# A Simple Way to Estimate Bid-Ask Spreads from Daily High and Low Prices

Shane A. Corwin

and

Paul Schultz*

*Both authors are from the Mendoza College of Business at the University of Notre Dame. We are grateful to seminar participants at the University of Notre Dame and to Hank Bessembinder, Joel Hasbrouck, and particularly Robert Battalio, for useful comments.

# A Simple Way to Estimate Bid-Ask Spreads from Daily High and Low Prices


**Abstract**

We develop a new way to estimate bid-ask spreads from daily high and low prices. Daily high (low) prices are almost always buy (sell) orders. Hence the ratio of high-to-low prices for a day reflects both the stock's variance and its bid-ask spread. When high-low price ratios are estimated over two days, the variance is twice as large but the bid-ask spread component is unchanged. This allows us to estimate bid-ask spreads by comparing high-low price ratios over one-day and two-day intervals. We compare the high-low estimator to alternative spread estimators and to spreads estimated from intraday TAQ data. We find that the estimator is accurate and easy to use, providing a useful measure of transaction costs in a wide variety of applications.

In this paper, we derive a simple way to estimate bid-ask spreads from daily high and low prices. The estimator is based on two uncontroversial ideas. First, daily high prices are almost always buy orders and daily low prices are almost always sell orders. Hence the ratio of high-to-low prices for a day reflects both the fundamental volatility of the stock and its bid-ask spread. Second, the component of the high-to-low price ratio that is due to volatility increases proportionately with the length of the trading interval, while the component due to bid-ask spreads is constant over different trading intervals.[1] This implies that the sum of the price ranges over two consecutive single days reflects two days' volatility and twice the spread, while the price range over one two-day period reflects two days' volatility and one spread. This allows us to derive an estimate of a stock's bid-ask spread as a function of the high-to-low price ratio for a single two-day period and the high-to-low ratios for two consecutive single days.

Our spread estimator should prove useful to researchers in a wide variety of applications. Even with intraday data now widely available, researchers make frequent use of the covariance estimator of Roll (1984) or its extensions in applications ranging from asset pricing, to corporate finance, to tests of efficient markets. In some cases, this is because the researcher is studying a time period that predates intraday data (see Bessembinder and Kalcheva (2008), Bharath, Pasquariello, and Wu (2008), Gehrig and Fohlin (2006), Kim, Lin, Singh, and Yu (2007, Lesmond, Schill and Zhou (2004), or Lipson and Mortal (2007)) or international markets without intraday data (Amihud, Lauterbach, and Mendelson (2002), Chakrabarti, Huang, Jayaraman, and Lee (2005) and Griffin, Kelly, and Nardari (2007)). In other cases, the Roll (1984) measure is used with intraday data when quotes and trades cannot be reliably matched (see Antunovich and Sarkar (2006) or Fink, Fink, and Weston (2006)). Other low-frequency spread measures based on the occurrence of zero returns are pioneered in Lesmond, Ogden, and Trzcinka (1999), and are used by Bekaert, Harvey, and Lundblad (2007), Lesmond, Schill, and Zhou (2004), Mei, Scheinkman, and Xiong (2005)).[2]

The high-low estimator derived here has a number of advantages over the daily estimators used in previous research. First, we show that it is much more accurate than the still popular Roll (1984) covariance estimator. Another advantage is that it is easy to use. We provide a closed-form solution for the spread which can be easily programmed, unlike measures that require an iterative process (Hasbrouck (2006)) or maximum likelihood estimation (Lesmond, Ogden, and Trzcinka (1999)). Third, unlike Hasbrouck's (2006) Gibbs estimator or the Holden (2006) measure, the high-low estimator is not

[1] In other words, we assume the spread and variance are constant over two days, and returns are serially uncorrelated except for microstructure noise.

[2] Amihud (2002), and Pástor and Stambaugh (2003) provide low frequency measures of liquidity that attempt to capture liquidity more generally . These measures tend to be highly correlated with low frequency spread estimates but incorporate both spreads and the price impact of trades.

computer-time intensive, making it ideal for large samples. A fourth advantage of our high-low estimator is that it can provide spread estimates for short windows of time, such as days or weeks. Finally, the high-low spread estimator is derived under very general conditions. It is not ad-hoc and does not depend on institutional quirks of a particular market for its accuracy.

We test the accuracy of the high-low spread estimates by comparing them with effective spreads from TAQ for 1993 through 2005. For comparison purposes, we also estimate effective spreads from daily data using the covariance spread estimator of Roll (1984) and the effective tick estimator of Holden (2006). Because researchers tackling different problems may care about different characteristics of the spread estimator, we provide several different tests of accuracy. We first examine the performance of the various spread estimators in the pooled sample of time-series and cross-sectional observations. Here we examine both the correlation of each spread measure with the TAQ spread and the deviations of spread estimates from the TAQ spread. We provide results based on both monthly and weekly spread estimates.

The results suggest that the high-low spread estimator is very accurate and dominates the alternative spread estimators. Across all stock-months, the correlation between TAQ effective spreads and high-low spreads is 0.873. The comparable correlations for the Roll spread and the effective tick spread are 0.694 and 0.720, respectively. The high-low estimator also does well matching the level of TAQ spreads. The mean absolute difference between the spread estimates and the TAQ effective spreads is 0.97% for the high-low estimator, 1.30% for the effective tick estimator, and 1.75% for the Roll spread estimator. Similar results are obtained when spreads are estimated weekly. Across all stock-weeks, the correlation between high-low spreads and TAQ effective spreads is 0.755. The comparable correlations for the Roll spread estimator and the effective tick estimator are 0.481 and 0.586.

We next calculate cross-sectional correlations between spread estimates and TAQ effective spreads on a month-by-month basis from 1993 through 2005. Examining cross-sectional correlations serves two purposes. First, in many cases, researchers care about the ability of the spread estimator to capture the cross-sectional distribution of spreads. Second, looking at the cross-sectional correlations on a month-by-month basis allows us to examine the performance of the estimators during different time periods. The three subperiods that we examine, 1993-1996, 1997-2000, and 2001-2005, correspond closely to periods when the minimum tick size in U.S. markets was one-eighth, one sixteenth, and one penny, respectively. In all subperiods, cross-sectional correlations between high-low spreads and TAQ effective spreads are higher than the cross-sectional correlations between TAQ effective spreads and either of the other estimators. As additional evidence, we examine cross-sectional correlations in monthly spread changes. For the entire period, the high-low estimator dominates with an average cross-sectional correlation of 0.447, compared to 0.224 for the Roll spread and 0.145 for the effective tick spread. In addition, for each of the tick-size subperiods, the high-low spread estimator does a better job of capturing the level of spreads than either the Roll spread or the effective tick spread.

We next calculate stock-by-stock time-series correlations between each of the spread estimators and TAQ effective spreads. This analysis serves two purposes. First, for some applications, researchers may be particularly interested in the ability of the estimator to capture the time-series of spreads. Second, this allows us to see how well the estimators perform for different types of stocks. For all size deciles and all exchange listings, we find that high-low spreads have much higher average correlations with TAQ effective spreads than do Roll spreads. For the great majority of stocks, we also demonstrate that high-low spreads have higher correlations with TAQ effective spreads than do effective tick spreads. For the very largest stocks, the effective tick estimator has a higher correlation with TAQ effective spreads than does the high-low estimator. However, we note that the quote clustering assumptions underlying the effective tick estimator are unlikely to hold under most market settings and most time periods, and the effective tick estimator is unlikely to perform as well for markets and time-periods when the minimum tick size is not binding.

We demonstrate the practical applications of the estimator with two additional analyses. First, we use the estimator to calculate bid-ask spreads for all NYSE stocks from 1926 through 2005. Among other things, we show that effective spreads were extremely high during the depression, and increased sharply in the 1974-1975 bear market and following the 1987 crash. Second, we document the potential application of the estimator to non-U.S. markets by calculating high-low spreads for securities in India and Hong Kong using data from datastream. We show that trading costs for Indian stocks dropped dramatically with the 1994 introduction of a new exchange with automated execution and spreads in Hong Kong increased significantly during the Asian currency crisis in 1997.

While these applications are interesting, the primary goal of this paper is to introduce the high-low spread estimator and document its performance. We show that it is accurate, making it a useful measure of execution costs for periods or markets without intraday data. We show that the high-low estimator can be estimated with a relatively short time series. This makes it useful for incorporating changes in spreads into event studies and measuring levels of information asymmetry over short intervals. Finally, we show that the high-low estimator is very easy to use. The estimator is ideal for researchers who want a quick and easily calculated spread measure to test hypotheses related to corporate finance or investment strategies. Its simplicity is particularly valuable for large samples.

The remainder of the paper is organized as follows. The high-low spread estimator is derived in Section 1. Section 2 discusses issues in estimating spreads using high and low prices. Section 3 discusses existing spread estimators that use daily data, and reviews empirical tests of these estimators. Spread estimates from the high-low estimators are compared with effective spread estimates from TAQ and from the Roll and effective spread estimator in Section 4. Section 5 provides examples of uses for the high-low spread estimator. Section 6 summarizes and concludes the paper.

## 1. A New Class of Spread Estimator: The High-Low Price Estimator

To estimate spreads, we assume that the true or actual value of the stock price follows a diffusion process. We also assume that there is a spread of S%, which is constant over the two-day estimation period. Because of the spread, observed prices of buy orders are higher than the actual values by (S/2)%, while observed prices of sell orders are lower than the actual value by (S/2)%. We assume further that the high price of the day is a buy order and is therefore grossed up by half of the spread, while the low price of the day is a sell order and is discounted by one half of the spread. Hence the observed high-low price range contains both the range of the true or actual prices and the bid-ask spread. With $H_t^A$ ($L_t^A$) as the actual high (low) stock price for day *t*, and $H_t^O$ ( $L_t^O$) as the observed high (low) stock price for day *t*, we can write

$$\left[\ln(H_t^O / L_t^O)\right]^2 = \left[\ln\left(\frac{H_t^A(1+S/2)}{L_t^A(1-S/2)}\right)\right]^2. \tag{1}$$

Rearranging (1) gives

$$\left[\ln\left(H_t^O / L_t^O\right)\right]^2 = \left[\ln\left(\frac{H_t^A}{L_t^A}\right)\right]^2 + 2\left[\ln\left(\frac{H_t^A}{L_t^A}\right)\right]\left[\ln\left(\frac{2+S}{2-S}\right)\right] + \left[\ln\left(\frac{2+S}{2-S}\right)\right]^2. \tag{2}$$

Equation (2) can be simplified by noting that the natural log of the ratio of high to low prices that appears as the first term in (2) is proportional to the stock's variance. Specifically, under the assumptions that stock prices follow the usual geometric Brownian motion and the price is observed continuously, Parkinson (1980) and Garman and Klass (1980) show that a powerful and unbiased variance estimator can be defined as

$$\sigma_{HL}^2 = \frac{1}{T}k\sum_{t=1}^{T}\left[\ln\left(\frac{H_t^A}{L_t^A}\right)\right]^2. \tag{3}$$

where $H_t$ is the high price on day *t*, $L_t$ is the low price on day *t*, and k = 1/(4 ln 2). Multiplying both sides of equation (2) by k, taking expectations, and substituting from (3) yields[3]

$$E\left(k\left[\ln\left(H_t^O / L_t^O\right)\right]^2\right) = \sigma_{Daily}^2 + 2\sqrt{k}\left[\ln\left(\frac{2+S}{2-S}\right)\right]\sigma_{Daily} + k\left[\ln\left(\frac{2+S}{2-S}\right)\right]^2. \tag{4}$$

There are two unknowns in (4), S, the spread, and σ, the stock's standard deviation.

The expectation of the sum of (4) over two single days is

$$E\left(\sum_{j=0}^{1} k\left[\ln\left(H_{t+j}^O / L_{t+j}^O\right)\right]^2\right) = 2\sigma_{Daily}^2 + 4\sqrt{k}\left[\ln\left(\frac{2+S}{2-S}\right)\right]\sigma_{Daily} + 2k\left[\ln\left(\frac{2+S}{2-S}\right)\right]^2. \tag{5}$$

To simplify the notation going forward, we set

$$\alpha = \left[\ln\left(\frac{2+S}{2-S}\right)\right], \qquad \beta = \sum_{j=0}^{1}\left[\ln\left(\frac{H_{t+j}^O}{L_{t+j}^O}\right)\right]^2 \tag{6}$$

This allows us to rewrite (5) as

$$0 = 2\sigma_{Daily}^2 + 4\sqrt{k}\alpha\sigma_{Daily} + 2k\alpha^2 - k\beta. \tag{7}$$

If the variance and spread are constant over two-day periods and true returns are serially uncorrelated, the variance component of the high-low price range is twice as large for a two-day period as it is for a single day, but the spread component of the price range is unaffected by the time interval. Thus, if we use the price range over one two-day period rather than summing the price ranges over two single days, we get

$$k\left[\ln\left(H_{t,t+1}^O / L_{t,t+1}^O\right)\right]^2 = k\left[\ln\left(\frac{H_{t,t+1}^A}{L_{t,t+1}^A}\right)\right]^2 + 2k\left[\ln\left(\frac{H_{t,t+1}^A}{L_{t,t+1}^A}\right)\right]\left[\ln\left(\frac{2+S}{2-S}\right)\right] + k\left[\ln\left(\frac{2+S}{2-S}\right)\right]^2, \tag{8}$$

where $H_{t,t+1}$ is the high price over the two days t and t+1 and $L_{t,t+1}$ is the low price over the same two-day period. To further simplify the notation we set

[3] In moving from (2) and (3) to (4), we ignore Jensen's inequality and assume that $E[\ln(H/L)] = \sqrt{\sigma^2/k} = \sqrt{2.77\sigma^2}$ . This allows us to simplify the estimator considerably - and we consider simplicity one of its chief virtues. With the simplification, we solve two quadratic equations for the spread rather than two higher-order polynomials. As shown in Parkinson (1980), the correct value is $E[\ln(H/L)] = \sqrt{2.55\sigma^2}$ . This difference appears to be unimportant in practice.

$$\gamma = \left[ \ln \left( \frac{H^O_{t,t+1}}{L^O_{t,t+1}} \right) \right]^2. \tag{9}$$

