Internet Appendix for
A Simple Way to Estimate Bid-Ask Spreads from Daily High and Low Prices*

SHANE A. CORWIN and PAUL SCHULTZ

This internet appendix provides additional results and robustness analyses that are not provided in the published text. The appendix also describes example applications of the high-low spread estimator to Datastream data, daily events studies, and intraday trade data.

*Citation Format: Corwin, Shane A., and Paul Schultz, 2012, Internet Appendix for "A Simple Way to Estimate Bid-Ask Spreads from Daily High and Low Prices," *Journal of Finance* 67, 719-759, http://www.afajof.org/supplements.asp. Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the authors of the article.

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

Contents

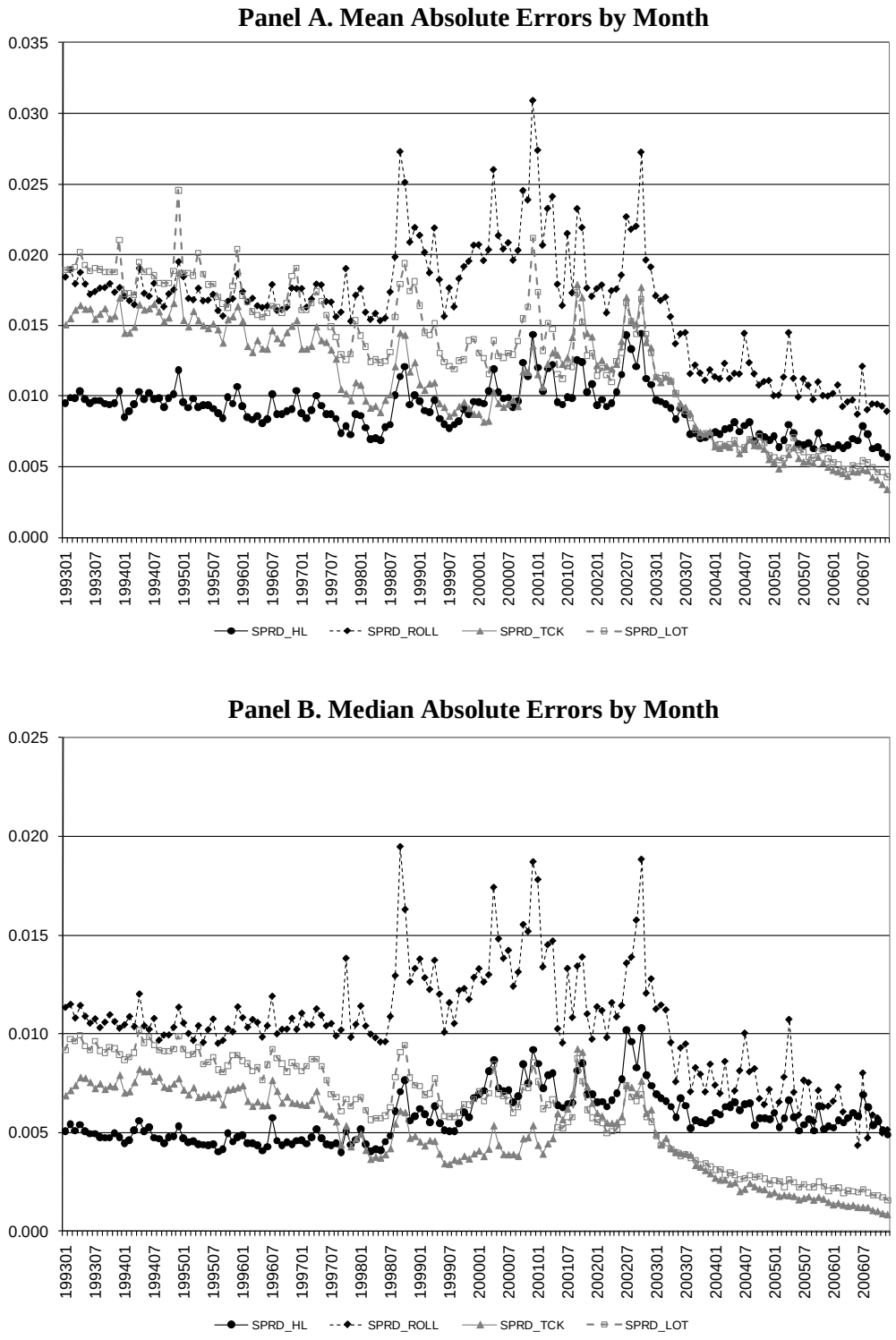


Figure IA.1. Cross-sectional mean and median absolute errors of spread estimates by month. The figure plots the mean and median absolute error across all securities by month from 1993 through 2006. The error and absolute error are defined for each stock-month based on the difference between the estimated spread measure and the effective spread from TAQ. The mean and median errors are then calculated across all stocks in a given month. The full sample includes all NYSE, Amex, and NASDAQ listed securities for which TAQ and CRSP data could be matched. Observations are then dropped if there are fewer than six monthly observations for the firm or if spread estimates are missing for the *Roll Spread*, *Tick Spread, LOT Measure,* or *HL Spread* (see Table III of the published text for definitions).

Table IA.I
Frequency of Daily Data Adjustments and Spread Characteristics

The table reports the fraction of daily observations for which specific data adjustments are made or for which specific data characteristics apply. The first three columns report the percentage of daily observations for which missing high and low prices are replaced based on the prior day's high and low prices. This can occur because there is zero volume on the day or because all trades take place at the same price (i.e., high=low). The fourth column reports the percentage of daily observations for which the high and low are adjusted due to overnight returns, based on the method described in Section II.A. The fifth column reports the percentage of daily observations for which the high and low are identical to the prior day's high and low. Finally, the last column lists the percentage of daily observations for which the high-low spread estimate is negative. The monthly high-low spread estimates in the paper are estimated after setting these negative daily values to zero. Frequencies are calculated based on daily observations within each month. The first row of the table then reports the average across all stock months. Subperiod frequencies are estimated by taking a cross-sectional average for each month and then averaging across all months within the subperiod. Frequencies for market capitalization quintiles are estimated by taking a time-series average for each stock and then averaging across all stocks within the specific market capitalization quintile. Stocks are assigned to market capitalization quintile based on NYSE breakpoints and firm market capitalization on the last date the firm is listed on CRSP.

	Reset Missing High and Low			Overnight Return Adjustment	High and Low Constant Across Days	Negative Daily Spread Estimate
	Zero Volume	High=Low	All Cases			
Full Sample	4.11	7.35	11.46	20.19	16.47	29.26
1993-1996	5.31	10.88	16.19	15.16	28.95	24.37
1997-2000	3.99	7.76	11.75	21.11	15.06	31.00
2001-2006	3.07	3.67	6.74	24.04	5.99	32.34
MV Quintile 1	7.30	12.05	19.35	19.92	25.05	24.44
MV Quintile 2	1.55	3.96	5.51	18.20	12.73	31.18
MV Quintile 3	0.48	1.60	2.08	18.93	7.96	35.35
MV Quintile 4	0.28	0.87	1.15	19.93	5.51	37.87
MV Quintile 5	0.06	0.12	0.18	22.67	2.07	40.18
Across All Daily Observations	5.99	8.10	14.08	22.55	16.34	29.97
Across All Monthly Observations (N>=12)	6.01	6.84	12.85	22.88	16.53	29.99

Table IA.II
High-Low Spread Estimates Based on Alternative Overnight Return Adjustments

The table provides characteristics for monthly high-low spread estimates without any adjustment for overnight returns and with two alternative overnight return adjustments. The first overnight return adjustment is the one used throughout the paper, based on the day *t* close and the day *t*+1 high-low range. Specifically, if the day *t*+1 low (high) is above (below) the day *t* close, we decrease (increase) the day *t*+1 high and low by the amount of the difference. The alternative overnight return adjustment is based on the day *t* close and the day *t*+1 open. In this case, if the day *t*+1 open is above (below) the day *t* close, we decrease (increase) the day *t*+1 high and low by the amount of the difference. The mean, median, standard deviation, and correlation are estimated using the full sample of pooled stock-months. The frequency of overnight return adjustments and negative daily estimates are calculated within each stock-month and then averaged across all stock-months.

