

An Application of the High-Low Spread Estimator to Intraday Data

Shane A. Corwin and Paul Schultz
February 2009

Corwin and Schultz (2009) derive an estimator for the bid-ask spread based on daily high and low prices. Their derivation of the high-low estimator is based on a comparison of one two-day period to two single days. However, the estimator is not limited to use with daily data. One potentially useful application of the estimator is with intraday trade data in cases where quote data is unavailable or when trades cannot be reliably matched with quotes. The use of the estimator with intraday data is also useful in cases where quote data are unwieldy. For example, recent TAQ quote files have become increasingly challenging to use, having grown to more than 10 times the size of the trade files.

To illustrate this potential application of the estimator, we use intraday trade data from TAQ to estimate high-low spreads at 15-minute intervals for all NYSE, Amex, and Nasdaq stocks. For comparison, we calculate high-low spreads for trading days in 1993 and 2006. The estimation follows the derivation of Corwin and Schultz (2009) except that it uses 15-minute rather than daily trading intervals. For example, the high-low spread estimate for the 15-minute interval starting at 9:45 a.m. is computed using the high and low prices from the 9:30-9:45 interval and the 9:45-10:00 interval.¹ We then calculate daily spreads for each stock by averaging across all 15-minute intervals within the day. We include only those stock-days with at least 10 intraday spread observations. In addition, following Corwin and Schultz (2009), we provide results using three alternative methods to account for negative spread estimates within the day: (1) set negative intraday spreads to zero, (2) leave negative two-day spreads unchanged, and (3) exclude negative two-day spreads. To test the accuracy of the high-low spread estimator as applied to intraday data, we compare the daily high-low spread estimates to daily trade-weighted effective spreads calculated directly from TAQ. The TAQ data include 1,544,655 stock-day observations for 1993 and 1,183,472 stock-day observations for 2006.² The analysis below is based on the subset of observations for which both a TAQ effective spread and a daily high-low spread could be calculated.

¹ Because spread estimates require two consecutive 15-minute intervals, no spread estimate is calculated for the 9:30-9:45 period on each day. In contrast to the daily analysis in Corwin and Schultz (2009), analysis based on intraday periods does not require an adjustment for overnight returns.

² The TAQ sample includes daily spread estimates for all common stocks that could be matched with a CRSP perm number. See Corwin and Schultz (2009) for a complete discussion of the effective spread estimates and associated data screens. The high-low estimator requires at least two consecutive 15-minute intervals with trades at multiple prices (i.e., high \neq low). As a result, requiring at least 10 intraday spread estimates eliminates a large fraction of stock-day observations. This data loss could be substantially reduced by requiring fewer intraday spread estimates or by using longer intraday intervals. For 1993, 24.9% of available stock-day observations meet this requirement, while 64.8% of stock-day observations have at least one intraday spread estimate. In 2006, 74.3% of available stock-day observations have at least 10 intraday spread estimates, while 89.8% of stock-day observations have at least one intraday estimate.

Summary statistics for TAQ effective spreads and daily high-low spreads are provided in Table 1. The TAQ effective spread averages 1.320% in 1993 and 0.272% in 2006. In comparison, the high-low spread for 1993 averages 0.837% if negative intraday values are included, 1.048% if negative values are set to zero, and 1.214% if negative values are excluded. For 2006, the three versions of the high-low spread average 0.056%, 0.181%, and 0.262%, respectively. While all three versions of the high-low estimator appear to underestimate spreads relative to the TAQ effective spread, the results suggest that eliminating negative intraday spread estimates prior to calculating the daily average produces results most comparable to the TAQ effective spread. This conclusion is confirmed by examining the mean absolute errors based on the difference between TAQ effective spreads and high-low spreads. The mean absolute error is lowest when negative intraday spreads are excluded, averaging 0.0026 for 1993 and 0.0013 for 2006. In the correlation analysis to follow, we focus on the version of the high-low spread estimator that excludes negative intraday spread estimates.

We note that the underestimation of high-low spreads relative to TAQ effective spreads may simply reflect intraday patterns in spreads and trading activity. While the high-low estimator weights each 15-minute period within the day equally, TAQ effective spreads are calculated as a trade-weighted average. As a result, the TAQ effective spread places greater weight on the relatively high spreads at the beginning and end of the trading day. A more detailed discussion of intraday patterns in spreads is provided below.

