Internet Appendix to
"A Simple Estimation of Bid-Ask Spreads from Daily Close, High, and Low Prices"

Farshid Abdi and Angelo Ranaldo *

University of St. Gallen

ABSTRACT

This supplemental appendix extends the main paper by presenting additional analyses and robustness checks. It also describes the procedure to construct the Monthly TAQ effective spread benchmark.

* University of St. Gallen. Address: Swiss Institute of Banking and Finance, University of St. Gallen, Unterer Graben 21, CH-9000 St. Gallen, Switzerland. E-mail addresses: farshid.abdi@unisg.ch (F. Abdi), angelo.ranaldo@unisg.ch (A. Ranaldo).

Contents

List of Tables

Appendix IA.A Chances of High (Low) Prices Being Buyer-(Seller-) Initiated

To empirically analyze the relevance of assumption of buyer-(seller-) initiated high (low) prices, we analyze the Daily TAQ data for the same stock-months that are used in the main sample of the paper: The sample period spans from October 2003 to December 2015. For every stock-day we find the high (low) prices and in line with the Lee and Ready (1991) algorithm, we assign trade directions by comparing the trade prices with quote midpoints. We clean up quote data and match the trades and quotes as explained in the paper. We consider no trade direction, i.e. zero sign, for high (low) prices at midquotes instead of using the tick test because, by construction, the tick test would identify high (low) prices as buyer- (seller-) initiated.

Table IA.1 shows the results for the entire sample, and for five quintiles of stock-days sorted by the number of trades. Because the high (low) prices for every stock-day are not unique, it is possible to report the results in different ways. Panel A shows the percent ratio of the days that include high (low) prices above (below) midquote. Panel B shows the percent ratio of the days in which high (low) prices are more often above (below) midquotes. In Panel C, we also count the days in which the high (low) prices are always at the midquotes or the high (low) prices happen to be above and below midquotes by the same number of times, in the numerators of the reported percent ratios.

Three clear results emerge from this analysis. First, overall the assumption of buyer- (seller-) initiated high (low) prices seems to hold on average. Second, the assumption holds better for more-frequently traded quintiles. Finally, the percent of high prices above midquotes is very close to the percent of low prices below midquote. This result holds true for the entire sample and for each of the quintiles suggesting that the assumption (iii) in appendix C of the paper is in line with the data.

Appendix IA.B Additional Numerical Simulations

We perform additional numerical simulations within the framework of section II of the paper. Table IA.2 shows the results of this analysis. To allow consistent comparisons with the near-ideal settings, panel A of the table repeats the near-ideal simulation results included in the paper.

In panel B, we relax the assumption of equal likelihood of buyer-initiated and seller-initiated trades and set the probability of a buyer-initiated trade to be 90%. As shown in the table, in this setting Roll estimates show a very considerable downward bias and higher estimation error. As an example, for the 8% spread case, the Roll measure shows a -3.8% bias and 4.6% RMSE. In contrast, both *CHL* and *HL* estimates are only marginally sensitive to this setting.

In panel C, we include overnight price changes corresponding to a half standard deviation of daily price returns. Comparing panel A and C of the table, *CHL* monthly-corrected estimates are generally the least affected ones, both in terms of the change in levels and in RMSEs. Moreover, *CHL* two-day corrected estimates show similar RMSEs to the monthly-corrected estimates.

Appendix IA.C Additional Comparisons to the Daily TAQ Benchmark

In addition to the time-series correlations of individual stocks, which we present in Table VI of the paper, here we assess the accuracy of the estimates in portfolio levels. For this purpose, we construct equally-weighted portfolios for different groups of stocks. Table IA.3 shows the results of this analysis. Panel A (B) of the table shows the time-series correlations for the levels (changes) of effective spreads for the entire market portfolio, while portfolios are constructed by exchanges, market capitalization, and spread size in panels C, D, and E, respectively. In addition to our estimates, labeled as *CHL*, we include the estimates introduced by Corwin and Schultz (*HL*; 2012), Roll (*Roll*; 1984), Hasbrouck (*Gibbs*; 2009), Holden (2009), jointly with Goyenko, Holden, and Trzcinka (*EffTick*; 2009), and Fong, Holden, and Trzcinka (*FHT*; 2017). As shown in Panel A of the table, the highest correlation belongs to the end-of-day spreads. In the absence of end-of-day quote data, the three highest correlations belong to *EffTick*, *CHL*, and *HL* estimates, with correlation coefficients above 0.959. The correlation coefficients of *CHL* and *HL* are not significantly different with the one for the *EffTick* estimates. Moreover, *CHL* estimates show the highest correlations for the small cap and less liquid portfolios.

We repeat the prediction error comparisons in the paper by using MAEs rather than RMSEs. Table IA.4 shows the results for this analysis. As shown in table, after removing the stock-months with zero estimates in Panel B, the results are fully consistent with the ones for RMSEs reported in Table VII of the paper.

We compare the accuracy of *CHL* with combinations of other estimates. Table IA.5 shows the results. We construct the combination of other estimates in two ways: First, as is common in the literature (e.g. Chordia, Roll, and Subrahmanyam 2000), we compute averages across estimates, as we expect doing so will reduce the noise due to estimation errors in different estimates. As all the estimates proxy the size of the effective spread, we are entitled to compute a simple average. Second, we use the first principal component across the estimates to capture the common behavior. As comparison criteria, we use cross-sectional correlation, time-series correlation, and RMSEs with respect to the Daily TAQ benchmark. As shows in the table, *CHL* generally outperforms all combinations, with only one exception out of 30 comparisons, which corresponds to the average cross-sectional correlation of NYSE stocks. Even for this subsample, *CHL* provides the highest average time-series correlation as well as the lowest average RMSE. The results of this analysis show that our method provides more accurate estimates of effective costs than do (overidentified) combinations of estimates obtained from other methods. This speaks in favor of the joint utilization of the close, high, and low prices in building estimates of transaction costs, like in our model, rather than using more limited information sets, like in other models.