Using this notation and taking expectations in (8) yields

$$0 = 2\sigma^2_{Daily} + 2\sqrt{2k}\alpha\sigma_{Daily} + k\alpha^2 - k\gamma. \tag{10}$$

This leaves two equations, (7) and (10), and two unknowns, σ and α. Solving equation (7) for σ gives

$$\sigma_{Daily} = -\sqrt{k}\alpha \pm \sqrt{\frac{k\beta}{2}}. \tag{11}$$

To ensure a positive estimate for σ, we choose the positive root,[4] or

$$\sigma_{Daily} = \sqrt{\frac{k\beta}{2}} - \sqrt{k}\alpha. \tag{12}$$

Substituting (12) into (10) and rearranging yields

$$0 = \alpha^2 + \alpha\left( \frac{2\sqrt{\beta} - 2\sqrt{2\beta}}{3 - 2\sqrt{2}} \right) + \left( \frac{\beta - \gamma}{3 - 2\sqrt{2}} \right). \tag{13}$$

The solution for α from the quadratic formula is then

$$\alpha = \frac{\sqrt{2\beta} - \sqrt{\beta}}{3 - 2\sqrt{2}} \pm \sqrt{\frac{\gamma}{3 - 2\sqrt{2}}}. \tag{14}$$

For a given β (the sum of the squared single day high-to-low ratios), greater γ (the squared two-day high-to-low ratio) implies smaller spreads. To ensure this negative relation, we use the solution for α in which the second term (involving γ) is subtracted from the first. That is,

$$\alpha = \frac{\sqrt{2\beta} - \sqrt{\beta}}{3 - 2\sqrt{2}} - \sqrt{\frac{\gamma}{3 - 2\sqrt{2}}}. \tag{15}$$

Using the definition of α in (15), we then solve (6) to obtain our spread estimator

[4] Note that equation (12) provides a spread-adjusted high-low estimator for the variance.

$$S = \frac{2\left(e^{\alpha} - 1\right)}{1 + e^{\alpha}}. \tag{16}$$

The spread estimator given in (16) is easy to compute and does not require the researcher to iterate through successive estimates of the spread to get the correct value.[5] Instead, the procedure we outline above produces an estimate of the spread and an estimate of the daily standard deviation using only the high and low prices from two consecutive days. To get spreads for longer periods like a month, we average the spread estimates from all overlapping two-day subperiods within the month.

## 2. Using the High-Low Spread Estimator in Practice

There are a number of implicit assumptions underlying the high-low estimator. One is that the stock trades continuously. Another is that stock values do not change while the market is closed. These assumptions are not true of course, raising some issues for the estimation of high-low spreads in practice.

### 2.1 Adjustment for Overnight Price Changes

Because markets are closed overnight, the ratio of high to low prices for the two-day period will reflect both the range of prices during each day and the overnight return. On the other hand, the two single-day high-low ranges reflect only the range of prices during trading hours. This causes the price range (and hence variance) estimated using one two-day period to be inflated relative to the variance estimated using two one-day periods. Hence, we need to adjust for overnight returns or the spread portion of the high-low price ratio will be underestimated.

To correct for overnight returns, we check to see if the close on day t is outside the range of prices for day t+1 for every pair of consecutive trading days. If the day t+1 low is above the day t close, we assume the price rose from the close to the low overnight and decrease both the high and low for day t+1 by the amount of the overnight change. Similarly, if the day t+1 high is below the day t close, we assume the overnight price change was from the close to the day t+1 high and increase the day t+1 high and low prices by the amount of this overnight decrease.

Alternatively, we could adjust for overnight returns using the difference between the day t close price and the day t+1 open price. There are three reasons why we do not use this adjustment. First, we want to adjust only those cases where the true value changes overnight. For many stocks, the change from close to open is more likely to occur as a result of bid-ask bounce than from an overnight return. Second, a primary use of this estimator is to estimate historic trading costs during periods when data on open prices

[5] We have also tested spread estimators derived from variance estimators in Garman and Klass (1980) that incorporate high, low, and closing prices. These estimators are more complex but failed to produce better spread estimates.

may not be available. For example, open prices are missing on CRSP from July, 1962 through June, 1992. Finally, we found a number of cases where the open price was outside the high-low price range reported by CRSP, suggesting that open price data may be unreliable.

### 2.2. True High and Low Prices are not Observed for Infrequently Traded Stocks

High and low prices are observed trade prices. Garman and Klass (1980) note that if a stock trades infrequently, the observed high price will be lower than the true high price during the day and the observed low price will be greater than the true low price for the day. In practice though, it seems likely that the probability of a trade will be especially high when prices are near their high and low values for the day. It is also unclear whether this will affect spread estimates that compare ratios of high to low prices for single days with multiple day ratios, since both estimates are subject to similar biases.

Infrequent trading is clearly a problem if a stock trades only once during a day or, more generally, if all trades occur at the same price. In this case, if the trade price is within the previous days price range, we assume the same high and low prices as the previous day. In those less common cases where the high and low are equal, but at a price outside the previous day's range, we use the same dollar range as the previous day assuming the high and low are increased or decreased by the amount the price lies outside the previous day's high-low price range. When a stock does not trade at all during a day, CRSP lists closing bid and ask prices in place of high and low prices. In these cases, we treat the bid and ask prices as if they are high and low prices.

### 2.3 High-Low Spread Estimates May Be Negative

The high-low estimator assumes that the expectation of a stock's variance over a two-day period is twice as large as the expectation of the variance over a single day. The impact of the bid-ask spread on the price range is the same regardless of the time interval. Hence, with a bid-ask spread, the expected variance over a two day period will be less than twice the single day observed variance. Even if this is true of the expectation though, the observed two-day variance may be more than twice as large as the single day variance during volatile periods or in cases with a large overnight price change. If the observed two-day variance is large enough, the high-low spread estimate will be negative.

Bayesian estimation techniques could be used to insure that two-day spread estimates are always positive.[6] For most of the analysis to follow, we incorporate the fact the spreads are never negative in a simpler way - we set all negative two-day spreads to zero before calculating monthly averages. As described in more detail below, this produces more accurate monthly spread estimates than either including

[6] Hasbrouck (2006) does this with Gibbs estimates of spreads.

or deleting negative two-day spread observations.

### 3. Other Classes of Spread Estimators that Use Daily Data

To our knowledge, this is the first use of high and low prices to estimate trading costs. Researchers have derived several other classes of spread estimators based on daily data. We describe several of these alternative estimators below.

### 3.1 Spread Estimators Derived from Return Covariances

Roll (1984) provides the first technique for estimating effective bid-ask spreads from daily data. He assumes that the value of a stock on day t, $V_t$ evolves as:

$$V_t = V_{t-1} + \varepsilon_t \tag{17}$$

where $\varepsilon_t$ is a serially uncorrelated innovation in the true value of the stock on day t. $P_t$, the observed closing price on day t, is equal to the stock's true value plus or minus half of the effective spread. That is,

$$P_t = V_t + \frac{1}{2} S Q_t \tag{18}$$

where $P_t$ is the observed closing price on day t, S is the effective spread, and $Q_t$ takes a value of +1 if the closing price is a buy order and -1 if the closing price is a sell order. Roll (1984) shows that if $Q_t$ is serially uncorrelated and uncorrelated with the innovation in the true value, the serial covariance of the change in price is

$$Cov(\Delta P_t, \Delta P_{t-1}) = -\frac{1}{4} S^2 \tag{19}$$

Solving for S yields Roll's spread estimator:

$$S = 2\sqrt{-Cov(\Delta P_t, \Delta P_{t-1})} \tag{20}$$

Roll's measure is simple, intuitive, and easy to compute. It can provide accurate spread estimates with intraday data if a researcher has trade prices but not quotes (Schultz (2000)). Even with a long time-series of daily data though, the covariance of price changes is frequently positive, forcing the researcher to arbitrarily convert an imaginary number into a spread estimate. In fact, Roll (1984) finds that cross-sectional average covariances are positive for some entire years. In these cases, researchers usually do one of three things: 1) treat the observation as missing, 2) set the Roll spread estimate to zero, or 3) multiply the covariance by negative one, estimate the spread, and multiply the spread by negative one. This last technique produces negative spread estimates, but these estimates may prove useful when average spreads are to be calculated and using only positive estimates may lead to an upward bias in the average. We provide more detail on the occurrence of negative spread estimates below.

Harris (1990) examines the small-sample properties of the Roll estimator. He demonstrates that the estimator is noisy even in relatively large samples and shows that the large number of positive autocovariance estimates is not surprising given the level of noise. He also shows that as a result of Jensen's inequality, spread estimates are significantly downward biased.

Researchers have proposed and tested a number of refinements to the Roll estimator. George, Kaul, and Nimalendran (1991) note that the Roll estimator will be downward biased if expected returns are time-varying and hence positively autocorrelated. They propose using a covariance estimator that is based on the residual of the regression of a stock's return on a measure of its expected return. They show that spreads estimated using this variation on the Roll estimator are less likely to be negative and are larger on average.

When a stock does not trade for a day, CRSP records the midpoint of its bid-ask range as its closing price. Holden (2006) observes that this results in Roll spread estimates that are too low. He proposes a revised version of the Roll estimator dividing the covariance of price changes by the percentage of days with trading.

Hasbrouck (2004, 2006) uses a Gibbs sampler and Bayesian estimation to improve the simple Roll estimator. As in Roll (1984), price changes are assumed to occur as a result of new, serially uncorrelated information, and as a result of shifts between bid and ask prices. The Gibbs estimator however, makes use of information in the series of prices to assign a posterior probability that each specific trade is a buy or sell order. The Gibbs estimation procedure involves iterations in which the distribution of all the parameters but one are taken as given. A drawing of the other variable is then made from the conditional distribution. This is repeated for each variable in succession for a number of iterations (called sweeps). As the number of sweeps grows large, the limiting distribution approaches the desired posterior.

Hasbrouck (2006) compares Gibbs estimates of annual effective spreads with TAQ estimates for 3,777 firm years over 1993-2005. The Gibbs estimator works well. The Pearson correlation between the Gibbs and TAQ effective spreads is 0.965. This is higher than the correlation of 0.853 between the high-low estimator and the TAQ effective spread that we report, but our estimates are monthly, not annual. When the estimators are compared across stocks within individual years, Hasbrouck (2006) finds that the Pearson correlation always exceeds 0.9. Hasbrouck also finds that a time-series cross-sectional regression of the TAQ effective spread estimate on the Gibbs estimate produces a slope coefficient of 0.935. Along with its advantages though, the Gibbs estimator has the disadvantage that it is, as Hasbrouck (2006) notes, "computationally intensive."

### 3.2. Spread Estimators Derived from Transaction Price Tick Size

The effective tick estimator for spreads, developed in Holden (2006) and Goyenko et al. (2009) is based on the idea that wider spreads are associated with larger effective tick sizes. For example, their model assumes that when the tick size is an eighth and the bid-ask spread is one-eighth all possible prices are used, but when the tick size is an eighth and the spread is a quarter, only prices ending on even-eighths, or quarters are used. Similarly, their model assumes that a spread of $0.50 implies that only prices ending in half or whole dollars are used. Christie and Schultz (1994) document a very strong relation between effective tick size and bid-ask spreads for Nasdaq stocks in the early 1990's, but the relation is much, much weaker for NYSE stocks.

Goyenko et al. (2009) show that their assumed relation between spreads and the effective tick size allows researchers to use price clustering to infer spreads. Suppose that there are four possible bid-ask spreads for a stock: $1/8, $1/4, $1/2 and $1. The number of quotes with odd-eighth price fractions, associated only with $1/8 spreads is given by $N_1$. The number of quotes with odd-quarter fractions, which occur with spreads of either $1/8 or $1/4, is $N_2$. The number of quotes with odd-half fractions, which can be due to spreads of $1/8, $1/4, or $1/2, is $N_3$. Finally, the number of whole-dollar quotes, which can occur with any spread width, is given by $N_4$.

To calculate an effective spread, the proportion of each price fraction observed over the estimation period is calculated as

$$F_j = \frac{N_j}{\sum_{j=1}^{J} N_j} \ for\ j = 1,\ldots,J. \tag{21}$$

The unconstrained probability $U_j$ of the $j^{th}$ spread (which corresponds to the $j^{th}$ price fraction) occurring is given by

$$U_j = \begin{array}{ll} 2F_j & j = 1 \\ 2F_j - F_{j-1} & j = 2,\ldots,J \\ F_j - F_{j-1} & j = J. \end{array} \tag{22}$$

The effective tick measure of the effective spread is a probability-weighted average of all possible spreads. However, using unconstrained probabilities can be problematic. When the number of observed prices on finer increments is high, the effective tick estimator’s unconstrained probability of a narrow spread can exceed one and the unconstrained probability of a wider spread may be negative. In the example above, if ten prices were observed and six had odd-eighth price fractions, the unconstrained probability of a one-eighth spread would be 1.2. If one of the ten prices had an odd-quarter fraction, the probability of a one-quarter spread would be .2 - .6 = -.4. Holden (2006) and Goyenko et al. (2009) constrain the probabilities of spreads estimated by the effective tick method to be non-negative and constrain the probability of an effective spread to be no more than one minus the probability of a finer spread, a practice we also adopt in the empirical work to follow.[7]

### 3.3 Spread Estimators Derived from the Frequency of Zero Returns

Lesmond, Ogden, and Trzcinka (1999) develop an effective spread estimator (the LOT estimator) based on the idea that a stock's true return is given by the market model, but that observed returns are only different from zero if true returns exceed the costs of trading. With $\alpha_1 < 0$ as the cost of selling and $\alpha_2 > 0$ as the cost of buying, the observed return on a stock on day t, $R_t^O$, is

$$
\begin{aligned}
R_t^O &= \beta R_{mt} + \varepsilon_t - \alpha_1 \quad if \;\; R_t^A < \alpha_1 \\
R_t^O &= 0 \qquad\qquad\qquad\;\; if \;\; \alpha_1 \le R_t^A \le \alpha_2 \\
R_t^O &= \beta R_{mt} + \varepsilon_t - \alpha_2 \quad if \;\; R_t^A > \alpha_2 .
\end{aligned}
\tag{23}
$$

Lesmond et al. (1999) use this relation between trading costs and observed returns to estimate trading costs. They maximize the likelihood function for a year of daily stock returns with respect to $\alpha_1$, $\alpha_2$, $\beta$ and $\sigma$. The estimate of the effective spread is then $\alpha_2 - \alpha_1$.

