The Journal of FINANCE
The Journal of THE AMERICAN FINANCE ASSOCIATION

THE JOURNAL OF FINANCE • VOL. LXVII, NO. 2 • APRIL 2012

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

SHANE A. CORWIN and PAUL SCHULTZ*

ABSTRACT

We develop a bid-ask spread estimator from daily high and low prices. Daily high (low) prices are almost always buy (sell) trades. Hence, the high–low ratio reflects both the stock's variance and its bid-ask spread. Although the variance component of the high–low ratio is proportional to the return interval, the spread component is not. This allows us to derive a spread estimator as a function of high–low ratios over 1-day and 2-day intervals. The estimator is easy to calculate, can be applied in a variety of research areas, and generally outperforms other low-frequency estimators.

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 buyer-initiated trades and daily low prices are almost always seller-initiated trades. The ratio of high-to-low prices for a day therefore 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 does not. This implies that the sum of the price ranges over 2 consecutive single days reflects 2 days' volatility and twice the spread, while the price range over one 2-day period reflects 2 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 2-day period and the high-to-low ratios for 2 consecutive single days. Simulations reveal that, under realistic conditions, the correlation between high–low spread estimates and true spreads is about 0.9 and the standard deviation of high–low spread estimates is one-fourth to one-half as large as the standard deviation of estimates from the popular Roll (1984) covariance spread estimator.

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

*Both authors are from the Mendoza College of Business at the University of Notre Dame. We are grateful to the Editor, Cam Harvey, the Associate Editor, and two anonymous referees for their valuable suggestions. We also thank seminar participants at the University of Notre Dame and the National Bureau of Economic Research Market Microstructure meeting, and Shmuel Baruch, Robert Battalio, Hank Bessembinder, Ryan Davies, Larry Harris, Joel Hasbrouck, Asani Sarkar, and Noah Stoffman for helpful comments.

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 or international markets without intraday data.[1] In other cases, these measures are used when intraday quotes and trades cannot be reliably matched.[2] Other low-frequency spread measures based on the frequency of zero returns are developed in Lesmond, Ogden, and Trzcinka (LOT; 1999) and have been applied in a number of recent papers.[3]

The high–low spread estimator derived here has a number of advantages over the daily estimators used in prior research. First, using TAQ data from 1993 to 2006, we show that it outperforms the still popular Roll (1984) covariance estimator or the LOT estimator of Lesmond et al. (1999). Another advantage is that it is easy to use. We provide a closed-form solution for the spread that can be easily programmed, unlike measures that require an iterative process (Hasbrouck (2009)) or maximum likelihood estimation (LOT). Third, unlike Hasbrouck's (2009) Gibbs estimator or the Holden (2009) measure, the high–low spread estimator is not computer-time intensive, making it ideal for large samples. Finally, the high–low spread estimator is derived under very general conditions. It is not ad hoc and can be applied to a variety of markets with different market structures.

We test the performance of the high–low spread estimator by comparing it to monthly TAQ effective spreads from 1993 through 2006. For comparison purposes, we also estimate spreads from daily data using the covariance spread estimator of Roll (1984), the effective tick estimator of Holden (2009), and the LOT measure of Lesmond et al. (1999). Because researchers tackling different problems may care about different characteristics of the spread estimator, we provide several different performance tests. We first calculate cross-sectional correlations between spread estimates and TAQ effective spreads on a month-by-month basis from 1993 through 2006. 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 cross-sectional correlations on a month-by-month basis allows us to examine the performance of the estimators in different trading environments. The three subperiods that we examine, 1993 to 1996, 1997 to 2000, and 2001 to 2006, 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

[1] See Asparouhova, Bessembinder, and Kalcheva (2010), Bharath, Pasquariello, and Wu (2009), Gehrig and Fohlin (2006), Kim et al. (2007), Lesmond, Schill, and Zhou (2004), or Lipson and Mortal (2009) for estimates of spreads for periods before the availability of intraday data. See Amihud, Lauterbach, and Mendelson (2003), Chakrabarti et al. (2005), or Griffin, Kelly, and Nardari (2010) for examples of the application of Roll spread estimators to international data.

[2] See Antunovich and Sarkar (2006), Fink, Fink, and Weston (2006), or Schultz (2000).

[3] See Bekaert, Harvey, and Lundblad (2007), Griffin et al. (2010), Lesmond et al. (2004), and Mei, Scheinkman, and Xiong (2009) for applications of the LOT measure. Amihud (2002) and Pástor and Stambaugh (2003) provide low frequency measures 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.

and TAQ effective spreads are higher than cross-sectional correlations between TAQ effective spreads and any of the other estimators. For the entire period, the average cross-sectional correlation of high–low spread estimates with TAQ effective spreads is 0.829, compared to 0.637 for the Roll spread, 0.683 for the effective tick spread, and 0.635 for the LOT measure. In additional tests, we examine cross-sectional correlations between monthly changes in high–low spreads and monthly changes in TAQ effective spreads. For the entire period, the high–low spread estimator dominates with an average cross-sectional correlation of 0.472, compared to 0.249 for the Roll spread, 0.183 for the effective tick spread, and 0.186 for the LOT measure. Thus, the high–low estimator outperforms the alternative measures in capturing the cross-sections of both spread levels and month-to-month changes in spreads. Notably, the high–low spread estimator performs particularly well during the 1993 to 1996 subperiod when the tick size was one-eighth, suggesting that it should perform well during earlier time periods when intraday data were not available.

Next, we 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 quintiles and all exchange listings, we find that high–low spreads have much higher average time-series correlations with TAQ effective spreads than do Roll spreads or spreads estimated from the LOT measure. For the great majority of stocks, high–low spreads also have higher time-series correlations with TAQ effective spreads than do effective tick spreads. However, effective tick spread estimates have higher correlations for the largest stocks, especially in the most recent time period.

The high–low spread estimator is derived under very general conditions. It is simple and easy to use. Both simulations and comparisons to TAQ data suggest that, for most applications, the high–low spread estimator outperforms alternative low-frequency spread estimators. It clearly dominates other low-frequency measures in capturing the cross-sections of bid-ask spreads and month-to-month changes in spreads. For small stocks, such as those listed on NASDAQ and Amex, the high–low spread also dominates other measures in capturing the time-series variation in individual stock spreads, especially within subperiods corresponding to different tick sizes. It is important to note that the high–low spread estimator captures liquidity more broadly than just the bid-ask spread. Price pressure from large orders will often lead to execution at daily high or low prices. Similarly, a succession of buy or sell orders in a shallow market may result in executions at daily high or low prices. The high–low spread estimator captures these transitory price effects in addition to the bid-ask spread.

To further demonstrate the potential uses of the high–low spread estimator, we provide several example applications. First, we use the high–low estimator to examine patterns in spreads for NYSE and Amex stocks for the period from 1926 through 2006. Using these data we then demonstrate that, despite its simplicity, the high–low estimator has similar power to the Amihud (2002) measure for capturing the relation between liquidity and stock returns. We also demonstrate the application of the high–low estimator to non-U.S. stocks by analyzing patterns in high–low spreads estimated from Datastream data for stocks listed in Hong Kong and India. These examples, as well as several additional applications provided in the Internet Appendix, demonstrate the potential uses of the high–low estimator in a wide variety of research areas across many types of markets.[4,5]

The remainder of the paper is organized as follows. The high–low spread estimator is derived in Section I. Section II discusses practical issues in estimating spreads using high and low prices. Section III discusses existing spread estimators that use daily data and reviews empirical tests of these estimators. We present simulation results for the high–low spread estimator in Section IV. In Section V, spread estimates from the high–low spread estimator are compared with TAQ effective spreads and with estimates based on the Roll spread, the effective tick spread, and the LOT measure. In Section VI, we provide examples of applications of the high–low spread estimator. Section VII concludes.

I. The High–Low Spread Estimator

The high–low spread estimator is based on a simple insight. The high–low price ratio reflects both the true variance of the stock price and the bid-ask spread. While the variance component grows proportionately with the time period, the spread component does not. This allows us to solve for both the spread and the variance by deriving two equations, the first a function of the high–low ratios on 2 consecutive single days and the second a function of the high–low ratio from a single 2-day period.

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 2-day estimation period. Because of the spread, observed prices for buys are higher than the actual values by (S/2)%, while observed prices for sells are lower than the actual value by (S/2)%. We assume further that the daily high price is a buyer-initiated trade and is therefore grossed up by half of the spread, while the daily low price is a seller-initiated trade and is discounted by half of the spread. Hence, the observed high–low price range contains both the range of the actual prices and the bid-ask spread. With H_t^A (L_t^A) denoting the actual high (low) stock price on day t and H_t^o (L_t^o) the observed high (low) stock price for day t, we can write

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

[4] In a recent paper, Deuskar, Gupta, and Subrahmanyam (2011) use the high–low estimator to measure transaction costs in OTC options markets.

[5] The Internet Appendix is available on the *Journal of Finance* website at http://www.afajof.org/supplements.asp.

Rearranging (1) gives

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

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

$$E\left\{\frac{1}{T}\sum_{t=1}^{T}\left[\ln\left(\frac{H_t}{L_t}\right)\right]^2\right\} = k_1\sigma_{HL}^2, \quad (3)$$

where H_t (L_t) is the high (low) on day t and $k_1 = 4\ln(2)$.[6] Similarly, Parkinson (1980) shows that

$$E\left\{\frac{1}{T}\sum_{t=1}^{T}\left[\ln\left(\frac{H_t}{L_t}\right)\right]\right\} = k_2\sigma_{HL}, \text{ where } k_2 = \sqrt{\frac{8}{\pi}}. \quad (4)$$

Taking expectations of (2) and substituting from (3) and (4) yields

$$E\left\{\left[\ln\left(\frac{H_t^o}{L_t^o}\right)\right]^2\right\} = k_1\sigma_{HL}^2 + 2k_2\sigma_{HL}\ln\left(\frac{2+S}{2-S}\right) + \left[\ln\left(\frac{2+S}{2-S}\right)\right]^2. \quad (5)$$

The expectation of the sum of (5) over 2 single days is

$$E\left\{\sum_{j=0}^{1}\left[\ln\left(\frac{H_{t+j}^o}{L_{t+j}^o}\right)\right]^2\right\} = 2k_1\sigma_{HL}^2 + 4k_2\sigma_{HL}\ln\left(\frac{2+S}{2-S}\right) + 2\left[\ln\left(\frac{2+S}{2-S}\right)\right]^2. \quad (6)$$

To simplify the notation going forward, we set

$$\alpha = \left[\ln\left(\frac{2+S}{2-S}\right)\right], \qquad \beta = E\left\{\sum_{j=0}^{1}\left[\ln\left(\frac{H_{t+j}^o}{L_{t+j}^o}\right)\right]^2\right\}. \quad (7)$$

This allows us to rewrite (6) as

$$2k_1\sigma_{HL}^2 + 4k_2\sigma_{HL}\alpha + 2\alpha^2 - \beta = 0. \quad (8)$$

Equation (8) links the high–low price ratios on 2 consecutive single days with two unknowns: α and σ. To solve for these unknowns, we define a second

[6] Using a sample of 208 stocks over 29 quarters from January 1973 through March 1980, Beckers (1983) demonstrates that the high–low variance estimator is more accurate than the traditional variance estimator based on closing prices. See Gallant, Hsu, and Tauchen (1999) for an application of high–low volatility estimators to the estimation of stochastic volatility.

equation that links the high–low ratio from the 2-day period and the same two unknowns. Squaring the log high–low ratio over a 2-day period yields

$$\left[\ln\left(\frac{H^{o}_{t,t+1}}{L^{o}_{t,t+1}}\right)\right]^{2} = \left[\ln\left(\frac{H^{A}_{t,t+1}}{L^{A}_{t,t+1}}\right)\right]^{2} + 2\left[\ln\left(\frac{H^{A}_{t,t+1}}{L^{A}_{t,t+1}}\right)\right]\left[\ln\left(\frac{2+S}{2-S}\right)\right] + \left[\ln\left(\frac{2+S}{2-S}\right)\right]^{2}, \tag{9}$$

where $H_{t,t+1}$ is the high price over the 2 days t and $t+1$ and $L_{t,t+1}$ is the low price over days t and $t+1$. To further simplify notation, we set

$$\gamma = \left[\ln\left(\frac{H^{o}_{t,t+1}}{L^{o}_{t,t+1}}\right)\right]^{2}. \tag{10}$$