We reiterate the measurement of systematic liquidity risk using 25 portfolios sorted by illiquidity level as in Acharya and Pedersen (2005). We create 25 portfolios sorted by the TAQ effective spread benchmark and the bid-ask spreads estimates from the *CHL*, *HL*, and *Roll* models. Then we measure how close the LCAPM systematic risk betas of the estimated portfolio spreads follow the betas of the TAQ benchmark. More precisely, we apply the following procedure: First, we sort stocks using the TAQ effective spreads of previous month and construct 25 gross-return-weighted portfolios using gross returns of the previous month as weights, in line with Asparouhova, Bessembinder, and Kalcheva (2010, 2013). Second, we calculate betas using the TAQ effective spread benchmark. Third, we repeat the same steps of sorting stocks, constructing portfolios, and calculating betas for the bid-ask spread estimates of *CHL*, *HL*, and *Roll*. Finally, we apply the previous framework explained in equation (23) to assess the overall quality of the estimates. As reported in Table IA.6, the results are fully consistent with the ones for individual stocks. Moreover, when AR(2) shocks are used, *CHL* liquidity-related betas, that is β_2^{CHL}, β_3^{CHL}, and , β_4^{CHL} each show a cross-sectional correlation exceeding 0.95 with the ones of the TAQ benchmark. This suggests that *CHL* betas are very accurate for sorted portfolios.

Appendix IA.D Comparison with the Amihud Illiquidity Measure

In the paper, we compare different estimates with the Daily TAQ effective spread benchmark. To see the capability of the estimation models in measuring market illiquidity in a more general framework, we compare the estimates with Amihud (2002) illiquidity measure. Table IA.7 shows the correlation coefficients between different bid-ask spread estimates and the Amihud measure. We also subsample the estimates into five quintiles based on the average number of trades per day for every stock-month, observed from the Daily TAQ data. As shown in the table, end-of-day spreads show the highest correlations with the ILLIQ measure. In the absence of quote data, *CHL* estimates show the highest correlation with the Amihud measure for the entire sample as well as for each of the five quintiles.

Appendix IA.E Comparisons with the Alternative Benchmark of Monthly TAQ (1993-2003)

To assess the accuracy of our estimator over other sample periods, we compare the bid-ask spread estimates with the Monthly TAQ data between January 1993 and September 2003. This time period complements the Daily TAQ data time period of the paper and it extends our analysis to 23 years, i.e. from the beginning of 1993 to the end of 2015.

We construct the Monthly TAQ benchmark by calculating the proportional effective spreads as shown in equation (IA.1).

$$ES_t = \frac{2|P_t - M_t|}{M_t} \quad , \qquad M_t = \frac{B_t + A_t}{2}. \tag{IA.1}$$

Following Corwin and Schultz (2012), we calculate the effective spread from Monthly TAQ data as follows: First, we match the trades during the market opening hours with the provided National Best Bid and Offers (NBBOs) that are valid one second prior to the trades, and calculate proportional effective spreads[1]. Second, we exclude the incidents in which (a) the bid-ask spread is above $5, (b) the proportional bid-ask spread or the proportional effective spread are above 40%, and (c) the effective spread exceeds four times the quoted spread. Third, we calculate the trade weighted average of intraday proportional effective spreads by using the trade dollar volumes as weights to construct daily proportional effective spreads. Finally, we take the average of the daily values to construct the monthly effective spread benchmark.

We calculate the spread estimates by following the same procedure described in the paper. Finally, we match the CRSP-based estimates with the TAQ benchmark, using CUSIPs. This matching strategy allows us to cover 94.7% of stock-months estimates.

[1] Given that the sample ends in 2003, we do not expect that the high-frequency quoting produces the measurement problems explained by Holden and Jacobsen (2014). Therefore, we match prices with midquotes that are in-force in the prior second instead of using their interpolated time approach.

Table IA.8 shows the summary statistics for different estimates. In line with the results reported in the paper, three main findings emerge from this table: First, two-day corrected estimates for *HL* and *CHL* show higher correlations with the TAQ benchmark compared with the monthly corrected versions. Second, the correlation between *CHL* estimates and the effective spread benchmark is 0.857, which is the highest one, even higher than the CRSP end-of-day quoted spreads. This confirms the results of the paper that *CHL* is generally more accurate than other estimates. It is worth underlining that in this sample period, *CHL* outperforms the CRSP spreads. This analysis extends the results in Chung and Zhang (2014) showing that it is possible to obtain the most accurate estimates of the bid-ask spreads without using end-of-day spreads, at least for Monthly TAQ data sample period. Finally, *CHL* two-day corrected estimates have the nearest mean and standard deviation to the effective spread benchmark, showing that our estimates follow the benchmark very closely.

We repeat the analysis of Section III of the paper in terms of time-series and cross-sectional correlations, as well as predictions errors of different estimates. Table IA.9 shows the average cross-sectional correlations with the Monthly TAQ effective spread for different estimates. Surpassing end-of-day spreads, *CHL* has the highest correlation for the entire sample, for two out of the three exchanges, and for the two subperiods before 2001, when stock market decimalization occurs. Besides, in the absence of quote data, *CHL* almost always shows the highest average cross-sectional correlations or average correlations not significantly different with the highest one. The only exception, out of 21 comparisons, is for the most liquid quintile of stock, in which *EffTick* estimates have the highest average cross-sectional correlation with the TAQ benchmark. Moreover, the correlation coefficients are in general higher than the ones for more recent data (i.e. Daily TAQ data), which we use in the paper. This confirms the usefulness of our estimates when researchers need a simple but accurate measure of trading costs over long periods. The results from the time-series correlations in Table IA.10 are fully consistent with those in the paper, confirming that in the absence of quote data, *CHL* estimates show the highest average time-series correlations with the TAQ benchmark, and they have the highest correlations for less-liquid liquid stocks. Finally, we compare prediction errors in terms of RMSEs and MAEs in panels A, and B of Table IA.11, respectively. As shown in the table, *CHL* estimates have the lowest prediction errors for the entire sample, as well as for two out of the three exchanges. Moreover, in this former sample period, *CHL* estimation errors, in both terms of MAEs and RMSEs, are lower than end-of-day spreads, for the entire period, two out of the three exchanges, and the two subperiods before 2001, confirming that end-of-day spreads are less accurate before stock market decimalization.