	No Overnight Return Adjustment	Current Overnight Return Adjustment	Alternative Overnight Return Adjustment
Pooled Correlation with TAQ Eff Spread	0.877	0.899	0.745
Mean (%)	1.95	2.10	0.88
Median (%)	1.22	1.32	0.50
Std. Dev. (%)	2.53	2.65	1.98
Daily Obs with Overnight Return Adjustments (%)	0.00	18.09	87.81
Negative Daily Estimates (%)	34.92	29.26	49.22
Negative Monthly Estimates when Negative Daily Estimates are Included (%)	43.15	24.00	83.70

Table IA.III
Pooled Correlations between TAQ Spreads and Estimated Liquidity Measures

The table lists correlations among the 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 six 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 of the published text for definitions).

	TAQ Effective Spread	TAQ Quoted Spread	Roll Spread	Effective Tick Spread	LOT Measure	Amihud Measure	High-Low Spread
Effective Spread	1.000						
Quoted Spread	0.979	1.000					
Roll Spread	0.707	0.695	1.000				
Tick Spread	0.734	0.724	0.526	1.000			
Lot Measure	0.699	0.676	0.506	0.677	1.000		
Amihud Measure	0.362	0.360	0.297	0.278	0.268	1.000	
HL Spread	0.899	0.883	0.702	0.719	0.692	0.359	1.000

Table IA.IV
Cross-Sectional Mean Errors and Absolute Errors

For each stock-month, errors and absolute errors are defined for each spread measure based on the difference between the spread measure and the TAQ effective spread. For each month from 1993 through 2006, we estimate the mean error and mean absolute error across all cross-sectional observations. The table then lists the mean across months. For reporting purposes, mean errors are multiplied by 100. The full sample includes all NYSE, Amex, and NASDAQ listed securities for which TAQ and CRSP data could be matched. Observations are then dropped if there are fewer than six monthly observations for the firm or if spread estimates are missing for the *Roll Spread*, *Tick Spread*, *LOT Measure*, or *HL Spread*. Panel A lists results based on monthly spreads and Panel B lists results based on first differences in monthly spreads (see Table III of the published text for definitions).

	N	Mean Error				Mean Absolute Error			
		Roll Spread	Eff. Tick Spread	LOT Measure	High-Low Spread	Roll Spread	Eff. Tick Spread	LOT Measure	High-Low Spread
Panel A. Errors Based on Monthly Spread Levels									
Full Period	168	0.0007	-0.0072	-0.0026	-0.0024	0.0166	0.0113	0.0132	0.0090
1993-1996	53	-0.0057	-0.0086	-0.0007	-0.0077	0.0173	0.0152	0.0180	0.0094
1997-2000	43	0.0020	-0.0043	-0.0004	-0.0027	0.0195	0.0104	0.0142	0.0092
2001-2006	72	0.0047	-0.0079	-0.0053	0.0016	0.0144	0.0088	0.0090	0.0086
Panel B. Errors Based on Changes in Monthly Spreads									
Full Period	167	0.0071	0.0021	0.0084	0.0012	0.0214	0.0086	0.0146	0.0067
1993-1996	52	0.0045	0.0008	0.0038	-0.0012	0.0219	0.0127	0.0214	0.0078
1997-2000	43	0.0307	0.0144	0.0300	0.0071	0.0255	0.0108	0.0175	0.0075
2001-2006	72	-0.0051	-0.0044	-0.0010	-0.0006	0.0187	0.0044	0.0081	0.0054

Table IA.V
Stock-by-Stock Time-Series Mean Errors and Absolute Errors

For each stock-month, errors and absolute errors are defined for each spread measure based on the difference between the spread measure and the TAQ effective spread. For each stock, we estimate the mean error and mean absolute error across all monthly time-series observations from 1993 through 2006. The table then lists the mean across all stocks. The full sample includes all NYSE, Amex, and NASDAQ listed securities for which TAQ and CRSP data could be matched. Observations are then dropped if there are fewer than six monthly observations for the firm or if spread estimates are missing for the *Roll Spread, Tick Spread, LOT Measure,* or *HL Spread* (see Table III of the published text for definitions). 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.

		Mean Error				Mean Absolute Error			
	N	Roll Spread	Eff. Tick Spread	LOT Measure	High-Low Spread	Roll Spread	Eff. Tick Spread	LOT Measure	High-Low Spread
Full Sample	12,507	-0.0014	-0.0086	-0.0016	-0.0048	0.0197	0.0145	0.0174	0.0108
NYSE	2,894	0.0021	0.0003	0.0043	0.0008	0.0100	0.0036	0.0080	0.0046
Amex	1,118	-0.0068	-0.0050	0.0051	-0.0100	0.0227	0.0168	0.0232	0.0129
NASDAQ	8,495	-0.0019	-0.0121	-0.0045	-0.0060	0.0225	0.0180	0.0199	0.0126
MV Quintile 1	8,555	-0.0032	-0.0112	-0.0022	-0.0076	0.0236	0.0192	0.0223	0.0131
MV Quintile 2	1,487	0.0016	-0.0043	-0.0009	0.0004	0.0123	0.0058	0.0082	0.0060
MV Quintile 3	975	0.0032	-0.0021	0.0001	0.0020	0.0106	0.0033	0.0056	0.0049
MV Quintile 4	803	0.0036	-0.0010	0.0004	0.0028	0.0091	0.0021	0.0042	0.0044
MV Quintile 5	574	0.0049	-0.0001	0.0004	0.0042	0.0079	0.0010	0.0025	0.0046

Table IA.VI
Summary Statistics for Spread Estimates (Alternative LOT Estimators)

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 six 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. *HL Spread* is the equally weighted average of the high-low spread estimator across all overlapping two-day periods within the month, where negative two-day spreads are set to zero. The *LOT Measure* is estimated based on two alternative methods for separating the positive and negative return regions in the likelihood function, as described in Goyenko, Holden, and Trzcinka (2009). The first method, *LOT Y-Split*, defines these regions based on stock returns only. The second method, *LOT-Mixed*, defines these regions based on both stock returns and market returns. Following the estimation in the paper, we define the market return for the LOT Y-Split measure based on the CRSP value-weighted index. For the LOT Mixed measure, we follow Goyenko et al. (2009) in defining the market return based on the CRSP equal-weighted index. The LOT measure is undefined if either the number of returns during the month is less than five or the percentage of zero returns during the month is greater than 80%. A firm-month is deleted if any of the spread estimates are missing or are greater than 50%.

