How well do adverse selection components measure adverse selection?
Van Ness, Bonnie F;Van Ness, Robert A;Warr, Richard S
Financial Management; Autumn 2001; 30, 3; ProQuest Central
pg. 77

How Well Do Adverse Selection Components Measure Adverse Selection?

Bonnie F. Van Ness, Robert A. Van Ness, and Richard S. Warr*

The performance of five adverse selection models are examined by comparing their component estimates to other measures of information asymmetry and informed trading. The models produce mixed results. Adverse selection components correlate with various volatility measures, but appear unrelated to other measures of asymmetric information. Only three of the five models have the expected relation with informed trader proxies, suggesting that the adverse selection models measure adverse selection weakly at best. Spread also relates to many of the volatility measures, suggesting that some adverse selection components might be measuring some other cost of trading.

One of the significant recent advancements in the market microstructure literature is the development of models that decompose the bid-ask spread into various components. In these models, the spread generally has three cost components: order processing, inventory holding, and adverse selection (or asymmetric information). Even though these microstructure models provide an important development in empirical finance, we know little about how well these models measure adverse selection and perform relative to each other.

In the corporate finance literature, variables such as market-to-book, volatility, and institutional ownership are often used to measure the asymmetric information present in a stock. Recent papers also use adverse selection components as a direct measure of information problems.[1] However, little is known about how well adverse selection components measure information asymmetries.

In this paper, we test the performance of five commonly used methods of computing adverse selection components. To determine the usefulness of the adverse selection models in measuring information problems, the relation between adverse selection components and measures of information asymmetries and various proxies for the presence of informed traders are examined. As a benchmark for the analysis, we also examine the performance of spread as a measure of information asymmetry.

Using a three-stage simultaneous equation framework, adverse selection models are found to relate inconsistently to the various information variables. The major determinant of adverse selection appears to be volatility. Other measures of information asymmetries, such as analyst forecast errors and market-to-book, are not related to adverse selection. The proxies for informed traders produce mixed results. Overall, the relation for four of the five models are similar to those

The authors would like to thank: Kee Chung, Rick Harris, Todd Houge, Ike Mathur, Tom McInish, Bob Wood, seminar participants at the University of Mississippi, University of Missouri, and University of Texas, Pan America, Lemma Senbet, Alex Triantis, and the referee for helpful comments and suggestions and FirstCall for generously providing earnings data.

Bonnie F. Van Ness, Robert A. Van Ness, and Richard S. Warr are Assistant Professors of Finance at Kansas State University, Manhattan, Kansas.

[1]For example, Singh, Zaman, and Krishnamurti (1994) examine adverse selection around repurchases. Fee and Thomas (1999) use market microstructure variables to examine information production on diversified compared to pure-play firms. Neal and Wheatley (1998) examine the information opacity of closed-end funds. Flannery, Kwan, and Nimalendran (2000) measure the relative opacity of bank assets, and Krinsky and Lee (1996) measure adverse selection around earnings announcements.

of the spread, which brings into question the added benefit of these models.[2]

The paper proceeds as follows: Section I discusses the background on adverse selection models. Section II lays out the motivation, while Section III discusses data and methodology. Section IV presents the results and analysis, and Section V concludes.

I. Background

Kyle (1985) argues that the presence of traders who possess superior knowledge of the value of a stock can impose adverse selection costs on liquidity traders and market makers. Market makers are compensated for bearing this cost by widening the bid-ask spread, and thus ultimately recoup the cost from liquidity traders.

Numerous researchers have tried to measure the size of these costs by decomposing the bid-ask spread into components, one of which is the adverse selection cost. Researchers often base their empirical estimates of adverse selection on transaction data, which explicitly ignores outside corporate finance factors that could cause the adverse selection. In this paper, we focus on the decomposition models of Glosten and Harris (1988), and George, Kaul, and Nimalendran (1991) (as modified by Neal and Wheatley, 1998 to accommodate transactions data); Huang and Stoll (1997); Lin, Sanger, and Booth (1995); and Madhavan, Richardson, and Roomans (1997).

If we look beyond the transaction data, adverse selection costs are likely to be a function of the degree of asymmetric information surrounding the stock's true value and the probability that informed traders can capitalize on this asymmetric information. For example, firms with large amounts of future growth options might be more likely to have greater asymmetric information about their true value. Due to random deviations from true value, asymmetric information can also cause greater stock volatility. Informed traders need these random deviations from true value to provide the opportunity to capitalize on their private information at the expense of the market maker. These broad stock characteristics do not measure adverse selection per se. Instead, they provide the circumstances under which we might expect that there will be more or less adverse selection. The great appeal of adverse selection components is that they attempt to measure information problems directly.

We build on the work of Neal and Wheatley (1998), who find that the ability of adverse selection components to measure information differences between closed-end funds and stocks is surprisingly weak. Neal and Wheatley (1998) represent a comprehensive attempt to measure the performance of these models. Their paper examines the relation between adverse selection components for closed-end funds and common stocks. Closed-end funds should be easy to value, because they report their net asset values weekly. Therefore, the adverse selection component for these assets should be smaller than that for similar common stocks. Neal and Wheatley (1998) find that although adverse selection components for closed-end funds are indeed smaller than for common stocks, the difference is not as great as hypothesized. They conclude that there is either an unknown source of asymmetric information for closed-end funds, or the adverse selection models are misspecified.

Flannery et al. (2000) use various market microstructure measures to examine the relative opacity of a sample of banks. They find that the adverse selection models of George et al. (1991) (as modified by Neal and Wheatley, 1998) and Lin et al. (1995) produce varied and

[2]An alternative and, we believe, less likely explanation for the results is that the corporate finance variables are poor proxies for asymmetric information and that the adverse selection components are accurately measuring information problems.

inconsistent estimates of adverse selection components. There is a varied, unreliable relation between balance sheet items that are hypothesized to reflect asymmetric information or asset opacity and the adverse selection components.

Currently, there is no prescription for researchers for which adverse selection models should be used to measure information asymmetry if such models are to be used at all. However, if adverse selection models are to be more than just a theoretical abstract and have a real use in empirical testing, such an analysis is essential. In this paper, the performance of adverse selection models and their relation to the corporate finance variables that measure the conditions that we believe are consistent with adverse selection are examined.

The performance of the bid-ask spread alone as a measure of adverse selection is also examined. We view the spread as the benchmark. If extracting the adverse selection component of the spread is to be worthwhile, then the adverse selection models are expected to perform better than the spread alone.

II. Data and Method

The basic method is to compare the adverse selection components to various corporate finance variables that are related directly or indirectly to asymmetric information.

The sample period examines the months of April, May, and June 1999. This period is used for two reasons. First, to have the freshest year-end financial statement data and because there is a delay in the publication of annual reports, we start in April. An April starting point ensures that the information was available to investors at the time of the analysis. Second, the tick size reduction on the NYSE occurred in June 1997. Because of the potential impact of tick size reduction on spreads,[3] we examine the performance of the adverse selection models after this rule change to increase the relevance of the findings to the market today.

Data is obtained for spread decomposition from the NYSE TAQ database. All other data is obtained from CRSP, Compustat, and FirstCall. The initial sample comprises 856 companies that are traded on the NYSE and have data available on TAQ, Compustat, and FirstCall. Financial service firms (SIC 6000-6999), regulated utilities (SIC 4800-4829 and 4910-4949), ADRs, REITS, and foreign firms are excluded. Firms whose stock prices are less than three dollars and all stocks that split during the component estimation period are also excluded. The sample is limited to NYSE stocks to ensure that the results are not affected by differences in the trading systems of the Nasdaq and NYSE (see Affleck-Graves, Hegde, and Miller, 1994). The goal is to give the adverse selection models the greatest possible opportunity to perform well.