All in all, analysis of the estimates' accuracy for the 1993-2003 period confirms the results of the paper and shows that *CHL* estimates are even more accurate in previous sample periods.

References

Acharya, Viral V., and Lasse H. Pedersen, 2005, Asset pricing with liquidity risk, Journal of Financial Economics 77, 375–410.

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

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

Asparouhova, Elena N., Hendrik Bessembinder, and Ivalina Kalcheva, 2013, Noisy prices and inference regarding returns, Journal of Finance 68, 665–714.

Chordia, Tarun, Richard Roll, and Avanidhar Subrahmanyam, 2000, Commonality in liquidity. Journal of Financial Economics 56, 3–28.

Chung, Kee H., and Hao Zhang, 2014, A simple approximation of intraday spreads using daily data, Journal of Financial Markets 17, 94–120.

Corwin, Shane A., and Paul Schultz, 2012, A simple way to estimate bid-ask spreads from daily high and low prices, Journal of Finance 67, 719–759.

Fong, Kingsley, Craig W. Holden, and Charles A. Trzcinka, 2017, What Are The Best Liquidity Proxies For Global Research?, forthcoming in the Review of Finance.

Goyenko, Ruslan Y., Craig W. Holden, and Charles A. Trzcinka, 2009, Do liquidity measures measure liquidity?, Journal of Financial Economics 92, 153–181.

Hasbrouck, Joel, 2009, Trading costs and returns for US equities: the evidence from daily data, Journal of Finance 64, 1445–1477.

Holden, Craig W., 2009, New low-frequency liquidity measures, Journal of Financial Markets 12, 778-813.

Holden, Craig W., and Stacey Jacobsen, 2014, Liquidity measurement problems in fast, competitive markets: expensive and cheap solutions, Journal of Finance 69, 1747–1785.

Lee, Charles M. C., and Mark J. Ready, 1991, Inferring trade direction from intraday data, Journal of Finance 46, 733–746.

Newey, Whitney K., and Kenneth D. West, 1987, A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, Econometrica 55, 703–708.

Roll, Richard, 1984, A simple implicit measure of the effective bid-ask spread in an efficient market, Journal of Finance 39, 1127–1139.

Table IA.1. Chance of High (Low) Prices Being Buyer-(Seller-) Initiated

For the Daily TAQ sample between October 2003 and December 2015, we find high and low prices per day and compare them with the quote midpoints. We do this for all stock-days and for five quintile groups sorted by the number of trades per day, from the lowest number of trades to the highest. Panel A shows percentage of stock-days that include high (low) prices above (below) midquotes. Panel B shows percentage of stock-days in which high (low) prices are more often above (below) midquotes. Panel C shows percentage of stock-days in which high (low) prices are more often (above) below midquotes, or the days with high (low) trades that are all at midquotes, or the days in which high (low) prices are above (below) midquote by the same number of times.

		Panel A		Panel B		Panel C	
	N	High	Low	High	Low	High	Low
All Stock-Days	12,027,382	90.0%	91.2%	86.0%	87.4%	89.1%	90.4%
NTD Quintile 1	2,405,476	78.8%	83.6%	74.4%	79.8%	77.8%	82.8%
NTD Quintile 2	2,405,476	89.7%	90.3%	85.7%	86.4%	89.1%	89.8%
NTD Quintile 3	2,405,476	91.8%	91.9%	87.6%	87.5%	91.1%	91.2%
NTD Quintile 4	2,405,476	93.9%	94.2%	90.0%	90.3%	93.0%	93.3%
NTD Quintile 5	2,405,478	95.7%	96.1%	92.2%	92.8%	94.6%	95.0%

Table IA.2. Estimated Bid-Ask Spreads from Additional Simulations

Each simulation consists of 10,000 21-day months of stock prices, and each day consists of 390 minutes. For each minute, the trajectory of a geometric Brownian motion with daily volatility of 3% and a constant relative spread with the values mentioned in the table is simulated. The labels in the first row refer to the estimators from the following models: ours (CHL), Corwin and Schultz's (HL; 2012), and Roll's (Roll; 1984). 2-Day and Monthly refer to the two-day corrected and monthly corrected versions, in which two-day or monthly negative estimates are set to zero. We run the simulations in five separate scenarios. Panel A shows the results for the near-ideal conditions. Panel B shows the results in a situation where 90% of trades are buyer-initiated. Panel C shows the results when there are "overnight" price movements and the standard deviation of the overnight price change is 50% of the standard deviation of the daily price change. The overnight adjustment procedure for HL estimates is exactly as in Corwin and Schultz (2012).

