ELSEVIER Journal of Financial Economics 38 (1995) 361–381

JOURNAL OF Financial ECONOMICS

Investment analysis and price formation in securities markets

Michael J. Brennan[a,b], Avanidhar Subrahmanyam*[,a,c]

[a]*Anderson Graduate School of Management, University of California, Los Angeles, CA 90024, USA*
[b]*London Business School, London NW1 4SA, UK*
[c]*Graduate School of Business, Columbia University, New York, NY 10027, USA*

(Received October 1993; final version received September 1994)

Abstract

This paper investigates the relation between the number of analysts following a security and the estimated adverse selection cost of transacting in the security, controlling for the effects of previously identified determinants of liquidity. Using intraday data for the year 1988, we find that greater analyst following tends to reduce adverse selection costs based on the Kyle (1985) notion of market depth. This result is consistent with the analysis of Admati and Pfleiderer (1988). Estimates of structural parameters of a version of the Admati and Pfleiderer model of endogenous information acquisition provide qualified support for the model.

Key words: Security analysis; Market depth; Asymmetric information
JEL classification: G14; D82

1. Introduction

The important paradigms of price formation in securities markets developed by Kyle (1984, 1985) and Admati and Pfleiderer (1988) suggest that trading by investors who possess superior information imposes significant liquidity costs on other market participants due to adverse selection, which we call *the adverse selection costs of transacting.* These theoretical models have stimulated the

*Corresponding author.

We are indebted to John Long (the editor) and an anonymous referee for their detailed and insightful suggestions. We also thank Larry Harris, Richard Roll, and participants at the UCLA finance seminar for helpful comments.

SSDI 0304405X9400811 E

development of empirical techniques for measuring the effect of informed trading on market liquidity in Glosten and Harris (1988), Madhavan and Smidt (1991), and Hasbrouck (1991), among others. Although these papers report significant evidence of adverse selection costs due to information-based trading in financial markets, for the most part they provide little empirical evidence on the cross-sectional determinants of the size of these costs. A notable exception is the paper by Glosten and Harris (1988), which reports a very weak association between an estimate of adverse selection costs and a measure of insider holdings, which is taken as a proxy for the intensity of informed trading.

We provide further evidence on the effect of information-based trading on liquidity costs by analyzing the empirical relation between the number of investment analysts following a stock and the estimated adverse selection cost of transacting in the stock, controlling for the effects of trading volume, price level, and return volatility. The number of investment analysts researching a firm is a simple proxy for the number of individuals producing information about the value of the firm, based on Brennan, Jegadeesh, and Swaminathan (1993), who find that stocks that are followed by many analysts react faster to common information than stocks that are followed by few analysts. The adverse selection cost is defined as the price impact of a marginal dollar of trade, and, apart from a price scale factor, is proportional to the inverse of the Kyle (1985) measure of market depth (in the Kyle, 1985, and Admati and Pfleiderer, 1988, models, depth is given by the reciprocal of the regression coefficient of the price change on the order flow). Recent theoretical work leads us to expect a relation between the adverse selection cost of transacting in a security and the number of individuals producing information about the security.

Thus the Admati–Pfleiderer model, which assumes that the information asymmetry is short-lived, predicts either a positive or a nonmonotone (Subrahmanyam, 1991) relation between the number of informed traders and market depth. However, when information is long-lived, an increase in the number of informed traders will tend to increase the rate at which private information comes to be reflected in price; consequently, market depth will be lower in the early rounds of trading, when the information disadvantage of the market maker is greatest, but the effect will be reversed in later rounds as the market maker gains more information from the order flow. Holden and Subrahmanyam (1992) present an explicit model of market depth with multiple informed traders and long-lived information.

While the number of analysts following a stock is an imperfect proxy for the number of informed traders, the influence of security analysis on market depth is an issue of interest in its own right, since security analysis is a costly activity whose social benefits remain largely unexplored (see, however, Brennan, Jegadeesh, and Swaminathan, 1993; Arbel, Carvell, and Strebel, 1983). The empirical relation between the number of analysts and market depth is particularly relevant in light of the positive relation between market illiquidity and required rates of return derived, for example, in Amihud and Mendelson (1986).

Our empirical results may be summarized as follows. Other things equal, an increase in the number of investment analysts tends to be associated with a reduction in the adverse selection costs of transacting, as would be predicted by the model of Admati and Pfleiderer (1988). Structural estimates of a non-linear simultaneous equation specification of the model of endogenous information acquisition developed by Admati and Pfleiderer are broadly supportive of the model. However, a more general specification which allows the number of analysts to depend on the degree of institutional participation, as in Bhushan (1989), performs better in characterizing the market for information.

In Section 2 we briefly summarize some of the recent theoretical literature on the effect of informed traders on market depth. In Section 3 we describe the data used in the empirical tests. Section 4 describes the estimation of the measures of market depth, while Section 5 presents the empirical results relating market depth to the number of investment analysts following a stock. Section 6 concludes.

2. Competition and market depth

To see the effect of the number of informed traders on market depth in a single-period setting, consider a special case of the model of Admati and Pfleiderer (1988), in which n risk-neutral traders receive a perfectly informative signal about the final payoff, u, on an asset. The informed traders, as well as uninformed noise traders, place market orders with a competitive, risk-neutral market maker who fills the orders at a single price, P, which depends on the total order flow q:

$$P = \mathrm{E}[u] + \lambda q\,, \tag{1}$$

where $\mathrm{E}[u]$ is the unconditional expectation of the asset payoff. Admati and Pfleiderer show that λ, *the inverse of* market depth, is given by

$$\lambda = \frac{\sqrt{n}}{n+1}\sqrt{\frac{\mathrm{var}(u)}{\mathrm{var}(z)}}\,, \tag{2}$$

where $\mathrm{var}(z)$ is the variance of the orders placed by noise traders.