#### 3.4 Combination Estimators

Holden (2006) makes use of both serial correlation and price clustering to derive an effective spread estimator. For a given time series of stock prices, Holden (2006) examines all triplets of prices from three consecutive days. He then maximizes a likelihood function across all price triplets in the sample with respect to seven variables: the probability of trading on a day, the probability of a $1/8 spread, the probability of a $1/4 spread, the weighted average of all possible spreads, the mean and standard deviation of returns, and the percentage of the spread due to adverse selection and inventory holding costs. As in the effective tick estimator, price fractions on odd-eighths are attributed to $1/8 spreads, price fractions on odd-quarters are attributed to both $1/8 and $1/4 spreads, an so on. As in the Roll measure, the covariance of successive price changes is attributed to shifts between bid and ask prices. The Holden estimator nests both the Roll covariance spread estimator and the effective tick

[7] During decimal pricing, we assume the effective tick can be 1¢, 5¢, 10¢, 25¢, 50¢, or $1.00

estimator as special cases. This estimator is computationally intensive and takes even more time to run than the Gibbs estimator of Hasbrouck (2006).

### 3.5 Prior Empirical Studies of Spread Estimates from Daily Data

All of the spread estimators that use daily data are derived from strong assumptions that are unlikely to be strictly true in practice. The Roll covariance estimator assumes bid and ask quotes are equally likely, but Harris (1989) and Porter (1992) show that closing trades are more likely to appear at ask prices. The Roll estimator also assumes that trade types are serially uncorrelated, but we know that stale limit orders and price continuity rules are likely to induce positive autocorrelation in stock prices. The effective tick estimator relies on a one-to-one mapping from odd-eighth prices to odd-eight quotes. That was true for some but not all Nasdaq stocks in the 1990's, but it has never been true for New York Stock Exchange stocks and is not likely to hold under decimalization. For example, even during the 1990s, Christie and Schultz (1994) report that spreads of $0.25 that use odd-eighth quotes are almost as common for NYSE stocks as spreads of $0.25 that use even-eighth quotes. The Lesmond, Ogden, and Trzcinka (LOT) model ignores the possibility that a positive or negative return could be the result of changes in the true value that accumulated over several days. Although many of these estimators are based on problematic assumptions, the standard for evaluating the estimators is how they perform in practice. This is addressed in a number of prior studies.

Goyenko et al. (2009) compare monthly effective spread estimates from all the commonly used low-frequency spread estimators with effective spreads from intraday TAQ data for 400 stocks over 1993-2005. On average, across all stocks and all months, the effective spread is 2.9%. The Roll estimator comes closest to this with a mean value of 2.7%. The effective tick, Holden, and Gibbs estimators are considerably lower on average, ranging from 1.6% to 1.8%. The LOT measure, as calculated in the original Lesmond et al. (1999) paper is much higher at 5.6%. If the LOT measure is adjusted so that the regions used in the maximum likelihood estimation are determined by the stocks actual return rather than its predicted return, the mean effective spread is much closer, at 2.3%.

Goyenko et al. (2009) also use their sample of 400 stocks to  calculate the cross-sectional correlation of the effective spread estimates with the TAQ effective spread for each month. On average, across the 156 months from 1993-2005, the Roll spread estimator has a cross-sectional correlation with the TAQ effective spread of 0.560. Other measures are above 0.6, with the highest correlations belonging to the Gibbs estimator (0.667) and the Holden estimator (0.682). As we will see, the high-low spread estimator does much better.

## 4. A Comparison of Spread Estimates from Daily CRSP Data with Estimates from TAQ

We compare the accuracy of monthly spread estimates generated by three estimators: the Roll spread estimator, the effective tick estimator, and the high-low spread estimator.[8] To assess the accuracy of the monthly spread measures, we compare them to trade-weighted effective spreads and time-weighted quoted spreads estimated for each security each month using the NYSE's TAQ data.

For each security and each trading day, we first determine the highest bid and lowest ask across all quoting venues at every point during the day.[9] At any time $t$, let $Bid_t$ equal the inside bid, $Ask_t$ equal the inside ask, and $Midpoint_t$ equal $(Bid_t + Ask_t)/2$. The percentage quoted spread at time $t$ is then defined as $(Ask_t\text{-}Bid_t)/Midpoint_t$. For each security, the average quoted spread for the day is defined as a weighted average across all spreads during the day, where each spread is weighted by the number of seconds it is in place. The monthly *Quoted Spread* for each security is obtained by averaging the daily estimates across all trading days within the month.

To estimate effective spreads, we compare each trade price during the day to the inside bid and ask posted at the time of the trade.[10] For each trade $I$, let $Price_i$ equal trade price and $Midpoint_i$ equal the bid-ask midpoint outstanding at the time of trade $I$. The percentage effective spread for trade $I$ is then defined as $2*|P_I - Midpoint_I|/Midpoint_I$. The average effective spread for each day is a trade-weighted average across all trades during the day. The monthly *Effective Spread* for each security is then defined as the average across all trading days within the month.

The Roll, effective tick, and high-low spreads are calculated using daily data from CRSP. For each spread estimator, we require at least 12 daily observations to calculate a monthly spread estimate. To match securities in the CRSP data to securities in the TAQ data, we follow a multi-step matching procedure. We first identify all unique cusip-ticker combinations in both the TAQ and CRSP datasets

[8] We do not compare the Gibbs estimates of Hasbrouck (2006) with the TAQ spreads. Hasbrouck makes monthly estimates available to researchers, but the individual stock estimates are obtained by a linear transformation of a market-wide spread factor. The intercept and slope coefficients are estimated for individual stocks annually, not monthly.

[9] For Nasdaq securities, we first establish the best bid and ask across all Nasdaq market makers. These inside quotes are then compared to the quotes on other venues. We apply several standard filters to the trade and quote data. We require quotes to have positive prices, positive depth, and a mode of 1, 2, 6, 10, or 12. We also exclude quotes if the ask is less than or equal to the bid or if either the bid or ask differs by more than 25% from the previous quote. We utilize only trades that occur during regular trading hours, have a positive price and quantity traded, have normal condition codes, and have trade correction codes less than two. We also exclude the first trade each day and trades for which the price differs by more than 25% from the preceding price. Finally, we exclude observations for which either the effective or quoted spread exceeds \$1 with a midpoint of \$5 or less, \$5 with a midpoint of \$100 or less, or \$10 with a midpoint greater than \$100.

[10] Lee and Ready (1991) suggest that effective spreads be measured by comparing trade prices to the quotes outstanding five seconds prior to the trade. However, later studies suggest that this lag be reduced or even eliminated in more recent data. Bessembinder (2003), for example, recommends using contemporaneous quotes for determining trade direction and price improvement, but using earlier quotes for measuring effective bid-ask spreads. To be consistent across time, we estimate effective spreads throughout the sample period based on the quotes outstanding one second prior to the trade.

from 1993 through 2005. We use eight-digit cusip numbers, where cusip numbers for TAQ securities are taken from the monthly TAQ Master Files. We then merge the TAQ and CRSP samples by cusip and ticker, assigning a CRSP perm number to each TAQ security. For those securities that cannot be matched in the first step, we then match based solely on the eight-digit cusip number. Finally, we attempt to match any remaining securities by either ticker symbol or six-digit cusip number. All securities matched solely by ticker or cusip are then hand verified for accuracy and corrections are made, where possible. As a final step, we hand verify any CRSP-TAQ matches where the number of daily observations in the two datasets differs by more than 10 days. The final sample includes only NYSE, Amex, and Nasdaq listed stocks with CRSP share codes equal to 10 or 11.

### 4.1 Pooled Time-Series and Cross-Section Estimates

Panel A of Table 1 provides summary statistics for spread estimates using the pooled sample across all stocks and all months from 1993 through 2005. For comparison purposes, data on effective spreads and quoted spreads from TAQ are presented first. The simple average effective spread from TAQ across all stock months is 2.60%, while the average quoted spread from TAQ is 3.62%.

Roll spread estimates are reported next. As discussed above, positive serial correlations in stock returns are common, forcing the researcher to make ad-hoc and arbitrary adjustments to the Roll estimator. For the full sample of stocks over 1993-2005, positive monthly serial correlation estimates occur for 37.6% of the stock months. We adopt the common ad-hoc adjustment of setting negative Roll spreads to zero. This yields a mean Roll spread of 2.51%, which is very close to the mean TAQ effective spread of 2.60%. If the positive correlations are instead omitted, more than a third of the observations are lost and the mean Roll spread is 4.02%, much greater than either the mean quoted or effective spread from TAQ. In the analysis to follow, we use the version of the Roll spread estimator in which positive correlations imply zero spreads.

Spread estimates obtained from the effective tick estimator are presented next. By construction, these estimates are always positive. The mean effective tick spread is 1.94%, slightly less than either the quoted or effective spread from TAQ. The median effective tick spread is 0.85%, again less than the median TAQ effective spread of 1.46%.

Results for three versions of the high-low spread estimator are reported next. The first high-low estimator sets all negative two-day spread estimates to zero before calculating the monthly average. This high-low estimator produces a mean spread of 2.65%, as compared to the mean TAQ effective spread of 2.60%. The median spread estimate from this version of the high-low estimator is 1.47%, which is very close to the median TAQ effective spread of 1.46%. When negative spreads are included, the mean high-low spread estimate drops to 1.85%. When the negative two-day spreads are omitted, the mean high-low spread rises to 3.66%, well above the mean TAQ effective spread of 2.60%. In addition, when the negative two-day spreads are omitted, remaining observations fall below 12 in many months, resulting in the loss of 20% of the total monthly observations. Throughout the remaining analysis, we use the version of the high-low spread estimator in which negative two-day spreads are set to zero.

There are slight differences in the stock months that are examined with different estimators in Panel A. In Panel B, we compare spread estimates for the 911,719 stock months for which we have Roll spread estimates, effective tick estimates, and high-low spread estimates. Throughout the remaining analysis, we refer to this as the restricted sample. The mean TAQ effective spread in the restricted sample is 2.42%. This compares to a mean spread of 2.24% for the high-low estimator, 2.48% for the Roll estimator, and 1.74% for the effective tick estimator. The median TAQ effective spread in the restricted sample is 1.38%. This compares to 1.38% for the high-low estimator, 1.27% for the Roll estimator, and 0.79% for the effective tick estimator.

One advantage of the high-low spread estimator is that it can provide estimates over relatively short intervals. To illustrate this, Panels C and D replicate Panels A and B but use *weekly* rather than monthly spread estimates. Weeks are defined as periods of five consecutive trading days. Spreads are estimated weekly for all stocks with three or more days of data. The restricted sample is again defined as the set of  observations for which we have all three spread estimators. All three estimators produce mean weekly spread estimates that are reasonably close to the mean effective spread. However, more than 10% of the Roll spread estimates equal zero. In addition, in the restricted sample of weekly estimates, the standard deviation of high-low spread estimates is 0.0287, compared to a of 0.0440 for both of the other spread estimators.

Table 2 reports correlations between the various spread estimates and the effective and quoted spreads from TAQ. As in Table 1, observations are pooled across all stocks and all months. Panel A reports monthly results for the full sample. These results demonstrate how well the high-low estimator works. The correlation between the high-low spread estimates and the TAQ effective spread is 0.853. The comparable correlations for the Roll spread and the effective tick spread are 0.694 and 0.693, respectively. The next column reports correlations between spread estimators and the TAQ quoted spread. For both the Roll and effective tick spreads, the correlation with TAQ quoted spreads is similar to the correlation with TAQ effective spreads. At 0.907, the correlation of high-low spreads with TAQ quoted spreads is even higher than its correlation with TAQ effective spreads. Results for the restricted sample (Panel B) are similar. High-low spread estimates continue to have much higher correlations with both effective and quoted spreads than do the other estimators.

Panels C and D provide correlations based on weekly spread estimates in the full and restricted samples, respectively. Weekly spread estimates are based on far fewer observations than monthly spread estimates and thus it is not surprising that correlations based on weekly estimates are lower than correlations based on monthly estimates. Nevertheless, the high-low spread estimator continues to perform well. At 0.755, the correlation between the weekly high-low spread estimate and the TAQ effective spread in the restricted sample is considerably higher than the correlation of the TAQ effective spread with either the Roll estimator (0.481) or the effective tick estimator (0.586). In fact, the correlation between the *weekly* high-low spread estimate and the TAQ effective tick estimate is higher than the correlation between the *monthly* TAQ effective spread and either the *monthly* Roll spread estimate or the *monthly* effective tick estimate. Again, these results suggest that the high-low estimator dominates the other two spread estimators.

The correlations between the spread estimates and the TAQ spreads show how much of the variation in TAQ spreads can be explained by the various estimators, but they don't show how close the estimates come to the true level of spreads. We define the spread estimate error as the difference between the spread estimate and the TAQ effective spread for a given week or month. In Table 3, we report the mean error and mean absolute error for weekly and monthly spread estimates for each of the three estimators. Panel A reports monthly results for the full sample and Panel B reports monthly results for the restricted sample. For the restricted sample, the mean errors show that the Roll estimator produces spreads that are biased upward by six basis points, while tick spreads are too low by 68 basis points on average, and high-low spread estimates are too low by an average of 18 basis points. The mean absolute error for the high-low spread estimates is 0.0089, compared to 0.0174 for the Roll estimates and 0.0116 for the effective tick estimates. On average then, the high-low spread estimates come closest to the TAQ effective spreads. Results based on weekly spread estimates, as reported in Panels C and D, are similar. The mean absolute error is smallest for the high-low estimator, larger for the effective tick estimator, and much larger for the Roll spread estimator.

Throughout the remaining analysis, we focus on the relation between the spread estimates and the TAQ effective spreads. To ensure comparability across estimators, all remaining tests are based on the restricted sample. In addition, because the results are similar for weekly and monthly spread estimates, we report results for the monthly spread estimates only.