	Bias					RMSE				
	CHL		HL		Roll	CHL		HL		Roll
	2-Day	Monthly	2-Day	Monthly		2-Day	Monthly	2-Day	Monthly	
Panel A: Near Ideal Conditions										
0.5% Spread	0.7%	0.2%	0.9%	0.1%	0.7%	0.8%	0.8%	1.0%	0.5%	1.5%
1.0% Spread	0.3%	0.0%	0.8%	0.0%	0.3%	0.5%	0.8%	0.8%	0.6%	1.5%
3.0% Spread	-0.6%	-0.1%	0.2%	-0.1%	-0.4%	0.8%	0.7%	0.5%	0.6%	1.9%
5.0% Spread	-0.7%	0.0%	0.0%	-0.1%	-0.4%	0.9%	0.6%	0.6%	0.6%	2.2%
8.0% Spread	-0.4%	0.0%	-0.2%	-0.2%	-0.5%	0.7%	0.5%	0.6%	0.6%	2.7%
Panel B: 90% Buyer-Initiated Trades										
0.5% Spread	0.7%	0.2%	0.9%	0.0%	0.6%	0.8%	0.8%	0.9%	0.5%	1.5%
1.0% Spread	0.3%	-0.2%	0.6%	-0.1%	0.2%	0.5%	0.8%	0.7%	0.6%	1.4%
3.0% Spread	-0.7%	-0.3%	0.1%	-0.3%	-1.3%	0.9%	0.7%	0.5%	0.7%	2.1%
5.0% Spread	-0.9%	-0.2%	-0.2%	-0.3%	-2.4%	1.0%	0.5%	0.6%	0.7%	3.1%
8.0% Spread	-0.7%	-0.2%	-0.3%	-0.3%	-3.8%	0.8%	0.5%	0.7%	0.7%	4.6%
Panel C: Overnight Price Movement										
0.5% Spread	0.9%	0.3%	0.9%	-0.1%	0.8%	1.0%	0.9%	1.0%	0.5%	1.7%
1.0% Spread	0.5%	0.0%	0.6%	-0.4%	0.4%	0.6%	1.0%	0.7%	0.7%	1.6%
3.0% Spread	-0.5%	-0.1%	0.0%	-0.6%	-0.4%	0.8%	0.9%	0.5%	1.0%	2.0%
5.0% Spread	-0.8%	-0.1%	-0.4%	-0.7%	-0.5%	1.0%	0.7%	0.8%	1.0%	2.4%
8.0% Spread	-0.6%	0.0%	-0.7%	-0.7%	-0.5%	0.9%	0.6%	1.0%	1.0%	2.8%

Table IA.3. Time-Series Correlations of Estimates for Equally Weighted Portfolio of Stocks Compared to the TAQ Benchmark

The labels in the first row refer to our estimator (CHL) and the estimators proposed by Corwin and Schultz (HL; 2012), Roll (Roll; 1984), Hasbrouck (Gibbs; 2009), Holden (EffTick; 2009), Fong, Holden, and Trzcinka (FHT; 2017), and Chung and Zhang (CRPS_S; 2014). N is the average number of stocks per month. The correlations are calculated between the spread estimate for each equally weighted portfolio and the high frequency benchmark. To compare estimators in the absence of quote data, we exclude the CRSP_S and an asterisk indicates numbers that are not significantly different from the estimator with the highest correlation, which is marked in bold in every row. We use a paired t-test for the statistical inferences. The size quintiles are sorted by increasing market capitalization at the last observed period for each individual stock. The spread quintiles are sorted by increasing average effective spreads during the whole sample period.

	N	CHL	HL	Roll	Gibbs	EffTick	FHT	CRSP_S
Panel A: Time-Series Correlations with Effective Spreads: Monthly Estimates								
Full Period	3’944.7	0.964*	0.959*	0.889	0.931	**0.968**	0.930	0.987
2003–2007	4’380.5	0.760	0.762	0.537	0.589	**0.893**	0.841*	0.936
2008–2011	3’870.8	0.978*	**0.980**	0.895	0.959*	0.969*	0.951	0.989
2012–2015	3’555.5	0.622	0.382	0.337	0.330	**0.925**	0.902*	0.965
Panel B: Time-Series Correlations with Changes in Effective Spreads: Monthly Estimates								
Full Period	3’895.7	0.812	**0.883**	0.530	0.561	0.845*	0.741	0.947
2003–2007	4’266.2	0.756*	0.773*	0.429	0.227	**0.794**	0.565	0.938
2008–2011	3’765.9	0.830*	**0.911**	0.558	0.712	0.884*	0.812*	0.950
2012–2015	3’461.3	0.639*	**0.713**	0.225	0.026	0.665*	0.144	0.874
Panel C: Time-Series Correlations for Separate Markets								
NYSE	1’337.1	0.816	0.798	0.700	0.798	**0.902**	0.863*	0.845
AMEX	297.1	**0.919**	0.869	0.821	0.896*	0.816	0.812	0.947
NASDAQ	2’310.5	0.959*	0.948*	0.903	0.923	**0.965**	0.935	0.989
Panel D: Time-Series Correlations for Five Market Capitalization Quintile Portfolios								
Size Quintile 1	595.4	**0.977**	0.974*	0.922	0.945	0.926	0.876	0.988
Size Quintile 2	678.3	0.938	0.928	0.909	0.901	**0.962**	0.921	0.977
Size Quintile 3	755.8	0.909	0.887	0.803	0.848	**0.943**	0.915*	0.965
Size Quintile 4	860.4	0.849*	0.844*	0.689	0.800	**0.891**	0.854*	0.953
Size Quintile 5	1’054.9	0.671	0.683	0.540	0.660	**0.853**	0.827*	0.893
Panel E: Time-Series Correlations for Five Effective Spread Quintile Portfolios								
ES Quintile 1	1’035.0	0.624	0.628	0.504	0.632	**0.814**	0.707	0.806
ES Quintile 2	841.4	0.728*	0.735*	0.579	0.671*	**0.762**	0.705*	0.835
ES Quintile 3	753.2	0.826*	0.820*	0.700	0.784*	**0.852**	0.826*	0.922
ES Quintile 4	715.9	**0.944**	0.936*	0.884	0.903	0.936*	0.911	0.977
ES Quintile 5	599.1	**0.972**	0.963*	0.953	0.956	0.936	0.818	0.986

Table IA.4. Mean-Absolute Errors

We measure the accuracy of different monthly estimates by computing their mean absolute errors (MAEs) with respect to the TAQ benchmark. Prediction errors are calculated every month and then averaged through the months in the sample. N is the average number of stocks per month. The labels refer to our estimator (CHL) and the estimators proposed by Corwin and Schultz (HL; 2012), Roll (Roll; 1984), Hasbrouck (Gibbs; 2009), Holden (EffTick; 2009), Fong, Holden, and Trzcinka (FHT; 2017), and Chung and Zhang (CRPS_S; 2014). To compare estimators in the absence of quote data, we exclude the CRSP_S and an asterisk indicates numbers not significantly different from the estimator with the lowest average prediction error marked in bold in every row. We test our hypotheses on the time series of pairwise difference in prediction errors for two estimators and assess whether the mean is significantly different from zero. We adjust for any potential time-series autocorrelation by using Newey-West (1987) standard errors with four lags autocorrelation.