A. Adverse Selection Components Models

The adverse selection component of the spread for each stock is calculated. To do so, we use the estimation procedures of Glosten and Harris (1988) and George et al. (1991) as modified by Neal Wheatley (1998) and Lin et al. (1995). For all the models, we initially compute the adverse selection components as a percentage of the spread. To compute the adverse selection cost of transacting, we express the dollar adverse selection component as a percentage of the stock price (as in Brennan and Subrahmanyam, 1995). Thus, we can control

[3] See Bessembinder (1999), Goldstein and Kavajecz (2000), and Jones and Lipson (2000) for discussion of the impact of the tick size reduction.

for stock price and measure the information cost of trading for a given dollar trade.[4]

1. Glosten and Harris (1988)

Glosten and Harris (1988) present one of the first trade indicator regression models for spread decomposition. A unique characteristic of their model is that they express the adverse selection component, Z_0, and the combined order processing and inventory holding component, C_0, as linear functions of transaction volume. The basic model is:

$$\Delta P_t = c_0 \Delta Q_t + c_1 \Delta Q_t V_t + z_0 Q_t + z_1 Q_t V_t + \varepsilon_t, \tag{1}$$

where the adverse selection component is $Z_0=2(z_0 + z_1 V_t)$ and the order processing/inventory holding component is $C_0=2(c_0 + c_1 V_t)$. P_t is the observed transaction price at time t, V_t is the number of shares traded in the transaction at time t, and e_t captures public information arrival and rounding error. Q_t is a trade indicator that is +1 if the transaction is buyer initiated and –1 if the transaction is seller initiated.

Glosten and Harris (1988) did not have quote data and, thus, were unable to observe Q_t. Having both trade and quote data, we use the Lee and Ready (1991) procedure for trade classification. OLS is used to obtain estimates for c_0, c_1, z_0, and z_1 for each stock in the sample.

The bid-ask spread in the Glosten and Harris (1988) model is the sum of the adverse selection and order processing/inventory holding components. In Equation (2), the average transaction volume for stock i is used to obtain an estimate of the adverse selection component as a proportion of the spread:

$$Z_i = \frac{2(z_{0,i} + z_{1,i}\bar{V}_i)}{2(c_{0,i} + c_{1,i}\bar{V}_i) + 2(z_{0,i} + z_{1,i}\bar{V}_i)} \tag{2}$$

2. George, Kaul, and Nimalendran (1991)

George, Kaul, and Nimalendran (1991) allow expected returns to be serially dependent. The serial dependence has the same impact on both transaction returns and quote midpoint returns. Hence, the difference between the two returns filters out the serial dependence. The transaction return is:

$$TR_t = E_t + \pi(s_q/2)(Q_t - Q_{t-1}) + (1\text{-}\pi)(s_q/2)Q_t + U_t, \tag{3}$$

where E_t is the expected return from time t-1 to t, π and (1-π) are the fractions of the spread due to order processing and adverse selection costs, respectively, s_q is the percentage bid-ask spread (assumed to be constant through time), Q_t is a +1/-1 buy-sell indicator, and U_t represents public information innovations.

---

[4]Consider the following example: a \$100 stock with a 1% spread and an adverse selection component of 30% of the spread has an adverse selection cost of 30 cents. A \$10 stock with a 2% spread, and an adverse selection component of 30% of the spread has a cost of 6 cents, but a \$100 trade in this stock will generate a total cost of 60 cents. If we just look at components as a percentage of the spread, then we would say that these firms have the same cost (30%). If we look at the cost per stock, then we would say that the \$10 stock has a lower cost (6 cents vs. 30 cents). As stock price is largely endogenous (to the extent that a firm can do a stock split at anytime), we seek to equally evaluate a firm that has a \$100 stock price or a firm that has recently undertaken a 10 for 1 split and has a \$10 stock price.

George et al. (1991) measure the quote midpoint immediately following the transaction at time t. An upper case T subscript is used to preserve the timing distinction for the quote midpoint. The midpoint return is:

$$MR_T = E_T + (1\text{-}\pi)(s_q/2)Q_T + U_T. \tag{4}$$

Subtracting the midpoint return from the transaction return and multiplying by two yields:

$$2RD_t = \pi s_q(Q_t - Q_{t-1}) + V_t, \tag{5}$$

where $V_t = 2(E_t - E_T) + 2(U_t - U_T)$. Relaxing the assumption that s_q is constant and including an intercept yields:

$$2RD_t = \pi_0 + \pi_1 s_q(Q_t - Q_{t-1}) + V_t. \tag{6}$$

The Lee and Ready (1991) procedure is used to determine trade classification. OLS is also used to estimate the order processing component, π_0, and an adverse selection component, $(1\text{-}\pi_1)$, for each stock in the sample.

3. Lin, Sanger, and Booth (1995)

Lin, Sanger, and Booth (1995) develop a method of estimating empirical components of the effective spread that follows Huang and Stoll (1994), Lin (1993), and Stoll (1989). Lin et al. (1995) define the signed effective half-spread, z_t, as the transaction price at time t, P_t, minus the spread midpoint, Q_t. The signed effective half spread is negative for sell orders and positive for buy orders. To reflect possible adverse information revealed by the trade at time t, Lin et al. (1995) add quote revisions of λz_t to both the bid and ask quotes. The proportion of the spread due to adverse information, λ, is bounded by zero and one. The dealer's gross profit as a fraction of the effective spread is defined as $\gamma = 1 - \lambda - \theta$, where θ reflects the extent of order persistence.

Since λ reflects the quote revision (in response to a trade) as a fraction of the effective spread, z_t, and since θ measures the pattern of order arrival, Lin et al. (1995) model the following:

$$Q_{t+1} - Q_t = \lambda z_t + \varepsilon_{t+1}, \tag{7}$$

$$Z_{t+1} = \theta Z_t + \eta_{t+1} \tag{8}$$

where the disturbance terms ε_{t+1} and η_{t+1} are assumed to be uncorrelated.

We follow Lin et al. (1995) by using OLS to estimate Equation (9) to obtain the adverse information component, λ, for each stock in the sample:

$$\Delta Q_{t+1} = \lambda z_t + e_{t+1} \tag{9}$$

The logarithms of the transaction price and the quote midpoint are used to yield a continuously compounded rate of return for the dependent variable and a relative effective spread for the independent variable.

4. Madhavan, Richardson, and Roomans (1997)

Madhavan, Richardson, and Roomans (1997) decompose the spread, using the idea that four parameters govern the behavior of transaction prices and quotes. The four parameters are θ, the asymmetric information parameter; ϕ, the cost of supplying liquidity; λ, the probability that a transaction takes place inside the spread; and, ρ, the autocorrelation of order flow. Madhavan et al. (1997) show that μ_t, the post-trade expected value of the stock, can be expressed as:

$$\mu_t = p_t - p_{t-1} - (\phi + \theta)x_t + (\phi + \rho\theta)x_{t-1} \tag{10}$$

where p_t is the transaction price at time t and x_t is an indicator variable for trade initiation. If a trade is buyer-initiated, then $x_t = 1$, and $x_t = -1$ if a trade is seller-initiated. Some trades, such as pre-negotiated crosses, are both buyer- and seller-initiated and take place between the prevailing bid and ask prices. $x_t = 0$ if a trade takes place within the prevailing bid and ask prices.

Madhavan et al. (1997) use generalized method of moments (GMM) to identify the parameter vector β=(θ, ϕ, λ, ρ) and a constant (drift) α implied by the model:

$$E\begin{pmatrix} x_t x_{t-1} - x_t^2 \rho \\ |x_t| - (1-\lambda) \\ u_t - \alpha \\ (u_t - \alpha)x_t \\ (u_t - \alpha)x_{t-1} \end{pmatrix} = 0 \tag{11}$$

The first equation defines the autocorrelation in trade initiation, the second equation is the crossing probability, the third equation defines the drift term, α, as the average pricing error, and the last two equations are OLS normal equations.

We estimate the Madhavan et al. (1997) asymmetric information parameter, θ, in dollar terms and further define the mean implied bid-ask spread at time t as 2(θ + ϕ). Hence, to obtain a proportional asymmetric information component, 2θ is divided by the mean implied spread for firm i during the sample period.

5. Huang and Stoll (1997)

Huang and Stoll (1997) develop a trade indicator model that "provides a flexible framework for examining a variety of microstructure issues." One of the objectives of their model is to reconcile the various component estimation models.[5] The three-way decomposition of the spread is based on induced serial correlation in trade flows. Huang and Stoll (1997) model this serial correlation by:

$$E(Q_{t-1} \mid Q_{t-2}) = (1-2\pi)Q_{t-2}, \tag{12}$$

$$\Delta M_t = (\alpha + \beta)\frac{S_{t-1}}{2}Q_{t-1} - \alpha(1-2\pi)\frac{S_{t-2}}{2}Q_{t-2} + \varepsilon_t \tag{13}$$

[5]For covariance models see Roll (1984), Choi, Salandro, and Shastri (1988), Stoll (1989), and George et al. (1991). For trade indicator models see Glosten and Harris (1988), and Madahavan et al. (1997).

where Q_t is the buy-sell indicator for the transaction price, P_t, and π is the probability that the trade at time t is opposite in sign to the trade at t-1. M_t is the midpoint of the quote that prevails just before the transaction at time t, S_t is the posted spread just prior to the transaction. α and β are the percentages of the half-spread attributable to adverse selection and inventory holding costs, respectively. Since α and β are stated as proportions, the order processing component is equal to $(1-\alpha-\beta)$.