	N	Mean (%)	Median (%)	Std. Dev. (%)	%<=0	Pooled Correlation with Eff. Spread
Effective Spread	966,333	2.32	1.28	3.05	0.00	1.000
HL Spread	966,333	2.10	1.32	2.65	0.00	0.888
Roll Spread	966,333	2.36	1.21	3.54	38.08	0.685
Eff. Tick Spread	966,333	1.60	0.71	2.83	0.00	0.732
Lot Y-Split	966,333	2.02	0.86	3.50	24.54	0.733
Lot Mixed	966,333	4.02	2.65	4.57	4.39	0.740

Table IA.VII
Average Cross-Sectional Correlations (Alternative LOT Estimators)

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, with results shown for the full sample period and for subperiods corresponding to different tick size regimes. Spread measures and sample restrictions are as described in Table IA.VI.

	N	High-Low Spread	Roll Spread	Effective Tick Spread	LOT Y-Split	LOT Mixed
Full Period	168	0.823	0.626	0.684	0.658	0.701
1993-1996	53	0.924	0.765	0.732	0.730	0.832
1997-2000	43	0.856	0.621	0.727	0.707	0.722
2001-2006	72	0.729	0.527	0.624	0.576	0.593

Table IA.VIII
Average Stock-by-Stock Time-Series Correlations (Alternative LOT Estimators)

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. Panels A, B, and C provide results for subperiods corresponding to different tick size regimes. Panel D provides results for the full sample period from 1993 to 2006. 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. Spread measures and sample restrictions are as described in Table IA.VI.

	N	High-Low Spread	Roll Spread	Effective Tick Spread	LOT Y-Split	LOT Mixed
Panel A. Average Correlation (1993-1996)						
Full Sample	9,031	0.650	0.345	0.477	0.338	0.404
NYSE	2,358	0.498	0.200	0.506	0.272	0.275
Amex	755	0.606	0.310	0.439	0.321	0.395
NASDAQ	5,918	0.716	0.408	0.471	0.367	0.456
MV Quintile 1	5,942	0.700	0.404	0.458	0.372	0.463
MV Quintile 2	1,111	0.655	0.312	0.496	0.279	0.324
MV Quintile 3	746	0.589	0.238	0.517	0.284	0.290
MV Quintile 4	642	0.499	0.171	0.525	0.254	0.260
MV Quintile 5	498	0.338	0.107	0.567	0.252	0.213
Panel B. Average Correlation (1997-2000)						
Full Sample	9,331	0.574	0.277	0.506	0.313	0.316
NYSE	2,316	0.496	0.163	0.517	0.247	0.222
Amex	770	0.630	0.325	0.507	0.396	0.397
NASDAQ	6,245	0.595	0.313	0.502	0.328	0.342
MV Quintile 1	6,292	0.637	0.341	0.509	0.362	0.379
MV Quintile 2	1,061	0.530	0.188	0.492	0.224	0.213
MV Quintile 3	758	0.441	0.139	0.509	0.210	0.168
MV Quintile 4	657	0.384	0.107	0.505	0.193	0.120
MV Quintile 5	489	0.307	0.080	0.513	0.192	0.168
Panel C. Average Correlation (2001-2006)						
Full Sample	7,419	0.563	0.272	0.514	0.287	0.347
NYSE	1,898	0.470	0.159	0.592	0.219	0.288
Amex	718	0.576	0.305	0.505	0.391	0.423
NASDAQ	4,803	0.598	0.312	0.485	0.298	0.359
MV Quintile 1	5,046	0.582	0.311	0.464	0.322	0.378
MV Quintile 2	825	0.473	0.180	0.597	0.252	0.283
MV Quintile 3	594	0.487	0.183	0.633	0.229	0.281
MV Quintile 4	508	0.564	0.183	0.647	0.179	0.269
MV Quintile 5	408	0.617	0.205	0.626	0.133	0.273

Table IA.VIII (continued)

	N	High-Low Spread	Roll Spread	Effective Tick Spread	LOT Y-Split	LOT Mixed
Panel D. Average Correlation (1993-2006)						
Full Sample	12,505	0.624	0.339	0.599	0.443	0.438
NYSE	2,894	0.412	0.158	0.705	0.443	0.356
Amex	1,118	0.665	0.381	0.547	0.447	0.474
NASDAQ	8,493	0.691	0.395	0.569	0.443	0.461
MV Quintile 1	8,553	0.700	0.413	0.549	0.451	0.482
MV Quintile 2	1,487	0.582	0.248	0.662	0.405	0.361
MV Quintile 3	975	0.477	0.172	0.719	0.435	0.341
MV Quintile 4	803	0.382	0.120	0.747	0.435	0.320
MV Quintile 5	574	0.184	0.047	0.789	0.445	0.267

Additional Applications of the High-Low Spread Estimator

The main text provides an analysis of the performance of the high-low spread estimator relative to monthly effective spreads from TAQ. The text also provides applications of the estimator to historical CRSP data and to Datastream data for stocks in Hong Kong and India. In this section, we describe several additional applications to demonstrate the potential usefulness of the high-low spread estimator to a variety of markets and research questions. The first application is an extension of the Datastream application in the text, incorporating data from nine additional countries. The second application describes the estimation of daily spreads in an event study around stock splits from 1926 through 1982. The third application describes the use of the estimator for intraday, rather than daily, periods.

A. An Application to Non-U.S. Markets Using Datastream Data

In this section, we demonstrate the application of the high-low spread estimator to non-U.S. markets using Datastream data for individual stocks in 11 countries. Results for Hong Kong and India are described in Section VI of the text. The nine additional markets include Korea, Japan, Italy, France, Belgium, Sweden, the U.K., Brazil, and New Zealand. As in the previous analyses, we calculate high-low spreads for each two-day interval following the derivation in Section I. We then calculate monthly spreads for each stock by averaging across all overlapping two-day intervals within the month. We include only those stock-months with at least 12 daily spread observations and we set all negative daily estimates to zero before taking the monthly average.[1] Finally, for each country, we calculate the cross-sectional average of high-low spreads by month using all stocks with sufficient data.

Table IA.IX provides summary statistics for the monthly high-low spreads in each of the 11 countries. The period of Datastream data coverage differs by country. We therefore provide results based on the period from January 1994 through December 2007, when data are available for all 11 countries. As the results show, the market-wide average spread varies substantially across countries, ranging from a low of 0.58% for New Zealand to a high of 1.47% for Korea. Among countries with at least 100 sample firms, the minimum monthly spread is lowest in the UK, at 0.90%. The monthly market-wide average spreads

[1] We adjust for overnight returns as described in the text based on a comparison of daily high and low prices to the prior day's close. Spreads are not estimated in cases in which the daily high and low are either equal or missing.

exhibit even more variation over time within individual countries. In India, for example, monthly average spreads range from a low of 0.96% to a high of 5.23%. Similarly, monthly average spreads in Italy range from a low of 0.58% to a high of 3.91%. The number of firms with sufficient data to allow spread estimation also varies widely across countries, with the average number of firms ranging from 54 for New Zealand to 2,481 for Japan.

To examine the relation between spreads across countries, we compute time-series correlations between the monthly market-wide spreads for each pair of countries. The results are provided in Table IA.X. The paired correlations range from a low of -0.441 for Hong Kong and India to a high of 0.736 for Sweden and France. The average time-series correlation across all country pairs is 0.214.