	N	CHL	HL	Roll	Gibbs	EffTick	FHT	CRSP_S
Panel A: MAEs, Breakdown for Different Periods, and Across Different Markets								
Full Period	3,944.7	0.0082	0.0082	0.0133	0.0146	0.0180	**0.0067**	0.0020
2003–2007	4,380.5	0.0067	0.0068	0.0112	0.0130	0.0154	**0.0055**	0.0015
2008–2011	3,870.8	0.0113	0.0110	0.0180	0.0177	0.0234	**0.0087**	0.0028
2012–2015	3,555.5	0.0067	0.0070	0.0108	0.0131	0.0154	**0.0059**	0.0018
NYSE	1,337.1	0.0076	0.0067	0.0098	0.0131	0.0074	**0.0016**	0.0006
AMEX	297.1	**0.0083**	0.0092	0.0190	0.0149	0.0511	0.0135	0.0041
NASDAQ	2,310.5	**0.0086**	0.0090	0.0146	0.0154	0.0201	0.0088*	0.0026
Panel B: MAEs, Stock-Months with Zero Estimates are Removed								
Full Period	648.4	**0.0087**	0.0095	0.0173	0.0136	0.0391	0.0103	0.0034
2003–2007	819.2	**0.0068**	0.0073	0.0142	0.0117	0.0300	0.0076	0.0022
2008–2011	617.8	**0.0120**	0.0128	0.0230	0.0172	0.0518	0.0141	0.0050
2012–2015	497.5	**0.0101**	0.0114	0.0226	0.0186	0.0733	0.0155	0.0029
NYSE	125.3	0.0085	0.0073	0.0168	0.0138	0.0186	**0.0025**	0.0009
AMEX	80.8	**0.0081**	0.0092	0.0188	0.0141	0.0853	0.0130	0.0044
NASDAQ	442.4	**0.0090**	0.0101	0.0173	0.0135	0.0368	0.0119	0.0038

Table IA.5. Comparison with Combinations of Models

We combine the bid-ask spread estimates of different models by computing simple averages and then compare the results with our estimates. The labels in the first row refer to our estimator (CHL) and the estimators proposed by Corwin and Schultz (HL; 2012), Roll (Roll; 1984), Hasbrouck (Gibbs; 2009), Holden (EffTick; 2009), and Fong, Holden, and Trzcinka (FHT; 2017). "All" refers to the last five mentioned estimators. In panels A and C, N refers to the average number of stocks per month. In panel B, N refers to the number of stocks in the subsamples with at least 6 months of estimates. The spread quintiles are sorted by increasing average effective spreads during the whole sample period. Cross-sectional correlations are calculated per month and averaged across the sample. Time-series correlations are calculated for each individual stock and then averaged across assets. RMSEs are calculated for every month and then averaged through time.

	N	CHL	Simple Average		First Principal Comp.	
			HL & Roll	All	HL & Roll	All
Panel A: Average Cross-Sectional Correlations with the TAQ Benchmark						
All Stocks, Levels	3,944.7	0.738	0.544	0.579	0.462	0.482
All Stocks, Changes	3,895.7	0.298	0.167	0.129	0.129	0.060
NYSE	1,337.1	0.495	0.316	0.478	0.243	0.524
AMEX	297.1	0.735	0.575	0.537	0.492	0.385
NASDAQ	2,310.5	0.710	0.510	0.531	0.437	0.442
ES Quintile 1	1,035.0	0.385	0.232	0.300	0.157	0.358
ES Quintile 2	841.4	0.375	0.233	0.323	0.166	0.363
ES Quintile 3	753.2	0.400	0.236	0.351	0.173	0.357
ES Quintile 4	715.9	0.497	0.326	0.350	0.271	0.288
ES Quintile 5	599.1	0.693	0.523	0.467	0.449	0.306
Panel B: Average Time-Series Correlations for Spread Estimates of Individual Stocks						
All Stocks, Levels	7,210	0.518	0.348	0.410	0.273	0.384
All Stocks, Changes	7,124	0.287	0.170	0.177	0.130	0.129
NYSE	2,174	0.430	0.275	0.358	0.208	0.371
AMEX	831	0.557	0.369	0.433	0.294	0.355
NASDAQ	4,587	0.540	0.367	0.421	0.290	0.386
ES Quintile 1	1,442	0.409	0.253	0.309	0.187	0.330
ES Quintile 2	1,442	0.420	0.262	0.323	0.189	0.344
ES Quintile 3	1,442	0.460	0.293	0.383	0.220	0.384
ES Quintile 4	1,442	0.588	0.394	0.468	0.315	0.419
ES Quintile 5	1,442	0.711	0.538	0.567	0.454	0.445
Panel C: Root-Mean-Squared Errors w.r.t TAQ Benchmark						
All Stocks, Levels	3,944.7	0.0104	0.0140	0.0148	0.0228	0.0477
All Stock, Changes	3,895.7	0.0082	0.0149	0.0148	0.0268	0.0475
NYSE	1,337.1	0.0089	0.0108	0.0098	0.0171	0.0219
AMEX	297.1	0.0115	0.0171	0.0235	0.0292	0.1014
NASDAQ	2,310.5	0.0111	0.0151	0.0157	0.0245	0.0479
ES Quintile 1	1,035.0	0.0076	0.0088	0.0074	0.0133	0.0115
ES Quintile 2	841.4	0.0098	0.0120	0.0108	0.0187	0.0216
ES Quintile 3	753.2	0.0105	0.0140	0.0139	0.0229	0.0371
ES Quintile 4	715.9	0.0103	0.0149	0.0177	0.0253	0.0621
ES Quintile 5	599.1	0.0145	0.0208	0.0233	0.0336	0.0873