Using this notation and taking expectations in (9) yields

$$2k_1\sigma^2_{HL} + 2\sqrt{2}k_2\sigma_{HL}\alpha + \alpha^2 - \gamma = 0. \tag{11}$$

This leaves two equations, (8) and (11), and two unknowns, σ and α. Because the spread is positive, α must also be positive. Hence, from (8), we choose the positive root for α

$$\alpha = -k_2\sigma_{HL} + \sqrt{\sigma^2_{HL}\left(k_2^2 - k_1\right) + {}^{\beta}\!/_{2}}. \tag{12}$$

Substituting from (12) into (11) and rearranging yields

$$\sigma^2_{HL}\left(k_2^2(2-2\sqrt{2}) + k_1\right) + \sigma_{HL}k_2(2\sqrt{2}-2)\sqrt{\sigma^2_{HL}\left(k_2^2 - k_1\right) + {}^{\beta}\!/_{2}} + \frac{\beta}{2} - \gamma = 0. \tag{13}$$

Equation (13) can be easily solved numerically for σ and the result inserted into (12) to obtain a value for α. A simple transformation of α in (7) then provides the high–low spread estimate

$$S = \frac{2\left(e^{\alpha} - 1\right)}{1 + e^{\alpha}}. \tag{14}$$

Inspection of either (7) or (14) reveals that, for small spreads, α and the spread are almost equal. This can simplify estimation in practice. A further simplifying assumption allows us to obtain closed-form solutions for σ and α. If we ignore Jensen's inequality in (4) and assume that

$$E\left\{\frac{1}{T}\sum_{t=1}^{T}\left[\ln\left(\frac{H_t}{L_t}\right)\right]\right\} = \sqrt{E\left\{\frac{1}{T}\sum_{t=1}^{T}\left[\ln\left(\frac{H_t}{L_t}\right)\right]^{2}\right\}} = \sqrt{k_1\sigma^2_{HL}} = \sqrt{k_1}\sigma_{HL}, \tag{15}$$

then $k_2^2 = k_1$ and (12) and (13) become

$$\alpha = -k_2\sigma_{HL} + \sqrt{\beta/2} \tag{12$'$}$$

$$\sigma_{HL}^2\left(k_2^2(3-2\sqrt{2})\right) + \sigma_{HL}k_2(2\sqrt{2}-2)\sqrt{\beta/2} + \frac{\beta}{2} - \gamma = 0. \tag{13$'$}$$

Rearranging yields

$$\sigma_{HL}^2 + \sigma_{HL}\frac{2\sqrt{\beta}-\sqrt{2\beta}}{k_2(3-2\sqrt{2})} + \frac{\beta/2-\gamma}{k_2^2(3-2\sqrt{2})} = 0. \tag{16}$$

Solving for σ and using the positive root to ensure a positive estimate yields

$$\sigma_{HL} = \frac{\sqrt{\beta/2}-\sqrt{\beta}}{k_2(3-2\sqrt{2})} + \sqrt{\frac{\gamma}{k_2^2(3-2\sqrt{2})}}. \tag{17}$$

Inserting the standard deviation from (17) into (12) provides an estimate of α

$$\alpha = \frac{\sqrt{2\beta}-\sqrt{\beta}}{3-2\sqrt{2}} - \sqrt{\frac{\gamma}{3-2\sqrt{2}}}. \tag{18}$$

This closed-form solution can be inserted into (14) as before to yield our "simple" high–low estimator.

The spread estimator given in (14) is easy to compute and does not require the researcher to iterate through successive estimates of the spread to get the correct value. 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 2 consecutive days. To get spreads for longer periods like a month, we average the spread estimates from all overlapping 2-day periods within the month.

One note of caution is needed here. In estimating spreads and variances, we use the observed ratio of high to low prices, while the estimator is derived using the expected ratio. Because the variance and the spread are nonlinear functions of the high–low price ratio, an average of spread estimates is not an unbiased estimate of the spread. However, our simulation results and empirical analysis both suggest that this is not a problem in practice.[7]

[7] To address the importance of this problem, we reestimate monthly spreads using an average of the high–low ratio parameters rather than an average of daily spread estimates over the month. We find in both our simulations and empirical tests that this method does not produce more accurate monthly spread estimates. We are grateful to an anonymous referee for this suggestion.

II. Using the High–Low Spread Estimator in Practice

There are a number of implicit assumptions underlying the high–low spread estimator. One is that the stock trades continuously while the market is open. 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.

A. Adjustment for Overnight Price Changes

Because markets are closed overnight, the ratio of high to low prices for the 2-day period reflects both the range of prices during each day and the overnight return. On the other hand, the two single-day high–low ratios reflect only the range of prices during trading hours. Though stock prices are more volatile during the trading day than at other times, stock prices often move significantly over nontrading periods (French and Roll (1986) and Harris (1986)). This causes the high–low price ratio (and hence variance) estimated using one 2-day period to be inflated relative to the variance estimated using two 1-day periods. Without an adjustment for overnight returns, the spread portion of the high–low price ratio will therefore be underestimated.

To correct for overnight returns, we determine whether 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 that the price rose overnight from the close to the day $t+1$ low and decrease both the high and low for day $t+1$ by the amount of the overnight change when calculating spreads. Similarly, if the day $t+1$ high is below the day t close, we assume the price fell overnight 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.

As an alternative, we could adjust for overnight returns using the difference between the day t close price and the day $t+1$ open price. We do not use this adjustment for three reasons. First, we want to adjust only those cases in which 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 change in the true value. Second, a primary use of this estimator is to estimate historic trading costs during periods when data on open prices may not be available. For example, open prices are missing on CRSP from July 1962 through June 1992. Finally, we found a small number of cases in which the open price was outside the high–low price range reported by CRSP, suggesting that open price data may be unreliable.[8]

[8] For completeness, we reestimated high–low spreads using no overnight adjustment and using an alternative adjustment based on the price change from the previous day's close to the current day's open. The results show that the overnight adjustment used in the paper clearly dominates either of the alternatives. These results are provided in the Internet Appendix. (An Internet Appendix for this article is available online in the "Supplements and Datasets" section at http://www.afajof.org/supplements.asp.)

B. 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 for the day and the observed low price will be greater than the true low price for the day. In practice, 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. 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 such cases, if the trade price is within the previous day's price range, we assume the same high and low prices as the previous day. In those less common cases in which 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 low and high prices. In practice, a researcher may benefit from using this information when estimating spreads. However, to provide a fair comparison with other estimators, we eliminate the bid and ask prices provided by CRSP in these cases and replace them with the most recent high and low trade prices available from a prior trading day.[9]

C. High–Low Spread Estimates May Be Negative

The high–low estimator assumes that the expectation of a stock's true variance over a 2-day period is twice as large as the expectation of the variance over a single day. Even if this is true in expectation, the observed 2-day variance may be more than twice as large as the single-day variance during volatile periods, in cases with a large overnight price change, or when the total return over the 2-day period is large relative to the intraday volatility. If the observed 2-day variance is large enough, the high–low spread estimate will be negative. For most of the analysis to follow, we set all negative 2-day spreads to zero before calculating monthly averages. As described in more detail later, this produces more accurate monthly estimates than either including or deleting negative 2-day spread observations.[10]

III. Other Classes of Spread Estimators That Use Daily Data

To our knowledge, this is the first paper to use high and low prices to estimate trading costs. Researchers have derived several other classes of spread

[9] We note that, in cases in which the current day's high and low are equal to the previous day's high and low, the variance component of the high–low estimator will be zero and the spread component is set to the high–low range. This may overstate the spread component in cases in which the variance component is small but is not observed.

[10] The Internet Appendix describes the frequency of each data adjustment described in Section II.

estimators based on daily data. We describe several of these alternative estimators below.

A. Spread Estimators Derived from Return Covariances

Roll (1984) assumes that the true value of a stock follows a random walk and that 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. Under these conditions, the expected autocorrelation of returns from observed prices will be negative and Roll derives the following simple estimator for the spread

$$S = 2\sqrt{-\text{Cov}(\Delta P_t, \Delta P_{t-1})}. \tag{19}$$

Roll's measure is intuitive and easy to compute. It provides accurate spread estimates with intraday data if a researcher has trade prices but not quotes (Schultz (2000)). However, even with a long time series of daily data, 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 such cases, researchers usually do one of three things: (1) treat the observation as missing, (2) set the Roll spread to zero, or (3) multiply the covariance by negative one, estimate the spread, and multiply the spread by negative one to produce a negative spread estimate.

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 is 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. Holden (2009) observes that, when a stock does not trade for a day, CRSP records the midpoint of its bid-ask range as its closing price. He proposes a revised version of the Roll estimator in which the covariance of price changes is divided by the percentage of days with trading. Hasbrouck (2004, 2009) 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 then uses information in the series of prices to assign a posterior probability that each specific trade is a buy or sell. Hasbrouck (2009) finds that Gibbs estimates of annual effective spreads are more accurate than spreads estimated with the basic Roll estimator, but that the procedure is "computationally intensive."

B. Spread Estimators Derived from Transaction Price Tick Size

The effective tick estimator, developed by Holden (2009) and Goyenko, Holden, and Trzcinka (2009), is based on the idea that wider spreads are associated with larger effective tick sizes. For example, their model assumes that, when both the tick size and the bid-ask spread are one-eighth, all possible prices are used, but when the tick size is one-eighth and the spread is one quarter, only prices ending on even-eighths 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 1990s, but the relation is much weaker for NYSE stocks.

Goyenko, Holden, and Trzcinka (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 prices observed at each price fraction is calculated as

$$F_j = \frac{N_j}{\sum_{j=1}^{J} N_j} \text{ for } j = 1, \ldots, J. \tag{20}$$

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

$$U_j = \begin{matrix} 2F_j & j = 1 \\ 2F_j - F_{j-1} & j = 2, \ldots, J-1 \\ F_j - F_{j-1} & j = J. \end{matrix} \tag{21}$$

The effective tick measure 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 10 prices were observed and 6 had odd-eighth price fractions, the unconstrained probability of a one-eighth spread would be 1.2. If 1 of the 10 prices had an odd-quarter fraction, the probability of a one-quarter spread would be 0.2 – 0.6 = –0.4. Holden (2009) 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 our examination of the effective tick estimator.[11,12]

C. Spread Estimators Derived from the Frequency of Zero Returns

Lesmond et al. (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 the observed return is only different from zero if the true return exceeds the costs of trading. With $\alpha_1 < 0$ denoting the cost of selling and $\alpha_2 > 0$ the cost of buying, the observed day t stock return R_t^o is

$$\begin{aligned} R_t^o &= \beta R_{mt} + \varepsilon_t - \alpha_1 && \text{if } R_t^A < \alpha_1 \\ R_t^o &= 0 && \text{if } \alpha_1 \le R_t^A \le \alpha_2 \\ R_t^o &= \beta R_{mt} + \varepsilon_t - \alpha_2 && \text{if } R_t^A > \alpha_2. \end{aligned} \tag{22}$$

Using this relation between trading costs and observed returns, Lesmond et al. (1999) estimate trading costs by maximizing the likelihood function for a year of daily stock returns with respect to α_1, α_2, β, and σ. The LOT estimate of the effective spread is then defined as $\alpha_2 - \alpha_1$.[13,14]

IV. Simulation Results for the High–Low Spread Estimator

To see how well the high–low spread estimator works under different conditions, we simulate 10,000 months of stock data. Each month contains 21 days and each day has 390 minutes. At the beginning of each month, the stock price is arbitrarily set to \$100. Then for each minute of each day, m, the true value

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

[12] Holden (2009) derives a spread estimator that nests both the Roll covariance spread estimator and the effective tick estimator as special cases. The estimator performs well, but is computationally intensive.

[13] Goyenko et al. (2009) propose two alternative methods to define the three regions of the likelihood function used to estimate the LOT measure. In the first method, referred to as LOT Mixed (as proposed in the original Lesmond et al. (1999) paper), the regions are defined based on both the stock return and the market return. In the second method, referred to as LOT Y-Split, the three regions are defined based on the stock return only. Goyenko et al. (2009) find that LOT Y-Split generally dominates LOT Mixed, and it is the LOT Y-Split that we use in this paper. As a robustness check, we reestimated the empirical results from Section V using the alternative LOT Mixed methodology. These results are provided in the Internet Appendix. Although the two measures can produce somewhat different estimates, our conclusions regarding the relative performance of the high–low spread and the LOT measure are robust to this estimation choice. The conclusions are also robust to the exclusion of outliers.