Eq. (2) implies that λ is decreasing in n, the number of informed traders, for $n > 1$, so that market depth is increasing in the number of informed traders. Admati and Pfleiderer also show that if there is a fixed cost of acquiring information, c, and the number of informed traders is determined endogenously, then n, the equilibrium number of informed traders, satisfies (except for the integer constraint):

$$\frac{\mathrm{var}(u)}{(1+n)^2\lambda} = c\,. \tag{3}$$

In general, λ depends on the sensitivity of the total order flow, q, to the information signal. Subrahmanyam (1991) shows that a risk-neutral market maker will set λ according to

$$\lambda = \frac{t\,\mathrm{var}(u)}{t^2\,\mathrm{var}(u) + \mathrm{var}(z)}, \tag{4}$$

where t, the intensity of informed trading, is equal to the coefficient of the perfectly informative signal of u in the informed traders' aggregate order function. While t is monotone increasing in n in the model of Subrahmanyam (1991), it can be seen from Eq. (4) that λ is a nonmonotonic function of t. In the model of Admati and Pfleiderer, in which the informed traders are risk-neutral, t is always sufficiently large that $\partial\lambda/\partial t < 0$, and therefore $\partial\lambda/\partial n < 0$. In Subrahmanyam (1991), however, risk aversion reduces the intensity of trading by the informed traders so that, for small n, $\partial\lambda/\partial t > 0$, and therefore $\partial\lambda/\partial n > 0$ for small n. In summary, the model of Admati and Pfleiderer predicts a negative relation between λ and the number of informed traders, while Subrahmanyam's model with risk-averse traders predicts that the relation will be negative only when the number of informed traders is sufficiently large. Eq. (3) shows that the number of informed traders cannot be taken as exogenous, but will be determined in equilibrium by the costs and benefits of becoming informed.

When information is long-lived, predictions regarding market depth are more ambiguous. Holden and Subrahmanyam (1992) show that in a model with a Kyle (1985) market maker and competing informed traders, the informed investors trade much more aggressively than in the monopoly case considered by Kyle. As a result, the price reflects their private information much more rapidly than in the monopoly case, which causes the market depth to increase with the number of informed traders, except possibly in the first few auctions if the number of informed traders is small. Overall, this analysis strongly suggests that market depth, though time-varying, will be higher on average the greater is the number of informed traders. It also suggests that the effect of an increase in the number of informed traders will be nonlinear, being greatest when the number of informed traders is small.

Thus, while models in which private information is short-lived suggest that an increase in the number of informed traders will increase market depth, models in which private information is long-lived have more ambiguous predictions, so that empirical evidence on the issue is of particular interest.

Our empirical work focuses on the relation between the number of analysts following a firm and estimates of the (inverse of) market depth, λ, holding constant factors which previous authors have found to be associated with market liquidity (see Benston and Hagerman, 1974; Branch and Freed, 1977). Glosten and Harris (1988) estimate λ for a small sample of NYSE securities and find that the estimated adverse selection trading costs for average size transactions are negatively related to the number of shareholders in the firm and (insignificantly) positively related to the concentration of insider holdings in the firm. They interpret their results as consistent with a Kyle-type model in which the adverse selection problem faced by the market maker is an increasing function of insider concentration, and the volume of noise trading is proxied by the number of shareholders. They suggest (p. 140) that the coefficient of the insider concentration variable may be insignificant because 'the 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'. We examine whether the information from which market makers must protect themselves is related in particular to the superior analytical ability and investment in information of security analysts.

In our empirical work, we rely primarily on the procedures developed by Glosten and Harris (1988) and Madhavan and Smidt (1991) to measure the adverse selection costs of transacting. However, to assess the robustness of the results to the estimation procedure for the Kyle λ, we also follow a procedure used by Foster and Viswanathan (1993) which is based on Hasbrouck (1991).

3. Data

The data employed in the empirical tests reported below were provided by the Institute for the Study of Security Markets and consist of intraday quotes as well as transaction prices and quantities for 1,550 common stocks that were listed continuously on the NYSE for the calendar year 1988. To minimize data errors, the data were screened as follows. First, quotations and transactions reported out of sequence were excluded. Second, the overnight price change and the closing quotes were omitted to eliminate price effects associated with opening and closing procedures, dividend payments, and overnight news arrival. Third, an error filter was used to screen out intraday reporting errors. The error filter discards a trade if the trade price is too far outside the price range defined as the minimum range that includes the *preceding* bid and ask quotations and the *immediately following* trade price or bid and ask quotations. If the price falls outside this range by more than four times the width of the range, the trade is discarded. This filter is conservative and discards fewer than one in 40,000 observations in the sample considered.

For each security, the number of investment analysts following the firm is defined as the number of analysts making an annual earnings forecast for that firm in December 1987, according to the Institutional Brokerage Earnings Estimates tape.[1] The daily return variance, the average daily trading volume in shares, and the average daily closing price are computed using data for 1988 from the Center for Research in Security Prices' New York Stock Exchange/American Stock Exchange daily tape. Finally, the *S&P Security Owners' Stock Guide* provides the number of institutions reported as owning shares in each company and the number of shares held by institutions as of December 1987. Forty-two companies in the sample did not have these data available, leaving 1,508 companies for which data on all variables were available.

4. Estimation of adverse selection costs

Before examining the relation between adverse selection costs and the number of analysts following a stock, we estimate λ, the inverse of market depth in a Kyle-type model. To facilitate comparison with earlier studies, we initially estimate λ using two different procedures developed by Glosten and Harris (1988) and Madhavan and Smidt (1991), respectively. The analysis of Glosten and Harris follows Kyle (1985) in assuming that investors can place only unconditional market orders. This assumption is implicit in the manner in which they model the adverse selection component of the spread [see Eq. (1a) of their paper]. Madhavan and Smidt, on the other hand, explicitly assume that informed investors condition their order flow on the price. Since the NYSE allows both market and limit orders, neither measure of (the inverse of) market depth is entirely appropriate, and by using both measures in our regressions we are able to assess the sensitivity of our results to the assumptions made about the order submission protocol.