Table IA.6. Estimating Systematic Liquidity Risks from the LCAPM Model for 25 Portfolios of Stocks Sorted by Illiquidity Level

We calculate the components of systematic risk implied by the LCAPM model (Acharya and Pedersen (2005)) for 25 portfolios sorted by illiquidity levels, constructed using the Daily TAQ effective spreads, Roll model estimates (Roll; 1984), the HL estimates (Corwin and Schultz; 2012), and the estimates from our model (labeled CHL). The table reports the cross-sectional correlation of portfolio betas based on Roll's, HL's, and CHL's estimates ($\beta_i^{Estimates}$), with portfolio betas based on the TAQ effective spreads ($\boldsymbol{\beta_i^{ES}}$). Portfolio spreads are in levels or as shocks, defined by residuals of AR(2) regressions. An asterisk indicates values not significantly different from that with the higher correlation marked in bold for every set of values. The statistical inferences are performed using Fisher's z-test.

Cross-Sectional Correlations of Betas with those Calculated using TAQ Effective Spreads (*continued)*

	$\rho(\beta_1^{ES}, \beta_1^{Estimates})$			$\rho(\beta_2^{ES}, \beta_2^{Estimates})$			$\rho(\beta_3^{ES}, \beta_3^{Estimates})$		
	CHL	HL	Roll	CHL	HL	Roll	CHL	HL	Roll
Levels	**0.663**	0.587*	0.240	**0.978**	0.975*	0.924	**0.257**	-0.378	0.087*
Shocks	**0.663**	0.587*	0.240	0.968*	**0.972**	0.213	**0.951**	0.939*	0.693

Table IA.6. Continued

Cross-Sectional Correlations of Betas with those Calculated using TAQ Effective Spreads

	$\rho(\beta_4^{ES}, \beta_4^{Estimates})$			$\rho(\beta_{Net}^{ES}, \beta_{Net}^{Estimates})$		
	CHL	HL	Roll	CHL	HL	Roll
Levels	**0.916**	0.862*	0.340	**0.723**	0.663*	0.309
Shocks	**0.956**	0.944*	0.683	**0.717**	0.652*	0.300

Table IA.7. Correlations with the Amihud Measure for Quintiles Based on Average Number of Trades (2003-2015)

The table shows the correlation coefficients between different monthly estimates and the Amihud illiquidity measure (Amihud, 2002). We group the stocks into five quintiles sorting them by their average number of trades per day during the sample period. The daily number of trades is counted using the Daily TAQ consolidated trades data for trades that occur between 9:30 and 16:00 and have a positive price and volume. The first four quintiles are constructed of 1,392 stocks, and the fifth is constructed of 1,393 stocks. The labels in the first row refer to our estimator (CHL) and the estimators proposed by Corwin and Schultz (HL; 2012), Roll (Roll; 1984), Hasbrouck (Gibbs; 2009), Holden (EffTick; 2009), Fong, Holden, and Trzcinka (FHT; 2017), and Chung and Zhang (CRPS_S; 2014). N refers to the number of stock-months of estimates for the entire sample, as well as for each quintile. To compare estimators in the absence of quote data, we exclude the CRSP_S and an asterisk indicates numbers not significantly different from the estimator with the highest correlation, using Fisher's z-test to compare the correlation coefficients.

	N	CHL	HL	Roll	Gibbs	EffTick	FHT	CRSP_S
Full Sample	579872	**0.406**	0.325	0.261	0.215	0.186	0.195	0.463
ANTD Quintile 1	77978	**0.515**	0.444	0.362	0.426	0.225	0.234	0.474
ANTD Quintile 2	103920	**0.409**	0.329	0.274	0.239	0.196	0.175	0.503
ANTD Quintile 3	110083	**0.320**	0.265	0.186	0.146	0.187	0.185	0.496
ANTD Quintile 4	130725	**0.181**	0.145	0.088	0.065	0.125	0.096	0.406
ANTD Quintile 5	157166	**0.161**	0.132	0.089	0.065	0.160*	0.146	0.479

Table IA.8. Summary Statistics for Different Estimators, January 1993-September 2003

This table provides the main summary statistics for the pooled sample of the main estimators considered in this paper. The column label N refers to the number of stock-months of estimates in the sample. The column label $\rho\ (\,.\,, ES_{i,t})$ refers to the correlation of different estimates with the Monthly TAQ effective spread benchmark. The row labels, refer to the TAQ effective spread benchmark (Effective Spread), our estimator (*CHL*), and the estimators proposed by Corwin and Schultz (HL; 2012), Roll (Roll; 1984), Hasbrouck (Gibbs; 2009), Holden (EffTick; 2009), and Fong, Holden, and Trzcinka (FHT; 2017). For the sake of completeness, we include the CRSP end-of-day bid-ask spreads (*CRSP_Spread*) as in Chung and Zhang (2014). For the calculation of *CHL* estimates, we replace the missing high, low, and close price with the previous days' values. We then discard monthly estimates for the months with fewer than 12 trading days (that is, days with positive high, low, and close price as well as positive volume). The *HL* estimates are calculated exactly as in Corwin and Schultz (2012); that is, (1) missing daily high and low prices are replaced with those of previous days, (2) overnight adjustments are applied, and (3) monthly estimates with fewer than 12 two-day estimates are discarded. We merge the results of different estimators and discard stock-months in which any of the estimates are missing. We compute two versions of the *HL* (*CHL*) estimator, that is, the two-day corrected and monthly corrected versions labeled Two-Day and Monthly. In the two-day corrected version for *HL* (*CHL*), we set each negative two-day spread (squared spread) to zero, and then the spreads (square roots of estimated squared spreads) are averaged within a month. The monthly corrected *HL* estimates are calculated by averaging all the two-day spreads within the month, and then setting negative monthly averages to zero. The monthly *CHL* estimates are calculated by inserting monthly averages in equation (9), then setting negative estimates of squared spreads to zero and finally, taking the square roots. The *Roll* estimates are calculated by setting positive monthly autocovariance estimates to zero. The zeros reported for *EffTick* estimates reflect the months in which none of the prices were divisible by the base-eight denomination increments. We consider a second variant of *EffTick* measure (*EffTick – Alt. Incr.*) using the tick sizes of 1¢, 5¢, 10¢, 25¢, 50¢, or $1.00 as our sample time span lies after the decimalization of stock markets.

