NICOLAS P.B. BOLLEN
TOM SMITH
ROBERT E. WHALEY*

MODELING THE BID/ASK SPREAD: On the Effects of Hedging Costs and Competition

ABSTRACT

The need to understand and measure market maker bid/ask spreads is crucial in evaluating the merits of competing market structures and security designs. Prior studies of bid/ask spreads suffer from several forms of misspecification, including inadvertent and erroneous use of weighted least squares regression. This study develops a simple, parsimonious model of the determinants of spread, and then tests it empirically on a sample of NASDAQ stocks. The model performs well and avoids the distortions of prior work. The study demonstrates the importance of proper model specification in providing meaningful inference regarding the determinants of spread.

August 7, 2001

* Bollen is from the Owen Graduate School of Management, Vanderbilt University; Smith is from the Australian Graduate School of Management, University of New South Wales; and Whaley is from the Fuqua School of Business, Duke University.

MODELING THE BID/ASK SPREAD:
On the Effects of Hedging Costs and Competition

The past thirty years have witnessed a proliferation of new financial products and exchanges worldwide. With each new product come decisions regarding market structure. Should the exchange assign the product to a single specialist responsible for making a market, or should it encourage competition among a number of willing market makers? Should trading take place at a specific location with face-to-face contact among traders and market makers, or should it take place in an anonymous, electronic format? And, with each new product introduced in these markets come decisions regarding optimal design. What should be the size of a standard trading unit? What should be the security's minimum price variation (i.e., tick size)?

Ultimately, informed decision-making regarding these issues must involve a careful balancing of the profit motives of key market participants. Investors (i.e., demanders of liquidity), on the one hand, want to maximize their profits after trading costs. Market makers (i.e., suppliers of liquidity), on the other, want to maximize profits from their business operations. A common focus of both of these sets of market participants is a security's bid/ask spread (i.e., the price of liquidity). The higher the spread, the greater the market maker profit per unit traded, but the fewer the number of units demanded. Understanding and measuring the determinants of the bid/ask spread are, therefore, critical in evaluating the merits of different market structures and products.

The purpose of this paper is to develop and test a new model of the market maker's bid/ask spread. As we show, the model is simple, parsimonious, and well grounded from a theoretical perspective. Using a sample of 1,689 NASDAQ stocks during November 1998, we also show that the model is strongly supported empirically. The outline of the paper is as follows. To begin, we describe the current state of the theoretical and empirical literature on market maker spreads. This discussion has two parts. In Section I, we provide a general framework for categorizing the costs associated with market making. Since much of our contribution consists of methodological improvements on the state of the art in bid/ask spread analysis, a detailed review of eight prior studies is provided in Section II. Section III contains the formal development of our theoretical model and contrasts its structure with the models used in earlier work. Section IV contains an empirical assessment of the model, and examines the importance of proper model specification in providing meaningful inference regarding the determinants of spread. We highlight problems of variable selection, model specification, and estimation that can distort inference. Section V contains a brief summary.

I. A FRAMEWORK OF ANALYSIS

A number of theoretical models specify the cost components of the market maker's bid/ask spread. Stoll (1978a) posits that market maker costs fall into three categories: order-processing costs, inventory-holding costs, and adverse information costs. Order-processing costs are those directly associated with providing the market making service and include items such as the exchange seat, floor space rent, computer costs, informational service costs, labor costs, and the opportunity cost of the market maker's time. Since these costs are largely fixed, at least in the short run, their contribution to the size of the bid/ask spread should fall with trading volume — the higher the trading volume, the lower the bid/ask spread. To some degree, however, this relation may be obfuscated by the fact that market makers often make markets in more than one security. In such cases, fixed order-processing costs can be amortized over total trading volume across securities. In addition, in a highly competitive market, bid/ask spreads should equal the expected marginal cost of supplying liquidity, in which case order-processing costs may be irrelevant in determining spread.[1]

Inventory-holding costs are the costs that a market maker incurs while carrying positions acquired in supplying investors with immediacy of exchange (i.e., liquidity). Here there are two obvious considerations: the opportunity cost of the funds that are tied up in carrying the market maker's inventory and the risk that the inventory value will change adversely as a result of security price movements. With respect to the cost of funds, however, it is important to recognize that market makers try to reduce or close out positions before the close of trading each day. If positions are opened and closed in the

[1] Anshuman and Kalay (1998) show that, if the startup costs to creating a competing exchange are significant, the tick size (i.e., the security's minimum price increment, can be set high enough that market makers can recoup their fixed costs as well as earn an economic profit.

same day, the marginal cost of financing is zero. But, even if inventory is carried overnight, it is not clear whether it represents a cost or a benefit. If during the day, the preponderance of customer orders are buys, the market maker may well be short inventory, in which case he will earn (rather than pay) interest overnight. Price-change volatility, on the other hand, appears to have an unambiguous effect on the bid/ask spread. Market makers can hedge the value of their inventory using derivative contracts written on the underlying securities or on other securities whose price movements are highly correlated with the price movements of the securities in the inventory. The hedge will not be costless, however, and will depend on, among other things, the price-change volatility of the securities in the market maker's inventory. The higher the price-change risk, the higher the bid/ask spread.

The third category is adverse selection costs. These costs arise from the fact that market makers, in supplying immediacy, may trade with individuals who are better informed about the expected price movement of the underlying security. For an individual stock, for example, it is easy to imagine that certain individuals possess "insider" information. Advance news of earnings, mergers, acquisitions, restructurings, spin-offs, and management changes are only a few examples that come to mind. While the intuition underlying why adverse selection may be an important determinant of spread is clear, the selection of an accurate measure of adverse selection costs is not. Probably the best proxy is stock price-change volatility.[2] Adverse selection is related to information flow, and the greater the information flow, the higher the price-change volatility, and hence the higher the bid/ask spread. It should also be noted that information flow may affect the bid/ask spread through trading volume, which is a proxy for the effects of order-processing costs. But, here the relation is direct—the higher the information flow (i.e., adverse selection), the higher the bid/ask spread.

---

[2] Other proxies for the effects of adverse selection costs have also been used. Branch and Freed (1977), for example, use the number of securities in which a dealer makes a market to proxy for adverse selection—the larger the number of securities managed, the less informed the dealer is, on average, about a particular stock. Stoll (1978b), on the other hand, uses a measure of turnover (i.e., dollar trading volume divided by market capitalization)—the higher the turnover, the greater the adverse selection. Glosten and Harris (1988) use the concentration of ownership by insiders—the higher the concentration, the greater the possibility of adverse selection. Finally, Harris (1994) uses the market value of shares outstanding—the larger the firm, the more well known, and hence the lower the possibility of adverse selection.

In addition to these costs, the level of the market maker's bid/spread is likely to be affected by the level of competition, particularly in an environment in which barriers to entry in the market for markets are being slowly but surely eliminated. As competition increases, the bid/ask spread approaches the expected marginal cost of supplying liquidity, that is, the sum of inventory-holding costs and adverse selection costs. The larger the number of market makers, the greater the competition, and the lower the bid/ask spread.