As recommended by Huang and Stoll (1997), sequential trades are treated at the same price without a change in the bid or ask quotes as only one trade. Each trade is paired with the last quote posted at least five seconds earlier, but within the same day. To obtain the adverse selection component, α, for each sample security, the GMM procedure is used to simultaneously estimate the two equation system.

B. Corporate Finance Variables

There are two factors necessary for the presence of adverse selection costs. These factors are conditions that contribute to asymmetric information on the true value of the stock and the presence of informed traders who exploit this asymmetric information. Variables that are proxies for each of these are computed. By grouping the following variables into these two factors, we can focus more precisely on the variables' contributions to adverse selection.

1. Asymmetric Information Variables

a. Volatility

We include the standard deviation of the quote mid point SDMID as a measure of intraday volatility to capture volatility in the true price of the stock. For return volatility, the standard deviation of the daily stock return, SIGR, and the standard deviation of daily volume SIGVOL are used.

b. Volume

Because less frequently traded stocks can have more information problems, we hypothesize that the average daily trading volume *VOL* might be related to adverse selection. However, Conrad and Niden (1992) find no detectable relation between trading volume and adverse selection around corporate acquisition announcements.

c. Leverage

Due to the presence of fixed charges, more highly levered firms can have greater volatility in their earnings. The value of these firms can be more volatile, thus exposing investors to potentially greater information asymmetries. In equilibrium, firms should trade off cash flow volatility against leverage. Although this variable is included, it may be unrelated to asymmetric information.

$$\text{LEVG} = (\text{Total long term debt} + \text{debt due in 1 year})/\text{Total Assets} \tag{14}$$

d. Analyst Earnings Forecast Errors

The fundamental value of a stock is the present value of the future cash flows. To the degree that earnings are a proportional proxy for cash flows, uncertainty about future earnings implies asymmetric information about fundamental value (for example, see D'Mello and Ferris, 2000).

To the extent that it is harder for analysts to accurately and consistently estimate future earnings, we measure forecast error as the median of the last earnings forecast for the sample period that are forecasting the next year-end EPS, divided by the actual EPS that was reported at the year-end. Firms for which the absolute value of analyst's errors is greater than 100% of the average stock price are excluded. (See Ali, Klein, and Rosenfeld, 1992; Frankel and Lee, 1998; and Thomas, 1999).

$$ERRE = |(EPS - \text{Median EPS forecast})/EPS| \tag{15}$$

e. Dispersion of Analyst Earnings Forecasts

The greater the asymmetric information about value, the greater the likelihood that analysts will have differing opinions on the future level of earnings. The dispersion of forecasts is measured as the standard deviation of forecasts divided by the median forecasts.

$$DISP = |\text{Standard deviation of Estimate/Median Estimate}| \tag{16}$$

All firms followed by fewer than three analysts during the estimation period and firms for which the standard deviation of analysts' forecasts exceeds 20% of the average stock price are excluded. The last forecasts in the sample period that forecast the next year-end EPS are used. We do not scale the analysts' forecast variables by price, because we are scaling adverse selection by price. Scaling both the earnings variables by price would cause price to obscure the true relation between the variables. Scaling by earnings is also problematic, especially for firms with earnings (or median estimates) very close to zero. However, although scaling by earnings might induce noise in the tests, there is no reason why this noise should bias the results in any particular way.

f. Market-to-Book

Corporate finance researchers often use market-to-book as a measure of the quality of the firm's growth opportunities. Due to the greater potential for asymmetric information about future cash flows, firms with more growth opportunities will be harder to value (see also Scherr and Hulbert, 2001).

$$MB = (CSHO*PRICE + TA - CEQ)/TA \tag{17}$$

where CSHO is the number of common shares outstanding, PRICE is the average month-end price for the estimation period, TA is total assets, and CEQ is common equity.

g. R&D and Intangibles

High research and development costs (R&D) as a percentage of revenues will be associated with greater information asymmetry. Because of the uncertainty of outcomes of R&D, firms with high R&D may be harder to value. For firms with no R&D expense, the R&D variable is set to zero.

$$RDSALES = R\&D\ Expense/Sales \tag{18}$$

Firms with large amounts of intangible assets (such as goodwill, patents, etc.) as a percentage of total assets are likely to be harder to value since there is inherent uncertainty about the value of these assets.

$$INTGTA = \text{Intangibles}/TA \tag{19}$$

2. Informed Trader Variables

The last group of variables that are associated with the presence of adverse selection are those that are proxies for the presence of informed traders. Three variables are used to measure the potential presence of informed traders.

a. Number of Analysts

Several papers document an inverse relation between analyst following and asymmetric information. However, there is also evidence of an opposite relation. Brennan and Subrahmanyam (1995) find that more analysts result in greater overall information production, and this information reduces the size of the adverse selection component. In their analysis, they use the Glosten and Harris (1988) and Madhavan and Smidt (1991) models to decompose the spread. For both models they find that the number of analysts following the firm is negatively related to the size of the adverse selection component, and this relation is significant. This relation suggests that the presence of more analysts results in greater information production and a reduction in the adverse selection component.

Brennan, Jegadeesh, and Swaminathan (1993) also find that the stock prices of firms that have larger analyst followings tend to react more rapidly to new information than do those firms with small analyst followings. Hong, Lim, and Stein (1999) find that investors who engage in a momentum strategy that involves buying firms with small analyst following will earn greater excess returns than they will if they buy firms with greater analyst following. This finding suggests that analysts reduce information uncertainty.

Chung, McInish, Wood, and Wyhowski (1995) examine the effect that the number of analysts following a firm has on the overall size of the spread. Recognizing the endogeneity of spread size and number of analysts, they use a simultaneous model to estimate the relation. They find that the number of analysts following a stock is increasing in the size of the spread. Their finding supports the hypothesis that analysts follow the stocks for which the marginal benefit of information production is greatest. To the degree that spreads may be positively correlated with information asymmetries, Chung et al.'s (1995) work appears to contradict the other papers.

b. Percentage of Stock Held by Institutions

In addition to analyst following, we use two institutional ownership variables, as in Brennan and Subrahmanyam (1995). LINST represents the log of the number of institutional owners, and LPINST, the log of the percentage of stock owned by institutions. The data for these variables are obtained from the Standard and Poor's *Stock Owners Guide* for March 1999. Because many institutions, in effect, compete with each other and the market maker to profit from their private information, a greater number of institutional owners could indicate less private information. Conversely, a higher proportion of institutional ownership (LPINST) could indicate more private information, since LPINST can measure the presence of (a few) large block holders.

3. Other Control Variables

a. Size and Industry

The market value of equity is an important determinant of the speed with which a stock price adjusts to new information, possibly because of the greater awareness of investors of larger firms (Merton, 1987). Additionally, if investors face some fixed cost in information production, then they tend to follow larger stocks in which they can take larger positions. To the extent that larger firms have more information surrounding them, we would expect larger firms to have smaller adverse selection components.

Industry dummies are assigned for the following two digit SIC code categories: Mining 10-14, Construction and Manufacturing 15-39, Transportation and Public Utilities 40-49, Wholesale and Retail Trade 50-59, Finance, Insurance and Real Estate 60-67, and Services 70-96 (see Bhushan, 1989).

III. Results and Analysis

In this section, we discuss the adverse selection component estimates and their relations to each other and the corporate finance variables. We then examine these relations in ordinary least squares and three-stage least squares frameworks.

A. Summary Statistics and Correlations

Table I presents summary statistics for the sample. Panel A contains the raw numbers for the adverse selection variables. The range of values that the Huang and Stoll (1997) model generates, -338 to 251, are noted. This range is theoretically impossible, so for this reason, in all future analysis, we will use only the components that are between zero and one (these constrained numbers are presented in Panel B).