To provide additional tests of the accuracy of the high-low spread estimator as applied to intraday data, we examine correlations between the high-low spread estimates and TAQ effective spreads. We provide three alternative types of tests. First, we examine the correlation between high-low spreads and TAQ effective spreads in the pooled sample of time-series and cross-sectional observations. Second, we estimate the cross-sectional correlations by day and calculate the mean cross-sectional correlation across all days in the year. Finally, we estimate stock-by-stock time-series correlations and calculate the mean time-series correlation across all stocks and across stocks in each market capitalization decile. For each test, we provide separate results for 1993 and 2006. The results are presented in Table 2.

The pooled correlation between high-low spreads and TAQ effective spreads is 0.912 in 1993 and 0.803 in 2006. Similarly, the average cross-sectional correlation between high-low spreads and TAQ effective spreads is 0.912 in 1993 and 0.815 in 2006. These results are generally consistent with the monthly analysis provided in Corwin and Schultz (2009) and suggest that the high-low estimator produces very accurate daily spread estimates.

Across all sample stocks, the average time-series correlation is 0.525 in 1993 and 0.458 in 2006. During 1993, high-low spreads have a higher correlation with TAQ effective spreads for small stocks than for large stocks. The average time-series correlation is 0.568 for decile one and 0.427 for decile ten. In

2006, the performance of the high-low spread estimator is more consistent across size deciles, with an average time-series correlation of 0.454 in for decile one and 0.438 for decile ten.

To further illustrate the accuracy of the high-low spread estimator, Figure 1 plots the cross-sectional average TAQ effective spread and the cross-sectional averages of three alternative high-low spread estimates by day. Results for 1993 are shown in Panel A and results for 2006 are shown in Panel B. As in Table 1, the results show that the high-low spread is slightly lower than the TAQ effective spread, on average, with the best results being evident when negative intraday spread estimates are excluded. The effects of negative spread estimates are most pronounced in the results for 2006. What is most striking in Figure 1 is the ability of the high-low estimator to capture the time-series patterns in average spreads, especially in 1993. When negative intraday spread estimates are excluded, the time-series correlation between market-wide high-low spreads and market-wide effective spreads is 0.98 in 1993 and 0.67 in 2006.

Previous research shows that bid-ask spreads exhibit pronounced intraday patterns. Using data from the late 1980s, McInish and Wood (1992) find that bid-ask spreads for NYSE stocks follow a U-shaped pattern, with high spreads at the beginning and end of the trading day. Using data from the early 1990s, Chan, Christie, and Schultz (1995) find that Nasdaq stocks exhibit high spreads at the beginning of the day, but a marked decrease in spreads at the end of the day. As an additional test of the high-low estimator, we examine whether estimated spreads follow the expected intraday patterns.

Figure 2 plots average spreads by intraday period for NYSE and Nasdaq stocks. Results for 1993 are shown in Panels A and B and results for 2006 are shown in Panels C and D. As expected, bid-ask spreads for NYSE stocks exhibit a U-shaped pattern with the highest spreads at the beginning of the trading day.³ This pattern is evident in both 1993 and 2006, though the time-series variation is strongest in 2006. The patterns are similar for Nasdaq stocks, with spreads being highest at the beginning of the trading day in both 1993 and 2006. While Nasdaq spreads remain relatively flat at the end of the trading day in 1993, they appear to increase at the end of the trading day in 2006. This suggests that intraday patterns in the spreads of Nasdaq stocks may have changed over time, becoming more like NYSE patterns in more recent years.

Take together, the results provided here suggest that the high-low estimator performs very well when applied to intraday data. Not only do the spread estimates follow the expected intraday patterns, but the aggregation of intraday estimates produces daily spreads that are very accurate in comparison to TAQ effective spreads. Thus, we expect the application of the high-low spread estimator to intraday data to

³ Recall that the high-low spread is not estimated for the 9:30-9:45 period. The beginning-of-day patterns illustrated in Figure 2 would likely be more pronounced if this period was included.

prove very useful in cases where intraday quote data are unavailable or cumbersome, and in cases where trades cannot be reliably matched with quotes.