To understand how Glosten and Harris relate λ to the time-series behavior of prices, let m_t denote the expected value of the security conditional on the market maker's information set at time t. Then, consistent with the Kyle (1985) model in which informed traders place market orders, the expectation will evolve according to

$$m_t = m_{t-1} + \lambda q_t + y_t, \tag{5}$$

where q_t is the (signed) order flow at time t and y_t is the public information innovation. It is standard in the empirical microstructure literature to allow for a fixed cost component of the price impact of a trade. This component compensates the market maker for the costs associated with operating a market. To

[1] We are grateful to Lynch, Jones, and Ryan for making these data available.

model the fixed cost component of the price response to a transaction, Glosten and Harris proceed as follows.[2] Let D_t denote the sign of the incoming order at time t ($+1$ for a buyer-initiated trade and -1 for a seller-initiated trade). Denoting the fixed-cost component by ψ, we can write

$$p_t = m_t + \psi D_t . \tag{6}$$

Substituting out m_t using (5), we have

$$p_t = m_{t-1} + \lambda q_t + \psi D_t + y_t . \tag{7}$$

However, since $p_{t-1} = m_{t-1} + \psi D_{t-1}$, we obtain

$$\Delta p_t = \lambda q_t + \psi [D_t - D_{t-1}] + y_t . \tag{8}$$

Eq. (8), which ignores the discrete nature of price quotes, is used to estimate the Glosten–Harris λ for each NYSE-listed stock for the year 1988. (Glosten and Harris find that their estimates of λ are not sensitive to the precise specification of the distribution of the equation error.)

To relate λ to the time-series behavior of prices in the Madhavan and Smidt (1991) model, let the μ_t denote mean of the private information. The order flow, q_t, which (contrary to the assumption of the Kyle model) depends on the price, p_t, can then be written as

$$q_t = \alpha(\mu_t - p_t) + z_t , \tag{9}$$

where z_t is the liquidity trading component. The risk-neutral market maker who sees $\tau_t \equiv p_t + \alpha^{-1} q_t$ as a noisy measure of μ_t will set the price according to

$$p_t = m_t + \psi D_t = \pi \gamma_t + (1 - \pi)\tau_t + \psi D_t , \tag{10}$$

where γ_t is the market maker's prior mean of the asset's value and π is the Bayesian weight placed on the prior observation. Now,

$$p_{t-1} = m_{t-1} + \psi D_{t-1} ,$$

which can be written as

$$\gamma_t = p_{t-1} - \psi D_{t-1} + \eta_t ,$$

where $\eta_t \equiv \gamma_t - m_{t-1}$. Substituting for γ_t in (10), we have

$$p_t = \pi(p_{t-1} - \psi D_{t-1} + \eta_t) + (1 - \pi)[p_t + \alpha^{-1} q_t] + \psi D_t . \tag{11}$$

Rewriting,

$$\Delta p_t = \lambda q_t + \frac{\psi}{\pi} D_t - \psi D_{t-1} + \eta_t , \tag{12}$$

[2]Glosten and Harris (1988) ignore inventory holding costs, which appear to be small in an intraday setting (see, for example, Stoll, 1989; George, Kaul, and Nimalendran, 1991; Madhavan and Smidt, 1991).

where $\lambda \equiv \alpha^{-1}(1-\pi)/\pi$. Madhavan and Smidt (1991) show that the error term η_t follows an MA(1) process in their framework. Further, Eq. (12) differs from Eq. (8) in the coefficients of D_t and D_{t-1}: the difference arises from the different assumptions about the dependence of the order flow on the price. Eq. (12) is used to obtain our estimate of the Madhavan–Smidt λ.

To estimate λ from either Eq. (8) or Eq. (12), it is necessary first to estimate D_t, the sign of the order quantity. We use the procedure suggested by Lee and Ready (1991): if a transaction occurs above the prevailing quote mid-point, it is regarded as a purchase, and if it occurs below the prevailing quote mid-point, it is regarded as a sale. If a transaction occurs exactly at the mid-point, it is signed using the 'tick' test, which assigns a positive sign to the trade if the price move from the previous transaction price is upward, and vice versa. If the price is the same as the previous transaction price, the test is applied using the last price following which there was a move.

Given the series of prices and signed order quantities, λ is estimated from both Eqs. (8) and (12). In conformance with the theoretical specifications of the Glosten and Harris (1988) and Madhavan and Smidt (1988) models, we assume an i.i.d. error process in the Glosten–Harris specification and an MA(1) error process in the Madhavan–Smidt specification. Further, to take account of possible misspecifications, we allow for intercepts in each of the two regression specifications. A significant assumption underlying empirical measures of market depth that rely on an analysis of the relation between price change and order flow is that the public information innovation [y_t in Eq. (8) and η_t in Eq. (12)] is uncorrelated with the order flow, q_t.[3] For example, if, contrary to the assumption of the Kyle (1985) model, market makers systematically 'lean against the wind' by contrarian trading, the estimated value of λ will be biased, which will affect our point estimates of trading costs. However, there is no reason to believe that the bias is related to the number of analysts following the stock, which would be necessary if it were to affect our inferences about the effect of investment analysis on market depth.

Of the 1,508 companies in the original sample, 87 yielded negative estimates for at least one of the λ's and were eliminated from the sample to facilitate estimation of the log-linear specifications we posit below. For each measure of λ, the adverse selection cost of transacting is estimated by dividing λ by the average daily closing price, *PRI*.

In the Kyle model, the total adverse selection cost of trading q shares is λq^2. Given the price, P, the marginal cost per dollar of transaction when q shares are traded is thus $2\lambda q/P$. Table 1 provides summary statistics on the marginal cost per dollar of transaction for the two measures of λ (in the table q is set equal to

[3]See Glosten and Harris (1988), Hasbrouck (1991), Madhavan and Smidt (1991), and Foster and Viswanathan (1993).

Table 1
Summary statistics of estimated market depth for the year 1988 and analyst following as of December 1987 for a sample of 1,421 stocks continuously listed on the NYSE for the year 1988 for which complete data were available on institutional holdings as of December 1987 and for which nonnegative estimates of market depth were obtained

λ_{GH} is the estimate of the inverse of market depth from the Glosten–Harris specification:

$$\Delta p_t = \lambda q_t + \psi [D_t - D_{t-1}] + y_t;$$

λ_{MS} is the estimate of the inverse of market depth from the Madhavan–Smidt specification:

$$\Delta p_t = \lambda q_t + \frac{\psi}{\pi} D_t - \psi D_{t-1} + \eta_t.$$

Δp_t is the price change at transaction t, q_t is the signed trade size, D_t is a dummy variable that is equal to $+1$ for a trade classified as a buy and -1 for a sell, and y_t and η_t are error terms. The Glosten–Harris model is estimated assuming that y_t is i.i.d., while the Madhavan–Smidt model is estimated assuming that η_t is MA(1). *PRI* is the average daily closing price for 1988.