For the Huang and Stoll (1997) model, implausible estimates are a non-trivial problem, as over half of the observations are lost.[6] Perhaps we can explain the poor performance of the Huang and Stoll model by the large number of probability of price reversals that are less than 0.5. The probability of price reversal for the Huang and Stoll model is examined in Panel D of Table I. It appears that although there seems to be some relation between implausible estimates and the frequency of price reversals of less than 0.5, the low probability of price reversals is not the sole explanation. The other models are not subject to this problem because most assume that the probability of price reversal is 0.5. Therefore, the great realism of the Huang and Stoll model comes at a cost.

The other models that produce values outside the theoretical range are the Madhavan et al. (1997) model, which loses 152 observations to implausible values and the Lin et al. (1995) model, which loses just four observations. The remaining models generate only plausible estimates that appear to be broadly in line with previous work.[7] Panel C shows the cost of adverse selection as a percentage of share price. It is apparent that (with the exception of the Huang and Stoll (1997) model) the average cost of adverse selection for all the models is about $0.03 to $0.04 for a $10 stock.

Table II presents the summary statistics for the other main variables. Firms in the sample

[6]Other authors report problems in obtaining plausible estimates using the Huang and Stoll (1997) model. Clarke and Shastri (2000), using the model on a random sample of 320 NYSE firms, find implausible estimates of the adverse selection component in approximately 60% of the cases. Krishnan (2000) states "...(a) the inventory carrying cost component turns out to be negative and (b) the adverse selection component is smaller after a sequence of trades in the same direction than after a sequence of trades in the opposite directions despite the probability of the latter event being higher than the former. Both these results are theoretically impermissible."

[7]For example, Glosten and Harris (1988) find an average adverse selection of 0.36 for a sample of 20 NYSE common stocks in 1981. Neal and Wheatley's (1998) modified version of the George et al. (1991) model finds 0.361 for 17 stocks in 1988. Huang and Stoll (1997) find 0.0959 for 20 firms from the Major Market Index in 1992. Lin et al. (1995) have a range of 0.198 to .626 for 150 NYSE firms from 1988. Madhavan et al. (1997) report 0.41 on a sample of 274 NYSE firms from 1990.

Table I. Summary Statistics of Adverse Selection Component

This table contains stocks listed on the NYSE from April 1999 to June 1999. It presents summary statistics of the 856 stocks in the sample. For each stock, the adverse selection component of the spread was computed using five different spread decomposition models. The models are: Glosten and Harris (1988) (GH), George et al. (1991) (GKN), Huang and Stoll (1997) (HS), Lin et al. (1995) (LSB), and Madhavan et al. (1997) (MRR).

	Mean	Median	Std. Deviation	Min.	Max.	Observations
Panel A. Components as a Proportion of the Spread						
GH	0.389	0.382	0.134	0.014	0.858	856
GKN	0.476	0.494	0.086	0.004	0.625	856
HS	1.934	-0.026	30.656	-338.064	251.458	856
LSB	0.454	0.438	0.166	-0.074	1.343	856
MRR	0.732	0.740	0.298	-1.672	1.854	856
Panel B. Adverse Selection Components >0 and <1 as a Proportion of the Spread						
GH	0.389	0.382	0.389	0.014	0.858	856
GK	0.476	0.495	0.086	0.004	0.625	856
HS	0.180	0.101	0.202	0.001	0.960	333
LSB	0.452	0.438	0.159	0.008	0.968	852
MRR	0.666	0.685	0.206	0.032	0.998	704
Panel C. Dollar Adverse Selection Components Divided by Price (Constrained as in Panel B) and Multiplied by 100						
GH	0.003	0.002	0.002	0.0004	0.017	856
GKN	0.003	0.003	0.002	0.0006	0.017	856
HS	0.001	0.0004	0.001	0.0000	0.013	333
LSB	0.003	0.002	0.002	0.0001	0.019	852
MRR	0.004	0.003	0.003	0.0002	0.030	704
Panel D. Probability of Price Reversal for Huang and Stoll (1997) Model[a]						
All Estimates	0.578	0.562	0.083	0.402	0.974	(15.9%) 856
Plausible AS Estimates	0.580	0.558	0.094	0.402	0.974	(12.9%) 333
Implausible AS Estimates	0.577	0.567	0.075	0.403	0.923	(17.8%) 523

[a]These numbers represent the percentage of estimates with a probability of price reversal less than 0.5, expressed as decimals.

have market capitalizations ranging between $61 million and $349 billion and a median of eight analysts estimating earnings.

Table III presents the Spearman rank correlations between the adverse selection components, and between these components and the spread. In Panel A, dollar terms are used to express the variables. There is a strong positive correlation between the estimates for the models. The adverse selection components also appear to have a strong positive correlation with the dollar spread.

A clear drawback of this analysis is that these correlations might be driven by the stock price. For example, higher-priced stocks are likely to have larger spreads. Panel B presents the Spearman correlations after scaling by price. Again, there is a strong correlation among all the components and the spread. This strong correlation indicates that at the very least, the adverse selection components of the models studied are measuring something similar.

Table IV presents the Spearman rank correlations between the adverse selection components and the analyst and informed trader measures. Rank correlations are used because we place greater weight on the ability of the components to measure ordinal information

Table II. Summary Statistics: Other Variables

This table presents summary statistics. NTS is the average number of trades per day. TSIZE is the average trade size. PRICE is the average trade price. SDMID is the standard deviation of the average daily spread midpoint. SPREAD is the average dollar bid-ask spread. PSPREAD is SPREAD divided by TPRICE. DISP is the standard deviation of analysts' earnings forecasts scaled by the median forecast. ERRE is the median absolute forecast error scaled by actual earnings. MVE is market value of equity. MB is the market value of equity plus book value of debt divided by total assets. RDSALES is R&D expense divided by sales. INTANGTA is intangibles divided by total assets. SIGR is the standard deviation of daily stock returns. ANLYST is the number of analysts who forecast stock earnings. PINST is the percentage of stock owned by institutions. INST is the number of institutions owning stock.

	Mean	Median	Std. Deviation	Min.	Max.	Observations
Panel A. Stock Trade Variables						
NTS	14025	7724	17469	282	166476	856
TSIZE	1904	1683	1006	565	9853	856
PRICE	31.76	27.10	20.33	3.377	128.52	856
SDMID	2.410	1.999	1.743	0.207	19.550	856
SPREAD	0.153	0.150	0.039	0.068	0.337	856
PSPREAD	0.007	0.006	0.005	0.001	0.036	856
Panel B. Volatility and Uncertainty Variables						
DISP	0.160	0.025	0.815	0	17	816
ERRE	0.327	0.075	0.763	0	8	782
LEVG	0.270	0.253	0.187	0	0.990	856
RDSALES	0.019	0	0.044	0	0.494	856
INTANGTA	0.125	0.061	0.156	0	0.768	856
MB	2.126	1.636	2.198	0.605	49.936	856
MVE	6919	1631	20160	61.655	349006	856
SIGR	0.030	0.028	0.011	0.006	0.076	856
Panel C. Informed Trader Variables						
ANLYST	9.500	8	5.994	0	34	856
PINST	0.515	0.534	0.224	0.014	0.981	658
INST	328	233	307	3	2030	662

asymmetries, rather than absolute information asymmetries.

All of the components are negatively correlated with ANLYST (the number of analysts issuing earnings forecasts). This negative correlation indicates that the presence of more analysts results in greater information production and less information asymmetry. However, due to the endogeneity of ANLYST and the adverse selection components, these results should be treated with caution. This relation could also be a function of the adverse selection components being scaled by price, and the extent to which high priced stocks attract greater interest from analysts.

The George et al. (1991) and Lin et al. (1995) adverse selection components are significantly and positively correlated with DISP, which supports the view that greater analyst uncertainty is related to adverse selection. However, this relation is not evident in the correlations with the analyst forecast error variable, ERRE. Here, all the components have negative correlations.

All of the adverse selection components models have a strong negative correlation with the number of institutions holding the stock. This result directly contradicts the hypothesis that greater institutional ownership should result in higher adverse selection. This result

Table III. Spearman Rank Correlation Coefficients for the Adverse Selection Components

This table presents the Spearman rank correlations for all the adverse selection models and the overall bid-ask spread. We use the models of Glosten and Harris (1998) (GH), George et al. (1991) (GKN), Huang and Stoll (1997) (HS), Lin et al. (1995) (LSB), and Madhavan et al. (1997) (MRR) to compute the adverse selection components. In Panel, A the adverse selection components and spread are expressed as raw dollar cost. In Panel B, the components and the spread are expressed as a percentage of the stock price. The values of the components as a percentage of the spread outside the theoretically plausible range of zero to one are excluded.