II. A RECONCILIATION OF PAST WORK

The model specifications used in prior studies to examine the empirical relation between the bid/ask spread and its determinants are nested in the following general regression equation:

$$SPRD_i = a_0 + a_1 OPC_i + a_2 IHC_i + a_3 ASC_i + a_4 COMP_i + e_i , \qquad (1)$$

where $SPRD_i$ is the difference between a security's bid and ask quotes (i.e., bid/ask spread), OPC_i is order-processing costs, IHC_i is inventory-holding costs, ASC_i is adverse selection costs, and $COMP_i$ is the degree of competition. We now review the regression specifications used in eight different studies and reconcile their models and results within our general framework. For ease of comparison and exposition, the model specifications and empirical results of past studies are summarized in Table 1. Also for expositional convenience, we now refer to the underlying security as being a share of common stock. We do this only because all of the studies discussed focus on bid/ask spreads in stock markets. It is not intended to diminish the importance of studies of spreads in other markets.[3]

Demsetz (1968) is the first empirical study to investigate the determinants of the market maker's bid/ask spread. He regresses a stock's spread on the logarithm of the number of trades/the logarithm of the number of shareholders, share price, and number of

---

[3] Neal (1987), for example, examines bid/ask spreads of individual stocks traded on the AMEX. George and Longstaff (1993) examine spreads of S&P 100 index options traded on the Chicago Board Options Exchange (CBOE), and Smith and Whaley (1994) examines the spread of the S&P 500 futures contract traded on the Chicago Mercantile Exchange (CME).

exchanges. The numbers of trades/shareholders, he argues, are direct proxies for the transaction cost rate. The higher the transaction cost rate, the lower the cost of waiting, and hence the lower the bid/ask spread. Under this line of argument, these trading frequency variables fall into the inventory-holding cost category. To see this, recall that price volatility is the only unambiguous determinant of inventory-holding costs. Price volatility is related not only to the variance of price changes but also to the amount of time that the market maker expects to hold an open position.

Demsetz' rationale for including share price as a determinant in the regression is as follows:

> "Spread per share will tend to increase in proportion to an increase in the price per share so as to equalize the cost of transacting per dollar exchanged. Otherwise, those who submit limit orders will find it profitable to narrow spreads on those securities for which spread per dollar exchanged is larger." (Demsetz (1968, p.45))

This line of reasoning suggests that relative spread (i.e., bid/ask spread divided by bid/ask midpoint) should be equal across stocks, holding other factors constant. This implies, of course, the higher the price, the higher the spread. This linkage between spread and price appears to be based on the notion that share price is a proxy for the market maker's capital investment. The higher the price, the greater the investment in inventory, the higher the carrying costs, and the higher the spread. But, as we have already noted, many market makers are in and out of positions during the trading day, in which case capital investment in inventory is not necessary let alone costly. Moreover, even if inventory is carried overnight, the market maker is not always in a net long position. A net short position generates cash that could be used to *earn* interest income.

That is not to say that we do not expect to find a positive relation between spread and price in Demsetz' regression, however. We do. The reason is that price may be acting as a proxy for inventory price risk in Demsetz' regression. To some degree, trading frequency captures one component of inventory price risk, that is, the amount of time the market maker expects the position to be open. But, another component, the variance of price changes, is not included. Since the standard deviation of price change is the product of the standard deviation of stock return and share price, a positive relation between spread and share price may appear in Demsetz' regression simply as a result of the fact that share price is is correlated with share price volatility.

Finally, Demsetz includes a competition variable, which he measures as the number of exchanges on which the stock was listed. To be sure, competition should reduce spreads. In Demsetz' investigation, however, we should not expect the relation to be very strong. The reason is that Demsetz' sample includes only New York Stock Exchange (NYSE) stocks. For these stocks, the lion's share of trading occurs on the NYSE and very little occurs on other exchanges. Consequently, the effects of competition, at least as measured by the number of exchanges making markets, will be undermined. A better measure of competition might be the number of market makers standing at the specialist's post, but such data are difficult to obtain.

Demsetz' sample includes 192 New York Stock Exchange (NYSE) stocks on two days during early 1965. More specifically, his cross-sectional regression is based on observations created by averaging the respective variables for each stock across January 5 and February 28, 1965. He finds that spread varies inversely and significantly with the log of the number of trades in one regression, and the log of the number of shareholders in another, and directly and significantly with price per share. The coefficient on the competition variable is negative, but insignificant.

Tinic (1972) argues that Demsetz' results are undermined by the fact that several stock characteristics affecting spread are either not considered or measured imprecisely. One important missing factor is inventory price risk. Tinic argues that a specialist is unable to hold a diversified position and must be compensated for bearing price risk. We agree. Unfortunately, he uses the standard deviation of price, not price change, as his measure of risk. This introduces an errors-in-the-variables problem that will tend to negate the importance of risk.

Tinic also provides an improved measure of competition. He uses a Herfindahl Index of concentration, which includes not only the number of markets but also the overall size and distribution of trading activity among markets. He also re-specifies the trading frequency variables that affect inventory-holding costs. In place of number of trades or number of shareholders, he uses the logarithm of number of shares traded, the number of institutions holding the stock, and the percentage of trading days on which at least one trade occurs. In Table 1, we include trading volume in the order-processing cost category. Holding trading frequency constant, spreads should fall with number of shares traded.

Tinic's sample contains 80 NYSE stocks over 19 trading days in March 1969. Like Demsetz, he averages his observations across days in the sample. Among other things, he finds that the relation between spread and the standard deviation of price is positive, but insignificant. One possible explanation is the errors-in-the-variables problem alluded to earlier. Another is that both standard deviation of price and price are included as explanatory variables. It is possible that the share price variable is better proxy for risk than the standard deviation of share price.

Tinic also finds that spreads increase significantly as trading in a particular security concentrates in one market. The use of the Herfindahl Index better captures the effects of competition. The logarithm of number of shares traded, the number of institutional trades holding the stock, and the percentage of trading days on which at least one trade occurs enter the regression with significant negative coefficients.

Tinic and West (1972) criticize the Demsetz study in a manner similar to Tinic (1972), but remedy the problems differently. They argue that price risk should be included and measure it as the ratio of the difference between high and low prices to the average share price. They argue that the number of exchanges making markets in NYSE stocks is a poor proxy for measuring competition, so they focus on the NASDAQ market, where barriers to entry are much lower. Based on a sample of 300 NASDAQ stocks on the first 5 trading days in November 1971, they find that the risk measure has a positive but insignificant coefficient. Again, errors-in-the-variables may be the problem. Risk measures based only on extreme points of a distribution are notoriously unreliable. [need a cite here] Moreover, the share price variable, which has a positive and significant coefficient, may be proxying for risk. The relation between spread and competition is negative and significant, re-affirming the importance of the level of competition in determining spreads.