To illustrate the time-series patterns in high-low spreads, we plot the monthly cross-sectional average by country in Panels A through I of Figure IA.2. Because data coverage in Datastream increases over time, the graphs also plot the number of firms used to compute the market-wide average in each month. Similar graphs for Hong Kong and India are provided in the main text.

Results for Korea and Japan are shown in Panels A and B of Figure IA.2. As is the case for Hong Kong, as discussed in the main text, average spreads in Korea and Japan exhibit a sharp increase at the time of the Asian currency crisis in October 1997. In Korea, mean high-low spreads range from 0.5% to 1.0% prior to the currency crisis, but jump to over 2% during much of 1998, 1999, and 2000. Spreads come down in early 2001, stabilizing at approximately 1.5%. While spreads appear to be high in 2002 to 2007 relative to the pre-crisis period, this may simply reflect the substantial increase in Datastream coverage in the later period. The results are similar for Japan, where high-low spreads exhibit a sharp increase in October 1997 and a subsequent decrease from late 2003 through early 2005. For both countries, the patterns in high-low spreads are consistent with a substantial increase in execution costs during the Asian currency crisis. For the period starting in January 1994, the average paired correlation between market-wide average spreads of Korea, Japan, and Hong Kong is 0.640.

Results for Italy, France, Belgium, and Sweden are provided in Panels C through F of Figure IA.2. For all four countries, the most striking feature is a significant increase in spreads in December 1994. In Italy, for example, spreads increased from an average of less than 1.0% to a peak of nearly 4.0%.

Similar increases are evident in the other three countries and all four countries experience a significant decrease in spreads in December of 1995. These effects correspond to the Mexican peso crisis in 1994 to 1995 and suggest that contagion effects or exposure to Mexican debt led to large increases in execution costs for these countries. For Sweden, the graph also points to an increase in spreads from late 1990 through 1993. This increase in execution costs may reflect the banking problems and subsequent kroner depreciation experienced in Sweden during this period. For the period starting in January 1994, the average paired correlation between market-wide average spreads in these four countries is 0.609.

Panel G of Figure IA.2 presents results for the UK. Several notable patterns are evident from the graph. First, there is a sharp increase in spreads from late 1989 through early 1991. This is followed by a significant drop in spreads during 1993. From late 1999 through early 2003, spreads in the UK roughly double, from less than 1.0% to over 1.8%. Finally, the graph shows a significant decrease in spreads in August 2003, coupled with a significant drop in the number of firms. While we cannot place economic meaning on all of these patterns, the decrease in spreads during 1993 coincides with the UK's exit from the European exchange rate mechanism.

Results for Brazil and New Zealand are provided in Panels H and I of Figure IA.2. There do not appear to be any systematic patterns in spreads for either of these countries. While spreads in Brazil are significantly higher in 1993 than in later years, this appears to reflect the small sample size during this period. For New Zealand, there are several spikes in spreads, including a general increase in spreads from mid-1998 through 2002. However, the number of stocks included in the sample for New Zealand is relatively small, making it difficult to draw firm conclusions.

Overall, the results presented in Figure IA.2 point to several economically important patterns in execution costs across multiple countries. These examples illustrate the potential use of the high-low spread estimator for analyzing execution costs in non-U.S. markets using Datastream data.

B. An Application to Daily Event Studies: Stock Splits from 1926 to 1982

One of the advantages of the high-low spread estimator is that it can be used to produce spread estimates over relatively short time intervals, such as days or weeks. This makes it ideal for analyzing

changes in spreads around specific events. Spreads, of course, provide information about the profitability of trading around events. In addition, spreads can be used as a measure of asymmetric information around events.

To illustrate this potential application of the estimator, we examine daily high-low spread estimates around stock splits over the period from 1926 to 1982. Notably, this period precedes the availability of intraday data through TAQ or ISSM. A number of previous studies, including Angel (1997), Conroy, Harris, and Benet (1990), and Schultz (2000), examine bid-ask spreads around stock splits and conclude that quoted and effective spreads rise following stock splits. All of these studies examine recent splits, because intraday trade and quote data are unavailable before 1983. Because the high-low spread estimator requires only daily high and low prices, it can be used to evaluate spread patterns around stock splits during earlier time periods.

To begin, we collect the full sample of stock splits for CRSP stocks from 1926 to 1982 (CRSP distribution code 5523). We include only those splits that increase shares outstanding by at least 20%. Using the high-low spread estimator, we calculate the cross-sectional average spread on each day from -10 to +10 around the split date. The high-low estimator involves comparing high and low price ranges over a single two-day period with the high-low price range over two one-day periods. To obtain daily spread estimates on particular days around the stock split, we take an average of the high-low spread estimates for the two overlapping intervals that include that day. For example, for day -5, we take the average of the spreads estimated over the two-day intervals from -6 to -5 and from -5 to -4. For the day before the split (day -1), we use just the spread estimated from days -2 to -1. For the first post-split day (day +1), we use just the spread estimated from days 0 to 1.[2]

If a stock trades at just one price over a two-day period, the high-low estimator for that period is not defined. We could restrict our sample to stocks with spread estimates every day around the split, but that would eliminate the less active stocks from our sample and we are interested in how their trading

[2] In the main text, we compute monthly spread estimates by taking an average across all two-day periods within the month. These monthly estimates are considerably less noisy than the daily estimates produced here. If intraday data are available, an alternative would be to calculate high-low spread estimates for intraday periods and then average across intraday periods to get a daily estimate. An application of the estimator to intraday trade data is provided below.

costs are affected by splits. Instead, we include stocks in our sample even if spreads cannot be estimated every day. Hence, the number of cross-sectional observations varies from 2,924 to 3,150 over the 21-day event period, with only one day having fewer than 3,000 observations. As in the main text, negative two-day estimates are set to zero. However, the conclusions are similar if negative spreads are included.

The results of the event study analysis are depicted in Figure IA.3. Mean spreads show little variation day-to-day but increase sharply the day following a split. Prior to splits, the mean high-low spread ranges from 2.92% to 3.09%. In the 10 days following the split, the mean spread ranges from 3.72% to 4.01%, with the maximum at day +1. Differences between post-split and pre-split spreads (measured from days -15 through -11) are highly statistically significant. These findings are consistent with studies of stock splits from more recent time periods.

The stock split analysis illustrated here demonstrates the potential usefulness of the high-low spread estimator in measuring changes in trading costs or uncertainty around events. In particular, this analysis demonstrates how the high-low estimator can be used to analyze spreads at the daily level.

C. An Application to Intraday Trade Data

The derivation of the high-low estimator described in the main text 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 in which quote data are unavailable or when trades cannot be reliably matched with quotes. The use of the estimator with intraday data is also useful in cases in which 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 calculate high-low spreads for 15-minute intraday periods in 1993 and 2006 for all NYSE, Amex, and NASDAQ stocks. The estimation follows the derivation in the main text, except that it uses consecutive intraday 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 to 9:45 interval and the 9:45 to 10:00 interval,

where the high and low are identified using intraday trade data from TAQ.[3] 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, we provide results using three alternative methods to account for negative spread estimates within the day: (1) setting negative intraday spreads to zero, (2) leaving negative two-day spreads unchanged, and (3) excluding negative two-day spreads. To test the performance of the high-low spread estimator as applied to intraday data, we compare the resulting high-low spread estimates to 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.[4] The analysis below is based on the subset of observations for which both a TAQ effective spread and a high-low spread could be calculated.