	GKN	HS	LSB	MRR	GH
			Panel A. Dollar Estimates		
HS	0.5735*				
LSB	0.7396*	0.5064*			
MRR	0.8481*	0.5770*	0.7452*		
GH	0.8681*	0.6216*	0.8285*	0.9096*	
SPREAD	0.9786*	0.5915*	0.7585*	0.8754*	0.8947*
			Panel B. Scaled by Price		
HS	0.3532*				
LSB	0.8399*	0.3893*			
MRR	0.8876*	0.3746*	0.8457*		
GH	0.9164*	0.4825*	0.8828*	0.9307*	
SPREAD	0.9171*	0.3069*	0.8168*	0.8158*	0.8240*

*Significant at the 0.10 level.

might be driven by the strong correlation between stock price and LINST (Spearman's rho=0.5582). That is, higher priced stocks tend to have a greater number of institutions following them, which will overwhelm any adverse selection effects. It is also possible that a large number of institutional owners serves to reduce adverse selection through competition.

Clearly, univariate analysis is subject to problems due to scaling, endogeneity and spurious correlations. For these reasons, we now focus on regression analysis.

B. Regressions of Adverse Selection on Corporate Finance Variables

Table V presents the results of ordinary least squares (OLS) regressions of the adverse selection components and the spread on the various corporate finance measures. Such regressions will control for some of the spurious correlations, but cannot control for the endogeneity of several of the variables.

One of the challenges in determining the performance of adverse selection models is that of establishing a valid benchmark. Throughout the regression analysis, we use spread, expressed as a percentage of the price, as a benchmark. Since the adverse selection variables are components of the spread, we expect that they will perform as well as spread in the regression analysis. This minimum level of performance is consistent with the components being fixed proportions of the spread. Some of the models are, by construction, closely related to the spread. For example, the Glosten and Harris (1988) model assumes that the spread is the sum of the adverse selection and inventory/order processing components. George et al. (1991) and Madhavan et al. (1997) explicitly include the magnitude of the spread in their estimation of the spread components. Lin et al. (1995) limit the adverse selection to be within zero and 100% of the effective spread.

Table IV. Spearman Rank Correlation Coefficients for Adverse Selection Components and Analyst Forecast Variables

This table shows the adverse selection components constrained to >0 and <1 and scaled by price. We use the models of Glosten and Harris (1998) (GH), George et al. (1991) (GKN), Huang and Stoll (1997) (HS), Lin et al. (1995) (LSB), and Madhavan et al. (1997) (MRR) to compute the adverse selection components. ANLYST is the number of analysts issuing earnings forecasts for the stock. ERRE is the error in analyst forecasts, defined as the absolute value of the difference between the median forecast and the actual earnings per share, scaled by actual earnings per share. DISP is dispersion in analyst forecasts, defined as the absolute value of the standard deviation of forecasts scaled by the median forecast. LINST is the log of the number of institutional owners of the stock. LPINST is the log of the percentage of shares held by institutions.

	GH	GKN	HS	LSB	MRR
ANLYST	-0.2139*	-0.2021*	-0.0764	-0.1585*	-0.1716*
DISP	0.0653	0.1308*	-0.0360	0.1552*	0.0772
ERRE	-0.0727	-0.0891*	-0.0100	-0.0630	-0.1117*
LINST	-0.8569*	-0.7661*	-0.5433*	-0.7723*	-0.8232*
LPINST	-0.0824*	-0.0572	-0.0345	-0.0048	0.0045

*Significant at the 0.10 level.

In examining the asymmetric information measures, we find that all adverse selection estimates (with the exception of Huang and Stoll, 1997) and the spread are negatively related to volume. This finding indicates that more highly traded stocks have fewer information problems. Both volatility measures (LSIGR and LVAR) are positively related to adverse selection component estimates of Glosten and Harris (1988), George et al. (1991), Lin et al. (1995), and Madhavan et al. (1997). LSIGR is also positively related to the spread. The analyst variables ERRE and DISP are significant only in the Huang and Stoll (1997) and the spread regressions. DISP, the dispersion of analysts' forecasts, is also positively related to spread. Also, spread is the only measure that is significantly related to RDSALES and LNMB (market-to-book). Finally, looking at the informed trader variables, LINST and LPINST, we find that adverse selection estimates from three models (Glosten and Harris, 1988; Huang and Stoll, 1997; and Lin et al. 1995) are negatively related to the number of institutional investors. Lin et al.'s (1995) estimate has a positive relation to the percentage of institutional ownership.

The results of Table V are mixed. First of all, the adverse selection regressions have some of the predicted relations. However, in many cases, spread also seems to capture certain firm characteristics that we hypothesize as indicating information problems. An important result of the regressions in Table V is that several of the models appear to be measuring some degree of information problems. However, there is significant dispersion in the coefficient estimates for the regressions, indicating that the models are picking up a lot of noise.

C. 3SLS Regression Models

As discussed above, we recognize the potential endogeneity problems that exist in the data. For example, does a higher analyst following result in lower adverse selection (because of greater information production) or higher adverse selection (because analysts follow stocks for which their potential gains from private information are greatest)?

Following Brennan and Subrahmanyam (1995), a three-equation system estimated using three-stage least squares is used. The system of equations is as follows and is a modification of Brennan and Subrahmanyam's (1995) equations 13, 14, and 15:

Table V. OLS Regressions of Adverse Selection Components Scaled by Price on Other Measures of Information Asymmetry

This table shows that the dependent variables are the natural log of the adverse selection component scaled by the stock price or the natural log of the spread scaled by the stock price. LANLYST is ln(1+ANLYST). LVOL is the log of daily average volume. LPRI is the log of the stock price. LVAR is ln($SDMID^2$). LSIGR is the log of the standard deviation of daily stock returns. LSIGVOL is the log of the standard deviation of daily volume. DISP is standard deviation of analyst forecasts scaled by the median forecast. ERRE is the absolute analyst forecast error, scaled by the median forecast. LNMVE is the log of market value of equity. LNMB is the log of the market-to-book ratio. LEVG is the debt ratio. RDSALES is research and development expense as a percentage of sales. LNINTGTA is the log of 1+ intangibles over total assets. LINST is the log of the number of institutions owning stock. LPINST is the log of the percentage of stock owned by institutions. We use the models of Glosten and Harris (1998) (GH), George et al. (1991) (GKN), Huang and Stoll (1997) (HS), Lin et al. (1995) (LSB), and Madhavan et al. (1997) (MRR) to compute the adverse selection components. White-corrected t-stats are in parenthesis.