In a comparative study, Tinic and West (1974) examine spreads of stocks listed on the Toronto Stock Exchange (TSE). Their purpose is to determine whether spreads set on the TSE are higher or lower, holding other factors constant, than spreads on the NYSE and NASDAQ given that the markets have different structures (i.e., rules and regulations). Unlike the NYSE, for example, the TSE dealers are regulated to be relatively passive participants and are not charged with the responsibility of maintaining price continuity. Policy issues aside, this study is important from a methodological standpoint because it is the first to introduce relative spread (i.e., the bid-ask spread divided by the bid-ask midpoint) as a dependent variable.

Tinic and West begin by estimating a model of absolute spread, with price per share, log of trading volume, price volatility (as measured by the high-low price range divided by price), trading continuity (as measured by the number of days the stock is traded during the sample period divided by the total number of days in the sample period), and the number of markets in which the security is traded being used as explanatory variables. Using a sample of 177 TSE stocks traded during the period December 1 through 13, 1971,[4] they find that the coefficients on share price and price volatility are positive and significant, the coefficient on trading volume is negative and significant, and the coefficients on trading continuity and number of exchanges are negative and insignificant. The adjusted R-squared in the regression is .499. They then proceed by estimating "an alternative functional form of the general relationship" that uses relative spread as the dependent variable and drops price per share as an explanatory variable, and find that the adjusted R-squared is a whopping .804! The most powerful explanatory variable turns out to be price volatility, whose coefficient is positive and highly significant. The coefficients on the log of trading volume and trading continuity are negative and significant, and the coefficient on the number of markets is negative and insignificant.

Two comments are in order here. First, if the regression equation for the absolute spread is correctly specified based on theoretical arguments, the regression equation for the relative spread is not. For the relative spread regression to be correctly specified, all of the explanatory variables must be deflated by share price. Second, given the high level

---

[4] The observations in the regression are, for the most part, simple averages of the variables over the nine trading days in the sample period. The exception is price volatility, which is measured by the high-low range over the period divided by the average share price over the period.

of the adjusted R-squared (i.e., .804), the relative spread regression gives the misleading impression of being a powerful explanatory model of bid/ask spread.[5] To understand how misleading the situation can be, assume that all stock spreads in the sample are equal to 1/8th. In such an environment, the absolute spread regression model would have no explanatory power. All of the coefficient values would equal zero, except the intercept term whose value would be 1/8th. Now, consider the relative spread regression. Since the absolute spread is assumed to be constant, the variation in the dependent variable is driven *only* by variation in the inverse of share price. It should not be surprising to find that explanatory variables such as price volatility are significant in a statistical sense, not because the regression is telling us anything meaningful about spreads, but rather because the price volatility variable has share price in its denominator. Indeed, as the variation in the share price range across becomes small, the goodness-of-fit of the relative spread regression will become perfect.

Benston and Hagerman (1974) examine month-end spreads of 314 NASDAQ stocks over a five-year period from January 1963 through December 1967. They specify their model in the log-form with the absolute spread being a function of the number of shareholders, price per share, idiosyncratic variance, and the number of dealers. They find that all of the spread determinants enter the regression with the expected sign and are statistically significant. The key innovation in the Benston/Hagerman study is in the measure of price risk—a stock's idiosyncratic risk. Their rationale is that the market maker needs to be rewarded for price risk only to the extent that he is not well diversified or is exposed to traders with superior information (i.e., adverse selection costs).[6] To measure this lack of diversification, they use the squared residual from a regression of stock returns on market returns over the 60-month period.

Given the state of the literature at the time, the introduction of a well-reasoned measure of price risk was certainly warranted. Earlier measures of risk lacked theoretical justification and were prone to measurement error. Nonetheless, the Benston/Hagerman

---

[5] For a lucid discussion of the pitfalls of using ratio variables in a regression framework, see Kronmal (1993).

[6] Benston and Hagerman were the first empirical investigators to include a specific discussion of the impact of adverse selection, a concept that had been introduced a few years earlier by Bagehot (1971).

measure is not beyond reproach. First, there is no particular reason for using the variance as opposed to the standard deviation of the residual in the market model regression. In both cases, the coefficient will measure risk aversion. Standard deviation, however, offers the advantage of being specified in more intuitive units of measurement—stock return rather than stock return squared. Second, the dollar spread of a stock is a function of price change volatility not return volatility. Consequently, the coefficient of a return volatility variable will not be constant across stocks as their regression model assumes, and will be directly proportional to share price or the square of share price depending upon whether the standard error of the residual or the variance of the residual is used as the risk measure.

Branch and Freed (1977) conduct a comparative analysis of spreads on the NYSE versus the American Stock Exchange (AMEX). From a methodological standpoint, the key innovation of this work lies in their advocacy of relative spread rather than absolute spread as the variable of interest. They point out that prior studies are divided in the use of absolute or percentage spreads, but then argue that the use of absolute spreads obscures any non-linearity in the relation between price and spread. Their model specifies that relative spread is a function of the trading volume, the number of exchanges making a market, price risk (as measured by the absolute price change from the previous day divided by share price), and the number of securities handled by the market maker. The reasons for including these variables have already been discussed.

Branch and Freed also include the inverse of share price as a determinant of relative spread. They argue that price discreteness, price clustering, the fixed costs of executing a transaction, and the desire of specialists to set their spreads high enough to permit room to maneuver place a minimum value on the spread, and that the minimum value is proportionately higher for low-priced shares than high-priced shares. The symptoms/problems associated with using a relative spread regression specification have already been noted. The situation is particularly egregious here given that the inverse of share price appears on both sides of the regression equation. Not surprisingly, the most significant variable in their regressions estimated using a sample of 1,734 NYSE stocks on January 24, 1974 and a sample of 943 AMEX stocks on May 28, 1974 is the inverse of share price. Moreover, the second most significant relation is their measure of price risk, which has share price in the denominator.

The models tested thus far were developed through economic reasoning rather than formal mathematical modeling. Given their *ad hoc* nature, they are open to criticisms regarding model specification and variable selection. In an attempt to analyze the supply of dealer services (i.e., liquidity) more rigorously, Stoll (1978a) develops an explicit theoretical model. The model shows that the relative bid/ask spread of a security equals the sum of inventory-holding costs, adverse selection costs, and order-processing costs, where each cost component has a precise definition. Interestingly, the inventory-holding cost expression includes a term that equals the product of return volatility and the expected time the market maker expects the position to be open, a distinction that had otherwise gone unnoticed in the literature.

Stoll (1978b) conducts empirical tests of this theoretical model of the pricing of dealer services. As in earlier work, assumptions regarding model specification and variable selection are necessary. As in Benston and Hagerman (1974), for example, Stoll specifies his regression model in log-linear form. And, because many of the variables in Stoll's theoretical model are not directly observable, proxy variables are substituted. Dollar trading volume, for example, is used to proxy for the length of the market maker's holding period (one of the components of inventory-holding costs), and turnover (as measured by dollar trading volume divided by dollar amount outstanding) is used to proxy for adverse selection costs. Price per share, Stoll argues, captures the effects of order-processing costs, that is, the fixed cost per trade is spread over more dollars for high priced shares.