Number of firms = 1,421	Mean	Median	Standard deviation
$(1{,}000)\lambda_{GH}/PRI$	0.0314	0.0082	0.0701
$(1{,}000)\lambda_{MS}/PRI$	0.0156	0.0046	0.0371
Number of analysts	8.90	5	9.99

500 shares). The mean of the marginal cost of purchasing 500 shares is 3.14% for the Glosten–Harris specification and 1.56% for the Madhavan–Smidt specification. The Glosten–Harris measure is approximately twice as variable in the cross-section as the Madhavan–Smidt measure. The correlation between the two measures of adverse selection cost is 0.93. To assess the sensitivity of our results to model specification, we use both estimates in our analysis.

Table 1 also provides summary statistics for the number of analysts following each firm. The distribution is highly skewed: the mean number is 8.9 while the median is 5, and 438 out of the total sample of 1,421 were not covered by the I/B/E/S service.

5. Empirical results

5.1. Cross-sectional determinants of the adverse selection cost of transacting

Ordinary least-squares (OLS) regressions in which the dependent variable is an estimate of the adverse selection cost are likely to be biased and inconsistent, because trading volume, a primary determinant of this cost, and the number of analysts, which is the key variable in our analysis, may both be affected by the cost of transacting. We therefore adopt a simultaneous equations approach. Following the earlier empirical work on the determinants of the bid–ask spread (Benston and Hagerman, 1974; Branch and Freed, 1977), the first equation explains the logarithm of the adverse selection cost, *LTC*, as a linear function of the logarithm of the volume of trading (*LVOL*), as measured by the average number of shares traded per day during the year, the logarithm of the stock price (*LPRI*) measured by the average daily closing price during the year, and the logarithm of the daily return variance measured over the year (*LVAR*). To introduce the number of analysts in a consistent manner, *LANAL* is defined as the logarithm of one plus the number of analysts allowing us to include in the regression firms for which no analyst is reported by I/B/E/S. This definition is also consistent with the notion that there is some informed trading even in the absence of security analysis reported on the I/B/E/S tape.

The second equation explains *LANAL*, the (log) number of analysts, in terms of the adverse selection cost variable, *LTC*, and the logarithms of variance, size, and price. Following Bhushan (1989), it also includes five industry dummies and *LNINST* and *LPINST*, the (log) number of institutions holding shares in the company and the (log) percentage of shares held by institutions, respectively. The third equation explains *LVOL*, the logarithm of trading volume, in terms of the trading cost variable, *LTC*, as well as *LANAL* and *LSIZE*. Thus the following equation system was estimated by two-stage least squares:

$$LTC = a_{S0} + a_{S1}LANAL + a_{S2}LVOL + a_{S3}LPRI + a_{S4}LVAR + e_{TC}\,, \tag{13}$$

$$\begin{aligned} LANAL = {} & a_{A0} + a_{A1}LTC + a_{A2}LVAR + a_{A3}LSIZE + a_{A4}LPRI \\ & + \sum_{i=1}^{5} a_{Ai+4}IND_i + a_{A10}LNINST + a_{A11}LPINST + e_{ANAL}\,, \end{aligned} \tag{14}$$

$$LVOL = a_{V0} + a_{V1}LTC + a_{V2}LANAL + a_{V3}LSIZE + e_{VOL}\,, \tag{15}$$

where IND_i is a dummy variable corresponding to one of five industry classifications; the industry classifications are obtained from the COMPUSTAT tapes and follow Bhushan (1989). The first and third equations in the above system are identified, while the second is underidentified. Table 2 reports the two-stage least-squares parameter estimates for the two identified equations of the system. The analysis is reported for the transaction cost variable, *LTC*, computed for each of the two measures of λ. The results for both the measures of λ are qualitatively similar.

Considering first the *LTC* regressions for the determinants of the adverse selection cost, the coefficient of the number of analysts is negative and significant for both measures of λ. This finding is consistent with the prediction of the

Table 2

Two-stage least-squares estimates of determinants of adverse selection cost of transacting, using two empirical measures of λ:

$$LTC = a_{S0} + a_{S1}LANAL + a_{S2}LVOL + a_{S3}LPRI + a_{S4}LVAR + e_{TC},$$

$$LANAL = a_{A0} + a_{A1}LTC + a_{A2}LVAR + a_{A3}LSIZE + a_{A4}LPRI$$

$$+ \sum_{i=1}^{5} a_{Ai+4}IND_i + a_{A10}LNINST + a_{A11}LPINST + e_{ANAL},$$

$$LVOL = a_{V0} + a_{V1}LTC + a_{V2}LANAL + a_{V3}LSIZE + e_{VOL}.$$

The equation for $LANAL$ is not identified.

$LTC_{GH} \equiv \log(\lambda_{GH}/PRI)$ and $LTC_{MS} \equiv \log(\lambda_{MS}/PRI)$ are the logs of the adverse selection costs of transacting. PRI is the average daily closing price.

λ_{GH} is the estimate of the inverse of market depth from the Glosten–Harris specification:

$$\Delta p_t = \lambda q_t + \psi[D_t - D_{t-1}] + y_t;$$

λ_{MS} is the estimate of the inverse of market depth from the Madhavan–Smidt model specification:

$$\Delta p_t = \lambda q_t + \frac{\psi}{\pi} D_t - \psi D_{t-1} + \eta_t.$$

Δp_t is the price change at transaction t, q_t is the signed trade size, D_t is a dummy variable that is equal to $+1$ for a trade classified as a buy and -1 for a sell, and y_t and η_t are error terms. The Glosten–Harris model is estimated assuming that y_t is i.i.d., while the Madhavan–Smidt model is estimated assuming that η_t is MA(1). The sample consists of 1,421 stocks continuously listed on the NYSE for the year 1988 for which complete data were available on institutional holdings as of December 1987 and for which nonnegative estimates of market depth were obtained.

The other variables are defined as follows: $LANAL$ is the logarithm of one plus the number of analysts as of December 1987, $LVOL$ is the logarithm of the average daily trading volume in 1988, $LPRI$ is the logarithm of the average daily closing price during 1988, $LVAR$ is the logarithm of the daily return variance during 1988, and $LSIZE$ is the logarithm of the average daily market value of equity in 1988. $LINST$ and $LPINST$ are logarithms of the number of institutions holding the stock and the percentage of shares held by institutions as of December 1987. IND_i is a dummy variable corresponding to one of five industry classifications, which are obtained using COMPUSTAT tapes and which follow Bhushan (1989).