[14] LOT estimation requires the choice of a market return proxy. The results reported in this paper are based on the CRSP value-weighted index, whereas Lesmond et al. (1999) and Goyenko et al. (2009) use the CRSP equal weighted index. We reestimated the LOT Y-Split using the CRSP equal weighted index as the market return proxy and find that the results are unchanged.

of the stock price, P_m, is simulated as

$$P_m = P_{m-1}e^{\sigma x}, \tag{23}$$

where σ is the stock standard deviation per minute and x is a random drawing from a unit normal distribution. The bid price for each minute is obtained by multiplying P_m by one minus half the bid-ask spread and the ask price is obtained by multiplying P_m by one plus half the bid-ask spread. We assume a 50% chance that the observed price at minute m is a bid, and a 50% chance that it is an ask. The high and low for the day are the highest and lowest observed prices, respectively, whether a bid or ask. The closing price for the day is the observed price for minute 390.

A. The Distribution of Simulated Spread Estimates

We first examine the performance of the high–low spread estimator under the near-ideal conditions of no overnight return and prices observed every minute. For comparison, we also present simulated results for the Roll spread estimator. For these simulations, we assume that the true value for the first minute of the day is equal to the true value for the last minute of the previous day. We assume the daily standard deviation of returns is 3% and repeat the simulations for spreads of 0.5%, 1%, 3%, 5%, and 8%. The simulated monthly spread estimates are described in Panel A of Table I, which reports the mean and standard deviation of spread estimates across the 10,000 simulations, as well as the percentage of simulated monthly spreads that are nonpositive.

Column 1 reports simulation results for the "simple" closed-form high–low spread estimator defined in equations (14) and (18). To estimate monthly spreads, we estimate spreads separately for each 2-day period and calculate the average across all overlapping 2-day periods in the month. For this spread estimator, the mean estimate across 10,000 simulations is very close to the assumed spread regardless of the spread width. For example, the mean estimate from the simple high–low spread estimator is 7.84% when the true spread is 8% and is 2.92% when the true spread is 3%. Regardless of the size of the true spread, the standard deviation of the high–low spread estimates is around 0.62%. For spreads of 3% or more, none of the monthly high–low spread estimates are negative. However, for spreads of 0.5%, almost 20% of monthly high–low spread estimates are negative.

The next two columns report simulation results when ad hoc adjustments are made for negative spreads. The first adjustment, shown in column 2, sets all negative 2-day estimates to zero before taking the monthly average. Under these near-ideal conditions, with no overnight returns and almost continuous observation of prices, this adjustment produces spread estimates that are too large for spreads of 0.5% to 3%. For wider spreads, setting negative 2-day spreads to zero before taking monthly averages leads to a slight improvement in the average spread estimates. The second adjustment, shown in the following

Table I

The Distribution of Estimated Spreads for Alternative Spread Estimators

Each simulation consists of 10,000 21-day stock-months and each day consists of 390 minutes. For each minute of the day, the true value of the stock price, P_m, is simulated as $P_m = P_{m-1}e^{\sigma x}$, where σ is the standard deviation per minute and x is a random draw from a unit normal distribution. The daily standard deviation equals 3% and the standard deviation per minute equals 3% divided by $\sqrt{390}$. Stock prices are assumed to be observed each minute, with a 50% chance that a bid (ask) is observed. The bid (ask) for each minute is defined as P_m multiplied by one minus (plus) half the assumed bid-ask spread. Daily high and low prices equal the highest and lowest observed prices during the day. Monthly high–low spreads are estimated either by taking an average of daily high–low spread estimates within the month, or by using the average β and γ parameters within the month. Results are shown both with and without an adjustment for Jensen's inequality. Negative high–low spread estimates are either left unadjusted or adjusted using one of two methods: (1) setting negative 2-day spread estimates to zero before taking the monthly average, or (2) setting negative monthly spread estimates to zero. Roll spreads are calculated as $2\sqrt{-\text{Cov}}$, where Cov is the autocovariance of daily returns obtained from simulated closing prices. Roll Spread estimates in months with positive autocorrelations are set to zero. For each assumed spread level, Panel A reports the mean spread estimate, the standard deviation of spread estimates, and the proportion of spread estimates that are nonpositive across the 10,000 simulations. Panel B reports results from simulations incorporating overnight returns and infrequent observation of prices. In these simulations, we assume a 10% chance of observing a trade at any given minute. To simulate overnight returns, we assume that the standard deviation of close-to-open returns equals 0.5 times the standard deviation of open-to-close returns and adjust for overnight returns as described in Section II.

Panel A: Simulated Spread Estimates under Near-Ideal Conditions

Aggregation:		Average 2-Day Spread Estimates				Average Parameters		Roll Spreads
Jensen's Inequality:		No Adj.	No Adj.	No Adj.	Adj.	No Adj.	Adj.	
Negative Set to Zero:		No	Daily	Monthly	No	No	No	Monthly
0.5% Spread	Mean	0.52%	1.43%	0.59%	0.43%	0.23%	0.24%	1.18%
	σ	0.62%	0.33%	0.50%	0.67%	0.71%	0.73%	1.37%
	$\% \leq 0$	19.62%	0.00%	19.62%	25.10%	36.41%	35.93%	49.01%
1.0% Spread	Mean	0.99%	1.74%	1.01%	0.93%	0.71%	0.74%	1.31%
	σ	0.62%	0.37%	0.58%	0.66%	0.70%	0.72%	1.44%
	$\% \leq 0$	6.01%	0.00%	6.01%	8.25%	15.92%	15.65%	45.83%
3.0% Spread	Mean	2.92%	3.21%	2.92%	2.93%	2.68%	2.74%	2.61%
	σ	0.62%	0.50%	0.62%	0.64%	0.70%	0.70%	1.90%
	$\% \leq 0$	0.00%	0.00%	0.00%	0.00%	0.02%	0.02%	23.72%
5.0% Spread	Mean	4.88%	4.96%	4.88%	4.91%	4.67%	4.73%	4.54%
	σ	0.63%	0.58%	0.63%	0.63%	0.69%	0.68%	2.24%
	$\% \leq 0$	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	9.02%
8.0% Spread	Mean	7.84%	7.85%	7.84%	7.89%	7.67%	7.73%	7.54%
	σ	0.63%	0.63%	0.63%	0.63%	0.68%	0.67%	2.79%
	$\% \leq 0$	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	3.14%

(continued)

Table I—*Continued*

Panel B: Simulated Spread Estimates with an Overnight Return and Only 10% of Prices Observed

Aggregation:		Average 2-Day Spread Estimates				Average Parameters		Roll Spreads
Jensen's Inequality:		No Adj.	No Adj.	No Adj.	Adj.	No Adj.	Adj.	
Negative Set to Zero:		No	Daily	Monthly	No	No	No	Monthly
0.5% Spread	Mean	−0.24%	1.03%	0.15%	−0.39%	−0.52%	−0.55%	1.32%
	σ	0.65%	0.29%	0.28%	0.72%	0.74%	0.79%	1.53%
	% ≤ 0	64.08%	0.00%	64.08%	70.01%	75.74%	75.37%	49.22%
1.0% Spread	Mean	0.05%	1.23%	0.30%	−0.07%	−0.23%	−0.24%	1.43%
	σ	0.67%	0.32%	0.40%	0.74%	0.76%	0.80%	1.60%
	% ≤ 0	45.74%	0.00%	45.74%	52.62%	60.96%	60.45%	46.94%
3.0% Spread	Mean	1.76%	2.45%	1.76%	1.70%	1.48%	1.53%	2.01%
	σ	0.74%	0.47%	0.73%	0.78%	0.82%	0.84%	2.05%
	% ≤ 0	1.21%	0.00%	1.21%	2.12%	4.21%	4.15%	27.42%
5.0% Spread	Mean	3.69%	4.02%	3.69%	3.69%	3.45%	3.51%	4.48%
	σ	0.75%	0.59%	0.75%	0.77%	0.82%	0.82%	2.43%
	% ≤ 0	0.01%	0.00%	0.01%	0.01%	0.03%	0.03%	11.78%
8.0% Spread	Mean	6.65%	6.74%	6.65%	6.68%	6.45%	6.52%	7.49%
	σ	0.75%	0.69%	0.75%	0.76%	0.81%	0.80%	2.95%
	% ≤ 0	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	4.12%

column, includes negative 2-day spread estimates in the monthly average but sets negative monthly spread estimates to zero. Using this alternative adjustment produces spread estimates that are comparable to those from the simple unadjusted high–low spread estimator. Thus, under these near-ideal conditions, it is not worthwhile to adjust for negative estimates when examining means.

The fourth column of Panel A provides the performance of the high–low spread estimator with the adjustment for Jensen's inequality in equations (12), (13), and (14). The performance of this alternative version of the estimator is very similar to, but slightly worse than, the performance of the simple high–low spread estimator reported in column 1. Under these near-ideal conditions, there is little benefit from incorporating the adjustment for Jensen's inequality.

As noted above, the spread estimate is a nonlinear function of the β and γ parameters. As a result, averaging 2-day spread estimates can produce a biased estimate of the spread. To address the importance of this Jensen's inequality problem, the estimators in the fifth and sixth columns take a slightly different approach to aggregating daily high and low prices into a monthly estimate. Rather than averaging spread estimates across 2-day periods within the month, we average the β and γ parameters across 2-day periods within the month. The fifth column reports results when these monthly parameter averages are plugged into the simple high–low spread estimator, while column 6 reports results from the more complicated estimator that incorporates an adjustment for the other Jensen's inequality complication, namely that the square root of the high–low variance estimate is not an unbiased estimate of the standard deviation. As the table shows, averaging the β and γ parameters over the month produces less accurate spread estimates and a larger proportion of nonpositive spreads than when 2-day spread estimates are averaged over the month.

The last column provides simulation results for the Roll spread estimator. As described in Section III, we set the Roll spread to zero in cases in which the serial covariance is positive. The results show that the Roll estimator performs far worse than the high–low spread estimator, especially when spreads are narrow. When the true spread is 8%, the mean high–low spread estimate is 7.84% and the mean Roll estimate is 7.54%. More importantly, the standard deviation of spread estimates is 0.0063 for the simple high–low spread estimator, compared to 0.0279 for the Roll estimator. When the true spread is 0.5%, the mean high–low spread estimate is 0.59% and the mean Roll estimate is 1.18%. Here again, the Roll estimates have a much larger standard deviation (0.0137) than the high–low estimates (0.0062).

Under the near-ideal conditions of these simulations, the simple version of the high–low spread estimator appears to work best, without adjustments for negative spreads or Jensen's inequality. Monthly spread estimates are also more accurate when 2-day spread estimates are averaged within the month, as opposed to using average parameter estimates within the month. The results also suggest that the high–low spread estimator performs significantly better than the Roll covariance estimator.

There are several ways in which the above simulation assumptions depart from market realities. We next examine two important complications that may affect the performance of the spread estimator. First, overnight returns affect the high–low ratio for a 2-day period, but do not affect the high–low ratios for either of the single days. This leads to an underestimate of spreads. Second, with infrequently observed prices, the observed high and low price may not reflect the true high and low, leading to a misestimation of the true spread.

We simulate infrequent observation of prices by assuming there is a 10% chance of observing a price at any given minute. This corresponds to an average of 39 trades per day—a realistic assumption for most NASDAQ and smaller NYSE stocks. We simulate overnight returns that are normally distributed with zero mean and standard deviation equal to 0.5 times the open-to-close standard deviation of returns.[15] We then adjust for the effects of overnight returns based on the method described in Section II.

Simulation results incorporating both overnight returns and infrequently observed prices are described in Panel B of Table I. Even after incorporating an adjustment for overnight returns, mean spread estimates decline for all versions of the high–low spread estimator. The simple high–low spread estimator provides mean spread estimates of 6.65% when the true spread is 8%,

[15] Lockwood and Linn (1990) (Table I) estimate the average to be 0.569 for the Dow Jones Industrials over 1964 to 1989. Oldfield and Rogalski (1980) (Table I) provide data that allow estimation of the ratio for five large individual stocks for October 1974 to December 1977. The average ratio across the five stocks is 0.502.