Stoll's sample includes 2,474 observations of NASDAQ stocks during six consecutive trading days (July 9 through 16) in 1973. The explanatory of the regression model is high, with an adjusted R-squared of .822. As in other studies that use relative spread as the dependent variable, the single, most powerful, explanatory variable is share price, which enters the regression with a negative coefficient. But, also as in these other studies, it is difficult to disentangle whether this means that differences in share price explain differences in spread or differences in the inverse of share price. A similar question can be asked about the dollar trading volume, whose coefficient is negative and significant. Is the coefficient's sign and significance driven by an inverse relation between bid/ask spread and trading frequency or by the correlation between price and the inverse of price?

Harris (1994) argues that the significance of the price variable in the relative spread regressions is driven by minimum price variation and discrete bid-ask spreads. To illustrate his point, he estimates two regressions. In the first, he specifies relative spread to be a function of share price, the standard deviation of stock return, the log of total market value, the inverse of the square root of the trading frequency, the log of the dollar trading volume, and the log of the ratio of the trading volume on the primary exchange to the trading volume across all exchanges. The motivations for including the different variables are as follows. The standard deviation of stock return is a proxy for inventory-holding costs, as shown in Stoll (1978a). The log of the stock's market capitalization serves as an inverse proxy for adverse selection costs. The larger the firm, the greater the degree of public information, the smaller the informational asymmetries among investors, and the smaller are the adverse selection costs. The log of dollar trading volume measures trading activity and should be inversely related to relative spread. The greater the trading activity, the lower are the fixed costs (i.e., order-processing costs) per share. The log of the ratio of trading volume on the primary exchange to consolidated trading volume measures competition.

Harris argues that the inverse of the square root of the number of trades on the primary exchange should be included as an independent variable because it serves as a proxy for the profitability market making in the stock. The fewer the number of trades, the less competitive the market. In the absence of competition, market makers can more easily quote a larger spread to cover their fixed costs and extract a surplus.[7] While this argument has merit, the inverse of trading frequency also acts as a proxy for the expected amount of time that the market maker expects his position to be open, one of the

[7] The motivation for applying the square root transformation is that the rate of market maker participation in trades declines with trading activity as more public orders cross. It can also be justified by information-theoretic considerations: If each transaction conveys a bit of information about value, the standard error of the dealer's inferred estimate of value will be proportional to the inverse square root of the transaction frequency.

components of inventory-holding costs. The number of minutes in a trading day, 390, times the inverse of the number of trades equals the average time between trades.

Harris estimates the regression model using a sample of 529 NYSE and AMEX stocks during the second quarter of 1989. Observations for each regression variable are averages across the days in the quarter. The results of the first regression show that share price, return volatility, the log of market capitalization, and the inverse of the square root of trading frequency have positive and significant coefficients. The coefficients of the log of dollar trading volume and the log of primary exchange volume are negative and significant. The adjusted R-squared of the regression fit is .804. In the second regression, Harris replaces the price variable with the inverse of price and finds that the results change dramatically. First, the goodness-of-fit (i.e., the adjusted R-squared) increases to a level of .987. Second, the coefficient on the inverse of share price is positive and highly significant. Third, the standard deviation of stock return and the inverse of the square root of the trading frequency have positive and significant effects on relative spread, and, fourth, all other variables become insignificant. These results appear to indicate that many stocks have spreads equal to the minimum price variation and, to the extent there is explainable variation in spread, it is driven by inventory-holding costs (e.g., the standard deviation of return and the inverse of the square root of trading frequency are the only significant determinants).

III. MODEL SPECIFICATION

Thus far, we have accomplished two goals. First, we developed a general framework for considering the factors that drive the level of market maker bid/ask spreads, and, second, we reconciled the results of past studies within this framework. We now turn to developing a formal model of the market maker's bid/ask spread.

To begin, we assume that the market maker sets his bid/ask spread in accordance with the general framework described by (1). Total order-processing costs are assumed fixed, so order-processing costs per share are directly proportional to the inverse of trading volume (denoted $InvTV_i$). Likewise, the competition variable is measured as the inverse of the number of dealers making a market in the security (denoted $InvND_i$).

Unlike past studies, we do not attempt to distinguish between inventory-holding costs and adverse selection costs. Both are compensation for the risk of an unfavorable price movement while the security is being held in inventory. In a competitive market, the amount of the compensation will equal the marginal cost of hedging price risk. This hedging cost (denoted HC_i), in turn, can be modeled as an option value.[8] If the market maker has no inventory and accommodates a customer order by buying at the bid, he needs protection against the price falling below his purchase price before he is able to unwind his position. Conversely, if the market maker has no inventory and accommodates a customer order to buy by selling at the ask, he needs protection against the price rising above his sales price. In the first case, the market maker needs to buy an at-the-money put written on the security, and, in the second, he needs to buy an at-the-money call. Under Black-Scholes (1973) option valuation assumptions, the hedging cost can be shown to be

$$HC_i = P_i\left[2N\left(.5s_i\sqrt{T_i}\right)-1\right], \qquad (2)$$

where P_i is the security price at which the market maker opens his position, $N(.)$ is the cumulative unit normal density function, s_i is the standard deviation of return, and T_i is the expected length of the market maker's holding period.[9] In the derivation of (2), the interest rate is assumed to be equal to zero since most open positions are closed by the end of the trading day. With the interest rate at zero, the cost of hedging an unanticipated price drop (i.e., the put option value) is the same as the cost of hedging an unanticipated price increase (i.e., the call option value).

Gathering terms, our model of the market maker's bid/ask spread is

$$SPRD_i = a_0 + a_1 InvTV_i + a_2 HC_i + a_3 InvND_i + e_i. \qquad (3)$$

The coefficient a_1 should be positive and may be quite large. After all, it represents the market maker's total order-processing costs. If the market is competitive, however, the

---

[8] Here we model the size of the bid/ask spread as the value of an at-the-money futures option with an exercise price equal to the price at which the market maker provides immediacy. In contrast, Copeland and Galai (1983) model the bid/ask spread as a straddle in which the market maker provides a prospective trader with an out-of-the-money call option to buy at the ask price and an out-of-the-money put option to sell at the bid price.

[9] The market maker's holding period is, of course, stochastic and depends on factors such as order flow.

market maker may not have the ability to recover fixed costs, in which case the coefficient will be indistinguishably different from zero. The coefficient a_2 should be positive—the higher the hedging costs, the greater the bid/ask spread. Indeed, if an accurate proxy for the expected length of market maker's holding period can be obtained, the coefficient value should be one. Perhaps most important, however, is that, unlike past studies, we have identified the theoretical structural relation between bid/ask spread and many of its determinants. As (2) shows, the marginal costs of inventory-holding and adverse selection are a specific function of share price, return volatility, and the time that the market maker expects the position to be open. Entering the variables separately on the right-hand-side of the regression equation, as has been done in past work, obfuscates their role. Finally, the coefficient a_3 should be positive—the greater the number of dealers, the lower the inverse of the number of dealers, and the lower the spread.

