liquidity is the ability to make large trades without affecting price, price–liquidity studies should examine not only the width of the spread for a typical trade, but also how this changes with trade size.

References

Amihud, Yakov and Haim Mendelson, 1986, Asset pricing and bid–ask spread, Journal of Financial Economics 17, 233–249.

Bagehot, Walter (Jack Treynor), 1971, The only game in town, Financial Analysts Journal 22, 12–14.

Benston, George J. and Robert L. Hagerman, 1974, Determinants of bid–asked spreads in the over-the-counter market, Journal of Financial Economics 1, 353–364.

Blume, Marshall E. and Robert F. Stambaugh, 1983, Biases in computed returns: An application to the size effect, Journal of Financial Economics 12, 387–404.

Branch, Ben and Walter Freed, 1977, Bid–asked spreads on the AMEX and the big board, Journal of Finance 32, 159–163.

Clark, Peter K., 1973, A subordinated stochastic process model with finite variance for speculative prices, Econometrica 41, 135–155.

Cohen, Kalman J., Steven F. Maier, Robert A. Schwartz, and David K. Whitcomb, 1979, On the existence of serial correlation in an efficient securities market, TIMS Studies in the Management Sciences 11, 151–168.

Constantinides, George M., 1986, Capital market equilibrium with transactions costs, Journal of Political Economy 94, 842–862.

Copeland, Thomas and Dan Galai, 1983, Information effects on the bid/ask spread, Journal of Finance 25, 383–417.

Easley, David and Maureen O'Hara, 1987, Price, trade size and information in securities markets, Journal of Financial Economics 19, 69–90.

French, K.R., 1980, Stock returns and the weekend effect, Journal of Financial Economics 8, 55–70.

French, K.R. and Richard Roll, 1986, Stock return variances: The arrival of information and the reaction of traders, Journal of Financial Economics 17, 5–26.

Glosten, Lawrence R., 1987a, Components of the bid/ask spread and the statistical properties of transaction prices, Journal of Finance 42, 1293–1308.

Glosten, Lawrence R., 1987b, Insider trading, liquidity and the role of the monopolist specialist, Working paper (Northwestern University, Evanston, IL).

Glosten, Lawrence R. and Paul R. Milgrom, 1985, Bid, ask and transaction prices in a specialist market with heterogeneously informed traders, Journal of Financial Economics 14, 71–100.

Gottlieb, Gary and Avner Kalay, 1985, Implications of the discreteness of observed stock prices, Journal of Finance 40, 135–153.

Harris, Lawrence, 1986, Estimation of true price variances and bid–ask spreads from discrete observations, Working paper (University of Southern California, Los Angeles, CA).

Harris, Lawrence, 1987, Transaction data tests of the mixture of distributions hypothesis, Journal of Financial and Quantitative Analysis 22, 127–141.

Hasbrouck, Joel and Thomas S.Y. Ho, 1987, Order arrival, quote behavior, and the return-generating process, Journal of Finance 42, 1035–1048.

Ho, Thomas S.Y. and Richard G. Macris, 1984, Dealer bid–ask quotes and transaction prices: An empirical study of some AMEX options, Journal of Finance 39, 23–45.

Ho, Thomas S.Y. and Hans Stoll, 1981, Optimal dealer pricing under transactions and return uncertainty, Journal of Financial Economics 9, 47–73.

Holthausen, Robert, Richard Leftwich, and David Mayers, 1987, The effect of large block transactions on security prices: A cross-sectional analysis, Journal of Financial Economics 19, 237–267.

Keim, Donald B., 1983, Size-related anomalies and stock return seasonality: Further empirical evidence, Journal of Financial Economics 12, 13–32.

Kyle, Albert S., 1985, Continuous auctions and insider trading, Econometrica 53, 1315–1335.

Niederhoffer, Victor and M.F.M. Osborne, 1966, Market making and reversals on the stock exchange, Journal of the American Statistical Association 61, 897–916.

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