Table 1 - Summary Statistics for High-Low Spreads and TAQ Effective Spreads

The High-Low Spread for each trading day is defined as the equally-weighted average of the High-Low Spread across all overlapping 15-minute intervals within the day. Results are shown based on three alternative methods to account for negative spread estimates within the day: (1) leave negative two-day spreads unchanged, (2) set negative two-period spreads to zero, or (3) exclude negative two-period spreads. We include only those trading days with at least 10 intraday High-Low Spread estimates. The TAQ effective spread for each trading day is defined as a trade-weighted average across all trades within the day. For each day and each stock, mean absolute errors are defined based on the difference between the High-Low Spread and the TAQ effective spread. The sample includes 384,557 stock-days IN 1993 and 878,782 stock-days in 2006 for which both the TAQ effective spread and the High-Low Spread could be estimated.

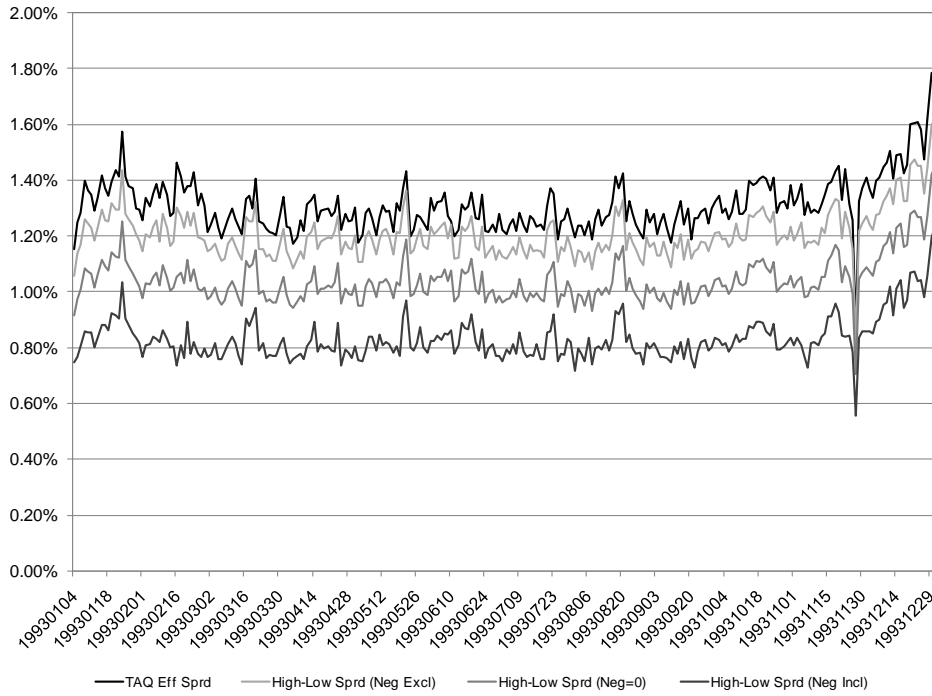
	1993	2006
Mean Spreads:		
TAQ Eff. Spread	1.320%	0.272%
High-Low Spread _{NegIncl}	0.837%	0.056%
High-Low Spread _{Neg=0}	1.048%	0.181%
High-Low Spread _{NegExcl}	1.214%	0.262%
Mean Absolute Errors:		
High-Low Spread _{NegIncl}	0.0054	0.0022
High-Low Spread _{Neg=0}	0.0035	0.0013
High-Low Spread _{NegExcl}	0.0026	0.0013

Table 2 - Correlations with TAQ Effective Spreads

The table lists correlations between the High-Low Spread estimates and TAQ effective spread. The High-Low Spread for each trading day is defined as the equally-weighted average of the High-Low Spread across all overlapping 15-minute intervals within the day, where negative spread estimates within the day are excluded prior to calculating the average. We include only those trading days with at least 10 intraday High-Low Spread estimates. The TAQ effective spread for each trading day is defined as a trade-weighted average across all trades within the day. The sample includes 384,557 stock-days in 1993 and 878,782 stock days in 2006 for which both the TAQ effective spread and the High-Low Spread could be estimated.