### 4.2 Cross-Sectional Comparisons of Spread Estimates with TAQ Effective Spreads

Next, we calculate the cross-sectional correlation between spread estimators and the TAQ effective spread each month from 1993 through 2005. This cross-sectional analysis serves two purposes.

First, in many applications, researchers may be particularly concerned with how well the estimator captures the cross-section of execution costs. Second, examining cross-sectional correlations on a month-by-month basis allows us to examined the performance of the estimator during different time periods. We calculate time-series averages of the cross-sectional correlations using the entire period and three subperiods: 1993-1996, 1997-2000, and 2001-2005. These subperiods correspond roughly to the periods when the regulatory minimum tick size and quoted spread were an eighth of a dollar, a sixteenth of a dollar, and one cent.[11]

We are particularly concerned with how the spread estimators perform during the 1993-1996 subperiod. An important use for all of the low frequency spread estimators is to estimate trading costs for periods before intraday data were available. During the 1993-1996 period, the tick size in U.S. markets was $0.125, just as it was during earlier periods. This suggest that the performance of spread estimators over 1993-1996 is a better predictor of their performance for earlier periods than is their performance during either 1997-2000 or 2001-2005.

Panel A of Table 4 reports the time-series means of the monthly correlations. For the entire period and for each subperiod, the high-low spread estimator produces higher correlations with TAQ effective spreads than either the Roll estimator or the effective tick estimator. For the entire period, the mean cross-sectional correlation of high-low spread estimates with TAQ effective spreads is 0.8259, while the Roll spread's correlation is 0.6444 and the effective tick estimates correlation is 0.6823. At 0.9121, the cross-sectional correlation of high-low spread estimates with the TAQ effective spread is particularly high during 1993-1996. This suggests that the estimator should work well for earlier periods.

Cross-sectional correlations between the spread estimators and TAQ effective spreads are shown month-by-month in Panel A of Figure 1. As the figure shows, the cross-sectional correlation between high-low spread estimates and TAQ effective spreads are consistently higher than the correlations based on either the Roll spread or the effective tick spread. The correlations based on the high-low spread estimates are particularly high during the 1993-1996 period. The Roll spread slightly outperforms the effective tick spread based on cross-sectional correlations during 1993-1996, but generally underperforms other measures in later periods.

For some applications, researchers may be interested in how well spread estimators capture changes in spreads. To address this issue, we estimate correlations between month-to-month changes in

[11] The minimum tick size on Amex changed from one-eighth to one-sixteenth during May 1997. The change occurred on both Nasdaq and the NYSE during June 1997. Both the NYSE and AMEX began to phase in decimal pricing in August 2000, with full implementation by January 2001. Nasdaq switched to decimal pricing during March and April of 2001. Hence we define the 1993-1996 period as all months from January 1993 through May 1996 and the 1997-2000 period as all months from June 1996 through December 2000. While not precise cutoffs, these breakpoints should capture the broad differences across the three tick-size regimes.

spread estimates and changes in TAQ effective spreads for each stock. Results are shown in Panel B of Table 4. Not surprisingly, correlations based on changes in spreads are lower than correlations based on spread levels. Still, the high-low spread estimator does a far better job of explaining changes in TAQ effective spreads than either the Roll spread or the effective tick spread. The mean cross-sectional correlation between changes in the high-low spread estimates and changes in the TAQ effective spread is 0.4476 for the entire period. The correlations for the Roll and effective tick estimators are 0.2241 and 0.1449, respectively.

The correlations between changes in TAQ spreads and estimated spreads are depicted month-by-month is Panel B of Figure 1. Correlations between changes in TAQ spreads and changes in estimated spreads are consistently higher with the high-low spread estimator than either the Roll estimator or the effective tick estimator. The effective tick estimator performs particularly poorly in capturing month-to-month changes in spreads.

To examine how accurately each spread estimator captures the level of spreads over time, we again examine mean errors and mean absolute errors. For each stock each month, we calculate the difference between each of the spread estimates and the TAQ effective spread. The cross-sectional mean error and mean absolute error are then calculated for each estimator each month. Panel A of Table 5 reports the time-series averages of the mean error and mean absolute error for the full sample period and for the three subperiods. For the full period, the Roll spread estimates tend to be slightly upward biased , while high-low spread estimates are downward biased and effective tick estimates are more severely downward biased. For the entire period and for each of the subperiods, mean absolute errors are smaller for high-low estimates than for either effective tick or Roll estimates. These results are generally consistent with those from the pooled sample. For the entire 1993-2005 period, the mean absolute error from the high-low estimate is 0.0089, while the mean absolute error for the effective tick estimator is 0.0114, and the mean absolute error for the Roll estimator is 0.0172. The high-low estimator does particularly well during the 1993 through 1996 subperiod, when the tick size was \$0.125. During this period, mean absolute errors were 0.0085 for the high-low estimator, 0.0142 for the effective tick estimator, and 0.0173 for the Roll estimator.

Panel B of Table 5 summarizes the time-series averages of mean errors and mean absolute errors based on monthly changes in spreads. For the entire period, the mean absolute error for spread changes averages 0.0070 for the high-low estimator, compared to 0.0219 for the Roll estimator and 0.0088 for the effective tick estimator. This again suggests that the high-low estimator does the best job of capturing changes in spreads. The only exception is during the decimal pricing period from 2001 through 2005, when the effective tick estimator has slightly lower mean absolute error for changes in spreads.

Panel A of Figure 2 plots mean absolute errors by month. Mean absolute errors are consistently highest for the Roll spread estimator. From 1993 through 2002, mean absolute errors for the Roll estimator always exceed 1.5%. For comparison, the mean effective spread over the entire sample period is 2.6%. The Roll estimator does particularly poorly in 1999 - 2002 with mean absolute errors that are often over 2.0% and sometimes over 2.5%. During 1993-1998, mean absolute errors are consistently smaller for the high-low estimator than for the effective tick estimator. Mean absolute errors are almost always less than 1.0% for the high-low estimator and typically range from 1.25% to 1.5% for the effective tick estimator. From 2000 on, mean absolute errors are similar for the effective tick and high-low estimator.

For comparability, Panel B depicts the median absolute errors by month. By this metric, the Roll estimator is consistently the worst performer. During 1993-1996, median absolute errors are always lower for the high-low spread estimator than the effective tick estimator. Beginning in 1998, however, median absolute errors for the effective tick estimator fall below those of the high-low estimator. This is particularly true during 2003 through 2005. During this period, median absolute errors for the effective tick estimator are substantially smaller than mean absolute errors. Again, it is significant that the high-low estimator produces the smallest absolute errors during the 1993-1996 period. This suggests that the high-low estimator is the one that is best suited for estimating historical U.S. trading costs.

### 4.3 The Time-Series of Spread Estimates for Individual Stocks

We next calculate stock-by-stock time-series correlations between the different spread estimates and the TAQ effective spread. These tests serve two purposes. First, they tell us how well the spread estimators work for different kinds of stocks. Second, for some applications, research may be concerned with how well the spread estimator captures the time series of spread. We summarize the time-series correlations across all stocks, by exchange, and by market capitalization decile. Decile breakpoints are based on NYSE stock capitalizations, so the smaller size deciles have a disproportionate number of stocks from Nasdaq and the Amex. The results provided in the table are based on the exchange and size deciles of each stock as of its last listing date on CRSP.

The stock-by-stock time-series correlations are summarized in Table 6. Panel A reports results for the entire 1993-2005 period. One clear result that emerges from Table 6 is that when it comes to explaining time-series variation in the spreads of individual stocks, the Roll estimator is dominated by both the effective tick estimator and the high-low estimators. The Roll spread estimates have a lower correlation with TAQ effective spreads than the other estimates for all stocks, for stocks on each of the exchanges, and for stocks in all size deciles.

A second clear result is that the high-low spread estimator outperforms the effective spread estimator on average because it does a better job with the spreads of smaller stocks. For both Nasdaq and Amex stocks, the high-low estimator has a higher correlation with TAQ effective spreads than does the effective tick estimator. However, the effective tick estimator has a higher correlation for NYSE stocks. Turning to size deciles, we see that the high-low estimator exhibits higher time-series correlations with TAQ effective spreads than the effective tick estimator for the two smallest deciles. These two deciles contain more than two-thirds of the sample stocks (8,559 out of 12,192). For larger size deciles though, the effective tick estimator produces a higher correlation with TAQ effective spreads. This is especially true for the largest decile, where the mean time-series correlation between the effective tick spread and the TAQ effective spread is 0.8036, compared to a correlation of only 0.1100 for the high-low spread. The high-low spread estimator has trouble with the largest stocks in part because the signal-to-noise ratio is so small for these stocks. Their trading costs are low. The effective tick estimator works well with the largest stocks because this estimator is similar to simply dividing the tick size by the stock price. This works well when the tick size places a binding lower bound on the spread width.

Panels B through D of Table 6 report results by subperiod. One notable result here is that the high-low estimator performs much better within subperiods than it does across the full sample period. This suggests that shifts in the tick size may drive the low correlation between high-low spread and TAQ spreads for large stocks in Panel A. During the 1993-1996 subperiod (Panel B), the effective tick and high-low spread estimators again dominate the Roll estimator. The high-low estimator also clearly outperform the effective tick estimator for the great majority of stocks. Across all stocks, the mean time-series correlation between the high-low spread and the TAQ effective spread is 0.6243. This compares to correlations of 0.4608 for the effective tick spread and only 0.3258 for the Roll spread. The mean time-series correlation based on the high-low estimator is higher than the mean correlation based on the effective tick estimator for the first five size deciles and the two are roughly equivalent in deciles six and seven. Thus, for the vast majority of stocks (92.7%), the high-low estimator produces time-series correlations that are lower than or roughly equal to those for the effective tick estimator.

Panels C and D of Table 6 report results for the 1997-2000 and 2001-2005 subperiods. For both subperiods, the Roll estimator is again dominated by both the high-low estimator and the effective tick estimator. The high-low estimator has the highest correlation with the TAQ effective spread for small stocks, while the effective tick estimator generally performs best for larger stocks. It is interesting though that in the 2001-2005 period, the superior performance of the effective tick estimator for large stocks is reduced or even reversed. During this time, the high-low spread estimator appears to work well for both small and large stocks.

Next, we examine how closely our spread estimator captures the level of TAQ effective spreads for different types of stocks. As in earlier tests, we first define the error for each spread estimator as the difference between the spread estimate and the TAQ effective spread in a given stock-month. We then calculate the mean time-series error and absolute error for each stock. Table 7 summarizes these time-series errors across all stocks, by exchange, and by size decile. To save space, we report only the full sample results. Across all stocks, the mean absolute error is 0.0103 for the high-low spread, compared to 0.0201 for the Roll spread, and 0.0145 for the effective tick spread. By this measure, the Roll estimator is again dominated by both the effective tick spread and the high-low spread for all types of stocks. For Nasdaq and Amex stocks, mean absolute errors are smaller for the high-low estimator than for the effective tick estimator. For NYSE stocks, mean absolute errors are somewhat higher for the high-low estimator than for the effective tick estimator, but the difference is relatively small. When results are compared by size decile, the high-low estimator again produces more accurate or equivalent spread estimates for the large number of stocks in deciles one through three, while the effective tick estimator produces lower mean absolute errors for larger stocks.

**4.4 Summary of the Estimator Comparisons**

In pooled and cross-sectional analyses, the high-low estimator dominates both the Roll spread estimator and the effective tick estimator. It has higher correlations with TAQ effective spreads, smaller deviations from TAQ effective spreads, and higher correlations with month-to-month changes in TAQ effective spreads than either of the other estimators. The high-low spread estimator does particularly well during the period from 1993 through 1996, when the minimum tick size was $0.125. Because this is the same tick size as in prior years, these results suggest that the high-low spread estimator may be superior to other estimators for historical analyses. In stock-by-stock time-series analyses, we again find that the high-low estimator dominates the Roll spread estimator. The high-low estimator is also superior to the effective tick estimator for the small stocks that are the most costly to trade. For the vast majority of stocks, the high-low estimator produces spread estimates that are closer to and more highly correlated with TAQ effective spreads than the effective tick estimator. For the very largest stocks, however, the effective tick estimator produces higher correlations with TAQ effective spreads. Thus, the effective tick estimator appears to work well when the tick size provides a binding lower bound on the spread.

Neither the Roll estimator nor the high-low estimator depend on institutional or regulatory features of specific markets. The effective tick estimator, on the other hand, is based on the assumption that there is a one-to-one correspondence between price fractions and spread widths. That was true on Nasdaq during the 1990's (see, for example, Christie and Schultz (1994)). It is also true trivially for stocks with spreads equal to one tick. For most markets and most time periods though, the correlation between spreads and price fractions is weak. This may limit the usefulness of the effective tick estimator.

## 5. Examples of Applications for the High-Low Spread Estimator

In most cases, intraday data is preferable for calculating trading costs. However, this data is only available for recent years and selected markets. For example, TAQ data is only available from 1993 on, while the harder to use ISSM data is unavailable for NYSE or Amex stocks before 1983 and for Nasdaq stocks before 1987. Researchers who wish to examine the impact of trading costs on asset pricing need a much longer time-series of data. Likewise, researchers who examine the profitability of trading strategies or the robustness of anomalies may need trading cost estimates for years or markets for which intraday data do not exist. In these cases, the high-low spread estimator should prove valuable to researchers.

To demonstrate the applicability of the high-low spread estimator, we provide two illustrative analyses. The first is an analysis of historical spreads on NYSE stocks from 1926 to the present. The second is an analysis of historical spreads on two non-U.S. markets based on daily data from Datastream.

### 5.1 Estimating Historical Spreads for U.S. Stocks using Daily CRSP Data

Using high and low price data from CRSP, we calculate bid-ask spreads for each NYSE stock each month from 1926 to 2005. As in the previous analyses, monthly spreads are defined as the average of all two-day spreads within the calendar month, negative two-day spread estimates are set to zero, and we require a minimum of 12 daily price ranges to calculate a monthly spread. The results are illustrated in Figure 3.