Summary statistics for daily estimates of TAQ effective spreads and high-low spreads are provided in Table IA.XI. 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.

[3] Because spread estimates require two consecutive 15-minute intervals, no spread estimate is calculated for the 9:30 to 9:45 period on each day. In contrast to the daily analysis in the main text, analysis based on intraday periods does not require an adjustment for overnight returns.

[4] The calculation of effective spreads and related TAQ data screens are discussed in more detail in Section V. 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 have at least 10 intraday spread estimates, 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.

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.

To provide additional tests of the performance 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 IA.XII.

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 Section V 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 for decile one and 0.438 for decile ten.

To further illustrate the accuracy of the high-low spread estimator, Figure IA.4 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 IA.XI, 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 IA.4 is the ability of the high-low estimator to capture the time-series patterns in average daily 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 IA.5 plots the average high-low spread 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.[5] 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 recent years.

Taken 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

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

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

Table IA.IX
Summary Statistics for High-Low Spreads by Country

The table provides summary statistics for the cross-section of high-low spread estimates for each of 11 countries covered by Datastream. We estimate monthly high-low spreads for each stock-month with at least 12 daily spread estimates. For each country, we then calculate the cross-sectional average high-low spread each month. The table reports the time-series properties of these country-specific cross-sectional averages, along with information on the number of firms included in each average each month. Data are from January 1994 through December 2007.

Country	Monthly High-Low Spread			Number of Firms in Cross-Sectional Avg		
	Mean	Min	Max	Mean	Minimum	Maximum
Korea	1.47%	0.73%	2.59%	1265	350	1865
Japan	0.76%	0.48%	1.12%	2481	1582	3623
Hong Kong	1.16%	0.50%	2.29%	475	198	957
India	1.65%	0.90%	5.23%	1083	406	1482
Italy	1.02%	0.55%	3.91%	264	147	315
France	0.95%	0.57%	2.01%	451	249	569
Belgium	0.83%	0.32%	2.26%	83	37	107
Sweden	1.10%	0.54%	2.95%	295	142	435
UK	0.90%	0.50%	1.85%	1189	673	1581
Brazil	1.31%	0.68%	3.73%	144	62	326
New Zealand	0.58%	0.34%	1.17%	54	15	79

Table IA.X
Correlations in High-Low Spreads Across Countries

The table provides paired time-series correlations between market-wide high-low spreads for each of 11 countries covered by Datastream. We estimate monthly high-low spreads for each stock-month with at least 12 daily spread estimates. For each country, we then calculate the cross-sectional average high-low spread each month. The table reports the paired time-series correlations between these country-specific cross-sectional averages. Data are from January 1994 through December 2007.

	Korea	Japan	Hong Kong	India	Italy	France	Belgium	Sweden	UK	Brazil
Japan	0.563									
Hong Kong	0.720	0.637								
India	-0.424	-0.220	-0.441							
Italy	-0.284	0.198	-0.258	0.146						
France	-0.067	0.494	0.097	-0.066	0.574					
Belgium	-0.050	0.305	0.051	0.007	0.542	0.628				
Sweden	0.061	0.496	0.213	0.004	0.589	0.736	0.583			
UK	0.274	0.471	0.425	-0.220	-0.005	0.542	0.266	0.323		
Brazil	0.168	0.295	0.230	0.362	-0.031	0.045	-0.105	0.096	0.176	
New Zealand	0.277	0.451	0.350	0.081	0.058	0.332	0.044	0.293	0.362	0.389

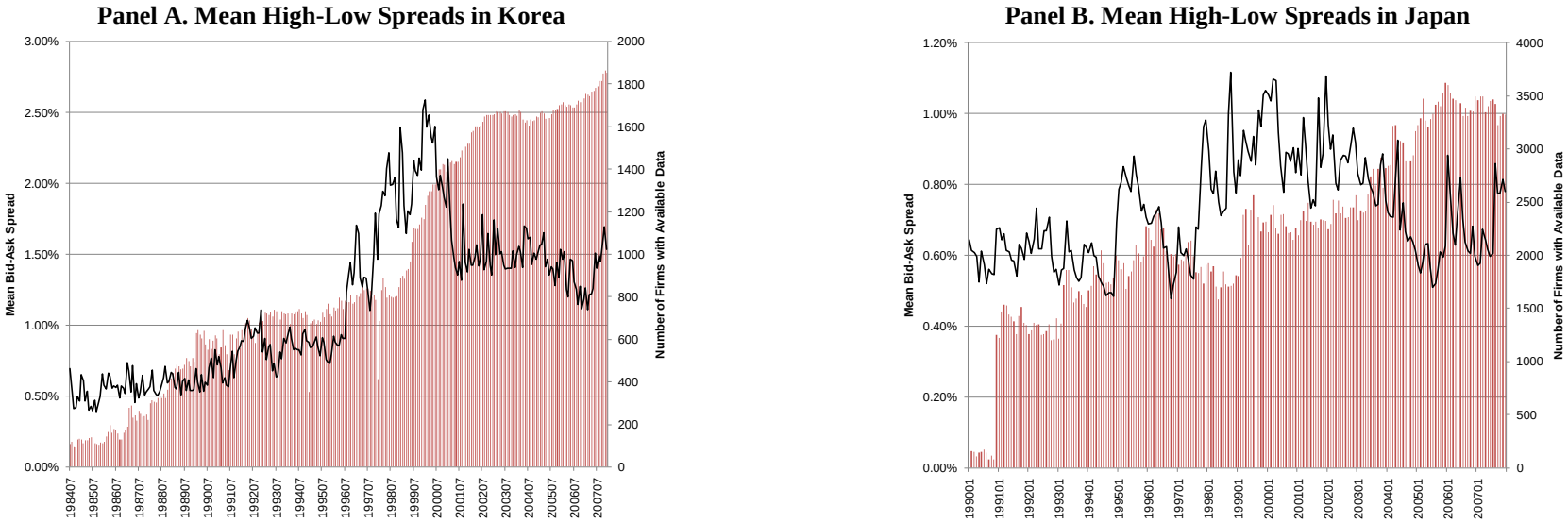
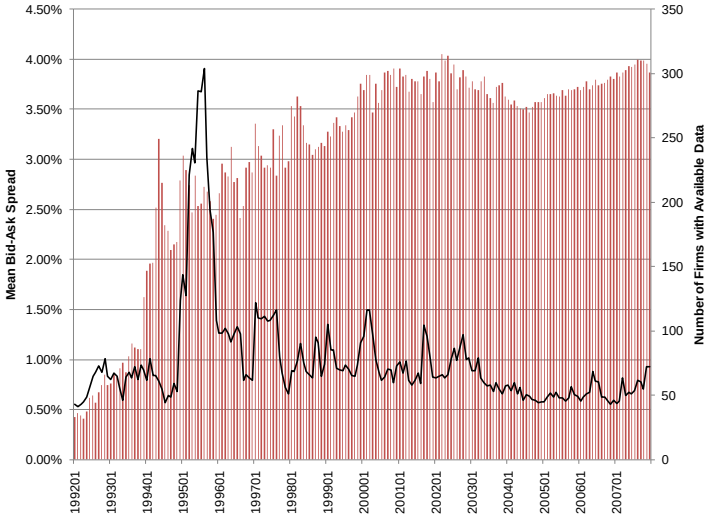
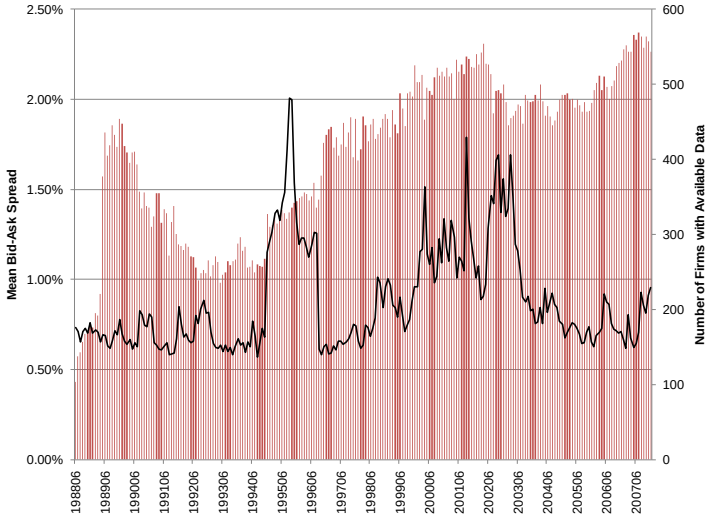