Another virtue of our theoretical model of the bid/ask spread is that, unlike the model's used in past studies, it is structurally consistent with the presence of an exchange-mandated tick size. The tick size of a security is its minimum allowable price increment. The importance of the tick size in this context is that it sets the lower bound of the market maker's bid/ask spread. In (3), the intercept term a_0 represents a stock's minimum price increment because the values of all three variables on the right hand-side of (3) are near or at zero for actively traded securities.

Finally, it is worth noting that, although our theoretical model is specified in terms of the absolute spread, it can also be estimated using relative spreads. But, care must be taken to preserve the underlying economic relation, that is, all variables in the regression must be deflated by price per share,

$$\frac{SPRD_i}{P_i} = a_0\left(\frac{1}{P_i}\right) + a_1\left(\frac{InvTV_i}{P_i}\right) + a_2\left(\frac{HC_i}{P_i}\right) + a_3\left(\frac{InvND_i}{P_i}\right) + u_i \,. \qquad (4)$$

This means that there is no intercept term in the relative spread regression (4) and that the inverse of share price must appear as an explanatory variable.[10] Note that estimating (4) is tantamount to running a weighted least squares (WLS) regression of (3), where the

[10] For a lucid examination of the problems encountered is using ratios as variables in an ordinary least squares regression, see Kronmal (1993).

respective weights are the inverse of share price. The choice between the two estimation methods must be based on the properties of the residuals.

IV. AN EMPIRICAL EVALUATION

The focus now turns to evaluating the empirical performance of our theoretical model (3). The first part of this section provides a description of the data used in our analyses, and the second part includes some summary statistics describing the sample. The third part contains the regression results.

A. Data

The trade and quote data used in this study were downloaded from NYSE's TAQ data files. Although the files contain information for all U.S. exchanges, our sample contains only NASDAQ stocks due to restrictions on the availability of information on the number of dealers.[11] The minimum price increment for NASDAQ stocks during November 1998 was 1/16th.

For all time-stamped trades, we matched the quotes prevailing immediately prior to the trade. From this matched file, we then computed six summary statistics for each stock each day: (a) the number of trades, (b) the end-of-day share price (i.e., the last bid/ask midpoint prior to 4:00PM EST), (c) the number of shares traded, (d) the equal-weighted quoted spread, (e) the volume-weighted effective spread, and (f) the average time between trades.

Thus far, we have said little about the types of spread measures that have been used in past studies. All of the studies cited in Section II use quoted spreads, where the *quoted spread* is defined as

$$\text{Quoted spread}_t = \text{ask price}_t - \text{bid price}_t\,, \tag{4}$$

where the subscript t represents the t-th trade of a particular stock during the trading day. The intuition for this measure is that, if a customer buys a stock and then immediately sells it, he would pay the quoted ask price and receive the quoted bid, thereby incurring a loss (i.e., a trading cost) equal to the bid/ask spread. This measure assumes that customers

[11] Information on the number of dealers making markets in NASDAQ stocks is available on NASDAQ'a website. Although NYSE's TAQ files breaks down trading volume by exchange, it does not break down trading volume by the different market makers standing at the specialist's post.

cannot trade within the quoted spread. It also assumes only market makers set the prevailing quotes and stand on the other side of customer trades. In general, past research has used the quoted spread at the end of the trading day as their variable of focus. We use an equal-weighted average of the quoted spreads (*EWQS*) appearing throughout the trading day.

The *effective* spread[12] circumvents the first of the two weaknesses of the quoted spread. It is based on the notion is that the trade is only costly to the investor to the extent that the trade price deviates from the "true" price, as proxied by the bid/ask price midpoint,

$$\text{Midpoint}_t = \frac{(\text{bid price}_t + \text{ask price}_t)}{2} \tag{5}$$

On a round-turn, the cost would be incurred twice, hence the measure of the effective spread is

$$\text{Effective spread}_t = 2\left|\text{trade price}_t - \text{midpoint}_t\right|. \tag{6}$$

Naturally, if all trades took place at the prevailing bid and ask quotes, the effective spread would be equal to the quoted spread. On the other hand, if some trades take place within the spread, the effective spread will be smaller than the quoted spread.

The effective spread measure assumes that, if a trade takes place above the bid/ask midpoint, it is a customer buy order, and, if it takes place below the bid/ask midpoint, it is a customer sell order. The absolute deviation of the trade price from the bid/ask midpoint, therefore, can be interpreted as the cost incurred by the customer and/or the revenue earned by the market maker. Furthermore, the product of one-half the effective spread times the trading volume can be interpreted as the market maker revenue from the trade. While the effective spread is a better measure for customer trader costs than the quoted spread, it remains overstated in the sense that it fails to account for the fact that trades may be executed between customers and may not involve the participation of the market maker at all. For such a trade, the effective spread equals zero, that is, the price concession conceded by one customer is awarded the other. Absent knowing the identity of both parties on each side of the trade, however, no better measure is possible. The volume-weighted effective spread (*VWES*) is a volume-weighted average of the effective spreads of the trades occurring throughout the day.

With the six summary statistics compiled for each stock each day, we computed average values for each stock across all days in the month. To mitigate the effects of outliers, we then constrained the sample to include only stocks whose average price was at least equal to $5 per share and whose shares traded at least five times each day during all twenty trading days during the month. The final sample contains 1,689 stocks.

Three additional measures were then appended to each stock trade record. First, the number of dealers making a market in the stock were gathered from the website www.nasdaqtrader.com. As a matter of routine, NASDAQ reports this information each month and makes it freely available of their website. Second, the rate of return volatility for each stock was computed using end-of-day share prices. The daily return standard deviation was annualized using the factor, $\sqrt{252}$. Finally, the hedging cost for each stock was computed using the option valuation equation (2), where P is the stock's average share price, s is the annualized return volatility, and T is the average time between trades[13]. Using the average time between trades as a proxy for the expected length of the market maker's holding period understates the hedging cost. Implicitly, it assumes that there is only one dealer making a market in the stock. If trading volume was uniformly distributed across all dealers, we could multiply the average time between trades by the number of dealers in measuring the expected holding period. But, this value would cause hedging costs to be overstated, since only a handful of dealers account for the lion's share of the trading volume of a stock. Thus, without a more refined breakdown of trading

---

[12] This measure of effective spread has been adopted in a number of studies in the stock market including Christie, Harris, and Schultz (1994) and Huang and Stoll (1994). Lightfoot *et al* (1986) use it in their examination of bid/ask spreads in the option market.