	1993	2006
Pooled Correlation:		
Full Sample	0.912	0.803
Mean Cross-Sectional Correlation:		
Full Sample	0.912	0.815
Mean Time-Series Correlation:		
Full Sample	0.525	0.458
Size Decile 1	0.568	0.454
Size Decile 2	0.543	0.450
Size Decile 3	0.492	0.459
Size Decile 4	0.461	0.466
Size Decile 5	0.452	0.463
Size Decile 6	0.452	0.467
Size Decile 7	0.485	0.483
Size Decile 8	0.389	0.470
Size Decile 8	0.419	0.481
Size Decile 10	0.427	0.438

Panel A - Average High-Low Spread by Day in 1993



Panel B - Average High-Low Spread by Day in 2006

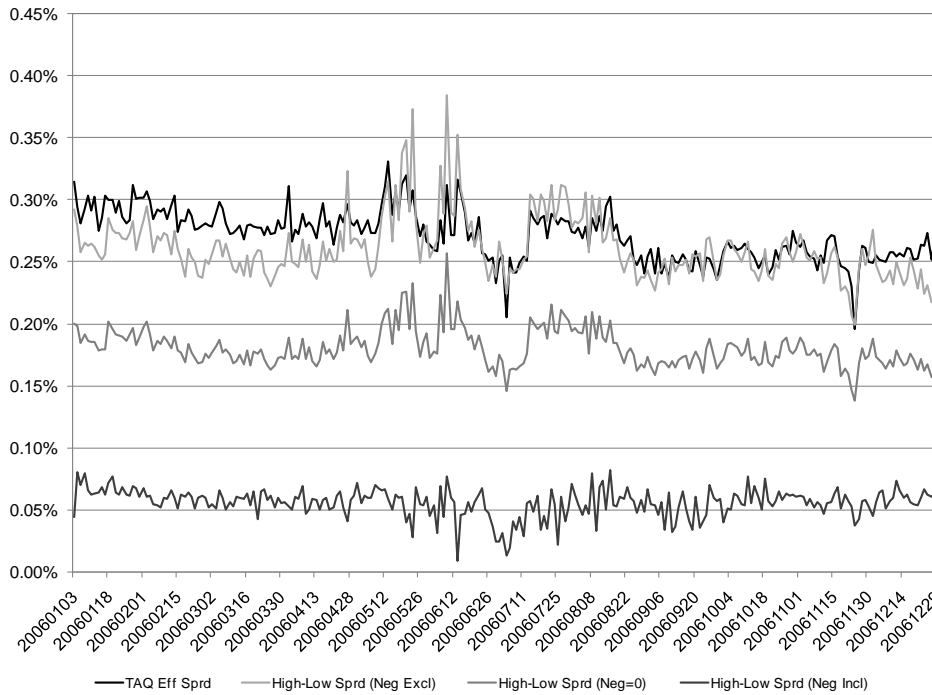


Figure 1 - Daily Patterns in High-Low Spreads

The figure plots cross-sectional average spreads by day. Results are shown for TAQ effective spreads as well as three alternative versions of the High-Low estimator, where negative spread estimates within the day are (1) excluded, (2) set to zero, or (3) included, prior to calculating the daily average. Results for 1993 are shown in Panel A and results for 2006 are shown in Panel B.