Figure IA.2. Average high-low spreads by month based on Datastream data. High-low spreads are estimated for each stock each month by averaging all overlapping two-day spread estimates within the month. The graph plots the equally weighted average spread by month across all stocks with at least 12 daily spread observations within the month. The graph also shows the number of firms included in the average each month. Panels A through I show results for stocks in Korea, Japan, Italy, France, Belgium, Sweden, the U.K., Brazil, and New Zealand, respectively. All data are from Datastream.

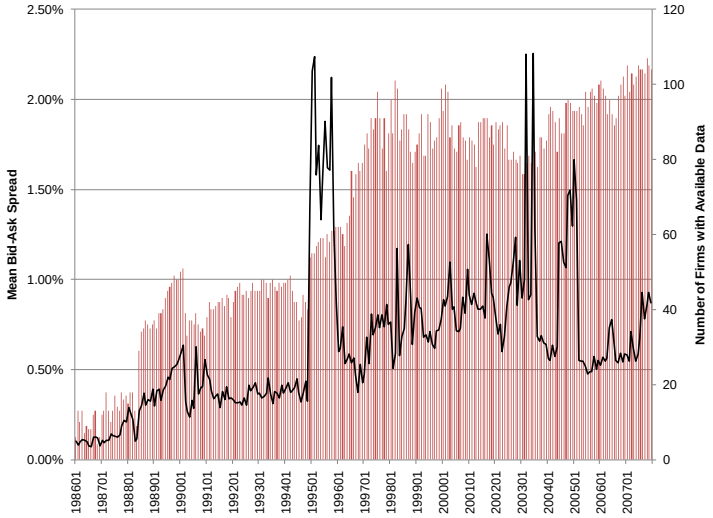
Panel C. Mean High-Low Spreads in Italy



Panel D. Mean High-Low Spreads in France



Panel E. Mean High-Low Spreads in Belgium



Panel F. Mean High-Low Spreads in Sweden

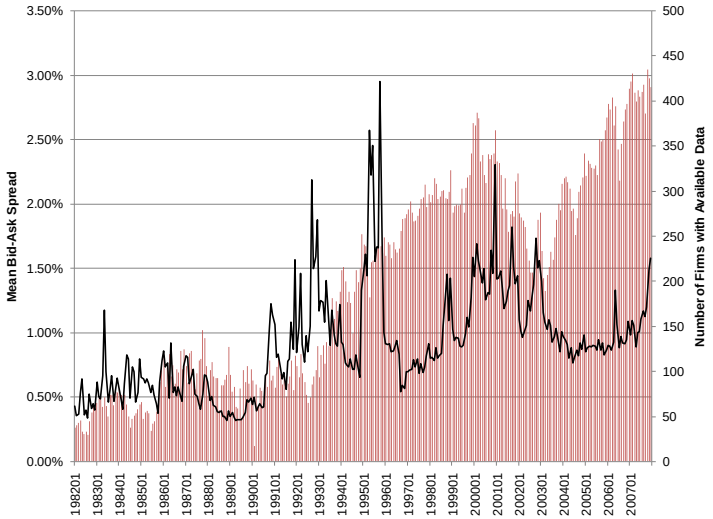
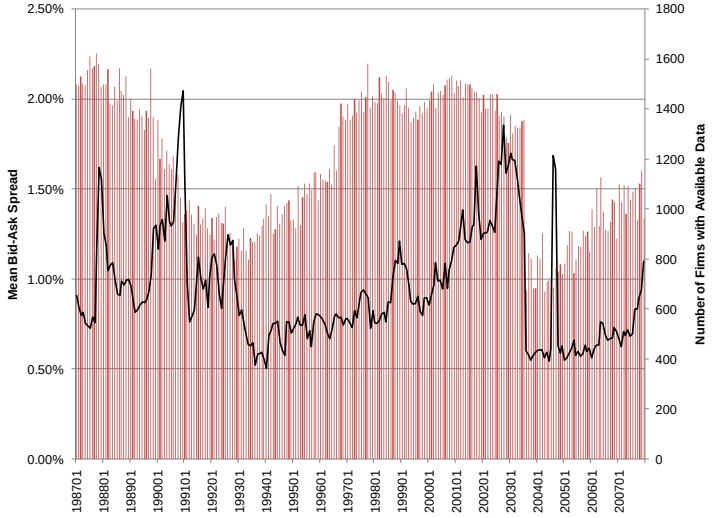
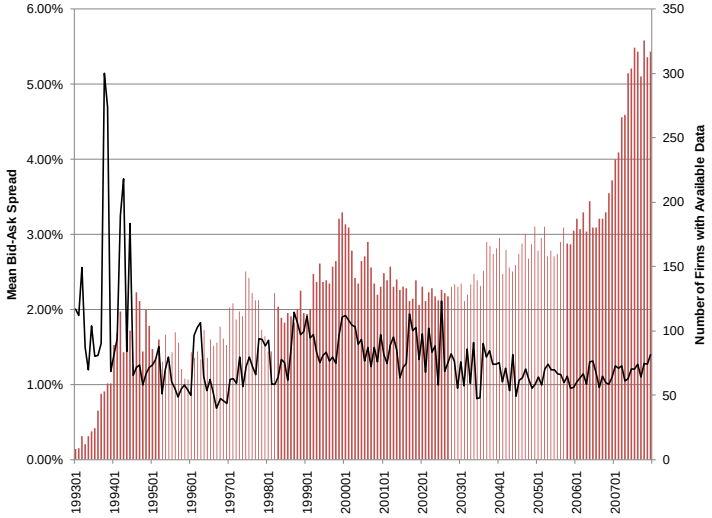


Figure IA.2 - continued

Panel G. Mean High-Low Spreads in the U.K.