3.69% when the true spread is 5%, 0.05% when the true spread is 1%, and –0.24% when the true spread is 0.50%. In contrast to Panel A, when prices are observed infrequently and there are overnight returns, ad hoc adjustments for negative spread estimates make the estimates more accurate. For spreads of 1% or greater, setting negative 2-day spreads to zero before taking monthly averages produces mean estimates that are much closer to the assumed spreads than the simple high–low spread estimator with no adjustment for negative spreads. Setting negative 2-day spreads to zero before taking monthly averages also results in a much smaller standard deviation of spread estimates than does the unadjusted estimator. As in Panel A, adjusting for negative spreads at the daily level appears to work better than adjusting for negative monthly spreads, and there appears to be little benefit to either incorporating an adjustment for Jensen's inequality or taking average parameter estimates rather than average 2-day spreads within the month.

The last column of the table shows that overnight returns and infrequently observed prices have little impact on the Roll spread estimator. In Panel B, mean spread estimates from the Roll estimator are slightly closer to true spreads than are high–low spread estimates for spreads of 3% or greater. However, the high–low spread estimator with negative 2-day spreads set to zero provides better mean estimates than the Roll estimator for spreads of 0.5% or 1%. More importantly, the standard deviation of spread estimates from the high–low spread estimator with negative 2-day spreads set to zero is only one-half to one-fourth as large as the standard deviation of Roll spread estimates. Even under these unfavorable conditions, the high–low spread estimator appears to outperform the Roll estimator.

B. The Cross-sectional Correlation of Simulated Spread Estimates

The simulations described in Table I illustrate the accuracy of the high–low spread estimator. Next, we examine how well the different versions of the estimator capture the cross-section of spreads under alternative assumptions about prices and spreads. Our simulations again consist of 10,000 stock-months with 21 days in each month and 390 minutes in each day, with stock prices simulated as in equation (23). However, for each of the 10,000 stock months, we randomly assign a spread from a uniform distribution with a range from 0% to 6%. As in Table I, there is a 50% chance that an observed price is a bid and a 50% chance that it is an ask.

Table II reports the correlations of spread estimates with simulated true spreads across the 10,000 stock months, where the standard deviation of daily returns is set to either 3% or 5%. The first two rows report simulations under the near-ideal conditions of no overnight return and prices observed each minute. Under these conditions, correlations between high–low spread estimates and simulated spreads are high for all versions of the high–low spread estimator, ranging from 0.926 to 0.940 when the daily return standard deviation is 3%. When the standard deviation of daily returns is 5%, the correlation between high–low spread estimates and simulated spreads ranges from 0.822 when β

Table II

Correlations between Estimated and Simulated Spreads

Each simulation consists of 10,000 21-day stock-months and each day consists of 390 minutes. Each minute, the true value of the stock price, P_m, is simulated as $P_m = P_{m-1}e^{\sigma x}$, where σ is the standard deviation per minute and x is a random draw from a unit normal distribution. When a stock price is observed, we assume a 50% chance that a bid (ask) is observed. The bid (ask) for each minute is defined as P_m multiplied by one minus (plus) half the assumed bid-ask spread. Daily high and low prices equal the highest and lowest observed prices during the day. Monthly high–low spreads are estimated by taking an average of daily high–low spread estimates within the month, or by using the average β and γ parameters within the month. Results are shown both with and without an adjustment for Jensen's inequality. Negative high–low spread estimates are either left unadjusted or adjusted using one of two methods: (1) setting negative 2-day spread estimates to zero before taking the monthly average, or (2) setting negative monthly spread estimates to zero. Roll spreads are calculated as $2\sqrt{-\text{Cov}}$, where Cov is the covariance of daily returns obtained from simulated closing prices. Roll spread estimates in months with positive autocorrelations are either defined as negative spreads or set to zero. In Panel A, the true spread for each stock-month is drawn from a uniform distribution with a range from 0% to 6%. We then perform simulations under various assumptions about prices and spreads. When an overnight return is incorporated, we assume that the standard deviation of close-to-open returns equals 0.5 times the standard deviation of open-to-close returns and adjust for overnight returns as described in Section II. When prices are observed infrequently, we assume a 10% chance of observing a price in any given minute. In simulations with autocorrelated returns, we assume that the innovation to the daily-expected return is normally distributed with a standard deviation of 1% per day. The daily expected return is then defined as the sum of the innovation plus 0.5 times the previous day's expected return, and the expected return for each 1-minute interval is defined as the daily expected return divided by 390. In simulations with random spreads, the initial spread is drawn from a uniform distribution ranging from 0% to 6%. Each day's spread is then obtained by multiplying the previous day's spread by e^{δ}, where δ is normally distributed with zero mean and standard deviation equal to 0.1. In this case, the simulated monthly spread is defined as the average spread across the 21 days within the month. In Panel B, the true spread for each stock-month is drawn from a uniform distribution with a specified range. These simulations incorporate both overnight returns and infrequent observation of prices.

Daily Standard Deviation	Aggregation:	Average 2-Day Spread Estimates				Average Parameters			
	Jensen's inequality:	No Adj.	No Adj.	No Adj.	Adjusted	No Adj.	Adjusted	Roll Spreads	
	Negative Set to 0:	No	Daily	Monthly	No	No	No	Monthly	No
No Overnight Returns, Prices Observed Each Minute, Spreads Constant over 21 Days, Returns Not Autocorrelated:									
3%		0.937	0.940	0.928	0.937	0.926	0.926	0.573	0.524
5%		0.848	0.865	0.825	0.847	0.822	0.822	0.338	0.297
With Overnight Return:									
3%		0.912	0.925	0.899	0.911	0.898	0.898	0.523	0.470
5%		0.788	0.835	0.753	0.784	0.755	0.755	0.290	0.248
With Overnight Return and Prices Observed for Only 10% of Minutes:									
3%		0.902	0.922	0.880	0.899	0.886	0.886	0.523	0.470
5%		0.757	0.828	0.697	0.749	0.720	0.719	0.289	0.248
With Overnight Return, Prices Observed for Only 10% of Minutes, and Autocorrelated Daily Returns:									
3%		0.890	0.914	0.859	0.886	0.870	0.869	0.504	0.449
5%		0.747	0.821	0.682	0.738	0.708	0.707	0.281	0.240
With Overnight Return, Prices Observed for Only 10% of Minutes, Autocorrelated Daily Returns, and Spreads That Follow a Random Walk:									
3%		0.923	0.941	0.908	0.919	0.910	0.910	0.602	0.531
5%		0.815	0.875	0.789	0.806	0.789	0.787	0.381	0.319

and γ are averaged over the month to 0.865 when spreads are calculated for 2-day periods and averaged over the month with negative 2-day spreads set to zero.

The correlations between Roll spread estimates and simulated spreads are much lower. When Roll spreads are set to zero for positive serial correlations, the correlation is 0.573 for a daily return standard deviation of 3% and 0.338 for a standard deviation of 5%. However, while setting Roll spreads to zero is the common ad hoc adjustment when the autocovariance is positive, it is not clear a priori whether this adjustment increases or decreases the correlation between Roll spreads and simulated spreads. As an alternative, the last column reports correlations between Roll spreads and simulated spreads when positive serial correlations are treated as negative spreads. Here, the correlation between Roll spreads and simulated spreads drops to 0.524 when the standard deviation is 3% and to 0.297 when the standard deviation is 5%.

The next two rows of the table report correlations from simulations that incorporate overnight returns and the overnight return adjustment described in Table I. As in Table I, the standard deviation of close-to-open returns is assumed to be 0.5 times the open-to-close return standard deviation. With overnight returns, correlations decline slightly for all versions of the high–low spread estimator. However, correlations remain highest for the simple version of the estimator in which negative 2-day spreads are set to zero before taking the monthly average. For this version of the estimator, the correlation falls from 0.940 to 0.925 for a standard deviation of 3% and from 0.865 to 0.835 for a standard deviation of 5%.

The middle two rows of Table II provide correlations between spread estimates and simulated spreads incorporating both overnight returns and infrequently observed prices. As in Table I, we assume there is a 10% chance of observing a price at any specific minute. Incorporating infrequent observation of prices reduces correlations slightly for all versions of the high–low spread estimator. As in Table I, however, infrequent observation of prices has little impact on the Roll spread estimator. Again, the correlations suggest that all versions of the high–low spread estimator dominate the Roll spread estimator, and the simple version of the high–low estimator in which negative 2-day estimates are set to zero outperforms other versions of the high–low spread estimator. This version of the high–low spread estimator produces a correlation of 0.922 for a daily return standard deviation of 3% and 0.828 for a standard deviation of 5%.

The next two rows report correlations when daily returns are positively autocorrelated. Specifically, we assume that the innovation to the expected return is normally distributed with a standard deviation of 1% per day. The daily expected return is then defined as the sum of the innovation plus 0.5 times the previous day's expected return, and the expected return for each 1-minute time interval is the daily expected return divided by 390. When returns are positively autocorrelated, the correlation with simulated spreads declines for every version of the high–low spread estimator but remains high. Again, the version of the high–low spread estimator in which negative 2-day spread estimates are set to zero produces the highest correlation and significantly outperforms the Roll estimator. For this version of the high–low spread estimator, the correlation between high–low spread estimates and simulated spreads is 0.914 for a daily standard deviation of 3% and 0.821 for a standard deviation of 5%.

Finally, the last two rows of Table II report correlations under the assumption that spreads themselves change randomly. In these simulations, we continue to assume that stock prices change overnight, that there is only a 10% chance that a price is observed at a particular minute, and that daily returns are positively autocorrelated. For each stock, an initial spread is drawn randomly from a uniform distribution over 0% to 6%. Next, each day's spread is obtained by multiplying the previous day's spread by e^{δ}, where δ is normally distributed with zero mean and standard deviation equal to 0.1. The simulated monthly spread is then defined as the average simulated spread across the 21 days within the month. Interestingly, the correlations between spread estimates and mean simulated spreads are higher for all estimators than the correlations when spreads are assumed to be constant within the month. This may reflect the fact that the range of mean spreads in these simulations is wider than the original 0% to 6%. More important, however, the relative performance of the spread estimators is unaffected by allowing spreads to vary over time.

The simulation results presented in Tables I and II provide several key findings. First, the simple high–low spread estimator in which 2-day spread estimates are averaged over the month works well and is far more accurate than the Roll spread. Second, adjusting for Jensen's inequality complicates the estimation but doesn't improve the accuracy of the estimates. Third, estimating spreads over 2-day periods and averaging the spreads over a month works better that averaging the parameters and calculating a single spread for the month. Finally, setting negative 2-day spread estimates to zero before calculating monthly spreads improves estimates, particularly when returns are generated overnight and stocks do not trade continuously. Throughout the remaining empirical analysis, we present results based on the simple high–low spread estimator, where monthly spreads are based on an average of 2-day spread estimates after setting negative 2-day estimates to zero.

V. A Comparison of Spread Estimates from Daily Data with TAQ Spreads

In this section, we compare the performance of monthly high–low spread estimates to estimates generated by three common alternative spread estimators: the Roll spread estimator, the effective tick estimator, and the LOT estimator. These alternatives provide estimates based on the autocovariance of returns, the price fraction of trade prices, and the frequency of zero returns, respectively. We focus on these specific estimators because they have been used as building blocks for other, more complex methods of estimating spreads (see, for example, Holden (2009)). Goyenko et al. (2009) provide a comprehensive study of the properties of these estimators and other estimators derived from them. Given its simplicity and accuracy, we believe that the high–low spread estimator may also serve as a foundation for more complex estimation techniques.

Monthly Roll, effective tick, LOT, and high–low spread estimates are calculated using daily data from CRSP for the period from January 1993 through December 2006. For each spread estimator, we require at least 12 daily observations to calculate a monthly spread estimate. The CRSP sample includes all NYSE, Amex, and NASDAQ stocks with CRSP share codes of 10 or 11 (i.e., U.S. common shares).