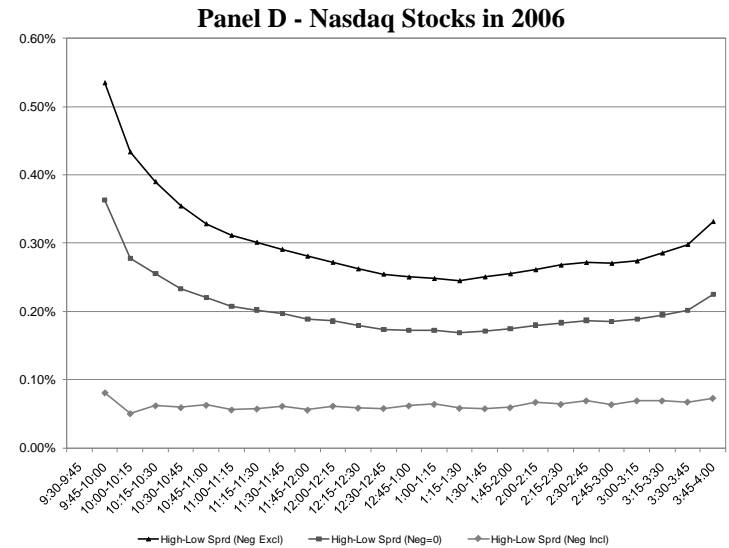
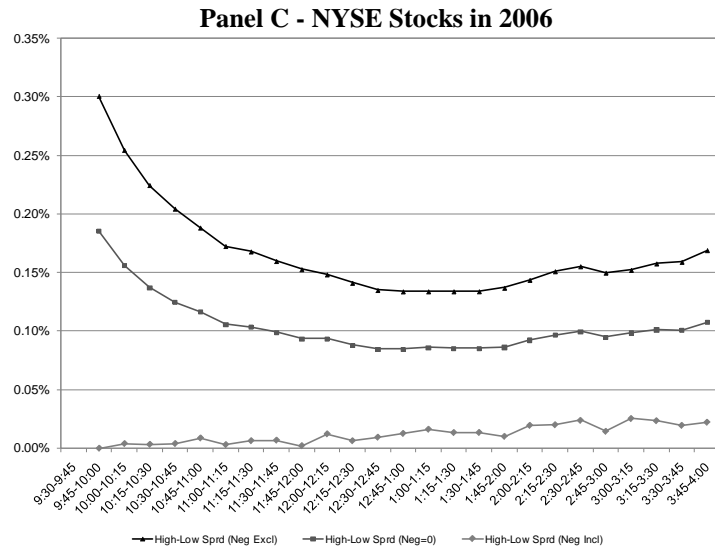
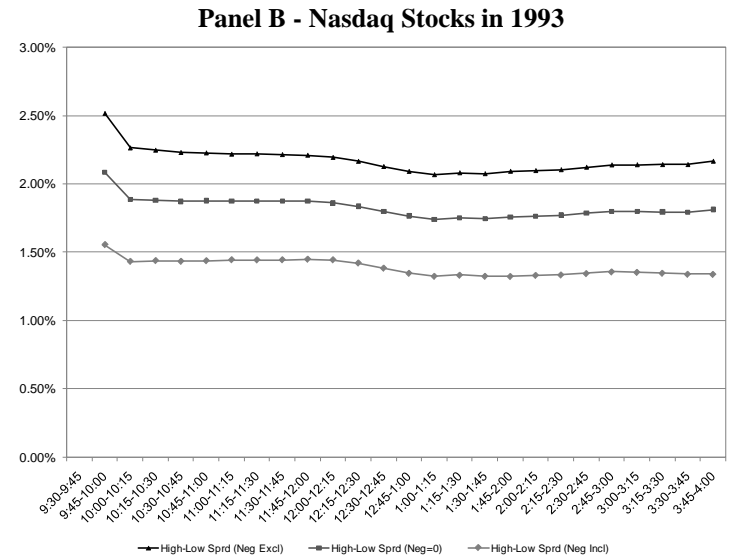
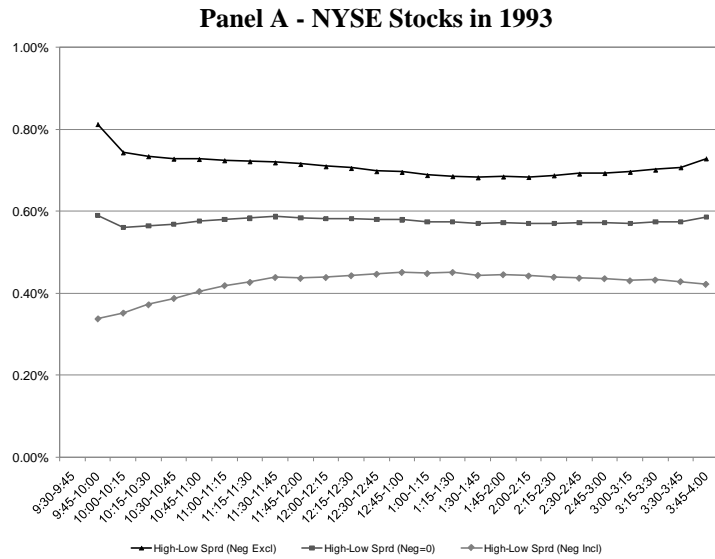


Figure 2 - Intraday Patterns in High-Low Spreads

The figure plots average High-Low Spreads across all days and stocks by 15-minute period. Results for NYSE and Nasdaq stocks in 1993 are shown in Panels A and B, respectively. Results for NYSE and Nasdaq stocks in 2006 are shown in Panels C and D, respectively. Results are shown for two alternative versions of the High-Low estimator, where negative spread estimates within the day are either (1) excluded, or (2) set to zero.

Corwin, Shane A., and Paul Schultz, 2009, A Simple Way to Estimate Bid-Ask Spreads from Daily High and Low Prices, working paper, University of Notre Dame.
http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1106193