The t-statistics are in parentheses.

Equation	(13)	(15)	(13)	(15)
Dependent variable	LTC_{GH}	$LVOL$	LTC_{MS}	$LVOL$
Constant	2.300	4.113	− 1.444	5.214
	(5.51)	(9.64)	(2.78)	(7.78)
LANAL	− 0.169	0.897	− 0.258	0.990
	(3.13)	(13.31)	(3.82)	(13.99)
LVOL	− 0.888		− 0.598	
	(26.05)		(14.08)	

Table 2 (continued)

Equation	(13)	(15)	(13)	(15)
Dependent variable	LTC_{GH}	$LVOL$	LTC_{MS}	$LVOL$
$LPRI$	0.275	– 0.907	0.049	– 0.919
	(6.42)	(15.80)	(0.92)	(18.84)
$LVAR$	0.638		0.528	
	(18.40)		(12.23)	
$LSIZE$		0.615		0.673
		(12.49)		(14.53)
LTC_{MS}				0.183
				(0.98)
LTC_{GH}		0.023		
		(0.32)		
R^2	0.75	0.68	0.58	0.63

Admati–Pfleiderer (1988) model and of the Subrahmanyam (1991) model (which incorporates risk aversion) when the number of informed traders is large, that market depth increases with the number of informed traders. The effect of trading volume on LTC is negative, which confirms the intuition that active markets will be deep, and is consistent with prior empirical findings on analyses of the determinants of the bid–ask spread (Branch and Freed, 1977; Stoll, 1978). The coefficient of the log of the stock price ($LPRI$) is positive for both regressions and significant for the LTC_{GH} regression. In interpreting these coefficients it is helpful to bear in mind that, while the marginal cost of transacting for a given number of shares, n, is proportional to λ/PRI, the marginal cost of transacting for a given *dollar* volume, $v \equiv n\ PRI$, is proportional to λ/PRI^2.[4] Thus, while the coefficient of $LTC \equiv \log(\lambda/PRI)$ is positive, it is less than one in all of the regressions, implying that while the marginal cost of transacting for a given *number of shares* is increasing in the share price, the marginal cost for a given *dollar* transaction is decreasing in the share price. Since the value of a transaction is a more natural measure of size than the number of shares involved, the coefficient estimates in the LTC regressions are consistent with the intuition that markets in high-priced stocks are more liquid. Finally, the coefficient of the log of the daily return variance ($LVAR$) is positive and highly

[4]The total cost for trading $v = qP$ dollars can be written as $\lambda(v/P)^2$. The marginal cost per dollar for trading v dollars in therefore $2\lambda v/P^2$.

significant in both regressions, consistent with the intuition that adverse selection costs will tend to be higher for stocks for which the flow of new information is higher, and also consistent with Glosten and Harris (1988).

In the *LVOL* regressions, the log trading volume is strongly positively related both to *LANAL*, the log number of analysts, and to *LSIZE*. The former relation suggests that security analysts are able to generate trading volume by their activities, consistent with the notion that security analysts tend to be employed by brokerage houses who benefit from the commissions from the additional trading generated by their analysts. The latter relation is consistent with intuition that the greater the size of the firm, the larger will tend to be the number of shareholders and the volume of noise trading. The coefficients of *LPRI* are close to (but significantly different from) -1; a value of -1 would imply that it is the *dollar* volume of trading that is determined by the other variables in the equation. The coefficient of *LTC* in these regressions is positive but it not strongly significant.

It is unclear *a priori* whether the institutional ownership variables that appear in the equation for *LANAL* should also be included in the equations for *LTC* and *LVOL*. (Note that inclusion of these variables in the *LVOL* equation alone will not influence the estimates of the *LTC* equation, as the ownership variables already appear as regressors in the system.) To check for robustness we reestimate the above equation system including *LNINST* and *LPINST* as explanatory variables in both the *LTC* and the *LVOL* equations. The coefficient of *LANAL* remains negative and significant in both the *LTC* and the *LVOL* equations. These results are not reported here for reasons of brevity.

To assess the robustness of the results to the empirical specification of λ we repeat the analysis with λ estimated by yet a third approach suggested by Foster and Viswanathan (1993) (based in turn on Hasbrouck, 1991). The Foster–Viswanathan approach estimates λ by measuring the price response to the *unexpected* component of the order flow. The idea is that if trades are autocorrelated or predictable from past price changes, then part of the current order flow is predictable and should not be included in measuring the information content of a trade. The approach involves first regressing the current order flow on lagged previous order flows and prices. The current price change is then modeled as a linear function of the residual from the order flow regression and the current trade sign minus the lagged trade sign. Finally, the parameter is measured as the coefficient (in the price change regression) of the residual from the order flow regression (see Foster and Viswanathan, 1993, for a detailed exposition and application of this approach). Table 3 provides the two-stage least-squares estimates of the parameters of the identified equations when *LTC* is computed using the Foster–Viswanathan estimator of λ (five lags of trades and prices are used, as in Foster and Viswanathan, 1993). The results are very similar to those for the Glosten–Harris and Madhavan–Smidt specifications; in particular, the magnitude of the *LANAL* coefficient is very close to the corresponding values in Table 2, and remains negative and significant. It thus appears that our results are robust across these three different empirical approaches to estimating λ. As the theoretical models predict, the number of analysts following a stock has a significantly negative effect on the adverse selection cost of transacting in the stock.

Table 3
Two-stage least-squares estimates of determinants of adverse selection cost of transacting, using the Foster–Viswanathan empirical measure of λ:

$$LTC = a_{S0} + a_{S1}LANAL + a_{S2}LVOL + a_{S3}LPRI + a_{S4}LVAR + e_{TC},$$

$$LANAL = a_{A0} + a_{A1}LTC + a_{A2}LVAR + a_{A3}LSIZE + a_{A4}LPRI + \sum_{i=1}^{5} a_{Ai+4}IND_i + a_{A10}LNINST + a_{A11}LPINST + e_{ANAL},$$

$$LVOL = a_{V0} + a_{V1}LTC + a_{V2}LANAL + a_{V3}LSIZE + e_{VOL}.$$

The equation for $LANAL$ is not identified.
$LTC_{FV} \equiv \log(\lambda_{FV}/PRI)$ is the log of the adverse selection costs of transacting. PRI is the average daily closing price.