	Percentage Spread	GH	GKN	HS	LSB	MRR
LANLYST	-0.005	0.053	0.183	-0.143	0.041	-0.054
	(-0.241)	(1.162)	(2.314)**	(-0.528)	(1.012)	(-1.079)
LVOL	-0.143	-0.386	-0.353	-0.202	-0.236	-0.350
	(-6.863)***	(-6.744)***	(-5.571)***	(-1.18)	(-5.641)***	(-5.403)***
LPRI	-0.666	-0.248	-0.374	0.020	-0.407	-0.269
	(-21.095)***	(-3.437)***	(-3.873)***	(0.058)	(-6.155)***	(-3.427)***
LVAR	0.019	0.093	0.077	0.007	0.068	0.070
	(1.404)	(2.652)***	(2.207)**	(0.057)	(2.483)**	(2.163)**
LSIGR	0.151	0.416	0.330	0.218	0.369	0.316
	(3.266)	(3.677)***	(3.412)***	(0.403)	(3.809)***	(2.573)**
LSIGVOL	-0.002	-0.043	-0.038	-0.048	-0.023	-0.001
	(-0.304)	(-3.243)***	(-2.463)***	(-0.535)	(-1.413)	(-0.066)
ERRE	-0.003	-0.011	-0.011	-0.050	-0.004	-0.009
	(-1.569)	(-1.567)	(-1.198)	(-3.584)***	(-1.076)	(-1.69)
DISP	0.002	-0.005	0.000	0.020	-0.006	0.004
	(3.067)***	(-1.369)	(-0.022)	(2.106)	(-1.194)	(0.57)
LEVG	0.046	0.186	0.222	0.193	0.085	0.252
	(1.251)	(2.061)***	(1.82)	(0.371)	(0.958)	(2.364)**
LNINTGTA	0.035	-0.044	0.061	-0.450	0.151	-0.087
	(0.628)	(-0.378)	(0.416)	(-0.617)	(1.306)	(-0.685)
RDSALES	0.245	0.279	0.429	2.619	-0.083	0.415
	(2.028)**	(1.199)	(1.647)	(1.575)	(-0.27)	(1.692)
LNMB	0.043	0.046	0.045	0.087	0.066	0.071
	(2.465)**	(1.063)	(0.783)	(0.362)	(1.479)	(1.458)
LNMVE	-0.028	-0.012	-0.036	0.313	-0.011	-0.049
	(-1.279)	(-0.234)	(-0.584)	(1.235)	(-0.225)	(-0.747)
LPINST	-0.018	0.057	0.019	0.301	0.118	-0.012
	(-1.089)	(1.475)	(0.382)	(1.025)	(2.516)**	(-0.249)
LINST	-0.017	-0.157	-0.020	-0.906	-0.162	-0.058
	(-0.851)	(-3.467)***	(-0.339)	(-3.003)***	(-3.331)***	(-1.061)
CONSTANT	0.185	3.765	2.618	-0.035	1.675	2.933
	(0.579)	(4.286)***	(2.863)***	(-0.011)	(2.497)**	(2.732)***
Adj. R^2	0.9587	0.8302	0.6889	0.3159	0.7988	0.8049
N	409	409	409	166	409	336

***Significant at the 0.01 level.
**Significant at the 0.05 level.

$$\mathrm{LTC} = \alpha_0 + \alpha_1\mathrm{LANLYST} + \alpha_2\mathrm{LVOL} + \alpha_3\mathrm{LPRI} + \alpha_4\mathrm{LVAR} + \alpha_5\mathrm{LSIGR} + \alpha_6\mathrm{LSIGVOL} + \alpha_7\mathrm{ERRE} + \alpha_8\mathrm{DISP} + \alpha_9\mathrm{LEVG} + \alpha_{10}\mathrm{LNINTGTA} + \alpha_{11}\mathrm{RDSALES} + \alpha_{12}\mathrm{LNMB} + \alpha_{13}\mathrm{LPINST} + \alpha_{14}\mathrm{LINST} + \varepsilon_{\mathrm{LTC}} \quad (20)$$

$$\mathrm{LANLYST} = \beta_0 + \beta_1\mathrm{LTC} + \beta_2\mathrm{LVAR} + \beta_3\mathrm{LNMVE} + \beta_4\mathrm{LPRI} + \beta_5\mathrm{IND}_1 + \beta_6\mathrm{IND}_2 + \beta_7\mathrm{IND}_3 + \beta_8\mathrm{IND}_4 + \beta_9\mathrm{LPINST} + \beta_{10}\mathrm{LINST} + \varepsilon_{\mathrm{LANLYST}} \quad (21)$$

$$\mathrm{LVOL} = \gamma_0 + \gamma_1\mathrm{LTC} + \gamma_2\mathrm{LANLYST} + \gamma_3\mathrm{LNMVE} + \gamma_4\mathrm{LINST} + \gamma_5\mathrm{LPINST} + \varepsilon_{\mathrm{LVOL}} \quad (22)$$

where LTC = ln(Adverse selection/price) or spread/price:

LANLYST = ln(number of analysts following)
LVOL = ln(volume)
LPRI = ln(price)
LVAR = ln(variance of spread midpoint)
LSIGR = ln(standard deviation of returns)
LSIGVOL = ln(standard deviation of daily volume)
ERRE = Analyst forecast error
DISP = Dispersion of analyst forecasts
LEVG = Debt/Total Assets
RDSALES = R&D expense/sales
LNINTGTA = ln(Intangibles/Total Assets)
LNMB = ln(market-to-book ratio)
LPINST = ln(percentage of institutional ownership)
LINST = ln(number of institutional owners)
$\mathrm{IND}_1 - \mathrm{IND}_4$ = Industry dummies

The original Brennan and Subrahmanyam (1995) model is modified to include additional variables to the LTC and the LVOL equations. Volatility measures (LEVG, LSIGR, and LSIGVOL), uncertainty variables (LNINTGTA, RDSALES, and LNMB), and informed trader proxies (LINST and LPINST) are added to the LTC equation. We add LINST and LPINST to the LVOL equation.

Table VI presents the results of the simultaneous equation model. The benchmark regression uses spread as the dependent variable. If the adverse selection models are any improvement over spread alone, we should expect them to have more of the predicted relations. Spread is positively related to number of analysts, a finding that supports the results of Chung et al. (1995). This relation also holds for the Huang and Stoll (1997) and Madhavan et al. (1997) models. The George et al. (1991) estimate is negatively related to LANLYST, confirming the result of Brennan and Subrahmanyam (1995). As noted earlier, a positive relation indicates that analysts are following stocks with greater information uncertainty and therefore greater profit potential. A negative relation could indicate that analyst information production reduces uncertainty about the stock's value. Because the models do not agree on the sign of this important variable, we have serious concerns about the ability of the models to measure adverse selection.

The other informed trader proxies present mixed results. LPINST is significant and positive for three models (Huang and Stoll, 1997; Lin et al., 1995; and marginally for Glosten and Harris, 1988). These coefficients support the expectation that a greater percentage of institutional ownership is likely to lead to greater adverse selection. However, the LINST

Table VI. Three-Stage Least Squares Models of Adverse Selection, Analyst Following, and Volume

Panels A, B, and C present the results from the following 3SLS regression model, where LTC is the log of the adverse selection component divided by price. In the first column, we use spread as measured by the log of spread divided by price to represent LTC.

$$LTC = \alpha_0 + \alpha_1 LANLYST + \alpha_2 LVOL + \alpha_3 LPRI + \alpha_4 LVAR + \alpha_5 LSIGR + \alpha_6 LSIGVOL + \alpha_7 ERRE + \alpha_8 DISP + \alpha_9 LEVG + \alpha_{10} LNINTGTA + \alpha_{11} RDSALES + \alpha_{12} LNMB + \alpha_{13} LPINST + \alpha_{14} LINST + \varepsilon_{LTC}$$

$$LANLYST = \beta_0 + \beta_1 LTC + \beta_2 LVAR + \beta_3 LNMVE + \beta_4 LPRI + \beta_5 IND_1 + \beta_6 IND_2 + \beta_7 IND_3 + \beta_8 IND4 + \beta_9 LPINST + \beta_{10} LINST + \varepsilon_{LANLYST}$$

$$LVOL = \gamma_0 + \gamma_1 LTC + \gamma_2 LANLYST + \gamma_3 LNMVE + \gamma_4 LINST + \gamma_5 LPINST + \varepsilon_{LVOL}$$

where *LANLYST* is the log of the number of analysts, *LVOL* is the log of volume, *LPRI* is the log of price, *LVAR* is the log of the variance of the spread midpoint, *LSIGR* is the log of the standard deviation of returns, *LSIGVOL* is the log of the standard deviation of daily volume, *ERRE* is the analyst forecast error, *DISP* is the dispersion of analyst forecasts, *LEVG* is Debt divided by Total Assets, *RDSALES* is the R&D cost divided by sales, *LNINTGTA* is the log of Intangibles divided by Total Assets, *LNMB* is the log of the market-to-book ratio, *LPINST* is the log of the percentage of institutional ownership, *LINST* is the log of the number of institutional owners, and $IND_1 - IND_4$ are industry dummies. We use the models of Glosten and Harris (1998) (GH), George et al. (1991) (GKN), Huang and Stoll (1997) (HS), Lin et al. (1995) (LSB), and Madhavan et al. (1997) (MRR) to compute the adverse selection components.