	N	Mean	Median	Standard Deviation	$\rho\ (\,.\,, ES_{i,t})$	$\% \leq 0$
Effective Spread	646'182	2.14%	1.27%	2.64%	1.000	0.00%
CHL - Two-Day	646'182	2.30%	1.59%	2.37%	0.857	0.00%
CHL - Monthly	646'182	2.27%	1.34%	3.09%	0.806	28.58%
HL - Two-Day	646'182	1.94%	1.33%	2.07%	0.840	0.00%
HL - Monthly	646'182	1.16%	0.47%	1.93%	0.792	28.13%
Roll	646'182	2.37%	1.11%	3.81%	0.599	40.86%
Gibbs	646'182	3.22%	2.35%	3.17%	0.695	0.00%
EffTick	646'182	2.76%	1.03%	5.90%	0.608	3.10%
EffTick – Alt. Incr.	646'182	1.20%	0.51%	2.16%	0.598	0.43%
FHT	646'182	1.25%	0.68%	1.98%	0.654	21.89%
CRSP_S	646'182	3.27%	2.15%	3.65%	0.836	0.03%

Table IA.9. Average Cross-Sectional Correlations with the TAQ Benchmark, January 1993-September 2003

This table shows the average cross-sectional correlations between estimates of transaction costs and the Monthly TAQ benchmark at each month. The monthly correlations are averaged over the specified sample periods. The labels in the first row refer to our estimator (CHL) and the estimators proposed by Corwin and Schultz (HL; 2012), Roll (Roll; 1984), Hasbrouck (Gibbs; 2009), Holden (EffTick; 2009), Fong, Holden, and Trzcinka (FHT; 2017), and Chung and Zhang (CRPS_S; 2014). N is the average number of stocks per month. To compare estimators in the absence of quote data, we exclude the CRSP_S and an asterisk indicates numbers that are not significantly different from the estimator with the highest correlation, which are marked in bold in every row. We test our hypotheses on the time-series of pairwise difference in correlations for two estimators and assess whether the mean is significantly different from zero. We adjust for any potential time-series autocorrelation by using Newey-West (1987) standard errors with four lags autocorrelation. The size quintiles are sorted by increasing market capitalization at the last observed period for each individual stock. The spread quintiles are sorted by increasing average effective spreads during the whole sample period.

	N	CHL	HL	Roll	Gibbs	EffTick	FHT	CRSP_S
Panel A: Average Cross-sectional Correlations with Effective Spreads for Monthly Estimates								
Full Period	5009.2	**0.861**	0.833	0.605	0.713	0.637	0.644	0.846
1993-1995	3922.4	**0.812**	0.808*	0.609	0.747	0.562	0.607	0.787
1996-2000	5830.4	**0.890**	0.869	0.620	0.737	0.728	0.684	0.836
2001-2003	4701.5	**0.860**	0.795	0.572	0.635	0.552	0.614	0.927
Panel B: Average Cross-sectional Correlations with Changes in Effective Spreads for Monthly Estimates								
Full Period	4925.6	**0.471**	0.460	0.206	0.266	0.194	0.153	0.578
1993-1995	3732.1	**0.328**	0.322*	0.151	0.226	0.135	0.110	0.387
1996-2000	5668.1	**0.524**	0.520*	0.227	0.300	0.268	0.185	0.611
2001-2003	4507.6	**0.513**	0.482	0.215	0.235	0.100	0.137	0.730
Panel C: Analysis across Different Markets								
NYSE	1578.4	0.810*	0.808*	0.453	0.629	**0.812**	0.755	0.856
AMEX	353.1	**0.929**	0.918	0.651	0.846	0.788	0.743	0.850
NASDAQ	4925.6	**0.471**	0.460	0.206	0.266	0.194	0.153	0.578
Panel D: Analysis across Market Capitalization								
Size Quintile 1	717.7	**0.836**	0.807	0.562	0.715	0.520	0.567	0.820
Size Quintile 2	851.2	**0.810**	0.769	0.532	0.645	0.537	0.519	0.790
Size Quintile 3	910.2	**0.761**	0.709	0.476	0.547	0.542	0.473	0.765
Size Quintile 4	1135.0	**0.706**	0.683	0.411	0.478	0.577	0.474	0.702
Size Quintile 5	1395	0.603*	0.613*	0.306	0.406	**0.621**	0.505	0.607
Panel E: Analysis across Effective Spread Size								
ES Quintile 1	1360.6	0.425	0.437	0.152	0.293	**0.657**	0.423	0.503
ES Quintile 2	1117.7	**0.579**	0.575*	0.273	0.337	0.525*	0.424	0.629
ES Quintile 3	962.6	**0.668**	0.627	0.359	0.430	0.482	0.413	0.689
ES Quintile 4	920.9	**0.749**	0.707	0.449	0.565	0.489	0.456	0.737
ES Quintile 5	647.4	**0.800**	0.776	0.536	0.696	0.481	0.510	0.777

Table IA.10. Average Time-Series Correlations for Spread Estimates of Individual Stocks Compared to the TAQ Benchmark, January 1993-September 2003

The labels in the first row refer to our estimator (*CHL*) and the estimators proposed by Corwin and Schultz (HL; 2012), Roll (Roll; 1984), Hasbrouck (Gibbs; 2009), Holden (EffTick; 2009), Fong, Holden, and Trzcinka (FHT; 2017), and Chung and Zhang (CRPS_S; 2014). N is the number of stocks in the subsamples with at least 6 months of estimates. The averages are computed across stocks. To compare estimators in the absence of quote data, we exclude the CRSP_S and an asterisk indicates numbers that are not significantly different from the estimator with the highest correlation, which is marked in bold in every row. We use a paired t-test for the statistical inferences. The size quintiles are sorted by increasing market capitalization at the last observed period for each individual stock. The spread quintiles are sorted by increasing average effective spreads during the whole sample period.