Panel A of Figure 3 plots the cross-sectional average of high-low spread estimates for NYSE stocks each month from 1926 through 2005. Results are shown for the full sample of NYSE stocks and for the smallest and largest market capitalization deciles. Examining the market-wide average, we see that spreads display considerable variation over time. They were very high in the early years of the depression, with mean spreads exceeding 10% for several months in 1932 and 1933. Spreads declined in 1935 and 1936 but increased sharply as the market performed poorly in 1937 and 1938. Spreads declined steadily until the early 1950's and remained relatively low through the early 1970's. The recession of 1974-1975 is clearly visible in the figure as a period of increased spreads. Spreads are also relatively high in the early 1990's and during the tech bubble of the late 1990's. As expected, the results show that small stocks tend to have higher execution costs than large stocks. However, the graph also illustrates that the difference between these groups is highly variable. For most months, spreads are 1% to 2% higher for small stocks than large stocks. During the depression, on the other hand, small stock spreads sometimes exceeded large stock spreads by 50%. So, at the time that spreads were 8% or 9% for large stocks, they were around 60% for small stocks. This shows that trading strategies involving small stocks were extremely expensive during the depression. It also indicates that if the returns to small stocks contain a premium to compensate for trading costs, that premium would have been especially high in the 1930's.

Panel B of Figure 3 provides a similar graph for the 1950-2005 subperiod. By omitting the depression and altering the scale of the graph, we get a clearer picture of the intertemporal variation in spreads over the last 50 years. Here, the impact of recessions and stock market declines in 1974-1975 and 1991-1992, the 1987 crash, and the “technology bubble” are clearly visible. The difference between spreads of small and large stocks was relatively large in the mid-1970's and also in the early 1990's. However, in recent years, the difference in spreads between small and large stocks has shrunk to almost nothing. Thus, while trading strategies involving small stocks may have been prohibitively expensive during the mid-1970s and early-1990s, these trading strategies may be more profitable today.

The point of this exercise is to illustrate how the high-low estimator can be used in practice. However, there is also a lesson in the analysis: trading costs prior to the early 1940s are too large to be ignored. The high-low estimator allows researchers who are studying this period to incorporate bid-ask spreads.

### 5.2 Estimating Historical Spreads for International Stocks Using Datastream Data

To demonstrate the applicability of the high-low estimator to non-U.S. markets, we estimate high-low spreads for individual stocks in Hong Kong and India using daily high and low prices from Datastream.[12] As discussed below, each of these markets provides a specific event around which we expect execution costs to change. Results for additional countries covered by Datastream are available from the authors upon request. Again, we include only those stock-months with at least 12 daily spread observations and we set all negative estimates to zero before taking the monthly average.

Hong Kong was significantly affected by the Asian Currency Crisis beginning in October 1997, when its currency came under pressure. During this period, the equity market in Hong Kong became more volatile, with the Hang Sang index falling 23% between October 20 and 23, 1997. We expect a significant increase in execution costs in the Hong Kong market during this period.

The cross-sectional average of high-low spread estimates for stocks in Hong Kong is plotted by month in Panel A of Figure 4. Because data coverage in Datastream increases over time, the graph also plots the number of firms used to compute the market-wide average in each month. As expected, average bid-ask spreads in Hong Kong increased sharply starting in October 1997. Average spreads increased

[12] Reuters provides intraday data for many international markets starting in 1996, but earlier intraday data is limited. Lesmond (2005) studies the ability of the Roll (1984), Amihud (2002), and Lesmond, Ogden, and Trzcinka (1999) measures to explain differences in bid-ask spreads within and across emerging markets.

from approximately 0.75% prior to 1997 to over 1.5% in late 1997, peaking at 2.3% in February 2000. This shift in spreads coincides with the Asian Currency Crisis and related turmoil in Hong Kong's equity markets in 1997-1998.

As of 1994, the Bombay Stock Exchange (BSE) was India's dominant market, accounting for 75% of equity volume. In November 1994, the National Stock Exchange (NSE) opened, providing Indian investors with an order-driven electronic limit order book, reduced tick sizes, satellite technology with links to sites all over India, and improved settlement and clearing standards (see Shaw and Thomas (2000)). By October 1995, NSE had surpassed the BSE, becoming the dominant equities market in India. We expect execution costs to decrease with the introduction of this new market structure.

Monthly high-low spread estimates for India are plotted in Panel B of Figure 4. Again, the graph shows the cross-sectional average across all stocks with available data in a given month, along with the number of firms used to compute the market-wide average each month. As we predicted, the average bid-ask spread across stocks in India decreased sharply in early 1995. Bid-ask spreads dropped from an average of approximately 4.5% in early 1994 to approximately 1.5% in early 1995. Spreads remain low after the introduction of the NSE, ranging from one to two percent from 1995 through 2006. This shift in spreads is consistent with the hypothesis that the change in market structure brought about by the introduction of the NSE led to a significant and permanent decrease in execution costs in India.

## 6. Summary and Conclusions

In this paper, we derive a new technique for estimating bid-ask spreads from high and low prices. The estimator is intuitive and easy to calculate. It is derived under very general conditions and does not rely on the characteristics of any particular market. We provide a closed-form solution for the spread, so it is easy to program and requires little computation time. The high-low estimator can be used with daily high and low prices when intraday trade and quote data are unavailable. It can also be used to estimate spreads from intraday trades when quotes are unavailable or are difficult to match with trades. It is ideal for researchers who need a simple but accurate measure of trading costs for work in corporate finance, asset pricing, or as part of a study of market efficiency.

We examine the performance of the high-low estimator by comparing effective spreads from TAQ with spread estimates from the high-low estimator, the Roll (1984) covariance estimator, and the effective tick estimator of Goyenko et al. (2009) and Holden (2006). The high-low estimator dominates, with higher time-series, cross-sectional, and pooled correlations with TAQ effective spreads. It also produces lower mean absolute differences between estimated spreads and TAQ spreads than either of the other estimators. It works particularly well with small stocks.

Undoubtedly, the performance of the high-low estimator can be improved. For example, we calculate the monthly spread by taking a simple average of all two-day spread estimates within the month. A different weighting of the observations or including estimates for periods other than two days might improve the monthly estimates. Bayesian techniques are also likely to improve the estimates when simpler techniques yield negative spreads. Finally, high-low spread estimates could be combined with the Roll estimator or other spread estimates as in Goyenko et al. (2009). Refinements like this would, however, make the estimation more complex - and simplicity is part of its appeal.

To illustrate the potential applications of the high-low estimator, we examine historical trading cost estimates for U.S. and international stocks. There are many other potential applications. The high-low estimator can produce spread estimates with relatively small amounts of data. This makes it ideal for analyzing changes in spreads around events. For example, we have used the high-low spread estimator to examine daily bid-ask spreads around stock splits over 1926-1982. Like studies that use recent intraday data, we find a sharp increase in spreads the day of the split. The estimator also allows researchers to estimate intraday trading costs without using quotes, which may become more important as quote data becomes more unwieldy. TAQ quote files are already challenging to use, having recently grown to more than 10 times the size of the trade files. We have used the high-low spread estimator and intraday trade data from TAQ to estimate spreads for fifteen minute periods. Our results suggest that the estimator performs well in this setting. Results from these additional analyses are available from the authors.

The high-low spread estimator can also be used to calculate trading costs for assets other than common stock. For example, trading costs could be estimated for futures markets from trade and sales data, which consist of trades only. Futures trades must be reported within the appropriate fifteen-minute period, but may be reported out of order. The high-low estimator is especially well-suited for these data.

The most important direction for further research, may not be with spread estimation at all. In deriving our spread estimator, we jointly derive an estimate of the spread and an estimate of the variance of a stock's true value - that is, the variance without microstructure noise. Bid-ask spreads can induce a significant upward bias in variance estimates for small stocks or even large stocks during periods with high trading costs. Hence a variance measure that is free from bid-ask bounce may prove very useful.[13] We leave a more detailed analysis of this high-low variance estimator to future work.

[13] Bandi and Russell (2006) use high-frequency data to separate the true variance from microstructure noise for S&P 100 stocks.

## References


Amihu, Yakov, 2002, Illiquidity and stock returns: Cross-section and time-series effects, *Journal of Financial Markets* 5, 31-56.

Amihud, Yakov, Beni Lauterbach, and Haim Mendelson, 2003, The Value of Trading Consolidation: Evidence from the Exercise of Warrants, *Journal of Financial and Quantitative Analysis* 38, 829-846.

Angel, James, 1997, Tick size, share prices, and stock splits, *Journal of Finance* 52, 655-681.

Antunovich, Peter, and Asani Sarkar, 2006, Fifteen Minutes of Fame? The Market Impact of Internet Stock Picks, *Journal of Business* 79, 3209-3251.

Bandi, Federico, and Jeffrey Russell, 2006, Separating microstructure noise from volatility, *Journal of Financial Economics* 79, 655-692.

Bekaert, Geert, Campbell Harvey, and Christian Lundblad, 2007, Liquidity and Expected Returns: Lessons from Emerging Markets, *Review of Financial Studies* 20, 1783-1831.

Bessembinder, Hendrik, and Ivalina Kalcheva, 2008, Liquidity Biases in Asset Pricing Tests, Working paper, University of Utah.

Bharath, Sreedhar, Paolo Pasquariello, and Guojun Wu, 2008, Does Asymmetric Information Drive Capital Structure Decisions?, Forthcoming, *Review of Financial Studies*.

Chakrabarti, Rajesh, Wei Huang, Narayanan Jayaraman, and Jinsoo Lee, 2005, Price and Volume Effects of Changes in MSCI Indices - Nature and Causes, *Journal of Banking and Finance* 29, 1237-1264.

Christie, William, and Paul Schultz, 1994, Why do Nasdaq market makers avoid odd-eighth quotes?, *Journal of Finance* 49, 1813-1840.

Chordia, Tarun, Sahn-Wook Huh, and Avanidhar Subrahmanyam, 2007, The cross-section of expected trading activity, Forthcoming, *Review of Financial Studies*.

Chordia, Tarun, Richard Roll, and Avanidhar Subrahmanyam, 2001, Market liquidity and trading activity, *Journal of Finance* 56, 501-530.

Conroy, Robert, Robert Harris, and Bruce Benet, 1990, The effects of stock splits on bid-ask spreads, *Journal of Finance* 45, 1285-1295.

Cremers, K. J. Martijn, and Jianping Mei, 2007, Turning over turnover, *Review of Financial Studies* 20, 1749-1782.

Fink, Jason, Kristin Fink, and James Weston, 2006, Competition on the Nasdaq and the Growth of Electronic Communication Networks, *Journal of Banking and Finance* 30, 2537-2559.

Garman, M. B., and M. J. Klass, 1980, On the estimation of security price volatilities from historical data, *Journal of Business* 53, 67-78.

Gehrig, Thomas, and Caroline Fohlin, 2006, Trading Costs in Early Securities Markets: The Case of the Berlin Stock Exchange 1880-1910, *Review of Finance* 10, 587-612.

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

Goyenko, Ruslan, Craig Holden, and Charles Trzcinka, 2009, Do liquidity measures measure liquidity, forthcoming, *Journal of Financial Economics*.

Griffin, John, Patrick Kelly, and Federico Nardari, 2007, Measuring Short-term International Stock Market Efficiency, Working paper, University of Texas.

Griffin, John, Federico Nardari, and René Stulz, 2007, Do investors trade more when stocks have performed well? Evidence from 46 countries, Forthcoming, *Review of Financial Studies*.

Harris, Lawrence, 1989,A day-end transaction price anomaly, *Journal of Financial and Quantitative Analysis* 24, 29-45.

Harris, Lawrence, 1990, Statistical properties of the Roll serial covariance bid/ask spread estimator, *Journal of Finance* 45, 579-590.

Hasbrouck, Joel, 2004, Liquidity in the futures pits: Inferring market dynamics from incomplete data, *Journal of Financial and Quantitative Analysis* 39, 305-326.

Hasbrouck, Joel, 2006, Trading costs and returns for U.S. equities: Estimating effective costs from daily data. Working paper, New York University, New York, NY.

Holden, Craig, 2006, New low-frequency liquidity measures. Unpublished working paper, Indiana University, Bloomington, IN.

Kim, Joonghyuk, Ji-Chai Lin, Ajai Singh, and Wen Yu, 2007, Dual-Class Splits and Stock Liquidity, Working paper, Case Western University.

Lesmond, David, 2005, Liquidity of emerging markets, *Journal of Financial Economics* 77, 411-452.

Lesmond, David, Joseph Ogden, and Charles Trzcinka, 1999, A new estimate of transactions costs, *Review of Financial Studies* 12, 1113-1141.

Lesmond, David, Michael Schill, and Chunsheng Zhou, 2004, The Illusory Nature of Momentum Profits, *Journal of Financial Economics* 71, 349-380.

Lipson, Marc, and Sandra Mortal, 2007, Capital Structure Decisions and Equity market Liquidity, Working paper, University of Virginia.

Mei, Jianping, José Scheinkman, and Wei Xiong, 2005, Speculative Trading and Stock Prices: Evidence from Chinese A-B Share Premia, Working paper, Princeton University.

Parkinson, M., 1980, The extreme value method for estimating the variance of the rate of return, *Journal of Business* 53, 61-65.

Pástor, Luboš, and Robert Stambaugh, 2003, Liquidity risk and expected stock returns, *Journal of Political Economy* 111, 642-685.

Porter, David, 1992, The probability of a trade at the ask: An examination of interday and intraday behavior, *Journal of Financial and Quantitative Analysis27*, 209-227.

Roll, Richard, 1984, A simple implicit measure of the effective bid-ask spread in an efficient market, *Journal of Finance* 39, 127-1139.

Schultz, Paul, 2000, Regulatory and legal pressures and the costs of NASDAQ trading, *Review of Financial Studies* 13, 917-957.

Schultz, Paul, 2000, Stock splits, tick size, and sponsorship, *Journal of Finance* 55, 429-450.

Shaw, A., and S. Thomas, 2000, David and Goliath: Displacing a primary market, *Global Financial Markets*, Spring 2000, 14-23.