λ_{FV} is the estimate of the inverse of market depth from the Foster–Viswanathan specification:

$$q_t = \alpha + \sum_{j=1}^{5} \beta_j \Delta p_{t-j} + \sum_{k=1}^{5} \gamma_k q_{t-k} + \tau_t, \qquad \Delta p_t = \zeta + \lambda\tau_t + \psi[D_t - D_{t-1}] + \varepsilon_t.$$

Δp_t is the price change at transaction t, q_t is the signed trade size, and D_t is a dummy variable that is equal to $+1$ for a trade classified as a buy and -1 for a sell. The sample consists of 1,421 stocks continuously listed on the NYSE for the year 1988 for which complete data were available on institutional holdings as of December 1987 and for which nonnegative estimates of market depth were obtained.

The other variables are defined as follows: $LANAL$ is the logarithm of one plus the number of analysts as of December 1987, $LVOL$ is the logarithm of the average daily trading volume in 1988, $LPRI$ is the logarithm of the average daily closing price during 1988, $LVAR$ is the logarithm of the daily return variance during 1988, $LSIZE$ is the logarithm of the average daily market value of equity in 1988. $LINST$ and $LPINST$ are logarithms of the number of institutions holding the stock and the percentage of shares held by institutions as of December 1987. IND_i is a dummy variable corresponding to one of five industry classifications, which are obtained using COMPUSTAT tapes and which follow Bhushan (1989).

The t-statistics are in parentheses.

Equation	(13)	(15)
Dependent variable	LTC_{FV}	$LVOL$
Constant	− 2.000	4.257
	(4.69)	(9.58)
LANAL	− 0.185	0.917
	(3.35)	(13.58)

Table 3 (continued)

Equation	(13)	(15)
Dependent variable	LTC_{FV}	$LVOL$
$LVOL$	– 0.857 (24.59)	
$LPRI$	0.279 (6.36)	– 0.920 (16.16)
$LVAR$	0.643 (18.16)	
$LSIZE$		0.630 (12.91)
LTC_{FV}		0.052 (0.70)
R^2	0.73	0.67

5.2. Structural estimation of the Admati–Pfleiderer model of endogenous information acquisition

While the two-stage least-squares parameter estimates reported above are consistent with the coefficient sign predictions of current theoretical models, they do not take into account the functional form of the equation for λ implied by the Admati–Pfleiderer Eq. (2) or the determinants of the number of analysts implied by the equilibrium condition (3). Therefore, we turn now to estimates of the structural parameters of the Admati–Pfleiderer model represented by Eqs. (2) and (3).

Dividing both sides of Eq. (2) by the stock price P, we obtain

$$\lambda/P = \frac{\sqrt{n}}{n+1}\sqrt{\frac{\mathrm{var}(R)}{\mathrm{var}(z)}}, \tag{16}$$

where R is the rate of return on the security. Transforming Eq. (3), we obtain

$$\frac{P^2\,\mathrm{var}(R)}{(1+n)^2\lambda} = c\,. \tag{17}$$

We use PRI, the average daily closing price, as a measure of P. Taking logarithms of (16) and (17), defining a new variable, $LINF$, as

$$LINF = \log[\sqrt{n}/(n+1)]\,,$$

recognizing that $LTC = \log(\lambda/PRI)$, and adding error terms to the two equations, the empirical version of the Admati–Pfleiderer model may be written as the following equation system:

$$LTC = a_0 + a_1 LINF + a_2 LSIGR + a_3 \log[\sigma(z)] + e_1, \tag{18}$$

$$\log(1 + n) = b_0 + b_1 LTC + b_2 LSIGR + b_3 LPRI - \log c + e_n, \tag{19}$$

where $LSIGR$ denotes the logarithm of the standard deviation of the rate of return. For empirical purposes, $\log c$, the logarithm of the cost of becoming informed, is assumed to be a function of firm size and industry classification:

$$\log c = k_0 + k_1 LSIZE + \sum_{i=1}^{5} IND_i .$$

The theoretical Admati–Pfleiderer specification implies that $a_0 = b_0 = 0$, $a_1 = a_2 = 1$, $a_3 = -1$, $b_1 = = -0.5$, $b_2 = 1$, and $b_3 = 0.5$.[5] Consistent with our previous specification, n is set equal to one plus the number of analysts reported by I/B/E/S. The log standard deviation of noise trading, $\log[\sigma(z)]$, is initially proxied by $LSIGVOL$, the logarithm of the standard deviation of daily trading volume.

The nonlinear two-stage least-squares estimates of Eq. (18) and (19) are presented in Table 4 for the two definitions of LTC. (The coefficients of the industry dummy variables are not reported to conserve space.) Considering first Eq. (18) for LTC, the coefficients of all three explanatory variables have the predicted sign and are significantly different from zero, except for $LINF$ in the LTC_{GH} specification which is only weakly significant. More importantly, the coefficients are of the predicted order of magnitude; thus the coefficients of $LSIGVOL$ and $LSIGR$ are within about 30% of their predicted values for both specifications and, while the coefficients of $LINF$ conform less well to the theoretical specification, they have high standard errors.

The signs of the coefficient estimates for Eq. (19), which explains the number of analysts, are also consistent with the model predictions and are significantly different from zero. Although the determinants of analyst following are not the primary focus of this paper, it is interesting to note that LTC has a significant negative influence on the number of analysts following a stock, as the Admati–Pfleiderer model predicts. While the estimates of b_1, b_2, and b_3 are significantly different from their theoretical values, they are of the right order of magnitude.

Since the standard deviation of daily volume is an imperfect proxy for $\sigma(z)$, the standard deviation of noise trading, Eq. (18) was reestimated by nonlinear

[5] Note that a positive value of a_1 implies that an increase in the number of analysis, holding constant the other explanatory variables in Eq. (18), will reduce the adverse selection cost variable $\log(\lambda/PRI)$.