Panel H. Mean High-Low Spreads in Brazil



Panel I. Mean High-Low Spreads in New Zealand

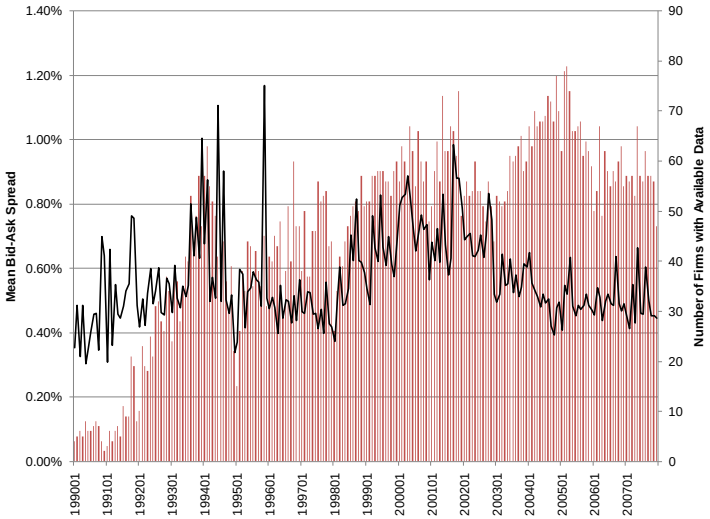


Figure IA.2 - continued

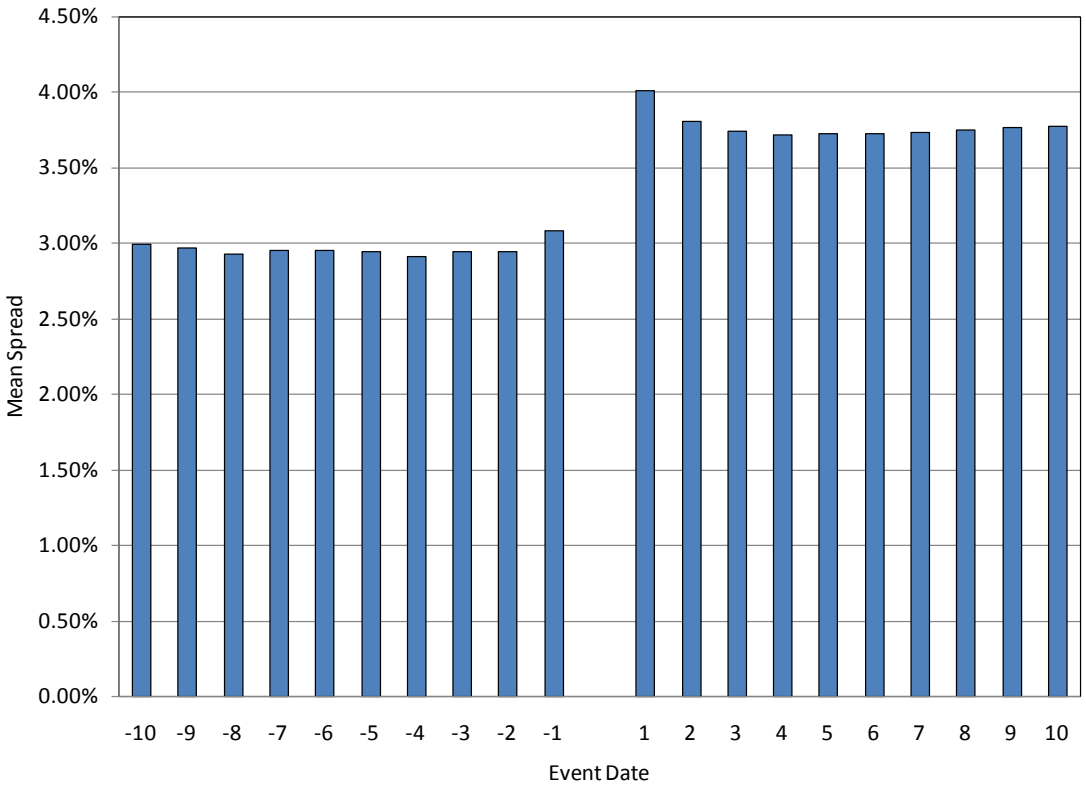


Figure IA.3. Average high-low spreads around stock splits from 1926 to 1982. The figure plots the cross-sectional average of daily high-low spreads for the 10 days prior to and 10 days after stock splits. The sample includes all stocks splits for NYSE, Amex, and NASDAQ firms from 1926 through 1982 that increase shares outstanding by at least 20%, where stock splits are identified by CRSP distribution code 5523. High-low spreads are estimated for each two-day period, as described in Section V. In general, daily estimates are then defined by averaging the high-low spread estimates for the two overlapping intervals that include that day. For the day before the split (day -1), we use just the spread estimated from days -2 to -1. For the first post-split day (day +1), we use just the spread estimated from days 0 to 1.

Table IA.XI
Summary Statistics for Daily 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 the 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.

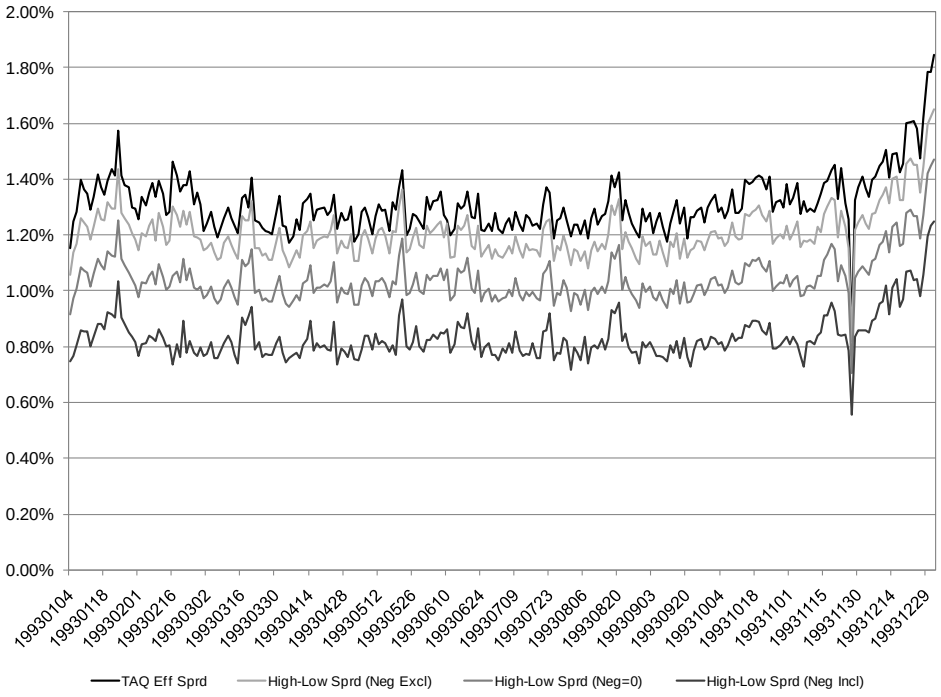
	1993	2006
Mean Spreads:		
TAQ Eff. Spread	1.320%	0.272%
High-Low Spread$_{NegIncl}$	0.837%	0.056%
High-Low Spread$_{Neg=0}$	1.048%	0.181%
High-Low Spread$_{NegExcl}$	1.214%	0.262%
Mean Absolute Errors:		
High-Low Spread$_{NegIncl}$	0.0054	0.0022
High-Low Spread$_{Neg=0}$	0.0035	0.0013
High-Low Spread$_{NegExcl}$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

Table IA.XII
Correlations of Daily High-Low Spreads 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 Spreads by Day in 1993