## Table 1 - Summary Statistics for Spreads based on Alternative Estimation Methods

The table provides summary statistics for spread estimates based on the pooled sample of monthly time-series and cross sectional observations. The full sample includes all NYSE, Amex, and Nasdaq listed securities for which TAQ and CRSP data could be matched. In the restricted sample, observations are dropped if there are fewer than 12 monthly observations for the firm or if spread estimates are missing for the *Roll Spread*, *Tick Spread,* or *HL Spread*. *Effective Spread* is the trade-weighted percentage effective spread estimated from TAQ and averaged across days within the month. *Quoted Spread* is the time-weighted percentage quoted spread estimated from TAQ and averaged across days within the month. The *Roll Spread* is two times the square root of the -1 x the autocovariance of daily returns. The *Effective Tick Spread* assumes that the spread is equal to the tick increment used in trade prices. The effective tick spread for the month is based on the average of the daily spreads implied by daily trade prices. *HL Spread* is the equally-weighted average of the high-low spread estimator across all overlapping two-day periods within the month. The table lists results for the *HL Spread* using three alternative methods to account for negative spread estimates within the month: (1) set negative two-day spreads to zero, (2) leave negative two-day spreads unchanged, and (3) exclude negative two-day spreads. Similarly, results for the Roll Spread are provided using two alternative methods to handle negative covariances: (1) setting spreads to zero when the covariance is negative and (2) exclude spreads when the covariance is negative.

| | N | Mean | Median | $10^{\%}$ | $90^{\%}$ | Std. Dev. |
|---|---|---|---|---|---|---|
| Panel A – Monthly Full Sample | | | | | | |
| Effective Spread | 960,457 | 0.0260 | 0.0146 | 0.0019 | 0.0617 | 0.0366 |
| Quoted Spread | 974,882 | 0.0362 | 0.0194 | 0.0025 | 0.0863 | 0.0527 |
| Roll Spread$_{\text{Neg=0}}$ | 919,083 | 0.0251 | 0.0128 | 0.0000 | 0.0655 | 0.0399 |
| Roll Spread$_{\text{Neg Dropped}}$ | 573,320 | 0.0402 | 0.0276 | 0.0084 | 0.0841 | 0.0441 |
| Eff. Tick Spread | 971,640 | 0.0194 | 0.0085 | 0.0006 | 0.0439 | 0.0416 |
| HL Spread$_{\text{Neg=0}}$ | 964,479 | 0.0265 | 0.0147 | 0.0044 | 0.0561 | 0.0424 |
| HL Spread$_{\text{NegIncluded}}$ | 964,479 | 0.0185 | 0.0068 | -0.0041 | 0.0489 | 0.0480 |
| HL Spread$_{\text{NegDropped}}$ | 771,828 | 0.0366 | 0.0234 | 0.0081 | 0.0739 | 0.0480 |
| Panel B – Monthly Restricted Sample | | | | | | |
| Effective Spread | 911,719 | 0.0242 | 0.0138 | 0.0019 | 0.0577 | 0.0326 |
| Quoted Spread | 911,719 | 0.0316 | 0.0178 | 0.0023 | 0.0760 | 0.0422 |
| Roll Spread$_{\text{Neg=0}}$ | 911,719 | 0.0248 | 0.0127 | 0.0000 | 0.0651 | 0.0389 |
| Roll Spread$_{\text{Neg Dropped}}$ | 567,965 | 0.0398 | 0.0275 | 0.0084 | 0.0835 | 0.0428 |
| Eff. Tick Spread | 911,719 | 0.0174 | 0.0079 | 0.0006 | 0.0404 | 0.0336 |
| HL Spread$_{\text{Neg=0}}$ | 911,719 | 0.0224 | 0.0138 | 0.0043 | 0.0476 | 0.0294 |
| HL Spread$_{\text{NegIncluded}}$ | 911,719 | 0.0140 | 0.0060 | -0.0044 | 0.0396 | 0.0294 |
| HL Spread$_{\text{NegDropped}}$ | 720,597 | 0.0318 | 0.0222 | 0.0079 | 0.0640 | 0.0347 |
| Panel C – Weekly Full Sample | | | | | | |
| Effective Spread | 3,836,911 | 0.0240 | 0.0129 | 0.0018 | 0.0572 | 0.0346 |
| Quoted Spread | 4,074,045 | 0.0360 | 0.0188 | 0.0024 | 0.0864 | 0.0535 |
| Roll Spread | 3,724,068 | 0.0249 | 0.0076 | 0.0000 | 0.0690 | 0.0459 |
| Eff. Tick Spread | 3,863,551 | 0.0210 | 0.0083 | 0.0006 | 0.0462 | 0.0495 |
| HL Spread$_{\text{Neg=0}}$ | 4,060,834 | 0.0266 | 0.0134 | 0.0027 | 0.0590 | 0.0464 |
| Panel D – Weekly Restricted Sample | | | | | | |
| Effective Spread | 3,632,287 | 0.0224 | 0.0122 | 0.0017 | 0.0534 | 0.0314 |
| Quoted Spread | 3,632,287 | 0.0289 | 0.0159 | 0.0022 | 0.0695 | 0.0393 |
| Roll Spread | 3,632,287 | 0.0242 | 0.0072 | 0.0000 | 0.0676 | 0.0440 |
| Eff. Tick Spread | 3,632,287 | 0.0192 | 0.0078 | 0.0006 | 0.0430 | 0.0440 |
| HL Spread$_{\text{Neg=0}}$ | 3,632,287 | 0.0202 | 0.0118 | 0.0025 | 0.0453 | 0.0287 |

## Table 2 - Pooled Correlations

The table lists correlations among the spread estimates based on the pooled sample of monthly time-series and cross-sectional observations. The full sample includes all NYSE, Amex, and Nasdaq listed securities for which TAQ and CRSP data could be matched. In the restricted sample, observations are dropped if there are fewer than six monthly observations for the firm or if spread estimates are missing for the *Roll Spread*, *Tick Spread,* or *HL Spread*.

| | Effective Spread | Quoted Spread | Roll Spread | Effective Tick Spread | High-Low Spread |
|---|---|---|---|---|---|
| Panel A – Monthly Full Sample | | | | | |
| Effective Spread | 1.000 | | | | |
| Quoted Spread | 0.949 | 1.000 | | | |
| Roll Spread | 0.694 | 0.696 | 1.000 | | |
| Tick Spread | 0.693 | 0.699 | 0.525 | 1.000 | |
| HL Spread | 0.853 | 0.907 | 0.681 | 0.675 | 1.000 |
| Panel B – Monthly Restricted Sample | | | | | |
| Effective Spread | 1.000 | | | | |
| Quoted Spread | 0.966 | 1.000 | | | |
| Roll Spread | 0.694 | 0.690 | 1.000 | | |
| Tick Spread | 0.720 | 0.718 | 0.517 | 1.000 | |
| HL Spread | 0.873 | 0.887 | 0.674 | 0.700 | 1.000 |
| Panel C – Weekly Full Sample | | | | | |
| Effective Spread | 1.000 | | | | |
| Quoted Spread | 0.934 | 1.000 | | | |
| Roll Spread | 0.481 | 0.500 | 1.000 | | |
| Tick Spread | 0.570 | 0.591 | 0.291 | 1.000 | |
| HL Spread | 0.748 | 0.845 | 0.505 | 0.538 | 1.000 |
| Panel D – Weekly Restricted Sample | | | | | |
| Effective Spread | 1.000 | | | | |
| Quoted Spread | 0.948 | 1.000 | | | |
| Roll Spread | 0.481 | 0.484 | 1.000 | | |
| Tick Spread | 0.586 | 0.590 | 0.273 | 1 | |
| HL Spread | 0.755 | 0.762 | 0.494 | 0.518 | 1 |

**Table 3 - Pooled Mean Errors and Mean Absolute Errors**

For each stock-month, errors are defined for each spread measure as the difference between the spread measure and the TAQ effective spread. The table lists the mean error and mean absolute error across all pooled cross-sectional and time-series observations. The full sample includes all NYSE, Amex, and Nasdaq listed securities for which TAQ and CRSP data could be matched. In the restricted sample, observations are dropped if there are fewer than six monthly observations for the firm or if spread estimates are missing for the *Roll Spread*, *Tick Spread,* or *HL Spread*.

| | N | Mean Error | Mean Absolute Error |
|---|---|---|---|
| | Panel A – Monthly Full Sample | | |
| Roll Spread | 915,913 | 0.0006 | 0.0175 |
| Effective Tick Spread | 957,365 | -0.0075 | 0.0130 |
| High Low Spread | 949,929 | -0.0012 | 0.0097 |
| | Panel B – Monthly Restricted Sample | | |
| Roll Spread | 911,719 | 0.0006 | 0.0174 |
| Effective Tick Spread | 911,719 | -0.0068 | 0.0116 |
| High-Low Spread | 911,719 | -0.0018 | 0.0089 |
| | Panel C – Weekly Full Sample | | |
| Roll Spread | 3,635,909 | 0.0019 | 0.0233 |
| Effective Tick Spread | 3,744,068 | -0.0035 | 0.0140 |
| High-Low Spread | 3,820,573 | -0.0021 | 0.0122 |
| | Panel D – Weekly Restricted Sample | | |
| Roll Spread | 3,632,287 | 0.0019 | 0.0233 |
| Effective Tick Spread | 3,632,287 | -0.0031 | 0.0132 |
| High-Low Spread | 3,632,287 | -0.0021 | 0.0115 |

## Table 4 - Average Cross-Sectional Correlations

For each spread measure and each month, we estimate the cross-sectional correlation between the spread measure and the effective spread from TAQ. The table lists the average cross-sectional correlation across all months. The full sample includes all NYSE, Amex, and Nasdaq listed securities for which TAQ and CRSP data could be matched. Observations are then dropped if there are fewer than six monthly observations for the firm or if spread estimates are missing for the *Roll Spread*, *Tick Spread,* or *HL Spread*. Panel A lists results based on monthly spreads and Panel B lists results based on first differences in monthly spreads.

| | N | Roll Spread | Effective Tick Spread | High-Low Spread |
|---|---|---|---|---|
| Panel A - Correlations with Effective Spread, Monthly Estimates | | | | |
| Full Period | 156 | 0.6444 | 0.6823 | 0.8259 |
| | | | | |
| 1993-1996 | 53 | 0.7649 | 0.7225 | 0.9121 |
| 1997-2000 | 43 | 0.6312 | 0.7199 | 0.8106 |
| 2001-2005 | 60 | 0.5473 | 0.6200 | 0.7607 |
| Panel B – Correlations with Changes in Effective Spreads, Monthly Estimates | | | | |
| Full Period | 155 | 0.2241 | 0.1449 | 0.4476 |
| | | | | |
| 1993-1996 | 52 | 0.2518 | 0.1360 | 0.4797 |
| 1997-2000 | 43 | 0.2354 | 0.1530 | 0.4456 |
| 2001-2005 | 60 | 0.1918 | 0.1468 | 0.4214 |

## Table 5 - Cross-Sectional Mean Absolute Errors

For each stock-month, errors are defined for each spread measure as the absolute value of the difference between the spread measure and the TAQ effective spread. For each month, we estimate the mean error and mean absolute error across all cross sectional observations. The table then lists the mean across months. For reporting purposes, mean errors are multiplied by 100. The full sample includes all NYSE, Amex, and Nasdaq listed securities for which TAQ and CRSP data could be matched. Observations are then dropped if there are fewer than six monthly observations for the firm or if spread estimates are missing for the *Roll Spread*, *Tick Spread,* or *HL Spread*. Panel A lists results based on monthly spreads and Panel B lists results based on first differences in monthly spreads.

| | N | Mean Error | | | Mean Absolute Error | | |
|---|---|---|---|---|---|---|---|
| | | Roll Spread | Eff. Tick Spread | High-Low Spread | Roll Spread | Eff. Tick Spread | High-Low Spread |
| Panel A – Spread Errors, Monthly Estimates | | | | | | | |
| Full Period | 156 | 0.0009 | -0.0070 | -0.0016 | 0.0172 | 0.0114 | 0.0089 |
| 1993-1996 | 53 | -0.0044 | -0.0073 | -0.0047 | 0.0173 | 0.0142 | 0.0085 |
| 1997-2000 | 43 | 0.0021 | -0.0042 | -0.0016 | 0.0195 | 0.0104 | 0.0093 |
| 2001-2005 | 60 | 0.0047 | -0.0086 | 0.0012 | 0.0153 | 0.0097 | 0.0089 |
| Panel B – Changes in Spread Errors, Monthly Estimates | | | | | | | |
| Full Period | 155 | 0.0079 | 0.0029 | -0.0001 | 0.0219 | 0.0088 | 0.0070 |
| 1993-1996 | 52 | 0.0085 | 0.0028 | 0.0022 | 0.0218 | 0.0123 | 0.0082 |
| 1997-2000 | 43 | 0.0290 | 0.0137 | 0.0002 | 0.0252 | 0.0105 | 0.0075 |
| 2001-2005 | 60 | -0.0078 | -0.0047 | -0.0024 | 0.0197 | 0.0046 | 0.0055 |

**Table 6 - Summary Statistics for Stock-by-Stock Time Series Correlations**

For spread measure and each stock, we estimate the time-series correlation between the estimated spread measure and the effective spread from TAQ. The table lists the average time-series correlation across all stocks. The full sample includes all NYSE, Amex, and Nasdaq listed securities for which TAQ and CRSP data could be matched. Observations are then dropped if there are fewer than six monthly observations for the firm or if spread estimates are missing for the *Roll Spread*, *Tick Spread,* or *HL Spread*. Panel A provides results for the full sample period and Panels B, C, and D provide subperiod results. Stocks are also separated by exchange and market capitalization decile based on the CRSP exchange code and market capitalization on the last date the firm is listed on CRSP. Market capitalization deciles are based on NYSE breakpoints.