Table 4

Nonlinear two-stage least-squares estimates of two-equation Admati–Pfleiderer model using standard deviation of daily trading volume as proxy for standard deviation of noise trading:

$$LTC = a_0 + a_1 LINF + a_2 LSIGR + a_3 LSIGVOL + e_1\,,$$

$$\log(1+n) = b_0 + b_1 LTC + b_2 LSIGR + b_3 LPRI + k_1 LSIZE + \sum_{i=1}^{5} IND_i + e_n\,.$$

LTC denotes the logarithm of the adverse selection cost of transacting, $\log(\lambda/PRI)$, where λ is estimated for a sample of 1,421 NYSE stocks using the Glosten–Harris and Madhavan–Smidt methods and *PRI* is the average daily closing price in 1988. *LINF* is defined as $LINF = \log[\sqrt{n}/(n+1)]$, where n is one plus the number of analysts as of December 1987. *LSIGR* is the logarithm of the standard deviation of the rate of return in 1988, *LSIGVOL* is the logarithm of the standard deviation of daily volume in 1988, *LPRI* is the logarithm of the average daily closing price in 1988, *LSIZE* is the logarithm of the average daily market value of equity in 1988. IND_i is a dummy variable corresponding to one of five industry classifications, which are obtained using COMPUSTAT tapes and which follow Bhushan (1989) (the coefficients on these variables are not reported for brevity).

The t-statistics in parentheses are asymptotic ones for testing whether the relevant coefficient is significantly different from zero. Theoretical values of the coefficients are shown in bold brackets.

Equation	(18)	(19)	(18)	(19)
Dependent variable	LTC_{GH}	$\log(1+n)$	LTC_{MS}	$\log(1+n)$
Constant	4.044	− 1.912	1.765	− 2.728
	(5.19)	(6.59)	(1.42)	(7.47)
	[0.0]		**[0.0]**	
LINF	1.168		4.276	
	(1.87)		(2.47)	
	[1.0]		**[1.0]**	
LTC_{MS}				− 0.358
				(13.36)
				[−0.5]
LTC_{GH}		− 0.307		
		(15.27)		
		[−0.5]		
LSIGR	0.869	0.431	0.720	0.467
	(12.08)	(6.33)	(9.39)	(5.94)
	[1.0]	**[1.0]**	**[1.0]**	**[1.0]**
LSIGVOL	− 1.028		− 0.657	
	(10.99)		(7.46)	
	[−1.0]		**[−1.0]**	
LPRI		0.322		0.210
		(7.39)		(4.44)
		[0.5]		**[0.5]**
LSIZE		0.081		0.120
		(4.11)		(5.88)
R^2	0.75	0.46	0.55	0.27

two-stage least squares with *LSIZE* replacing *LSIGVOL* as a proxy for $\log[\sigma(z)]$. [Note that Eq. (19) is now underidentified.] The results, which are given in Table 5, show that the coefficient of *LINF* is now significantly different from zero and of the correct sign for both empirical *LTC* specifications. This finding is consistent with the results reported in Section 5.1 in that an increase in the number of analysts reduces the adverse selection cost of transacting, other things equal. In accordance with the role of *LSIZE* as a proxy for noise trading, its coefficient is negative and significantly different from zero. Though the

Table 5
Nonlinear two-stage least-squares estimates of the first equation of the Admati–Pfleiderer model using firm size as a proxy for variance of noise trading:

$$LTC = a_0 + a_1 LINF + a_2 LSIGR + a_3 LSIZE + e_1\,,$$

$$\log(1+n) = b_0 + b_1 LTC + b_2 LSIGR + b_3 LPRI + k_1 LSIZE + \sum_{i=1}^{5} IND_i + e_n\,.$$

LTC denotes the logarithm of the adverse selection cost of transacting, $\log(\lambda/PRI)$, where λ is estimated for a sample of 1,421 NYSE stocks using the Glosten–Harris and Madhavan–Smidt methods and *PRI* is the average daily closing price in 1988. *LINF* is defined as $LINF = \log[\sqrt{n}/(n+1)]$, as where n is one plus the number of analysts as of December 1987. *LSIGR* is the logarithm of the standard deviation of the rate of return in 1988, *LPRI* is the logarithm of the average daily closing price in 1988, *LSIZE* is the logarithm of the average daily market value of equity in 1988. IND_i is a dummy variable corresponding to one of five industry classifications, which are obtained using COMPUSTAT tapes and which follow Bhushan (1989) (the coefficients on these variables are not reported for brevity).

Parameter estimates are given for only the first equation since the second is underidentified. The t-statistics in parentheses are asymptotic ones for testing whether the relevant coefficient is significantly different from zero. Theoretical values of the coefficients are shown in bold brackets.

	LTC_{GH}	LTC_{MS}
Constant	− 4.033	− 5.309
	(12.87)	(15.64)
	[0.0]	**[0.0]**
LINF	1.139	2.208
	(2.90)	(5.19)
	[1.0]	**[1.0]**
LSIGR	0.136	0.321
	(1.69)	(3.68)
	[1.0]	**[1.0]**
LSIZE	− 0.539	− 0.251
	(8.69)	(3.74)
	[−0.0]	**[−0.0]**
R^2	0.59	0.45

Table 6

Nonlinear two-stage least-squares robustness test of two-equation Admati–Pfleiderer model using standard deviation of daily trading volume as a proxy for standard deviation of noise trading:

$$LTC = a_0 + a_1 LINF + a_2 LSIGR + a_3 LSIGVOL + e_1\,,$$

$$\log(1+n) = b_0 + b_1 LTC + b_2 LSIGR + b_3 LPRI + b_4 LNINST$$

$$+ b_5 LPINST + k_1 LSIZE + \sum_{i=1}^{5} IND_i + e_n\,.$$

LTC denotes the logarithm of the adverse selection cost of transacting, $\log(\lambda/PRI)$, where λ is estimated for a sample of 1,421 NYSE stocks using the Glosten–Harris and Madhavan–Smidt methods and *PRI* is the average daily closing price in 1988. *LINF* is defined as $LINF = \log[\sqrt{n}(n+1)]$, where n is one plus the number of analysts as of December 1987. *LSIGR* is the logarithm of the standard deviation of the rate of return in 1988, *LSIGVOL* is the logarithm of the standard deviation of daily volume in 1988, *LPRI* is the logarithm of the average daily closing price in 1988, *LNINST* is the logarithm of number of insitutions holding the stock as of December 1987, *LPINST* is the logarithm of the percentage of the firm held by institutions as of December 1987, *LSIZE* is the logarithm of the average daily market value of equity in 1988. IND_i is a dummy variable corresponding to one of five industry classifications, which are obtained using COMPUSTAT tapes and which follow Bhushan (1989) (the coefficients on these variables are not reported for brevity).