To assess the performance of these monthly spread measures, we compare them to trade-weighted effective spreads estimated for each security each month using NYSE TAQ data.[16] 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.[17] At any time t, let Bid_t equal the inside bid, Ask_t equal the inside ask, and Midpoint_t equal $(\text{Bid}_t + \text{Ask}_t)/2$. To estimate effective spreads, we compare each trade price during the day to the inside bid and ask posted one second prior to the trade. 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 $2^*|P_I - \text{Midpoint}_I|/\text{Midpoint}_I$. 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.

Any comparison of alternative spread estimators must be qualified, because the estimators may capture different components of liquidity or transitory volatility. As noted earlier, the high–low estimator captures transitory volatility over 2 trading days, which may include temporary price pressure from large orders in addition to bid-ask spreads. At the daily frequency, we expect the estimator to closely approximate effective spreads. We therefore test the

[16] To match securities in the CRSP data to securities in the TAQ data, we first identify all unique CUSIP-ticker combinations in both the TAQ and CRSP data sets from 1993 through 2006. 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 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 needed. Finally, we hand verify any CRSP-TAQ matches for which the number of daily observations in the two data sets differs by more than 10 days.

[17] 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 include only regular NBBO-eligible quotes with positive prices and positive depth. 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 use 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 the 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.

Table III

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 from 1993 through 2006. The sample includes all NYSE, Amex, and NASDAQ listed securities with at least 6 months of data and for which TAQ and CRSP data could be matched. Monthly observations are dropped if based on fewer than 12 daily observations or if spread estimates are missing for the Roll Spread, Tick Spread, LOT Measure, or HL Spread. Effective Spread is the trade-weighted percentage effective spread estimated from TAQ and averaged across days within the month. The Roll Spread is two times the square root of –1 times the autocovariance of daily returns. The Effective Tick Spread assumes that the spread is a function of the tick increment used in trade prices. The effective tick spread for the month is based on the average of the spreads implied by daily trade prices. HL Spread is the equally weighted average of the high–low spread estimator across all overlapping 2-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 2-day spreads to zero, (2) leave negative 2-day spreads unchanged, and (3) exclude negative 2-day spreads. Similarly, results for the Roll spread are provided using two alternative methods to handle positive covariances: (1) set spreads to zero when the covariance is positive and (2) exclude spreads when the covariance is positive.

	N	Mean (%)	Median (%)	Standard Deviation (%)	% $\leq$0
Effective Spread	973,229	2.38	1.29	3.37	0.00
Roll Spread$_{\text{Neg=0}}$	973,229	2.42	1.21	3.88	38.04
Roll Spread$_{\text{NegDropped}}$	603,032	3.90	2.66	4.30	0.00
Eff. Tick Spread	973,229	1.67	0.72	3.39	0.00
Lot Measure	973,229	2.15	0.86	4.94	24.43
HL Spread$_{\text{Neg=0}}$	973,229	2.10	1.32	2.65	0.00
HL Spread$_{\text{NegIncluded}}$	973,229	1.26	0.56	2.62	24.00
HL Spread$_{\text{NegDropped}}$	973,227	2.76	1.91	2.97	0.00

performance of the high–low estimator at capturing effective bid-ask spreads as measured by intraday TAQ data and compare this performance to that of several alternative low-frequency spread estimators. We note, however, that one inherent benefit of the high–low estimator is that it may capture other forms of transitory volatility, and therefore liquidity costs, that are not reflected in the effective spread.[18]

A. Summary Statistics

Table III provides summary statistics for spread estimates using the pooled sample of all stocks and all months from 1993 through 2006 for which all

[18] To examine the relation between high–low spreads and more general measures of liquidity, we provide results based on the Amihud illiquidity measure, defined as the average ratio of absolute return to dollar volume across all trading days during the month. As expected, the high–low spread measure has a higher correlation with the Amihud measure than any of the other spread estimators. The correlation between the high–low spread and the Amihud measure is 0.359, compared to correlations of 0.297, 0.278, and 0.268 for the Roll spread, effective tick spread, and LOT measure, respectively. These results are provided in the Internet Appendix.

four spread estimators could be calculated. For comparison purposes, data on effective spreads from TAQ are presented first. The simple average effective spread from TAQ across all stock-months is 2.38%.

Roll spread estimates are reported next. For the full sample of stocks over 1993 to 2006, positive monthly serial correlation estimates occur for 38.0% of the stock-months. We adopt the common ad hoc adjustment of setting Roll spreads to zero in these cases. This yields a mean Roll spread of 2.42%, which is very close to the mean TAQ effective spread. If the positive correlations are instead omitted, more than a third of the observations are lost and the mean Roll spread is 3.90%, much greater than the mean effective spread from TAQ. In the analysis to follow, we use the version of the Roll spread estimator in which positive autocorrelations are defined as zero spreads.

Spread estimates obtained from the effective tick estimator and the LOT estimator are presented next. By construction, effective tick estimates are always positive. The mean effective tick spread is 1.67% and the median is 0.72%, both of which are less than the comparable effective spread estimate from TAQ. The mean and median LOT estimates are 2.15% and 0.86%, with 24.4% of monthly LOT estimates being nonpositive.[19]

Results for three versions of the high–low spread estimator are reported next. The first high–low spread estimator sets all negative 2-day spread estimates to zero before calculating the monthly average. This high–low spread estimator produces a mean spread of 2.10%, compared to the mean TAQ effective spread of 2.60%. The median spread estimate from this version of the high–low spread estimator is 1.32%, which is very close to the median TAQ effective spread of 1.29%. When negative spreads are included, the mean high–low spread estimate equals 1.26%, with 24.0% of monthly spread estimates being nonpositive. When negative 2-day spreads are omitted, the mean high–low spread increases to 2.76%. These findings are consistent with the simulation results and suggest that the high–low spread estimator performs best when negative 2-day estimates are set to zero before taking the monthly average. The results throughout the remainder of the paper are therefore based on the simple version of the high–low spread estimator in which negative 2-day spreads are set to zero.

B. Cross-sectional Comparisons of Spread Estimates with TAQ Effective Spreads

To analyze the performance of the high–low spread estimator, we first calculate the cross-sectional correlation between spread estimates and the TAQ

[19] The frequency of zero estimates is sensitive to the LOT estimation method. As discussed earlier, we report results based on LOT Y-Split. In robustness tests reported in the Internet Appendix, we find that LOT Mixed produces a significantly lower frequency of zero estimates. In our full sample, the frequency of zero estimates based on the LOT Mixed estimator is 4.4%. As noted earlier, the conclusions regarding the relative performance of the high–low spread and the LOT measure are unaffected by the choice of LOT estimation method. See Goyenko et al. (2009) for a description of these alternative estimation methods.

effective spread each month from 1993 through 2006. 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 examine 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 to 1996, 1997 to 2000, and 2001 to 2006. These subperiods correspond roughly to the periods in which the regulatory minimum tick size and quoted spread were an eighth of a dollar, a sixteenth of a dollar, and one cent.[20]

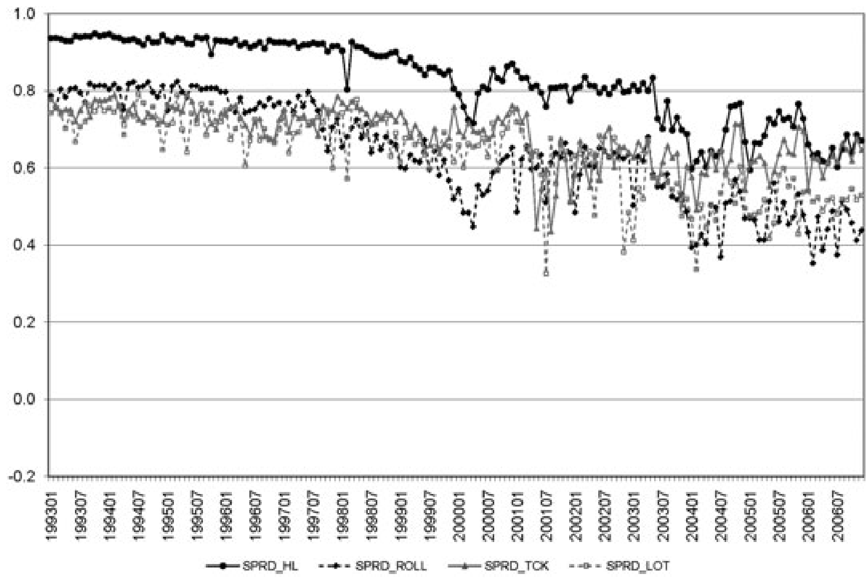
An important use for all of the low frequency spread estimators is to estimate trading costs for periods before intraday data were available. Thus, we are particularly concerned with how the spread estimators perform during the 1993 to 1996 subperiod. During this period, the tick size in U.S. markets was $0.125, just as it was during earlier periods. Hence, the performance of spread estimators over 1993 to 1996 is likely to be a better predictor of their performance for earlier periods than is their performance during either 1997 to 2000 or 2001 to 2006.

Panel A of Table IV reports time-series means of the monthly cross-sectional correlations. For the entire period and for each subperiod, the high–low spread estimator produces higher cross-sectional correlations with TAQ effective spreads than any of the alternative spread estimators. For the entire period, the mean cross-sectional correlation of TAQ effective spreads with high–low spreads is 0.829, compared to a correlation of 0.637 for the Roll spread, 0.683 for the effective tick spread, and 0.635 for the LOT measure. Based on these results, the high–low spread estimator appears to dominate the alternative spread estimators at capturing the cross-section of TAQ effective spreads. At 0.930, the cross-sectional correlation of high–low spread estimates with the TAQ effective spread is particularly high during the 1993 to 1996 subperiod. This suggests that the estimator should work well for earlier periods.

The month-by-month cross-sectional correlations between the various spread estimators and TAQ effective spreads are plotted in Panel A of Figure 1. As the figure shows, the cross-sectional correlation between high–low spread estimates and TAQ effective spreads is consistently higher than the correlations based on the Roll spread, the effective tick spread, or the LOT measure. Again, the correlations based on the high–low spread estimates are particularly high during the 1993 to 1996 period. The Roll spread slightly outperforms the effective tick spread and the LOT measure based on cross-sectional correlations

[20] 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 to 1996 period as all months from January 1993 through May 1996 and the 1997 to 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.

Panel A. Correlations Based on Percentage Spreads



Panel B. Correlations Based on First Differences in Percentage Spreads

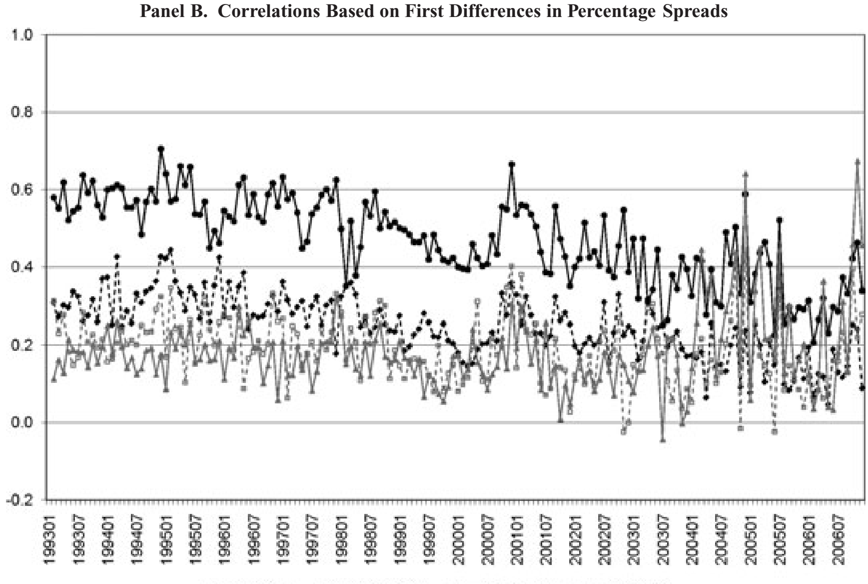


Figure 1. Cross-sectional correlations of spread estimates with TAQ effective spreads by month. The figure plots monthly cross-sectional correlations between three estimated spread measures and the effective spread from TAQ. The correlations shown in Panel A are estimated from monthly spread estimates. The correlations shown in Panel B are estimated from first differences in monthly spread estimates. The sample includes all NYSE, Amex, and NASDAQ listed securities with at least 6 months of data and for which TAQ and CRSP data could be matched. Monthly observations are dropped if based on fewer than 12 daily observations or if spread estimates are missing for the Roll Spread, Tick Spread, LOT Measure, or HL Spread (see Table III for definitions).