[13] Since volatility is expressed on an annualized basis, the time between trades must be measured in years. To accomplish this task, we divided the average number of minutes between trades by 390 (i.e., the number of minutes in a trading day) and then by 252 (i.e., the number of trading days in a year).

volume across market makers, we opt for the more conservation measure.[14]

B. Summary statistics

Table 2 contains summary statistics across the stocks in sample. In the top panel, we report spreads measures. The equal-weighted quoted spread (*EWQS*) for the stocks in the sample has a mean value of .2375 or nearly four ticks. The volume-weighted effective spread (*VWES*), on the other hand, is about 2/3rds the size at .1624. The difference between these values indicates that a large number of trades are being executed at prices within the prevailing bid/ask quotes. Summary statistics for relative spreads are also provided in the table. The relative volume-weighted effective spread (*RVWES*), for example, averages 1.14%, and has a range from .06% (MSFT) to 6.06% (BOSHQ). Clearly, trading less active stocks can be quite costly.

The second panel contains summary statistics for the variables used as determinants of the spread. The average share price is $19.72 per share, and ranges from $5 (ESFT) to $177.91 (YHOO). The average number of shares traded per day is about 350,000, with a minimum of 441 (BOSHQ) and a maximum of 19,592,491 (DELL). The average number of dealers for each stock is 45, reflecting the high degree of competition among market makers on the NASDAQ trading system. The number of dealers (*ND*) ranges from 5 (DXCPO) to 456 (DELL). The average annualized return volatility is .6991, reflecting the high degree of risk of the technology-laden NASDAQ market. The average time between trades (*T*) is 7.4 minutes, and ranges from 0.02 (DELL) to 36.82 minutes (CTBC). Finally, the hedging cost (*HC*) is .0305, indicating that a premium of at least three cents is necessary for market makers to cover their marginal costs of operation. The cost ranges from .08 cents (SHVA) to 28.30 cents (ABBBY).

Table 3 contains estimates of the correlation among the variables used in the analysis. A number of interesting results appear. First, the correlations between absolute spreads and relative spreads are extremely small—.247 for the equal-weighted quoted spreads and .244 for the volume-weighted effective spreads. Recall that past studies of market maker spreads are split in the use of absolute spread and relative spread as the variable of focus. The low level of correlation between the two measures indicates that

---

[14] One way to attack the problem is to set the coefficient a_2 equal to one in equation (3) and use non-linear regression to estimate of the average holding period across stocks. Assuming a constant holding period

the two variables are describing different phenomenon. Since the objective of the studies has been to explain the level of spread, focusing on absolute spread seems the most sensible approach.

Second, the correlations among the regressors in (3) are small, providing assurance that multicollinearity is not affecting our regression estimates in any serious way. The correlation between the inverse of trading volume and hedging cost is .086; the correlation between the inverse of trading volume and the inverse of the number of dealers is .239; and the correlation between hedging cost and the inverse of the number of dealers is .189.

Finally, it is interesting to note that hedging cost is highly correlated with both absolute spread measures, .750 and .745 for the *EWQS* and *VWES*, respectively, while it is virtually uncorrelated with the relative spread measures, .083 and .046 for *REWQS* and *RVWES*, respectively. This underscores the importance of proper model specification. Hedging costs are clearly important in the determination of the bid/ask spread, but the relation gets masked when only of the two variables (i.e., spread) is scaled by share price.

C. Regression results

Table 4 contains a summary of the regression results. All of the *t*-ratios are corrected for heteroscedasticity in the residuals. In all, five different regressions are estimated. The first regression uses the equal-weighted quoted spread as the dependent variable. As the table shows, all of the coefficients are positive and significant in a statistical sense. The single most important explanatory variable appears to be hedging cost. Its coefficient estimate is considerably larger than one, indicating that, as expected, the average time between trades is a downward biased estimate of the expected length of the market maker's expected holding period. The significance of the coefficient a_3 indicates that the inverse of the number of dealers also plays an important role in explaining the absolute level of the bid/ask spread. Clearly increased competition drives spreads to lower levels. The effects of competition are also likely contributing to the fact that the inverse of trading volume has the lowest significance of the variables in the regression. Recall that in a highly competitive market, market makers can recover only their marginal costs. Finally, recall that our model (3) is structured in a way that the level

across stocks, however, is probably unrealistic.

of the intercept term equals the minimum tick size. The estimate of the intercept term a_0 is .0578, and is not significantly different from the minimum tick size in the NASDAQ market, .0625.

The second regression uses effective spread rather than quoted spread as the dependent variable. Since many trades take place within the prevailing price quotes, effective spread is a more accurate measure of market maker revenue. Like in the first regression, hedging cost and the inverse of the number of dealers have the strongest explanatory power. Interestingly, the coefficient of hedging cost drops to a level of 2.6035. Apparently, the average time between trades is not as poor a measure of holding period as we expected, given the large number of dealers making markets. Note also that the intercept term is significantly less than the minimum tick size. This is not surprising since the effective spread can have values as low as zero.[15] The estimate, .0403, represents the level of revenue per share that the market maker can expect to earn for providing liquidity in an extremely active stock.

The third regression is the same as the second, except that all variables are scaled by share price and the intercept term is suppressed. The choice between using model (3) or model (4) should be based primarily on which regression has the most well behaved residuals. Since we are correcting the standard errors for heteroscedasticity in both cases, however, the inferences should not be strikingly different. Indeed, the results of the third regression are very similar to those of the second. They should be. They are the same model.

The fourth regression illustrates what can happen if one inadvertently includes an intercept term in the relative spread regression. While the coefficient estimates of the explanatory variables are about the same order of magnitude as those in the third, the meaning of the coefficient of the inverse of share price is lost. Neither that coefficient nor the intercept term provides any meaningful information about the minimum bid/ask spread. It is also worthwhile to note the adjusted R-squared is considerably larger in the relative spread regression than in the absolute spread regression (i.e., .8119 versus .6540). This comparison is meaningless, however, and does not in any way suggest that the

[15] The effective spread equals zero in instances in which the quoted spread is an even number of ticks and the trade takes place at the midpoint.

relative spread performs better. Our objective is to explain the variance of spread, not the variance of the ratio of spread to price.

The fifth regression is illustrates what can happen when the determinants of spread are specified in an *ad hoc* fashion. The studies reviewed in Section II fall into this category. While each study provides well-reasoned arguments regarding the choice of determinants, the decision regarding the structural relation between the spread and its determinants was arbitrary. Some used a linear model; others a non-linear model. Some used absolute spread; others used relative spread. In Section III, we developed a simple parsimonious theoretical model of the market maker's bid/ask spread. Its structure accounts for the minimum price variation of the stock and provides an explicit measure of the dollar cost of inventory-holding and adverse selection. Suppose we had not developed a theoretical model but had made well-reasoned arguments for including share price, return volatility, and the average time between trades (i.e., the determinants of hedging cost) together with the inverse of trading volume and the inverse of the number of dealers as the determinants of spread. Furthermore, like in some of the past studies, we assume the relation is linear. The results are reported in the fifth panel of Table 4. All of the variables enter the model with their expected signs. Moreover, each variable is significant in the statistical sense. Of course, the intercept term no longer can be interpreted as being the minimum spread. The most important result is that the adjusted R-squared for this regression is only .5175, when the adjusted R-squared for our simpler model (3) is .6540. Although both regressions contain the same independent variables, knowing the proper variable definitions and model structure substantially improves performance.