Panel A. LTC (Adverse Selection or Spread)

	Spread		GH		GKN		HS		LSB		MRR	
	Coef.	Z	Coef.	Z	Coef.	Z	Coef.	Z	Coef.	Z	Coef.	Z
LANLYST	0.205	*2.890****	0.019	*0.117*	-0.674	*-3.568****	2.543	*2.714****	0.013	*0.084*	0.637	*3.369****
LVOL	-0.217	*-9.311****	-0.365	*-6.937****	-0.132	*-2.055***	-0.666	*-1.814*	-0.226	*-4.438****	-0.591	*-9.018****
LPRI	-0.706	*-24.815****	-0.338	*-5.245****	-0.295	*-4.034****	0.136	*0.362*	-0.335	*-5.702****	-0.474	*-5.899****
LVAR	0.020	*1.895*	0.123	*4.994****	0.055	*2.025***	-0.032	*-0.234*	0.037	*1.658*	0.123	*4.429****
LSIGR	0.178	*4.626****	0.293	*3.366****	0.425	*4.260****	0.007	*0.017*	0.547	*6.932****	0.154	*1.544*
LSIGVOL	-0.003	*-0.388*	-0.046	*-2.833****	-0.010	*-0.536*	-0.007	*-0.120*	0.000	*-0.021*	-0.011	*-0.566*
ERRE	-0.003	*-1.201*	-0.012	*-2.115***	-0.005	*-0.776*	-0.010	*-0.532*	0.000	*-0.019*	-0.007	*-1.206*
DISP	0.002	*0.491*	-0.007	*-0.878*	0.001	*0.153*	0.003	*0.170*	-0.002	*-0.207*	-0.002	*-0.292*
LEVG	0.018	*0.450*	0.137	*1.528*	0.109	*1.039*	0.015	*0.049*	0.063	*0.802*	0.084	*0.773*
LNINTGTA	0.051	*0.915*	0.011	*0.089*	0.005	*0.037*	-0.045	*-0.089*	0.067	*0.611*	0.056	*0.397*
RDSALES	0.156	*1.071*	0.076	*0.234*	0.204	*0.548*	0.521	*0.528*	0.008	*0.027*	-0.012	*-0.036*
LNMB	0.041	*2.155****	0.059	*1.393*	0.007	*0.144*	-0.005	*-0.021*	0.028	*0.723*	0.034	*0.696*
LPINST	0.001	*0.049*	0.064	*1.887*	0.018	*0.409*	0.418	*2.165***	0.130	*3.847****	-0.006	*-0.138*
LINST	-0.034	*-1.584*	-0.183	*-3.820****	-0.053	*-0.853*	-0.884	*-2.414***	-0.188	*-3.989****	-0.060	*-1.058*
CONSTANT	1.041	*3.974****	3.321	*5.617****	0.540	*0.747*	2.571	*0.700*	1.769	*3.069****	5.096	*7.180****
Adj. R^2	0.9410		0.8240		0.4754		-0.2865		0.9410		0.6648	
N	409		409		409		166		409		336	

***Significant at the 0.01 level.
**Significant at the 0.05 level.

Table VI. Three-Stage Least Squares Models of Adverse Selection, Analyst Following, and Volume (*Continued*)

Panel B. ANLYST (Analyst Following)

	Spread		GH		GKN		HS		LSB		MRR	
	Coef.	Z	Coef.	Z	Coef.	Z	Coef.	Z	Coef.	Z	Coef.	Z
LTC	-0.949	-2.137**	-0.415	-2.353**	-0.629	-3.439***	0.278	2.284**	-0.145	-0.578	-0.467	-2.103**
LVAR	0.114	6.420***	0.124	7.442***	0.100	5.610***	0.114	4.290***	0.083	3.249***	0.126	7.348***
LNMVE	0.069	0.930	0.159	2.442**	0.056	0.830	0.194	2.730***	0.183	2.900***	0.189	2.166**
LPRI	-0.794	-3.092***	-0.409	-10.590***	-0.355	-7.323***	-0.399	-6.037***	-0.238	-2.573**	-0.467	-10.348***
IND1	0.079	1.252	0.011	0.199	0.033	0.465	0.068	0.648	0.192	2.728***	0.052	0.786
IND2	-0.137	-3.260***	-0.075	-2.075**	-0.121	-2.876***	-0.032	-0.479	-0.183	-3.789***	-0.101	-2.441**
IND3	-0.102	-1.193	-0.028	-0.406	-0.054	-0.683	-0.049	-0.485	-0.052	-0.585	-0.042	-0.580
IND4	0.006	0.111	-0.014	-0.305	-0.012	-0.227	0.000	-0.006	-0.022	-0.375	-0.039	-0.712
LINST	0.054	0.960	-0.071	-1.087	0.010	0.181	0.374	2.286**	0.051	0.707	-0.115	-1.664
LPINST	-0.055	-1.000	0.053	1.136	-0.003	-0.059	-0.108	-1.096	0.024	0.473	0.042	0.717
CONSTANT	-0.977	-1.057	0.139	0.287	-0.867	-1.514	1.934	3.467***	0.514	0.803	0.301	0.526
Adj. R^2	0.3899		0.3506		0.1182		-0.1742		0.4083		0.3780	
N	409		409		409		166		409		336	

Panel C. VOL (Volume)

	Spread		GH		GKN		HS		LSB		MRR	
	Coef.	Z	Coef.	Z	Coef.	Z	Coef.	Z	Coef.	Z	Coef.	Z
LTC	0.835	5.165***	0.308	0.851	1.693	3.937***	-1.134	-4.178***	1.573	4.084***	-0.326	-0.749
LANLYST	2.445	13.314***	2.939	16.193***	3.114	12.841***	3.498	11.053***	1.975	7.484***	2.629	15.214***
LNMVE	0.483	4.645***	0.105	0.862	0.562	3.254***	0.041	0.251	0.662	4.074***	-0.134	-0.825
LINST	-0.040	-0.361	0.163	1.138	0.153	1.046	-0.962	-2.381**	0.378	2.665***	0.203	1.450
LPINST	0.270	2.592***	0.023	0.233	0.170	1.257	0.532	2.145**	0.036	0.313	-0.072	-0.620
CONSTANT	12.038	20.238**	10.078	10.201***	14.582	10.176***	5.391	4.574***	14.439	11.385***	8.623	7.708***
Adj. R^2	0.9609		0.1334		-0.4810		-0.6205		0.3803		0.2632	
N	409		409		409		166		409		336	

***Significant at the 0.01 level.
**Significant at the 0.05 level.

variable is negative for these models, which suggests that a greater number of institutional owners reduces adverse selection. This latter finding could result from greater competition between informed traders, reducing the overall level of adverse selection. Both of the institutional variables are insignificant in the spread regression. This result provides some evidence for the superiority of the adverse selection models in capturing information problems attributable to the presence of informed traders.

One or more of the volatility measures, LVAR, LSIGR, and LSIGVOL, generates significant relations with the adverse selection estimates from Glosten and Harris (1988), George et al. (1991), Lin et al. (1995) and Madhavan et al. (1997). LSIGR is also positively related to spread. We must be careful about inferring that the relations between adverse selection and the volatility measures are caused by information uncertainties. An alternative explanation is that inventory costs are higher because of higher volatility, and these costs are reflected in a higher spread.[8] Since these models (Glosten and Harris, 1988; George et al., 1991; Lin et al., 1995; and Madhavan et al., 1997) are highly correlated with the spread, they may be picking up this correlation between spread and inventory costs. The Huang and Stoll (1997) model estimate also has the lowest correlation with spread (see Table II, Panel A).

Measures of information asymmetries, such as ERRE, DISP, LNINTGTA, RDSALES, and LNMB, are surprisingly insignificant for all the models (except Glosten and Harris, 1988, which is negatively related to ERRE). It could be that these variables are captured to some degree in the volatility variables. However, spread is positively related to market-to-book. This finding supports the argument that firms with greater growth opportunities are more opaque.

The results of the three-stage least squares analysis paint a mixed picture. The adverse selection models seem to capture the volatility of the stock, thus supporting the hypothesis. However, most of the variables that measure information asymmetries are not related to adverse selection. The informed trader proxies have some impact on adverse selection, but this impact is not uniform across the models. No single model appears to perform significantly better than the others. Although the Glosten and Harris (1988) model has the largest number of predicted relations, the LANLYST variable for the Glosten and Harris component is insignificant. The Madhavan et al. (1997) and George et al. (1991) models perform in a similar fashion. This is not surprising, since these models are closely related. The Huang and Stoll (1997) model has the least number of predicted relations but is the only model that has the hypothesized relation for all the informed trader variables. We consider that this is an important quality of this model. However, the smaller sample size that results from the large number of implausible estimates is a drawback that severely reduces the usefulness of the Huang and Stoll (1997) model.