| | N | Roll Spread | Effective Tick Spread | High - Low Spread |
|---|---|---|---|---|
| Panel A – Correlations with Effective Spread | | | | |
| Full Sample | 12,192 | 0.3364 | 0.5946 | 0.6237 |
| NYSE | 2,836 | 0.1574 | 0.6991 | 0.4064 |
| Amex | 1,095 | 0.3826 | 0.5443 | 0.6349 |
| Nasdaq | 8,261 | 0.3917 | 0.5654 | 0.6968 |
| MV Decile 1 | 7,054 | 0.4327 | 0.5350 | 0.7198 |
| MV Decile 2 | 1,505 | 0.3029 | 0.6007 | 0.6282 |
| MV Decile 3 | 865 | 0.2492 | 0.6434 | 0.5846 |
| MV Decile 4 | 640 | 0.1827 | 0.7080 | 0.5066 |
| MV Decile 5 | 481 | 0.1671 | 0.7118 | 0.4674 |
| MV Decile 6 | 393 | 0.1334 | 0.7209 | 0.3783 |
| MV Decile 7 | 400 | 0.1297 | 0.7571 | 0.3960 |
| MV Decile 8 | 309 | 0.0734 | 0.7868 | 0.2506 |
| MV Decile 9 | 241 | 0.0352 | 0.7733 | 0.1694 |
| MV Decile 10 | 199 | 0.0279 | 0.8036 | 0.1100 |
| Panel B – Correlations with Effective Spread (1993-1996) | | | | |
| Full Sample | 9,036 | 0.3258 | 0.4608 | 0.6243 |
| NYSE | 2,360 | 0.1956 | 0.5041 | 0.4816 |
| Amex | 764 | 0.3050 | 0.4335 | 0.5649 |
| Nasdaq | 5,912 | 0.3803 | 0.4471 | 0.6889 |
| MV Decile 1 | 5,078 | 0.3869 | 0.4337 | 0.6756 |
| MV Decile 2 | 1,075 | 0.3380 | 0.4633 | 0.6471 |
| MV Decile 3 | 639 | 0.3031 | 0.4877 | 0.6318 |
| MV Decile 4 | 499 | 0.2268 | 0.4848 | 0.5816 |
| MV Decile 5 | 371 | 0.2404 | 0.5007 | 0.5676 |
| MV Decile 6 | 308 | 0.1846 | 0.5070 | 0.4966 |
| MV Decile 7 | 330 | 0.1828 | 0.5412 | 0.5138 |
| MV Decile 8 | 266 | 0.1394 | 0.5299 | 0.4270 |
| MV Decile 9 | 208 | 0.0966 | 0.5543 | 0.3405 |
| MV Decile 10 | 179 | 0.1043 | 0.5915 | 0.2966 |

**Table 6 (continued)**

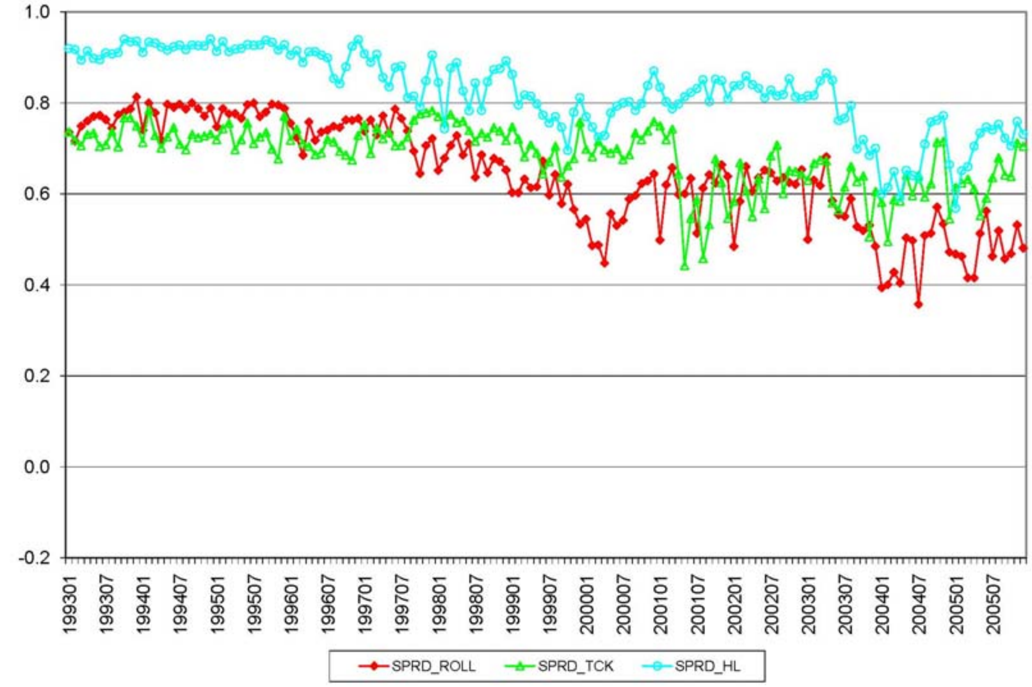
| | N | Roll Spread | Effective Tick Spread | High-Low Spread |
|---|---|---|---|---|
| Panel C – Correlations with Effective Spread (1997-2000) | | | | |
| Full Sample | 9,341 | 0.2782 | 0.5051 | 0.5794 |
| | | | | |
| NYSE | 2,319 | 0.1642 | 0.5167 | 0.4873 |
| Amex | 781 | 0.3217 | 0.5070 | 0.5702 |
| Nasdaq | 6,241 | 0.3152 | 0.5006 | 0.6148 |
| | | | | |
| MV Decile 1 | 5,347 | 0.3606 | 0.5112 | 0.6655 |
| MV Decile 2 | 1,093 | 0.2442 | 0.4896 | 0.5596 |
| MV Decile 3 | 633 | 0.1874 | 0.4892 | 0.5216 |
| MV Decile 4 | 514 | 0.1627 | 0.4952 | 0.4829 |
| MV Decile 5 | 385 | 0.1218 | 0.5077 | 0.4354 |
| MV Decile 6 | 317 | 0.1249 | 0.4942 | 0.3992 |
| MV Decile 7 | 331 | 0.1137 | 0.5002 | 0.3757 |
| MV Decile 8 | 275 | 0.0813 | 0.5309 | 0.3142 |
| MV Decile 9 | 198 | 0.0803 | 0.5225 | 0.2886 |
| MV Decile 10 | 181 | 0.0895 | 0.5035 | 0.3117 |
| Panel D – Correlations with Effective Spread (2001-2005) | | | | |
| Full Sample | 7,114 | 0.2747 | 0.5160 | 0.5816 |
| | | | | |
| NYSE | 1,836 | 0.1592 | 0.5829 | 0.4786 |
| Amex | 695 | 0.3064 | 0.5010 | 0.5854 |
| Nasdaq | 4,583 | 0.3161 | 0.4914 | 0.6222 |
| | | | | |
| MV Decile 1 | 4,056 | 0.3351 | 0.4500 | 0.6263 |
| MV Decile 2 | 818 | 0.2182 | 0.5665 | 0.5228 |
| MV Decile 3 | 474 | 0.1827 | 0.5785 | 0.4786 |
| MV Decile 4 | 398 | 0.1713 | 0.6222 | 0.4580 |
| MV Decile 5 | 296 | 0.1672 | 0.6216 | 0.4539 |
| MV Decile 6 | 235 | 0.1782 | 0.6412 | 0.5107 |
| MV Decile 7 | 268 | 0.1949 | 0.6385 | 0.5768 |
| MV Decile 8 | 219 | 0.1879 | 0.6637 | 0.5962 |
| MV Decile 9 | 163 | 0.1706 | 0.6247 | 0.5975 |
| MV Decile 10 | 152 | 0.2334 | 0.5904 | 0.6518 |

## Table 7 - Average Stock-by-Stock Mean Absolute Errors

For each stock-month, errors are defined for each spread measure as the absolute value of the difference between the spread measure and the TAQ effective spread. For each stock, we estimate the mean error and mean absolute error across all time series observations. The table then lists the mean across all stocks. The full sample includes all NYSE, Amex, and Nasdaq listed securities for which TAQ and CRSP data could be matched. Observations are then dropped if there are fewer than six monthly observations for the firm or if spread estimates are missing for the *Roll Spread*, *Tick Spread*, or *HL Spread*. Stocks are also separated by exchange and market capitalization decile based on the CRSP exchange code and market capitalization on the last date the firm is listed on CRSP. Market capitalization deciles are based on NYSE breakpoints.

| | | Mean Error | | | Mean Absolute Error | | |
|---|---|---|---|---|---|---|---|
| | N | Roll Spread | Eff. Tick Spread | High-Low Spread | Roll Spread | Eff. Tick Spread | High-Low Spread |
| Full Sample | 12,192 | -0.0009 | -0.0081 | -0.0028 | 0.0201 | 0.0145 | 0.0103 |
| | | | | | | | |
| NYSE | 2,836 | 0.0020 | 0.0004 | 0.0009 | 0.0102 | 0.0037 | 0.0046 |
| Amex | 1,095 | -0.0072 | -0.0050 | -0.0055 | 0.0233 | 0.0172 | 0.0143 |
| Nasdaq | 8,261 | -0.0010 | -0.0114 | -0.0037 | 0.0231 | 0.0178 | 0.0118 |
| | | | | | | | |
| MV Decile 1 | 7,055 | -0.0031 | -0.0113 | -0.0058 | 0.0259 | 0.0210 | 0.0136 |
| MV Decile 2 | 1,505 | 0.0005 | -0.0062 | -0.0006 | 0.0144 | 0.0084 | 0.0070 |
| MV Decile 3 | 865 | 0.0017 | -0.0040 | 0.0007 | 0.0123 | 0.0056 | 0.0056 |
| MV Decile 4 | 640 | 0.0030 | -0.0029 | 0.0018 | 0.0118 | 0.0043 | 0.0054 |
| MV Decile 5 | 481 | 0.0034 | -0.0021 | 0.0020 | 0.0108 | 0.0034 | 0.0048 |
| MV Decile 6 | 393 | 0.0035 | -0.0012 | 0.0025 | 0.0098 | 0.0025 | 0.0044 |
| MV Decile 7 | 400 | 0.0037 | -0.0009 | 0.0028 | 0.0092 | 0.0021 | 0.0043 |
| MV Decile 8 | 309 | 0.0044 | -0.0006 | 0.0037 | 0.0089 | 0.0016 | 0.0046 |
| MV Decile 9 | 241 | 0.0049 | -0.0001 | 0.0040 | 0.0082 | 0.0010 | 0.0044 |
| MV Decile 10 | 199 | 0.0053 | 0.0001 | 0.0045 | 0.0075 | 0.0007 | 0.0046 |

**Panel A - Correlations based on Percentage Spreads**




**Panel B - Correlations based on First Differences in Percentage Spreads**

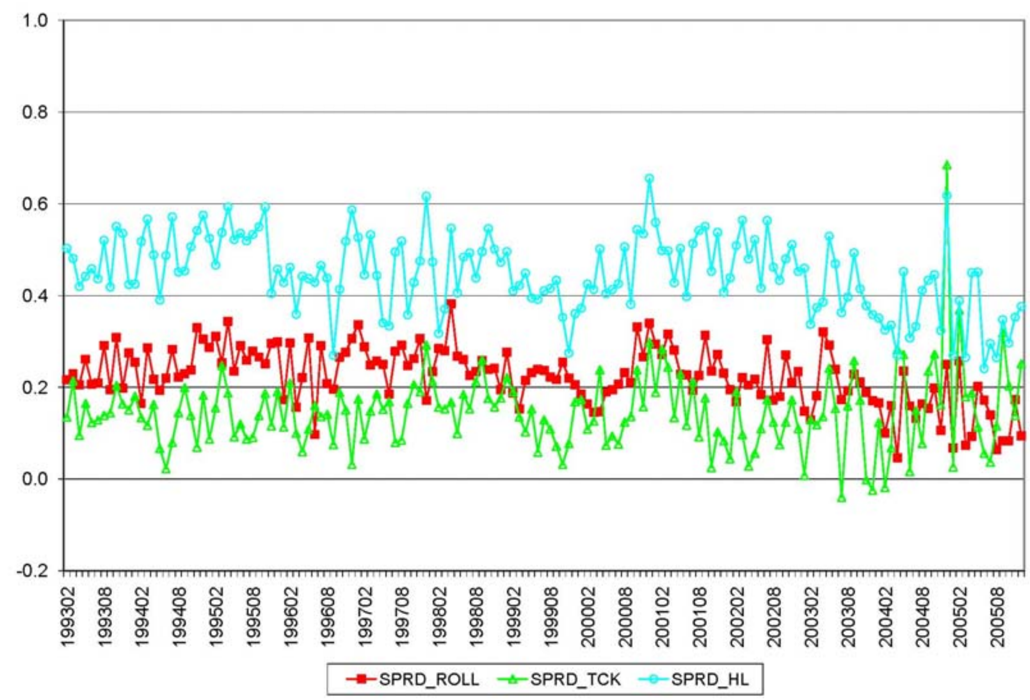
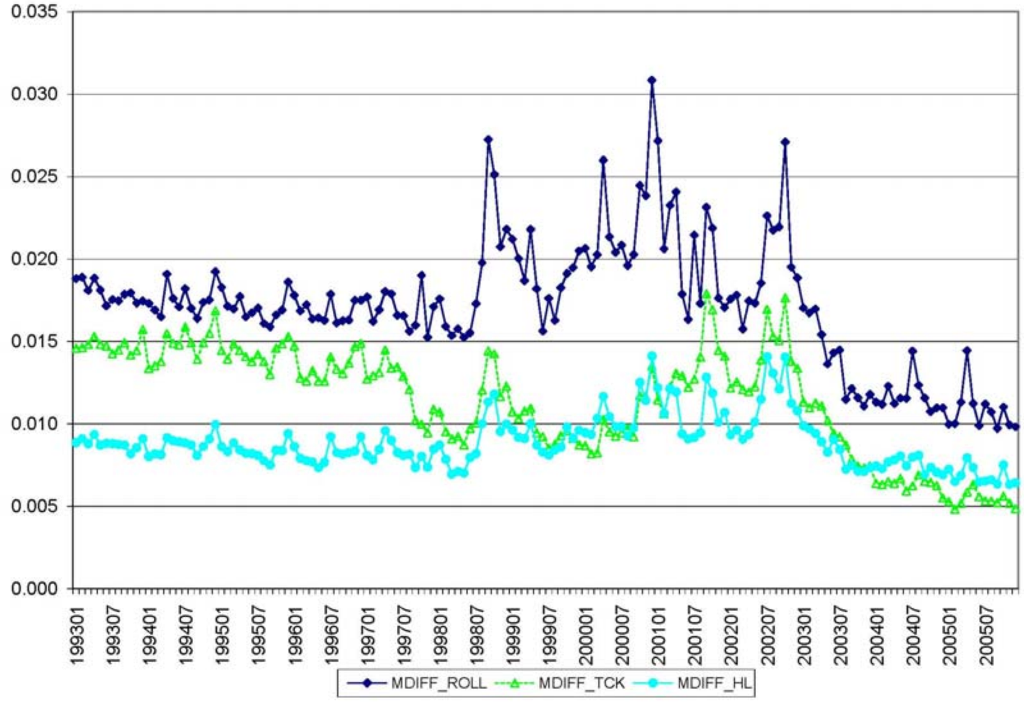



**Figure 1 - Cross Sectional Correlations of Spread Estimates with TAQ Effective Spreads by Month**
The figure plots monthly cross-sectional correlations between three estimated spread measures and the effective spread from TAQ. The correlations shown in Panel A are estimated from monthly spread estimates. The correlations shown in Panel B are estimated from first differences in monthly spread estimates. The full sample includes all NYSE, Amex, and Nasdaq listed securities for which TAQ and CRSP data could be matched. Observations are dropped if there are fewer than six monthly observations for the firm or if spread estimates are missing for the *Roll Spread*, *Tick Spread,* or *HL Spread*.