V. SUMMARY

This study develops and tests a new model of the market maker's bid/ask spread. The model is simple and parsimonious, showing that the market maker's bid/ask spread is only a function of the minimum tick size, the inverse of trading volume, expected hedging cost, and the inverse of the number of dealers. The inverse of trading volume helps isolate the market maker's order processing costs, and the expected hedging cost accounts for the market maker's exposure to inventory price risk and adverse selection. The expected hedging cost is modeled as an at-the-money option and is a function of share price, return volatility, and the length of time that the market maker expects his position to remain open. The inverse of the number of dealers acts as a proxy for competition. The model is tested using a sample of 1,689 NASDAQ stocks during November 1998, and is shown to perform well. The effects of model misspecification are also discussed.

REFERENCES

Anshuman, V. Ravi and Kalay, Avner, 1998, Market making with discrete prices, *Review of Financial Economics* 11 (1), 81-109.

Bagehot, W., 1971, The only game in town, *Financial Analysts Journal*, 12-22.

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

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

Christie, William, Harris, Jeffrey, and Schultz, Paul, 1994, Why did NASDAQ market makers stop avoiding odd-eighth quotes?, *Journal of Finance* 49, 1841-1860.

Copeland, Thomas E. and Galai, Dan, 1983, Information effects on the bid-ask spread, *Journal of Finance* 38, 1457-1469.

Demsetz, Harold, 1968, The cost of transacting, *Quarterly Journal of Economics* 82 (February), 33-53.

George, Thomas J. and Longstaff, Francis A., 1993, Bid/ask spreads and trading activity in the S&P 100 index options market, *Journal of Financial and Quantitative Analysis* 28 (September), 381-397.

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

Harris, Lawrence, 1994, Minimum price variations, discrete bid-ask spreads, and quotation sizes, *Review of Financial Studies* 7, 149-178.

Huang, Roger D. and Stoll, Hans R., 1994, Market microstructure and stock return predictions, *Review of Financial Studies* 7, 179-213.

Kronmal, Richard A., 1993, Spurious correlation and the fallacy of the ratio standard revisited, *Journal of the Royal Statistical Society, Series A* 156, 379-392.

Lightfoot, Lois, Martin, Peter, Atkinson, William, and Davis, Jeffrey, 1986. The effects of multiple trading on the market for OTC options. SEC, Directionate of Economic and Policy Analysis.

Neal, Robert, 1987, Potential competition and actual competition in equity options, *Journal of Finance* 60, 511-531.

Smith, Tom and Whaley, Robert E., 1994, Assessing the costs of regulation: The case of dual trading, *Journal of Law and Economics* 37 (April), 215-246.

Stoll, Hans R., 1978b, The pricing of security dealer services: An empirical study of NASDAQ stocks, *Journal of Finance* 33, 1153-1172.

Stoll, Hans R., 1978a, The supply of dealer services in security markets, *Journal of Finance* 33, 1133-1151.

Tinic, Seha, 1972, The economics of liquidity services, *Quarterly Journal of Economics* 86, 79-83.

Tinic, Seha and West, Richard, 1972, Competition and the pricing of dealer services in the over-the-counter market, *Journal of Financial and Quantitative Analysis* 8, 1707-1727.

Tinic, Seha and West, Richard, 1974, Marketability of common stocks in Canada and the U.S.A.: A comparison of agent versus dealer dominated markets, *Journal of Finance* 29, 729-746.

Table 1. Model specifications used in eight empirical studies of quoted bid/ask spreads: Each specifies spread as a linear function of several variables and are nested in the following general regression equation:

$$S = a_0 + a_1 OPC + a_2 IHC + a_3 ASC + a_4 COMP + \varepsilon$$

Proxy variable definitions are: *S*=quoted bid/ask spread, *P*=share price, *NS*=number of shareholders, *TV*=trading volume, *DTV*=dollar trading volume, *NT*=number of trades, *MC*=market capitalization, *NI*=number of institutional shareholders, *%TD*=percent of trading days with at least one trade, σ_e^2=idiosyncratic risk, *B*=absolute lagged price change, σ_P=standard deviation of price, *HL*=high/low price range, σ_R=standard deviation of return, *PP*=relative purchasing power, *SS*=number of specialist stocks, *ND*=number of dealers, *NX*=number of exchanges, *CONC*=ratio of primary exchange trading volume to total volume, and *HI*=Herfindahl Index. The first panel of the table summarizes the sample, and the second panel specifies the spread measure and goodness-of-fit. The remaining panels summarize the regression specifications/results. The symbol '+' ('-') is used to signify a positive (negative) but insignificant relation, and '++' ('--') a positive (negative) and significant relation.

Market legend: N = NYSE, A = AMEX, Q = NASDAQ, T = TSE		Study											
Category	**Variable definition**	**Demsetz (1968)**		**Tinic (1972)**	**Tinic-West (1972)**	**Tinic-West (1974)**		**Benston-Hagerman (1974)**	**Branch-Freed (1977)**		**Stoll (1978b)**	**Harris (1994)**	
Market		N	N	N	O	T		O	N	A	O	N. A	
No. of observations		192	192	80	300	177		314	1,734	943	2,474	529	529
Spread estimate		*S*	*S*	*S*	*S*	*S*	*S/P*	ln *S*	*S/P*	*S/P*	ln *S/P*	*S/P*	*S/P*
Adjusted R-squared		.569	.535	.836	.490	.498	.804	.777	.486	.687	.822	.804	.987
Order-processing costs (OPC)													
	TV								--	--			
	ln *TV*			--	--	--	--				--	--	-
	ln *DTV*												
	$1/NT^{1/2}$											++	++
Inventory-holding costs (IHC)													
	ln *NS*		--					--					
	ln *NT*	--											
	NI			--									
	%TD			--		-	--						
	ln *P*							++			--		
	1/*P*								++	++			++
	P	++	++	++	++	++						++	
	ln σ_e^2							++					
	B/P								++	++			
	σ_P			+									
	HL/P				+	++	++						
	σ_R											++	++
	ln σ_R^2										++		
	PP			-									
Adverse selection costs (ASC)													
	SS			++					++	+			
	ln *DTV/MC*										++		
	ln *MC*											++	+
Competition (COMP)													
	NX	-	-			-	-		--	-			
	HI			++									
	ln *ND*							--			--		
	ln *CONC*										++	--	-
	ND				--								

Table 2. Summary of descriptive statistics of variables used in the cross-sectional regression analysis for NASDAQ stocks during the month of November 1998. The notation used in the table below is defined as follows: $EWQS$ is the equal-weighted quoted bid/ask spread, $VWES$ is the volume-weighted effective bid/ask spread, $REWQS$ and $RVWES$ are the equal-weighted quoted and volume-weighted effective bid/ask spreads divided by share price, respectively, $P(InvP)$ is the (inverse of) share price, $TV(InvTV)$ is the (inverse of the) number of shares traded, $ND(InvND)$ is the (inverse of the) number of dealers, s is the annualized close-to-close return volatility, T is the average time between trades, and HC is hedging cost as defined by $OV = P\left[2N\left(.5s\sqrt{T}\right)-1\right]$. To be included in the sample, the stock must have an average share price of at least $5 and must have traded at least five times each day during the 20 trading days in November 1998. The number of observations is 1,689.