	N	CHL	HL	Roll	Gibbs	EffTick	FHT	CRSP_S
Panel A: Average Time-Series Correlations with Effective Spreads: Monthly Estimates								
Full Period	10783	**0.586**	0.580	0.280	0.445	0.464	0.402	0.778
1993-1995	6137	0.496	0.504	0.229	0.379	**0.527**	0.243	0.634
1996-2000	9130	0.558	0.575	0.252	0.415	**0.590**	0.350	0.777
2001-2003	5805	**0.568**	0.539	0.211	0.446	0.273	0.249	0.761
Panel B: Average Time-Series Correlations with Changes in Effective Spreads: Monthly Estimates								
Full Period	10676	**0.401**	0.400*	0.155	0.285	0.221	0.116	0.586
1993-1995	5976	**0.300**	0.294	0.119	0.218	0.256	0.090	0.464
1996-2000	8974	0.403	**0.413**	0.150	0.289	0.326	0.111	0.619
2001-2003	5679	**0.420**	0.414	0.143	0.317	0.098	0.091	0.577
Panel C: Analysis across Different Markets								
NYSE	2754	0.343	0.353	0.095	0.286	**0.460**	0.408	0.584
AMEX	1078	**0.644**	0.622	0.285	0.505	0.600	0.350	0.649
NASDAQ	7744	**0.645**	0.632	0.326	0.482	0.450	0.383	0.863
Panel D: Analysis across Market Capitalization								
Size Quintile 1	2155	**0.796**	0.776	0.475	0.686	0.582	0.489	0.867
Size Quintile 2	2157	**0.708**	0.682	0.371	0.555	0.480	0.357	0.810
Size Quintile 3	2157	**0.653**	0.637	0.302	0.457	0.431	0.327	0.807
Size Quintile 4	2157	0.519	**0.534**	0.201	0.340	0.420	0.390	0.763
Size Quintile 5	2157	0.252	0.270	0.054	0.188	0.405	**0.447**	0.646
Panel E: Analysis across Effective Spread Size								
ES Quintile 1	2157	0.200	0.209	0.014	0.171	0.335	**0.402**	0.600
ES Quintile 2	2157	0.526	**0.553**	0.189	0.325	0.458	0.400	0.778
ES Quintile 3	2157	**0.686**	0.674	0.325	0.483	0.484	0.376	0.838
ES Quintile 4	2157	**0.743**	0.716	0.406	0.574	0.524	0.401	0.842
ES Quintile 5	2155	**0.773**	0.746	0.468	0.673	0.519	0.431	0.835

Table IA.11. Prediction Errors, January 1993-September 2003

We measure the accuracy of different monthly estimates by computing their root-mean-squared errors (RMSEs) as well as mean absolute errors (MAEs) with respect to the TAQ benchmark. Prediction errors are calculated every month and then averaged through the months in the sample. N is the average number of stocks per month. The labels refer to our estimator (CHL) and the estimators proposed by Corwin and Schultz (HL; 2012), Roll (Roll; 1984), Hasbrouck (Gibbs; 2009), Holden (EffTick; 2009), Fong, Holden, and Trzcinka (FHT; 2017), and Chung and Zhang (CRPS_S; 2014). To compare estimators in the absence of quote data, we exclude the CRSP_S and an asterisk indicates numbers that are not significantly different from the estimator with the lowest average prediction error, which is marked in bold in every row. We test our hypotheses on the time-series of pairwise difference in prediction errors for two estimators and assess whether the mean is significantly different from zero. We adjust for any potential time-series autocorrelation by using Newey-West (1987) standard errors with four lags autocorrelation.

	N	CHL	HL	Roll	Gibbs	EffTick	FHT	CRSP_S
Panel A: RMSEs, Breakdown for different periods, and across different markets								
Full Period	5009.2	**0.0142**	0.0151	0.0309	0.0252	0.0466	0.0225	0.0224
1993-1995	3922.4	**0.0205**	0.0217	0.0313	0.0233	0.0420	0.0288	0.0266
1996-2000	5830.4	**0.0113**	0.0121	0.0296	0.0237	0.0398	0.0198	0.0250
2001-2003	4701.5	**0.0126**	0.0135	0.0328	0.0299	0.0639	0.0205	0.0133
NYSE	1578.4	0.0065	**0.0059**	0.0168	0.0161	0.0182	0.0074	0.0213
AMEX	353.1	**0.0118**	0.0156	0.0319	0.0186	0.0522	0.0238	0.0466
NASDAQ	3077.7	**0.0223**	0.0234	0.0408	0.0341	0.0591	0.0326	0.0208
Panel B: MAEs, Breakdown for different periods, and across different markets								
Full Period	5009.2	**0.0080**	0.0085	0.0183	0.0146	0.0173	0.0120	0.0124
1993-1995	3922.4	**0.0080**	0.0088	0.0166	0.0113	0.0140	0.0133	0.0156
1996-2000	5830.4	**0.0074**	0.0077	0.0188	0.0149	0.0138	0.0115	0.0137
2001-2003	4701.5	0.0093*	**0.0095**	0.0195	0.0176	0.0274	0.0115	0.0066
NYSE	1578.4	0.0048	0.0042*	0.0101	0.0104	0.0048*	**0.0040**	0.0144
AMEX	353.1	**0.0079**	0.0110	0.0207	0.0111	0.0237	0.0154	0.0314
NASDAQ	3077.7	**0.0145**	0.0152	0.0267	0.0211	0.0281	0.0210	0.0133