**Panel A - Mean Absolute Errors by Month**




**Panel B - Median Absolute Errors by Month**

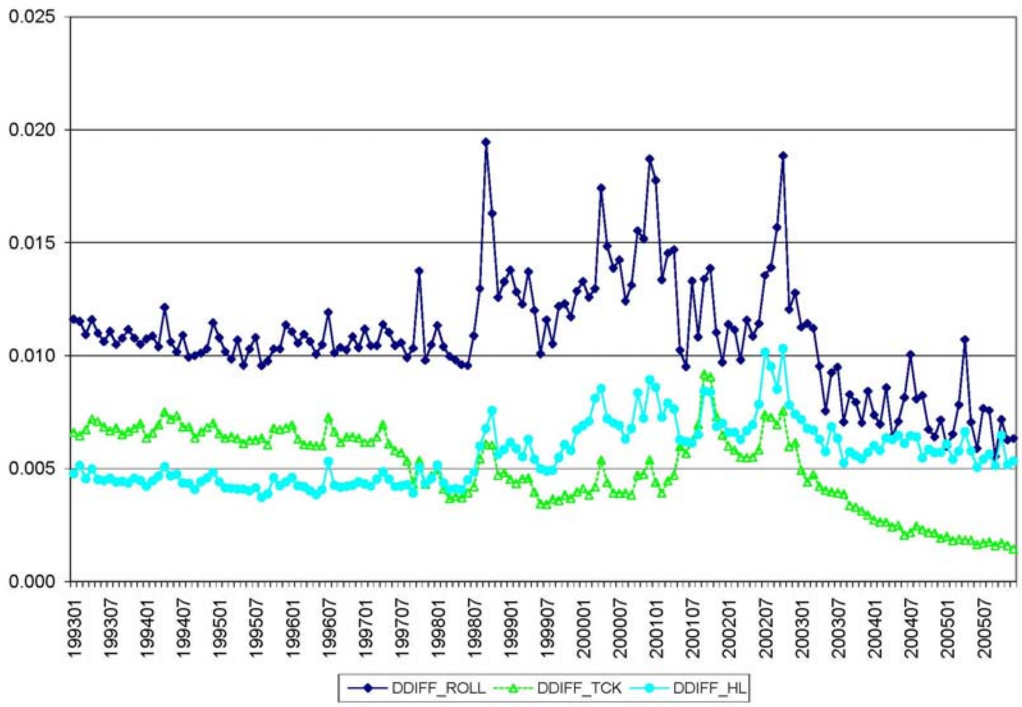
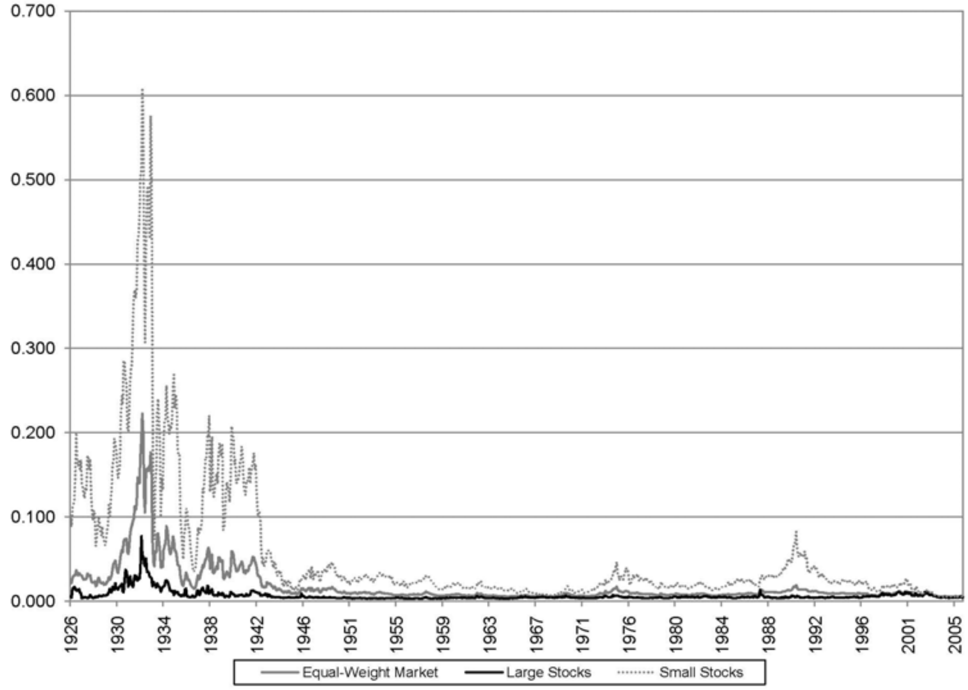


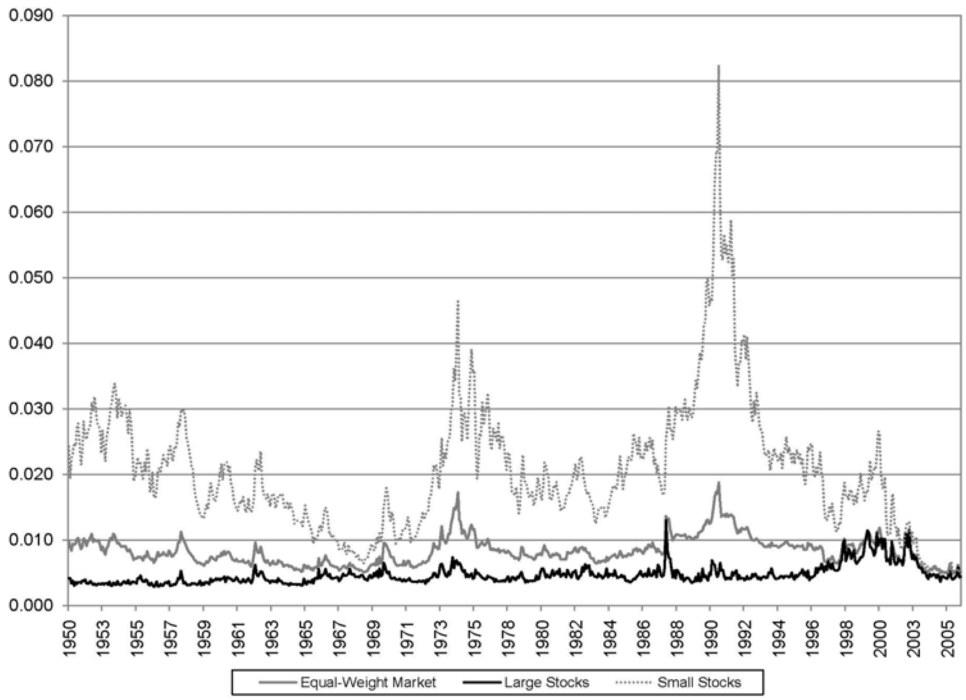

**Figure 2 - Cross Sectional Mean and Median Absolute Errors of Spread Estimates by Month**

The figure plots the mean absolute error across all securities by month. The mean absolute error is defined for each spread measure as the absolute value of the difference between the estimated spread measure and the effective spread from TAQ. The full sample includes all NYSE, Amex, and Nasdaq listed securities for which TAQ and CRSP data could be matched. Observations are then dropped if there are fewer than six monthly observations for the firm or if spread estimates are missing for the *Roll Spread*, *Tick Spread,* or *HL Spread*.

**Panel A - Average High-Low Spreads for NYSE Stocks by Month, 1926-2006**



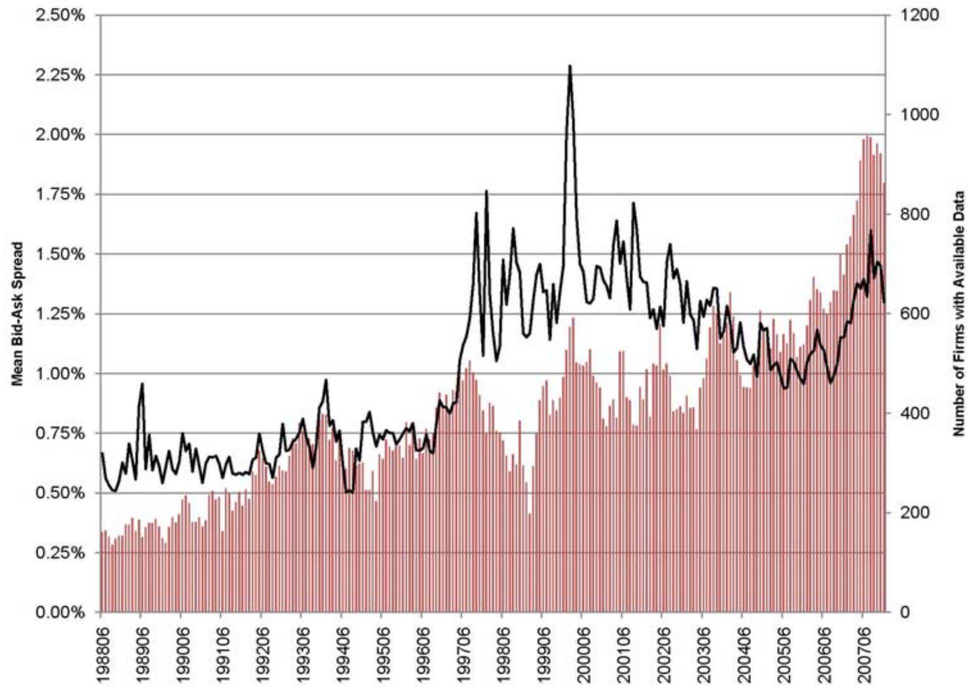

**Panel B - Average High-Low Spreads for NYSE Stocks by Month, 1950-2006**




## Figure 3 - Historical High-Low Spread Estimates based on CRSP Data

High-low spreads are estimated for each stock each month by averaging two-day spread estimates within the month. The graph plots the equally weighted average spread by month across all stocks with at least 13 daily spread observations within the month. Results are shown for the full sample of NYSE stocks, and for the smallest and largest deciles by market capitalization. The graph also shows the number of firms included in the average each month. Panel A shows results from 1926-2005 and while Panel B shows results from 1950-2005. All data are from CRSP.

**Panel A - Average High-Low Spreads for Stocks in Hong Kong by Month, 1988-2007**




**Panel B - Average High-Low Spreads for Stocks in India by Month, 1990-2007**

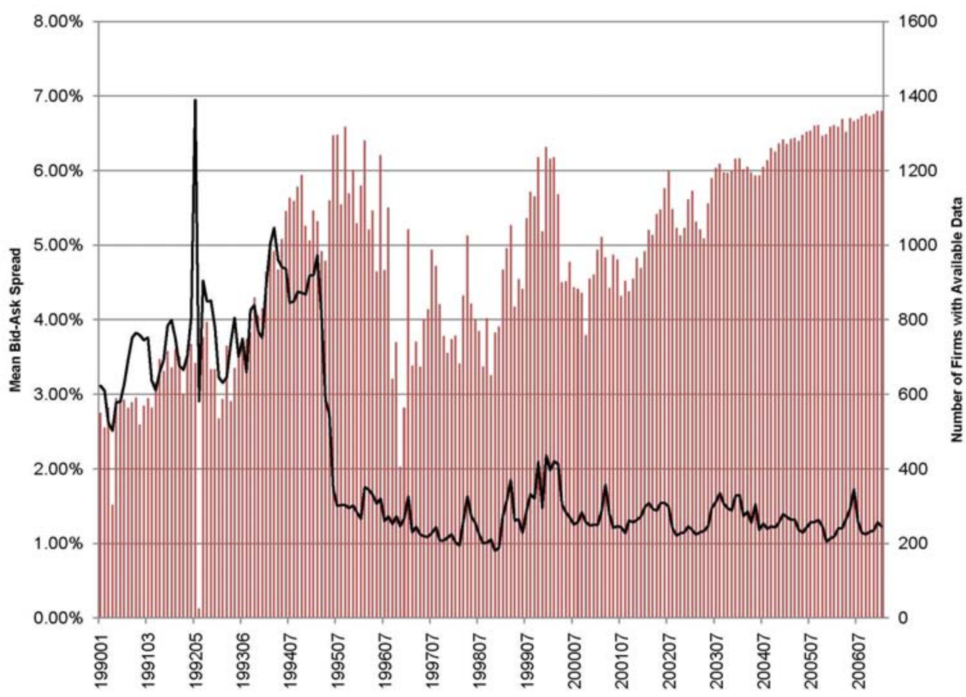



**Figure 4 - Historical High-Low Spread Estimates based on Datastream Data**

High-low spreads are estimated for each stock each month by averaging two-day spread estimates within the month. The graph plots the equally weighted average spread by month across all stocks with at least 13 daily spread observations within the month. The graph also shows the number of firms included in the average each month. Panel A shows results for stocks in Hong Kong and Panel B shows results for stocks in India. All data are from Datastream.