IV. Conclusion

In this paper, the ability of five frequently used adverse selection models to measure asymmetric information is analyzed. We believe that such an analysis is essential if these models are to have a practical use in empirical finance.

As a measure of adverse selection, we benchmark the performance of the adverse selection components against the spread. The tests are joint tests of whether the adverse selection components measure adverse selection and whether the other variables correctly measure the conditions that might lead to information asymmetries. This issue is addressed by examining variables that have been frequently used in other studies as proxies for information problems. A

[8]The authors thank the referee for suggesting this explanation.

three-stage least squares model is used to control for potential endogeneity problems.

Overall, the results indicate that the adverse selection models are related to stock volatility and the presence of informed traders, although the significance of these relations varies from model to model. However, these variables are also related to spread. Therefore, we cannot say whether these models are capturing uncertainty due to information problems, or some other costs. For example, higher volatility stocks and stocks with greater institutional ownership could require the specialist to carry greater inventory. There are also no relations between adverse selection component estimates from four of the five models we study and several other variables that are designed to capture uncertainty (analyst forecast errors, market-to-book, R&D expenses, and intangibles). We find these insignificant relations puzzling. However, perhaps they show that these adverse selection components are merely noisy measures of spread.

The Huang and Stoll (1997) model stands out from the other models in that its only significant relations are those with the informed trader variables, raising the possibility that their model is superior at measuring adverse selection. However, these results are tempered by the problems that arise in generating reliable estimates for this model.

Overall, the results suggest that we cannot conclusively rule out that the adverse selection models we study might not be capturing other costs of trading. Furthermore, because of the high correlation of spread and adverse selection, several of the models appear to be noisy transformations of the spread. Therefore, we believe that researchers should use care when implementing adverse selection models to use as proxies for information problems.

An alternative explanation for the results is that the corporate finance proxies are poor measures of asymmetric information relative to the adverse selection components. However, given the long history of usage of variables such as market to book and R&D expenses as proxies for asymmetric information, this is a less likely explanation. Furthermore, the finding that the adverse selection components perform in a similar manner to the spread alone suggests that the weakness lies with the estimated adverse selection components rather than the corporate finance variables.■

References

Affleck-Graves, J., S.P. Hegde, and R.E. Miller, 1994, "Trading Mechanisms and the Components of the Bid-Ask Spread," *Journal of Finance* 49, 1471-1488.

Ali, A., A. Klein, and J. Rosenfeld, 1992, "Analysts' Use of Information About Permanent and Transitory Earnings Components in Forecasting Annual EPS," *The Accounting Review* 67, 183-198.

Bessembinder, H., 1999, "Trade Execution Costs on Nasdaq and the NYSE: A Post-Reform Comparison," *Journal of Financial and Quantitative Analysis* 34, 387-407.

Bhushan R., 1989, "Firm Characteristics and Analyst Following," *Journal of Accounting and Economics* 11, 255-274.

Brennan, M.J., and A. Subrahmanyam, 1995, "Investment Analysis and Price Formation in Securities Markets," *Journal of Financial Economics* 38, 361-381.

Brennan, M.J., N. Jegadeesh, and B. Swaminathan, 1993, "Investment Analysis and the Adjustment of Stock Prices to Common Information," *Review of Financial Studies* 6, 799-824.

Choi, J., D. Salandro, and K. Shastri, 1988, "On the Estimation of Bid-Ask Spreads: Theory and Evidence," *Journal of Financial and Quantitative Analysis* 23, 219-230.

Chung, K.H., T.H. McInish, R.A. Wood, and D.J. Wyhowski, 1995, "Production of Information, Information Asymmetry and the Bid-Ask Spread: Empirical Evidence From Analysts' Forecasts," *Journal of Banking and Finance* 19, 1025-1046.

Clarke J. and K. Shastri, 2000, "On Information Asymmetry Metrics," University of Pittsburgh Working Paper.

Conrad, J. and C.M. Niden, 1992, "Order Flow, Trading Costs, and Corporate Acquisition Announcements," *Financial Management* 21, 22-31.

D'Mello, R. and S. Ferris, 2000, "The Information Effects of Analyst Activity as the Announcement of New Equity Issues," *Financial Management* 29, 78-95.

Fee, C.E. and S. Thomas, 1999, "Corporate Diversification, Asymmetric Information, and Firm Value: Evidence From Stock Market Trading Characteristics," University of Pittsburgh Working Paper.

Flannery, M.J., S.H. Kwan, and M. Nimalendran, 2000, "Market Evidence on the Opaqueness of Banking Firm's Assets," University of Florida Working Paper.

Frankel R. and C. Lee, 1998, "Accounting Valuation, Market Expectation, and Cross-Sectional Stock Returns," *Journal of Accounting and Economics* 25, 283-319.

George, T.J., G. Kaul, and M. Nimalendran, 1991, "Estimation of the Bid-Ask Spread and its Components: A New Approach," *Review of Financial Studies* 4, 623-656.

Glosten, L.R. and L.E. Harris, 1988, "Estimating the Components of the Bid/Ask Spread," *Journal of Financial Economic,* 21, 123-42.

Goldstein, M. and K. Kavajecz, 2000, "Eighths, Sixteenths, and Market Depth," *Journal of Financial Economics* 56, 125-149.

Hong, H., T. Lim, and J. Stein, 1999, "Bad News Travels Slowly: Size, Analyst Coverage and the Profitability of Momentum Strategies," *Journal of Finance* 55, 265-295.

Huang, R.D. and H.R. Stoll, 1994, "Market Microstructure and Stock Return Predictions," *Review of Financial Studies* 7, 179-213.

Huang, R.D. and H.R. Stoll, 1997, "The Components of the Bid-Ask Spread: A General Approach," *Review of Financial Studies* 10, 995-1034.

Jones, C. and M. Lipson, 2000, "Sixteenths: Direct Evidence on Institutional Trading Costs," *Journal of Financial Economics* 59, 253-278.

Krinsky, I., and J. Lee, 1996, "Earnings Announcements and the Components of the Bid-Ask Spread," *Journal of Finance* 51, 1523-1535.

Krishnan, C., 2000, "Determinants of the Pricing Process: Re-examining the Huang-Stoll (1997) Model," University of Wisconsin-Madison Working Paper.

Kyle, A.S., 1985, "Continuous Auctions and Insider Trading," *Econometrica* 53, 1315-1336.

Lee, C.M.C. and M.A. Ready, 1991, "Inferring Trade Direction from Intraday Data," *Journal of Finance* 46, 733-746.

Lin, J., G. Sanger, and G.G. Booth, 1995, "Trade Size and Components of the Bid-Ask Spread," *Review of Financial Studies* 8, 1153-1183.

Lin, J., 1993, Order Persistence, Adverse Selection, and Gross Profits Earned by NYSE Specialists," *Journal of Finance* Proceedings (Abstract) (July), 1108-09.

Madhavan, A., M. Richardson, and M. Roomans, 1997, "Why Do Security Prices Change? A Transaction-Level Analysis of NYSE Stocks," *Review of Financial Studies* 10, 1035-1064.

Madhavan A. and S. Smidt, 1991, "A Bayesian Model of Intraday Specialist Pricing," *Journal of Financial Economics* 30, 99-134.

Merton, R., 1987, "A Simple Model of Capital Market Equilibrium with Incomplete Information," *Journal of Finance* 42, 483-510.

Neal, R. and S.M. Wheatley, 1998, "Adverse Selection and Bid-Ask Spreads: Evidence from Closed-End Funds," *Journal of Financial Markets* 1, 121-149.

Roll, R., 1984, "A Simple Implicit Measure of the Effective Bid-Ask Spread in an Efficient Market," *Journal of Finance* 39, 1127-1139.

Scherr, F.C. and H.M. Hulburt, 2001, "The Debt Maturity Structure of Small Firms," *Financial Management* 30, 85–111.

Singh A.K., M.A. Zaman, and C. Krishnamurti, 1994, "Liquidity Changes Associated with Open Market Repurchases," *Financial Management* 23, 47-55.

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

Thomas, S., 1999, "Firm Diversification and Asymmetric Information: Evidence from Analysts' Forecasts and Earnings Announcements," *Journal of Financial Economics* (Forthcoming).