Table IV

Average Cross-sectional Correlations

For each spread measure and each month from 1993 through 2006, 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 sample includes all NYSE, Amex, and NASDAQ listed securities with at least 6 months of data and for which TAQ and CRSP data could be matched. Monthly observations are dropped if based on fewer than 12 daily observations or if spread estimates are missing for the Roll Spread, Tick Spread, LOT Measure, or HL Spread (Table III for definitions). 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	Lot Measure	High–Low Spread
Panel A: Correlations with Effective Spread, Monthly Estimates					
Full Period	168	0.637	0.683	0.635	0.829
1993–1996	53	0.789	0.741	0.718	0.930
1997–2000	43	0.632	0.718	0.680	0.862
2001–2006	72	0.528	0.619	0.546	0.732
Panel B: Correlations with Changes in Effective Spreads, Monthly Estimates					
Full Period	167	0.249	0.183	0.186	0.472
1993–1996	52	0.320	0.177	0.221	0.570
1997–2000	43	0.249	0.162	0.191	0.446
2001–2006	72	0.199	0.147	0.158	0.391

during 1993 to 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 cross-sectional correlations between month-to-month changes in spread estimates and changes in TAQ effective spreads. Results are shown in Panel B of Table IV. Not surprisingly, correlations based on changes in spreads are lower than those based on spread levels. Still, the high–low spread estimator does a far better job of explaining the cross-section of changes in TAQ effective spreads than any of the alternative measures. For the entire sample period, the mean cross-sectional correlation between changes in high–low spreads and changes in TAQ effective spreads is 0.472. The comparable correlations for the Roll spread, the effective tick spread, and the LOT measure are 0.249, 0.183, and 0.186, respectively. As in Panel A, the high–low spread estimator performs best during the 1993 to 1996 subperiod, with an average cross-sectional correlation of 0.570.

The month-by-month correlations between changes in TAQ spreads and changes in estimated spreads are plotted in Panel B of Figure 1. The correlations are consistently higher for the high–low spread estimator than any of the alternative spread estimators. The effective tick estimator and LOT measure perform particularly poorly in capturing the cross-section of month-to-month changes in spreads, especially during the earlier part of the sample period.

Taken together, the results in Table IV and Figure 1 provide clear evidence that the high–low spread estimator dominates the alternative estimators at capturing the cross-section of both spread levels and changes in spreads. As noted above, the high–low spread estimator performs particularly well during the 1993 to 1996 subperiod, suggesting that it should work well for earlier time periods.

C. 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 spreads. We summarize the time-series correlations across all stocks, by exchange, and by market capitalization quintile. Quintile breakpoints are based on NYSE stock capitalizations, so the smaller size quintiles have a disproportionate number of NASDAQ and Amex stocks. The results provided in the table are based on the exchange listing and size quintiles of each stock as of its last listing date on CRSP.

The time-series correlations are summarized in Table V. Panels A, B, and C report results for the three tick-size subperiods, while Panel D reports results for the full sample period. Time-series correlations between each of the spread estimates and effective spreads are lower than pooled estimates. This is not surprising as there is a lot of variation in spreads across securities. One clear result that emerges from Table V is that the high–low spread estimator and the effective tick estimator dominate both the Roll spread and the LOT measure in explaining time-series variation in the spreads of individual stocks. Estimates from the Roll spread and the LOT measure produce the lowest time-series correlations for the full sample of stocks, for stocks on each of the exchanges, and for stocks in all size quintiles.

A second clear result is that the high–low spread estimator outperforms the effective tick estimator on average because it does a better job with the spreads of smaller stocks. For each of the three subperiods, the high–low spread estimator has significantly higher time-series correlations with TAQ effective spreads for NASDAQ and Amex stocks than does the effective tick estimator. However, for NYSE stocks, the correlation for the high–low spread estimator is slightly lower than that for the effective tick estimator in the first two subperiods, and substantially lower in the 2001 to 2006 subperiod. Turning to size quintiles, we see that the high–low spread estimator exhibits higher time-series correlations than the effective tick estimator for the first three quintiles in the 1993 to 1996 subperiod, for the first two quintiles in the 1997 to 2000 subperiod, and only for the first quintile in the 2001 to 2006 subperiod. Notably, these quintiles contain the majority of the sample stocks. In the 1993 to 1996 subperiod, for example, the first quintile includes approximately two-thirds of the sample stocks and the first three quintiles include approximately 87% of the sample stocks. For the largest firms, however, the effective tick estimator produces higher

Table V

Summary Statistics for Stock-by-Stock Time Series Correlations

For each 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 sample includes all NYSE, Amex, and NASDAQ listed securities with at least 6 months of data and for which TAQ and CRSP data could be matched. Monthly observations are dropped if based on fewer than 12 daily observations or if spread estimates are missing for the Roll Spread, Tick Spread, LOT Measure, or HL Spread (Table III for definitions). Panels A, B, and C provide results for subperiods corresponding to different tick size regimes. Panel D provides results for the full sample period. Stocks are also separated by exchange and market capitalization quintile based on the CRSP exchange code and market capitalization on the last date the firm is listed on CRSP. Market capitalization quintiles are based on NYSE breakpoints.

	N	Roll Spread	Effective Tick Spread	Lot Measure	High–Low Spread
Panel A: Correlations with Effective Spread (1993–1996)					
Full Sample	9,056	0.348	0.476	0.338	0.649
NYSE	2,360	0.199	0.506	0.268	0.495
Amex	757	0.313	0.436	0.313	0.602
NASDAQ	5,939	0.412	0.469	0.368	0.716
MV Quintile 1	5,968	0.408	0.456	0.372	0.700
MV Quintile 2	1,107	0.307	0.495	0.275	0.652
MV Quintile 3	749	0.240	0.514	0.280	0.586
MV Quintile 4	644	0.172	0.523	0.255	0.497
MV Quintile 5	497	0.106	0.567	0.253	0.338
Panel B: Correlations with Effective Spread (1997–2000)					
Full Sample	9,349	0.279	0.507	0.312	0.574
NYSE	2,317	0.163	0.516	0.246	0.496
Amex	771	0.321	0.507	0.392	0.628
NASDAQ	6,261	0.317	0.503	0.327	0.596
MV Quintile 1	6,311	0.344	0.509	0.361	0.639
MV Quintile 2	1,058	0.186	0.492	0.220	0.525
MV Quintile 3	760	0.139	0.510	0.211	0.441
MV Quintile 4	658	0.106	0.507	0.193	0.382
MV Quintile 5	488	0.081	0.512	0.191	0.305
Panel C: Correlations with Effective Spread (2001–2006)					
Full Sample	7,427	0.273	0.515	0.288	0.564
NYSE	1,898	0.160	0.592	0.221	0.470
Amex	718	0.306	0.506	0.387	0.579
NASDAQ	4,811	0.313	0.487	0.299	0.599
MV Quintile 1	5,056	0.312	0.466	0.322	0.582
MV Quintile 2	819	0.181	0.599	0.257	0.474
MV Quintile 3	598	0.183	0.631	0.229	0.490
MV Quintile 4	508	0.188	0.648	0.176	0.567
MV Quintile 5	408	0.208	0.626	0.133	0.616

(*continued*)

Table VI—*Continued*

	N	Roll Spread	Effective Tick Spread	Lot Measure	High–Low Spread
Panel D: Correlations with Effective Spread (1993–2006)					
Full Sample	12,507	0.343	0.599	0.443	0.626
NYSE	2,894	0.159	0.704	0.441	0.412
Amex	1,118	0.381	0.547	0.441	0.667
NASDAQ	8,495	0.400	0.571	0.444	0.693
MV Quintile 1	8,559	0.418	0.550	0.451	0.702
MV Quintile 2	1,481	0.247	0.662	0.405	0.581
MV Quintile 3	977	0.173	0.719	0.437	0.480
MV Quintile 4	804	0.122	0.746	0.431	0.385
MV Quintile 5	573	0.050	0.789	0.444	0.182

time-series correlations with TAQ effective spreads than does the high–low spread estimator. The high–low spread estimator has trouble with the largest stocks in part because their trading costs are low, resulting in a small signal-to-noise ratio. In contrast, the effective tick estimator works well for the largest stocks because this estimator is similar to simply dividing the minimum tick size by the stock price. This works well when the tick size places a binding lower bound on the spread width.

The subperiod results in Panels A, B, and C suggest that the high–low estimator dominates the Roll and LOT estimators in capturing time-series variation in spreads. It does better than the effective tick estimator for most stocks, but not for the largest stocks. In addition, the estimator performs particularly well in the 1993 to 1996 subperiod. In Panel D, we present results for the full sample period. On average, the high–low estimator continues to dominate the alternative estimators. However, the variation in performance between small and large stocks is magnified in the full sample period results. The average time-series correlation between high–low spreads and TAQ effective spreads drops from 0.702 for the smallest size quintile to 0.182 for the largest quintile. In contrast, the average correlation based on the effective tick estimator increases from 0.550 for the smallest quintile to 0.789 for the largest quintile. In addition, the average correlation based on the LOT measure averages between 0.40 and 0.45 for all categories of stocks. When combined with the subperiod results, these findings suggest that the improved performance of the effective tick estimator and the LOT measure for large stocks in the full sample period may be driven by the ability of these estimators to capture changes in the minimum tick size and the overall downward trend in spreads over this time period. During historical periods with a constant tick size, however, we expect the high–low estimator to dominate for all but the largest stocks.

D. Summary of the Estimator Comparisons

In pooled and cross-sectional analyses, the high–low spread estimator dominates the Roll spread estimator, the effective tick estimator, and the LOT measure. It has significantly higher cross-sectional correlations with TAQ effective spreads and with month-to-month changes in TAQ effective spreads than the alternative estimators. The high–low spread estimator does particularly well during the period from 1993 through 1996, when the minimum tick size was $0.125. These results suggest that the high–low spread estimator may be superior to other estimators for historical analyses.[21]

In stock-by-stock time-series analyses, we again find that the high–low spread estimator dominates the Roll spread estimator and the LOT measure. The effective tick estimator appears to work well for the very largest stocks, where the tick size provides a binding lower bound on the spread. However, for the vast majority of stocks, the high–low spread estimator is superior to the effective tick estimator, especially during the 1993 to 1996 subperiod when the tick size was one-eighth.

VI. Example Applications of the High–Low Spread Estimator

To demonstrate the potential uses of the high–low spread estimator, we provide several example applications. The first is a description of historical spreads for NYSE and Amex stocks from 1926 through 2006. The second is an illustration of the potential use of the estimator in asset pricing tests. The third is an application to non-U.S. markets using data from Datastream.

A. Estimating Historical Spreads for U.S. Stocks Using Daily CRSP Data

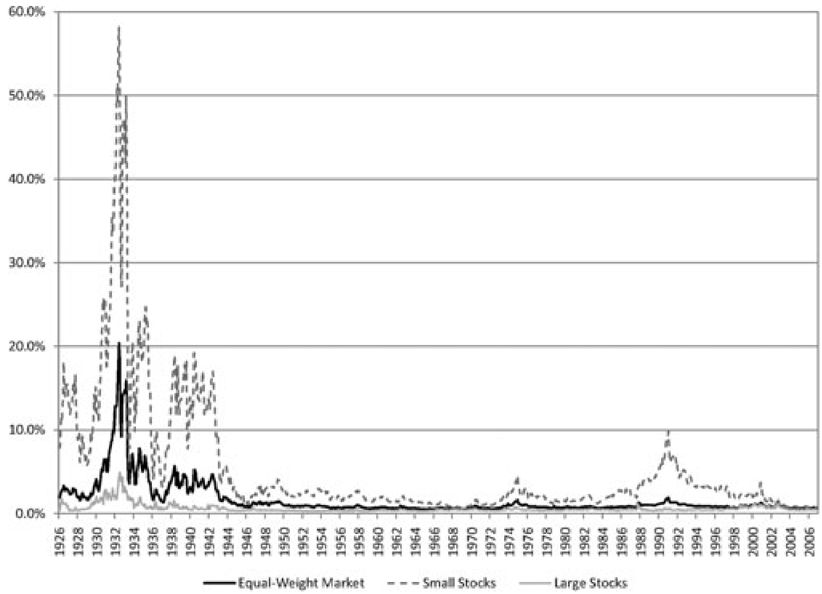
Using high and low price data from CRSP, we calculate bid-ask spreads based on the high–low estimator for each NYSE/Amex stock each month from 1926 to 2006. As before, monthly spreads are defined as the average of all 2-day spreads within the calendar month, negative 2-day spread estimates are set to zero, and we require a minimum of 12 daily price ranges to calculate a monthly spread.[22] The results are illustrated in Figure 2.