Panel B. Average High-Low Spreads by Day in 2006

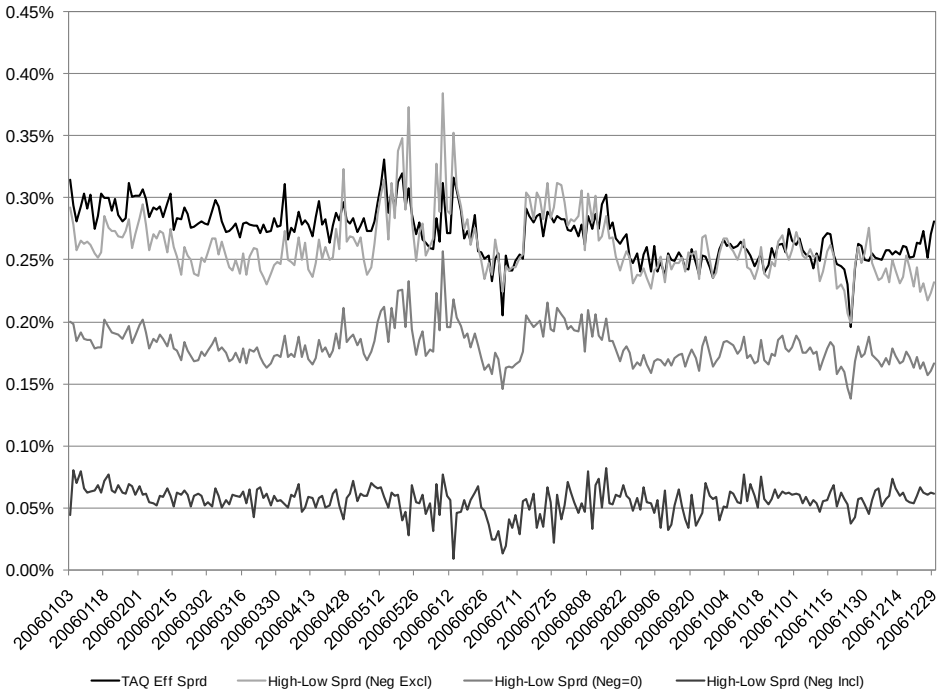


Figure IA.4. Daily patterns in high-low spreads. The figure plots cross-sectional average spreads by day. Results are shown for TAQ effective spreads as well as three alternative versions of the high-low estimator, where negative spread estimates within the day are (1) excluded, (2) set to zero, or (3) included, prior to calculating the daily average. Results for 1993 are shown in Panel A and results for 2006 are shown in Panel B.

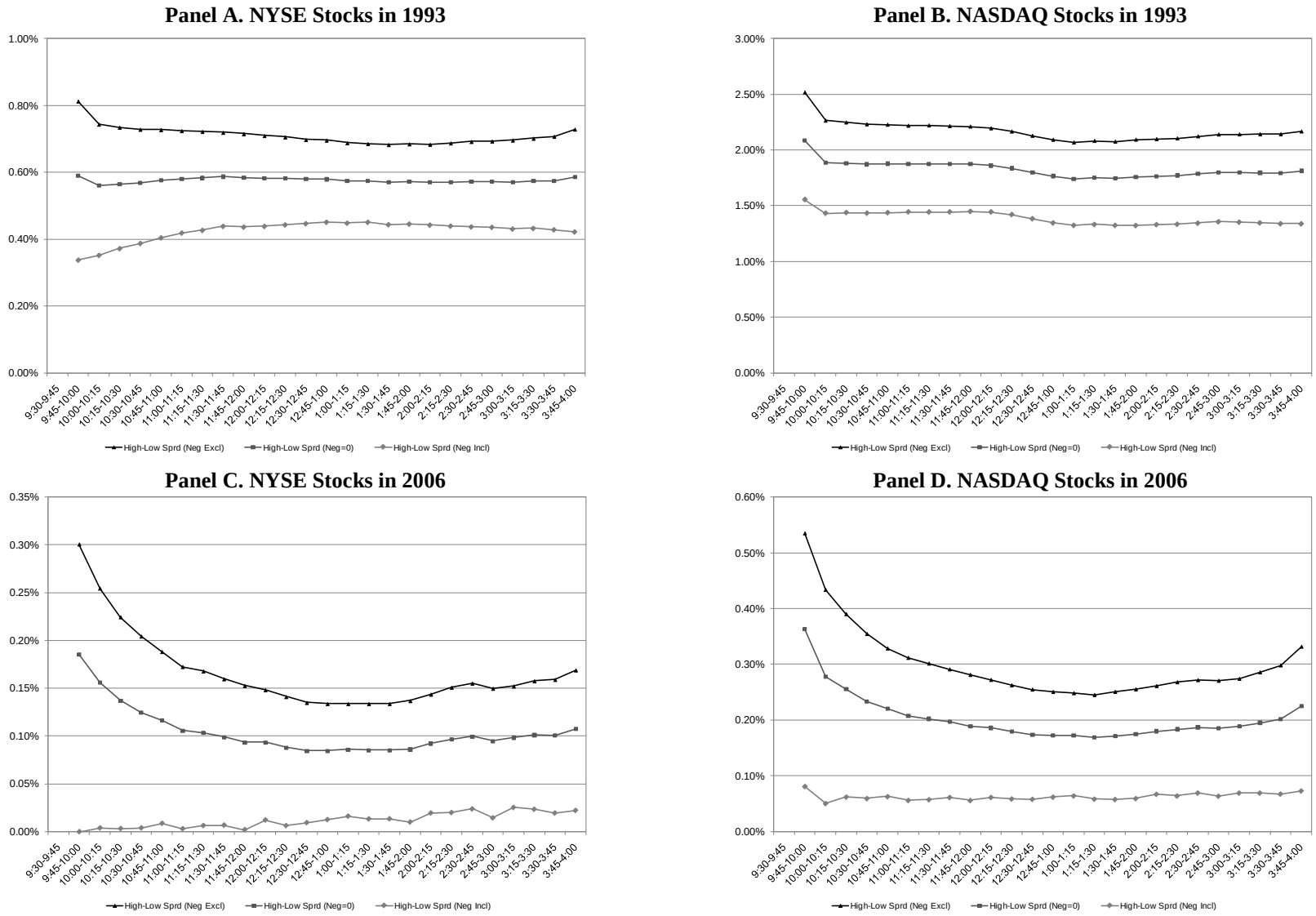


Figure IA.5. Intraday patterns in high-low spreads. The figure plots average high-low spreads across all days and stocks by 15-minute period. Results for NYSE and NASDAQ stocks in 1993 are shown in Panels A and B, respectively. Results for NYSE and NASDAQ stocks in 2006 are shown in Panels C and D, respectively. Results are shown for two alternative versions of the High-Low estimator, where negative spread estimates within the day are (1) excluded or (2) set to zero.

REFERENCES

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

Chan, K. C., W. G. Christie, and P. H. Schultz, 1994, Market structure and the intraday evolution of bid-ask spreads for NASDAQ securities, *Journal of Business* 68, 35-60.

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

Goyenko, Ruslan, Craig Holden, and Charles Trzcinka, 2009, Do liquidity measures measure liquidity? *Journal of Financial Economics* 92, 153-181.

McInish, Thomas H., and Robert A. Wood, 1992, An analysis of intraday patterns in bid/ask spreads for NYSE stocks, *Journal of Finance* 47, 753-764.

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