Variable	Mean	Median	Minimum	Maximum	StdDev
Spread measures					
EWQS	0.2375	0.2110	0.0371	1.3593	0.1270
VWQS	0.1624	0.1442	0.0292	1.1167	0.0837
REWQS	0.01635	0.01479	0.00067	0.07699	0.00949
RVWQS	0.01141	0.01006	0.00057	0.06062	0.00682
Determinants of spread					
P	19.72	15.18	5.00	177.91	16.02
Inv P	0.0798	0.0659	0.0056	0.2000	0.0489
TV	349,292	113,545	441	19,592,491	973,425
Inv TV	0.0000158	0.0000088	0.0000001	0.0022676	0.0000574
ND	45	31	5	456	41
Inv ND	0.0342	0.0323	0.0022	0.2000	0.0198
s	0.6991	0.6035	0.0242	5.6677	0.4418
T	7.404	5.560	0.020	36.820	6.531
HC	0.0305	0.0258	0.0008	0.2830	0.0220

Table 3. Summary of cross-correlations between variables used in the cross-sectional regression analysis for NASDAQ stocks during the month of November 1998. The notation used in the table below is defined as follows: $EWQS$ is the equal-weighted quoted bid/ask spread, $VWES$ is the volume-weighted effective bid/ask spread, $REWQS$ and $RVWES$ are the equal-weighted quoted and volume-weighted effective bid/ask spreads divided by share price, respectively, $P(InvP)$ is the (inverse of) share price, $TV(InvTV)$ is the (inverse of the) number of shares traded, $ND(InvND)$ is the (inverse of the) number of dealers, s is the annualized close-to-close return volatility, T is the average time between trades, and HC is hedging cost as defined by $OV = P\left[2N\left(.5s\sqrt{T}\right)-1\right]$. To be included in the sample, the stock must have an average share price of at least $5 and must have traded at least five times each day during the 20 trading days in November 1998. The number of observations is 1,689.

	EWQS	***VWQS***	***REWQS***	***RVWQS***	***P***	***Inv P***	***TV***	***Inv TV***	***ND***	***Inv ND***	***s***	***T***	***HC***
EWQS	1												
VWQS	0.979	1											
REWQS	0.247	0.275	1										
RVWQS	0.192	0.244	0.988	1									
P	0.375	0.355	-0.560	-0.573	1								
Inv P	-0.420	-0.381	0.656	0.691	-0.733	1							
TV	-0.268	-0.252	-0.329	-0.307	0.307	-0.167	1						
Inv TV	0.202	0.237	0.307	0.326	-0.057	0.063	-0.081	1					
ND	-0.358	-0.346	-0.533	-0.503	0.407	-0.261	0.853	-0.142	1				
Inv ND	0.397	0.432	0.728	0.721	-0.329	0.321	-0.371	0.239	-0.649	1			
s	-0.168	-0.150	0.180	0.195	-0.210	0.304	0.144	-0.051	0.272	-0.170	1		
T	0.446	0.476	0.677	0.663	-0.256	0.245	-0.306	0.295	-0.539	0.825	-0.175	1	
HC	0.750	0.745	0.083	0.046	0.397	-0.368	-0.153	0.086	-0.165	0.189	0.241	0.270	1

Table 4. Summary of cross-sectional regression results for NASDAQ stocks during the month of November 1998. The notation is defined as follows: $EWQS_i$ is the equal-weighted quoted spread of stock i, $VWES_i$ is the volume-weighted effective spread, $RVWES_i$ is the volume-weighted effective spread divided by share price, $InvTV_i$ is the inverse of the number of shares traded, HC_i is the expected cost of hedging computed using an option valuation equation, and $InvND_i$ is the inverse of the number of dealers making a market. The value of each variable, except HC_i and $InvND_i$, is computed each trading day and then the values are averaged across days during November 1998. The value of HC_i is computed using $HC_i = P_i\left[2N\left(.5s_i\sqrt{T_i}\right)-1\right]$, where P_i is the average share price, s_i is the close-to-close return volatility during November 1998, and T_i is the average time between trades. The value of ND_i is the number of dealers that made markets at least one day during November 1998. To be included in the sample, the stock must have an average share price of at least \$5 and must have traded at least five times each day during the 20 trading days in November 1998. The number of observations is 1,689.

Model specification	R^2 / $\bar{R}^2$	Coefficient estimates / *t*-ratios							
$EWQS_i = a_0 + a_1 InvTV_i + a_2 HC_i + a_3 InvND_i + e_i$			$\hat{a}_0 / t(\hat{a}_0)$	$\hat{a}_1 / t(\hat{a}_1)$	$\hat{a}_2 / t(\hat{a}_2)$	$\hat{a}_3 / t(\hat{a}_3)$			
	0.6362		0.0578	186.3241	4.0264	1.5719			
	0.6356		8.94	2.82	13.63	11.63			
$VWES_i = a_0 + a_1 InvTV_i + a_2 HC_i + a_3 InvND_i + e_i$									
	0.6547		0.0403	163.7910	2.6035	1.1680			
	0.6540		8.46	2.96	12.80	12.65			
$RVWES_i = a_0(1/P_i) + a_1 InvTV_i / P_i + a_2 OV_i / P_i + a_3 InvND_i / P_i + e_i$									
			0.0376	136.7866	2.0959	1.0492			
			13.87	8.07	11.22	12.20			
$RVWES_i = b_0 + b_1(1/P_i) + b_2 InvTV_i / P_i + b_3 OV_i / P_i + b_4 InvND_i / P_i + e_i$		$\hat{b}_0 / t(\hat{b}_0)$	$\hat{b}_1 / t(\hat{b}_1)$	$\hat{b}_2 / t(\hat{b}_2)$	$\hat{b}_3 / t(\hat{b}_3)$	$\hat{b}_4 / t(\hat{b}_4)$			
	0.8124	0.0032	0.0101	139.4017	1.6028	1.2771			
	0.8119	14.51	3.45	9.19	10.07	15.21			
$VWES_i = g_0 + g_1 InvTV_i + g_2 InvND_i + g_3 P_i + g_4 s_i + g_5 T_i + e_i$			$\hat{g}_0 / t(\hat{g}_0)$	$\hat{g}_1 / t(\hat{g}_1)$		$\hat{g}_2 / t(\hat{g}_2)$	$\hat{g}_3 / t(\hat{g}_3)$	$\hat{g}_4 / t(\hat{g}_4)$	$\hat{g}_5 / t(\hat{g}_5)$
	0.5175		0.0078	139.1847		1.4557	0.0030	0.0174	412.7884
	0.5161		0.64	3.53		7.48	8.75	3.84	7.03