Panel A of Figure 2 plots the cross-sectional average of high–low spread estimates for NYSE/Amex stocks each month from 1926 through 2006. Results are shown for the full sample of NYSE/Amex 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

[21] We also calculate mean absolute errors based on the difference between monthly spread estimates for each of the low-frequency estimators and monthly effective spreads from TAQ. Across all sample months, the cross-sectional mean absolute error averages 0.0090 for the high-low spread estimator, compared to 0.0166 for the Roll spread, 0.0113 for the effective tick spread, and 0.0132 for the LOT measure. Results based on month-to-month changes in spreads provide similar conclusions. These results suggest that the high–low spread estimator is more accurate than the three alternative estimators in capturing both the level of spreads and changes in spreads. These results are provided in the Internet Appendix.

[22] As noted in Section II, CRSP provides closing bid and ask prices in place of low and high prices on those days when a stock does not trade. In this section, we make use of this information and include these prices when calculating the high-low spread estimator.

Panel A. Mean Monthly Spread Estimates from 1926 to 2006



Panel B. Mean Monthly Spread Estimates from 1950 to 2006

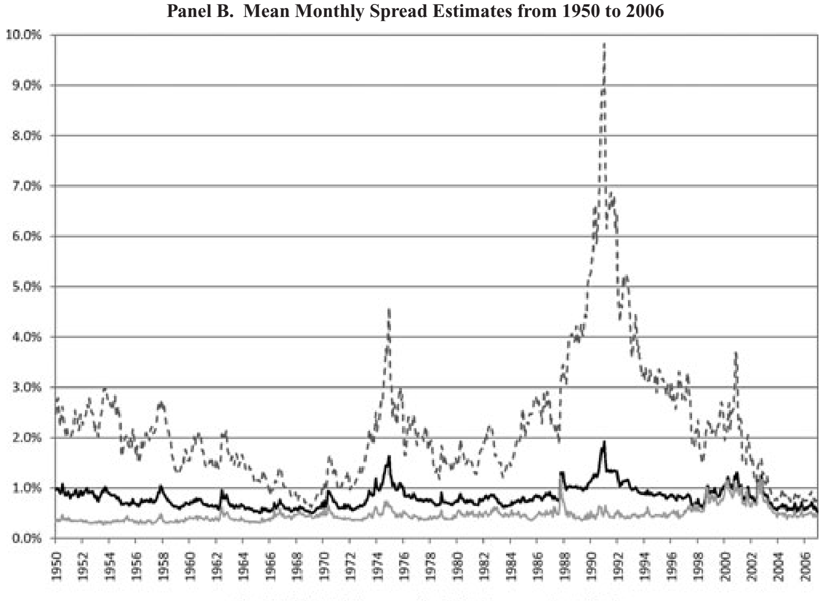


Figure 2. Historical high–low spread estimates based on CRSP data. High–low spreads are estimated for each stock each month by averaging 2-day spread estimates within the month. The graph plots the equally weighted average spread by month across all stocks with at least 12 daily spread observations within 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 to 2006 and Panel B shows results from 1950 to 2006. All data are from CRSP.

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 1950s and remained relatively low through the early 1970s. The recession of 1974 to 1975 is clearly visible in the figure as a period of increased spreads. Spreads are also relatively high in the early 1990s and during the tech bubble of the late 1990s.

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 1930s.

Panel B of Figure 2 provides a similar graph for the 1950 to 2006 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 to 1975 and 1991 to 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-1970s and also in the early 1990s. 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 studying this period to incorporate bid-ask spreads.

B. Using High–Low Spread Estimates in Asset Pricing Tests

In recent years, there has been a great deal of interest in the effects of liquidity and liquidity risk on asset pricing. We leave it to another paper to thoroughly study the impact of liquidity, as measured by our estimator, on asset prices. However, to demonstrate the potential usefulness of the estimator in asset pricing studies, we provide an analysis of abnormal returns on portfolios sorted by liquidity, where liquidity is defined by either the Amihud (2002) measure or the high–low spread estimator.

The Amihud measure, like the high–low spread estimator, is well suited to asset pricing studies because it can be estimated for a long time series using daily data. However, it is important to note that the two estimators may capture very different things. The Amihud illiquidity measure captures how much a given trading volume moves prices. The high–low spread estimator captures transitory volatility at the daily level and will closely approximate the effective spread, or the cost of immediacy. The high–low estimator also has the advantage that it does not require data on trading volume. It can therefore be applied in settings such as emerging markets, where the quality or availability of volume data may be suspect (e.g., Bekaert et al. (2007)), and for comparisons across markets such as the NYSE and NASDAQ, where volume measures may not be comparable.

For each month from June 1926 through December 2006, we calculate Amihud illiquidity measures and high–low spread estimates for each NYSE/Amex stock with a price of $5 or more. We sort stocks each month into 10 portfolios based on their average Amihud measure and, separately, into 10 portfolios based on their average high–low spread, using data from the prior 6 months. One-month-ahead and 6-month-ahead returns are then calculated for each portfolio. To avoid the biases in return estimates that come from using equal-weighted portfolios, we weight each stock by its prior-month return as suggested by Asparouhova et al. (2010). Abnormal returns are calculated for each portfolio by regressing the time series of monthly returns on the Fama-French factors, as obtained from Ken French's website.

Panel A of Table VI provides results for the month after portfolio formation, where portfolios are formed based on the Amihud measure. Looking across each row, we see that, as we go from liquid to illiquid stocks, coefficients on SMB and HML increase sharply. The intercept coefficient is small and insignificant in general, but is 0.0094 and highly significant for the portfolio of least liquid stocks. The last column shows the coefficients from a regression of the difference in returns between the least and most liquid stocks on the Fama-French factors. This long-short portfolio loads strongly on SMB and HML. The intercept indicates that, after adjusting for the Fama-French factors, this portfolio would earn a statistically significant abnormal return of 1.03% per month. These findings are generally consistent with the results from prior studies that show a significant relation between liquidity, as measured by the Amihud measure, and stock returns.

Results for 1-month returns on portfolios sorted based on the high–low spread estimator are provided in Panel B of Table VI. Despite its simplicity and other advantages, the results are very similar to those based on the Amihud measure. Again, coefficients on SMB and HML increase as we go from portfolios formed of more liquid stocks to portfolios formed of less liquid stocks. More importantly, the intercept for the illiquid portfolio is 0.0120 and the intercept from the regression based on the difference between the least liquid and most liquid portfolios is a statistically significant 1.05% per month.

Panel C reports regression results for 6-month returns, where portfolios are formed based on the Amihud measure. Because the 6-month periods overlap, all t-statistics are calculated based on Newey-West standard errors with five lags. As in Panels A and B, the less liquid portfolios have larger coefficients on SMB and HML than do the more liquid portfolios. The regression intercept for the least liquid portfolio suggests that these stocks earn a statistically significant

Table VI

Abnormal Returns on Liquidity-Sorted Portfolios

For each month over 1927 to 2006, we calculate the high–low spread estimate and Amihud illiquidity measure for all NYSE/Amex stocks with month-end prices of $5 or more. We discard all stocks without at least 12 days of positive volume (needed for the Amihud measure) or fewer than 12 CRSP returns. Stocks are then sorted into 10 portfolios by high–low spreads, and separately, into 10 portfolios by Amihud measure. To minimize biases in calculated returns, stocks are weighted by formation month gross returns. Portfolio returns are calculated for 1 and 6 months ahead. Abnormal returns are calculated by regressing the time series of portfolio returns on the Fama-French factors. The table reports intercepts and slope coefficients from these regressions, with t-statistics reported in parentheses below the coefficients. Panels A and B report results for 1-month-ahead returns for portfolios sorted based on the Amihud measure and the high–low spread measure, respectively. Panels C and D report results for 6-month ahead returns for portfolios sorted based on the Amihud measure and the high–low spread measure, respectively. In Panels C and D, where overlapping 6-month ahead returns are examined, t-statistics are based on Newey-West standard errors with five lags.

	Liquid	2	3	4	5	6	7	8	9	Illiquid	10–1
Panel A: Amihud Measure, 1-Month-Ahead Returns											
R_t^{Mkt}–R_t^F	1.0510	1.0359	1.0496	1.0354	1.0215	1.0035	1.0450	1.0237	1.0587	1.0843	0.0333
	(143.2)	(109.5)	(112.2)	(105.0)	(90.20)	(95.64)	(86.82)	(84.21)	(78.94)	(60.02)	(1.75)
SMB	−0.1142	0.0111	0.1478	0.2740	0.3378	0.4628	0.5500	0.6917	0.8317	1.1393	1.2535
	(− 9.82)	(0.74)	(9.97)	(17.53)	(18.82)	(27.83)	(28.83)	(35.90)	(39.13)	(39.79)	(41.47)
HML	0.0609	0.1031	0.2065	0.1981	0.2005	0.2909	0.4217	0.4739	0.6127	0.8924	0.8315
	(5.78)	(7.59)	(15.38)	(14.00)	(12.33)	(19.31)	(24.40)	(27.15)	(31.83)	(34.41)	(30.37)
Intercept	−0.0009	0.0001	−0.0003	0.0005	0.0001	0.0002	0.0007	0.0016	0.0024	0.0094	0.0103
	(− 2.39)	(0.22)	(− 0.73)	(0.97)	(0.20)	(0.30)	(1.22)	(2.58)	(3.48)	(10.26)	(10.64)
Panel B: High–Low Spreads, 1-Month-Ahead Returns											
R_t^{Mkt}–R_t^F	0.8277	0.9052	0.9683	1.0446	1.0918	1.1233	1.1328	1.1095	1.1167	1.0959	0.2682
	(81.50)	(98.06)	(103.5)	(99.68)	(103.5)	(98.50)	(98.20)	(85.42)	(81.18)	(63.73)	(12.78)
SMB	0.0091	0.1027	0.1703	0.2384	0.3225	0.3898	0.4856	0.6004	0.8194	1.2229	1.2138
	(0.57)	(7.02)	(11.48)	(14.35)	(19.28)	(21.57)	(26.56)	(29.17)	(37.58)	(44.87)	(36.50)
HML	0.0069	0.1083	0.1724	0.2585	0.3910	0.4078	0.4572	0.4468	0.5452	0.6802	0.6732
	(0.48)	(8.17)	(12.83)	(17.18)	(25.81)	(24.91)	(27.61)	(23.96)	(27.61)	(27.55)	(22.35)
Intercept	0.0014	0.0008	0.0005	0.0001	−0.0004	−0.0007	−0.0012	−0.0002	0.0014	0.0120	0.0105
	(2.80)	(1.75)	(1.04)	(0.27)	(− 0.82)	(− 1.27)	(− 2.05)	(− 0.36)	(1.95)	(13.69)	(9.87)

(continued)