The t-statistics in parentheses are asymptotic ones for testing whether the relevant coefficient is significantly different from zero.

Equation	(18)	(19)	(18)	(19)
Dependent variable	LTC_{GH}	$\log(1+n)$	LTC_{MS}	$\log(1+n)$
Constant	4.785	− 0.701	3.402	− 0.858
	(6.77)	(2.52)	(3.03)	(2.94)
LINF	1.594		3.820	
	(2.26)		(2.84)	
LTC		− 0.099		− 0.100
		(4.80)		(4.78)
LSIGR	0.880	0.130	0.831	0.140
	(16.57)	(2.14)	(13.63)	(2.26)
LSIGVOL	− 1.072		− 0.834	
	(15.49)		(11.35)	
LPRI		− 0.001		− 0.034
		(0.02)		(0.81)
LSIZE		0.017		0.032
		(0.38)		(0.73)
LNINST		0.464		0.474
		(6.78)		(6.84)
LPINST		0.119		0.122
		(2.43)		(2.43)
R^2	0.76	0.67	0.59	0.66

coefficient of *LSIGR* is of the right sign in both regressions, it is significantly different from zero only in the LTC_{MS} regression.

As a further check on the robustness of the results, the two institutional ownership variables, *LNINST* and *LPINST*, are included in Eq. (19) for analyst following, with the results shown in Table 6.[6] While there is no substantive change in the estimated coefficients for Eq. (18) for *LTC*, both institutional ownership variables enter significantly in Eq. (19) for the number of analysts, and the R^2 for this equation is considerably higher than the corresponding R^2 in Table 4. Thus, the Admati–Pfleiderer model, although it fares surprisingly well when taken as a literal description of the market for information, and in particular as a model of the adverse selection cost of trading, appears to be too simple to capture all the institutional features of the market for information that affect the number of analysts following a stock; some of these are considered in the informal model of Bhushan (1989).

6. Summary and conclusions

Identification of the cross-sectional determinants of the depth of securities markets is of importance from both an academic and a practical standpoint. Recent models of price formation predict that an important determinant of market depth, which is inversely related to the adverse selection costs of trading, is the number of investors with superior information about the security. The analysis of Admati and Pfleiderer (1988), which assumes short-lived information, suggests that market depth will improve with an increase in the number of informed traders. When information is long-lived, the dynamic model of Holden and Subrahmanyam (1992) implies that the effect of the number of informed investors on market depth will be time-varying, and that depth will improve on average with the number of informed investors.

A simple measure of the number of informed investors in a stock is the number of security analysts who are following the company. In this paper we advance the empirical literature on the determinants of market depth by using intraday data to investigate the relation between the number of analysts and the estimated adverse selection costs of transacting, holding constant previously identified determinants of market liquidity. Consistent with the analysis of Admati and Pfleiderer (1988), the estimated adverse selection cost decreases with the number of analysts, other things equal. In addition, structural estimates of

[6] We do not report the estimates of the parameters in Eqs. (18) and (19) for the Foster–Viswanathan specification as the manner in which their measure is estimated is not within the spirit of the Kyle–Admati–Pfleiderer framework. However, qualitatively similar results were obtained using the Foster–Viswanathan specification as well.

the Admati–Pfleiderer model of endogenous information acquisition are consistent with the model. The results support the notion that an increase in analyst coverage leads to deeper markets because of enhanced competition between informed agents.

References

Admati, Anat R. and Paul Pfleiderer, 1988, A theory of intraday patterns: Volume and price variability, Review of Financial Studies 1, 3–40.

Amihud, Yakov and Haim Mendelson, 1986, Asset pricing and the bid–ask spread, Journal of Financial Economics 17, 223–249.

Arbel, Avner, Steven Carvell, and Paul Strebel, 1983, Giraffes, institutions, and neglected firms, Financial Analysts Journal 39, 57–63.

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

Bhushan, Ravi, 1989, Firm characteristics and analyst following, Journal of Accounting and Economics 11, 255–274.

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

Brennan, Michael J., Narasimhan Jegadeesh, and Bhaskaran Swaminathan, 1993, Investment analysis and the adjustment of stock prices to common information, Review of Financial Studies 6, 799–824.

Foster, F. Douglas and S. Viswanathan, 1993, Variations in trading volume, return volatility, and trading costs: Evidence on recent price formation models, Journal of Finance 48, 187–211.

George, Thomas, Gautam Kaul, and M. Nimalendran, 1991, Estimating the components of the bid–ask spread: A new approach, Review of Financial Studies 4, 623–656.

Glosten, Lawrence and Lawrence Harris, 1988, Estimating the components of the bid–ask spread, Journal of Financial Economics 21, 123–142.

Hasbrouck, Joel, 1991, Measuring the information content of stock trades, Journal of Finance 46, 179–207.

Holden, Craig W. and Avanidhar Subrahmanyam, 1992, Long-lived private information and imperfect competition, Journal of Finance 47, 247–270.

Kyle, Albert S., 1984, Market structure, information, futures markets, and price formation, in: Gary G. Storey, Andrew Schmitz, and Alexander S. Sarris, International agricultural trade: Advanced readings in price formation, market structure, and price instability (Westview Press, Boulder, CT) 45–64.

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

Lee, Charles M.C. and Mark Ready, 1991, Inferring trade direction from intradaily data, Journal of Finance 46, 733–746.

Madhavan, Ananth and Seymour Smidt, 1991, A Bayesian model of intraday specialist pricing, Journal of Financial Economics 30, 99–134.

Stoll, Hans R., 1978, The pricing of dealer services: An empirical study of NASDAQ stocks, Journal of Finance 33, 1152–1173.

Stoll, Hans R., 1989, Inferring the components of the bid–ask spread, Journal of Finance 44, 115–134.

Subrahmanyam, Avanidhar, 1991, Risk aversion, market liquidity, and price efficiency, Review of Financial Studies 3, 417–441.