Table VI—*Continued*

	Liquid	2	3	4	5	6	7	8	9	Illiquid	10–1
Panel C: Amihud Measure, 6-Month-Ahead Returns											
R_t^{Mkt}–R_t^F	0.9825	0.9269	0.9260	0.9330	0.9263	0.9337	0.9704	0.9949	0.9980	1.0664	0.0839
	(39.08)	(30.95)	(33.27)	(34.52)	(32.28)	(33.11)	(30.15)	(25.21)	(25.82)	(24.57)	(2.05)
SMB	−0.0762	0.1033	0.2552	0.3797	0.4788	0.5441	0.7126	0.8444	1.0071	1.5051	1.5812
	(− 1.69)	(1.64)	(4.94)	(8.06)	(8.52)	(8.44)	(14.20)	(13.46)	(16.70)	(16.49)	(18.51)
HML	0.0560	0.1387	0.1493	0.1737	0.2012	0.2432	0.3239	0.4044	0.4859	0.5748	0.5188
	(1.19)	(2.77)	(2.72)	(2.93)	(3.44)	(3.90)	(5.03)	(5.06)	(8.42)	(7.22)	(7.58)
Intercept	−0.0103	−0.0055	−0.0068	−0.0045	−0.0047	−0.0033	−0.0023	0.0021	0.0082	0.0550	0.0653
	(− 3.74)	(− 2.02)	(− 2.44)	(− 1.56)	(− 1.60)	(− 1.15)	(− 0.74)	(0.57)	(2.19)	(10.22)	(12.34)
Panel D: High–Low Spreads, 6-Month-Ahead Returns											
R_t^{Mkt}–R_t^F	0.7969	0.8715	0.9229	0.9474	0.9733	0.9984	1.0162	1.0081	1.0116	1.1174	0.3206
	(24.88)	(32.37)	(32.40)	(33.30)	(33.93)	(34.22)	(33.95)	(35.81)	(28.56)	(27.02)	(6.51)
SMB	0.0440	0.1652	0.2342	0.3214	0.4510	0.5360	0.6346	0.8299	0.9838	1.5598	1.5158
	(0.82)	(3.92)	(4.05)	(6.63)	(9.54)	(11.19)	(11.48)	(13.93)	(15.69)	(17.52)	(16.38)
HML	−0.0021	0.1053	0.1529	0.2244	0.3116	0.3186	0.3514	0.4072	0.4303	0.4557	0.4579
	(− 0.04)	(1.88)	(2.85)	(3.93)	(5.59)	(5.78)	(5.68)	(7.22)	(7.12)	(4.95)	(5.13)
Intercept	0.0038	−0.0016	−0.0044	−0.0042	−0.0078	−0.0088	−0.0097	−0.0072	0.0011	0.0664	0.0626
	(1.15)	(− 0.61)	(− 1.55)	(− 1.56)	(− 2.91)	(− 3.15)	(− 3.05)	(− 2.05)	(0.30)	(12.51)	(10.08)

abnormal return of 5.50% in the 6 months after portfolio formation. Again, the last column shows regression results when the differences in returns between the least liquid and most liquid portfolios are regressed on the Fama-French factors. The intercept is a highly significant 6.53% per 6 months.

Results for 6-month returns on portfolios formed based on the high–low estimator are provided in Panel D of Table VI. Again, the results are very similar to those based on the Amihud measure. The intercept for the least liquid portfolio is now 6.64%. Further, when the difference in returns between the least liquid portfolio and the most liquid portfolio is regressed on the Fama-French factors, the intercept is 0.0626 and is highly significant. This difference in abnormal returns is slightly smaller than but very similar to that for portfolios sorted based on the Amihud measure.

Overall, the results in Table VI suggest that the power of the high–low spread estimator to predict cross-sectional differences in returns is very similar to that of the Amihud illiquidity measure. Given its simplicity and potential application in settings where the Amihud measure cannot be used, this suggests that the high–low spread estimator may have many potential applications in asset pricing and other related research areas.

C. Estimating Spreads for Non-U.S. Markets 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.[23] As discussed later, each of these markets provides a specific event around which we expect execution costs to change. Results for additional countries covered by Datastream are provided in the Internet Appendix. 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 3. Because Datastream coverage increases over time, the graph also plots the number of firms used to compute

[23] Datastream data definitions suggest that closing bid and ask data are available for a small number of countries beginning in the mid-1990s, while data on daily high and low prices are available for more than 50 countries and for longer time periods. While data coverage in Datastream differs by country and improves over time, this suggests that the high-low spread estimator will be a useful tool for researchers who require a measure of liquidity or transaction costs for these markets. Intraday quote data are available for many international markets through Thomson Reuters and Bloomberg starting in the mid-1990s, but earlier intraday data are limited. Lesmond (2005) studies the ability of the Roll (1984), Amihud (2002), and Lesmond et al. (1999) measures to explain differences in bid-ask spreads within and across emerging markets.

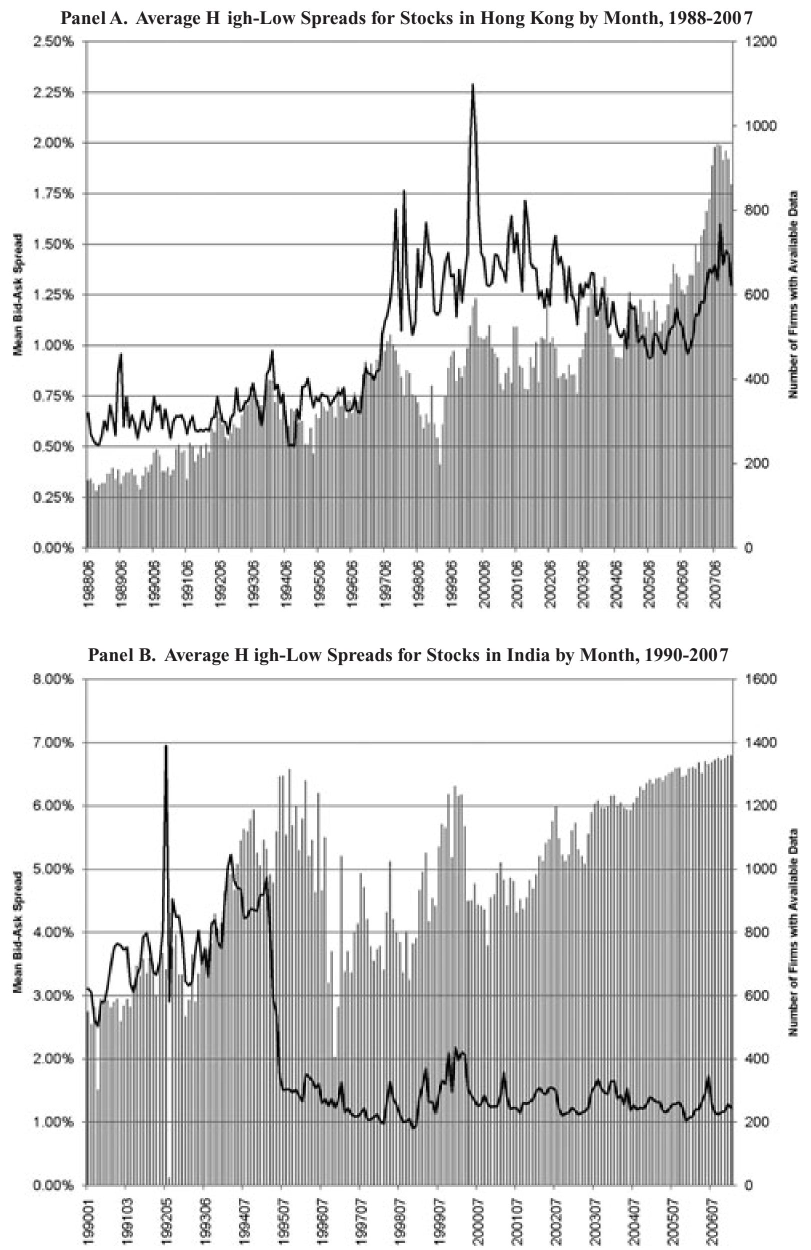


Figure 3. Historical high–low spread estimates based on Datastream data. High–low spreads are estimated for each stock each month by averaging 2-day spread estimates within the month. The graph plots the equally weighted average spread by month across all stocks with at least 12 daily spread observations 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.

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 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 to 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 Shah 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 3. 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 expected, 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 1% to 2% 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.

VII. 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. Further, 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 spread 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 useful 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.

Simulations reveal that the high–low spread estimator is very accurate under ideal conditions. When there are significant overnight returns and prices are observed sporadically, the high–low spread estimator tends to underestimate spreads. Even under these more general conditions, however, high–low spread estimates produce a correlation with simulated spreads of approximately 0.9. The simulations also suggest that the high–low spread estimator is far more accurate than the Roll estimator.

The simulation results are borne out in the data. We examine the performance of the high–low estimator by comparing effective spreads from TAQ with spread estimates from the high–low spread estimator, the Roll (1984) covariance estimator, the effective tick estimator of Goyenko et al. (2009) and Holden (2009), and the LOT measure of Lesmond et al. (1999). In cross-sectional tests, the high–low spread estimator clearly dominates, providing higher correlations with TAQ effective spreads and with month-to-month changes in spreads than any of the alternative spread estimators. The high–low spread estimator works particularly well in the 1993 to 1996 subperiod when the minimum tick was one-eighth. This suggests that the estimator should perform well in applications involving data from earlier time periods. In time-series tests, the high–low spread estimator dominates other measures for smaller stocks, such as those listed on NASDAQ and Amex, and performs particularly well during subperiods defined by different tick sizes.

To illustrate the potential applications of the high–low spread estimator, we apply the estimator to several other settings. First, we use the estimator to calculate bid-ask spreads for all NYSE/Amex stocks from 1926 through 2006. Among other things, we show that effective spreads were extremely high during the Depression, and increased sharply in the 1974 to 1975 bear market and following the 1987 crash. Using these same data, we then examine the performance of the high–low spread estimator in simple asset pricing tests. Despite its simplicity and other advantages, we find that the power of the high–low spread estimator to predict cross-sectional differences in returns is very similar to that of the Amihud illiquidity measure. We also 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. Several additional applications are provided in the Internet Appendix. Together, these examples demonstrate the wide range of applications for which the high–low spread estimator can be used. The high–low spread estimator can also be used to calculate trading costs for assets other than common stock or in settings in which quote data are either unavailable or difficult to use. For example, Deuskar et al. (2011) apply the estimator to OTC options markets and the estimator could also be applied to trade and sales data from the futures markets.

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.[24] We leave a more detailed analysis of this high–low variance estimator to future work.

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

REFERENCES

Amihud, 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.

Antunovich, Peter, and Asani Sarkar, 2006, Fifteen minutes of fame? The market impact of internet stock picks, *Journal of Business* 79, 3209–3251.

Asparouhova, Elena, Hendrik Bessembinder, and Ivalina Kalcheva, 2010, Liquidity biases in asset pricing tests, *Journal of Financial Economics* 96, 215–237.

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

Beckers, Stan, 1983, Variances of security price returns based on high, low, and closing prices, *Journal of Business* 56, 97–112.

Bekaert, Geert, Campbell Harvey, and Christian Lundblad, 2007, Liquidity and expected returns: Lessons from emerging markets, *Review of Financial Studies* 20, 1783–1831.

Bharath, Sreedhar, Paolo Pasquariello, and Guojun Wu, 2009, Does asymmetric information drive capital structure decisions? *Review of Financial Studies* 22, 3211–3243.

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.

Deuskar, Prachi, Anurag Gupta, and Marti G. Subrahmanyam, 2011, Liquidity effect in OTC options markets: Premium or discount? *Journal of Financial Markets* 14, 127–160.

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.

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

Gallant, A. Ronald, Chien-Te Hsu, and George Tauchen, 1999, Using daily data to calibrate volatility diffusions and extract the forward integrated variance, *Review of Economics and Statistics* 81, 617–631.

Garman, Mark B., and Michael 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? *Journal of Financial Economics* 92, 153–181.

Griffin, John, Patrick Kelly, and Federico Nardari, 2010, Do market efficiency measures yield correct inferences? A comparison of developed and emerging markets, *Review of Financial Studies* 23, 3225–3277.

Harris, Lawrence, 1986, A transaction data study of weekly and intradaily patterns in stock returns, *Journal of Financial Economics* 16, 99–117.

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, 2009, Trading costs and returns for U.S. equities: Estimating effective costs from daily data, *Journal of Finance* 64, 1445–1477.

Holden, Craig, 2009, New low-frequency spread measures, *Journal of Financial Markets* 12, 778–813.

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, 2009, Liquidity and capital structure, *Journal of Financial Markets* 12, 611–644.

Lockwood, Larry J., and Scott C. Linn, 1990, An examination of stock market return volatility during overnight and intraday periods, 1964–1989, *Journal of Finance* 45, 591–601.

Mei, Jianping, José Scheinkman, and Wei Xiong, 2009, Speculative trading and stock prices: Evidence from Chinese A-B share premia, *Annals of Economics and Finance* 10, 225–255.

Oldfield, George, and Richard Rogalski, 1980, A theory of stock returns over trading and non-trading periods, *Journal of Finance* 35, 729–751.

Parkinson, Michael, 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.

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.

Shah, Ajay, and Susan Thomas, 2000, David and Goliath: Displacing a primary market, *Journal of Global Financial Markets* 1, 14–21.