Bid–Ask Spread Estimator from High and Low Daily Prices: A Note on its Practical Implementation for Corporate Bonds

Belén Nieto
University of Alicante

ABSTRACT

Corwin and Schultz (2012) have proposed an estimator of transaction costs for assets with unobservable bid–ask spreads. The estimator is based on high and low daily prices and needs the asset to trade in two consecutive days. Recently, this measure has been estimated in the corporate bond market, where most bonds are infrequently traded. I show that this practice can produce an important upward bias, even when bonds with high activity requirements are selected. I propose a modification of the original measure that allows its practical estimation for non-continuously traded assets and to quantify the magnitude of the bias due to discontinuities.

Keywords: Transaction cost; Corporate bonds; Liquidity

JEL: G12

1. Introduction

This paper suggests a more general version of the bid–ask spread estimator proposed by Corwin and Schultz (2012; hereafter CS) that can be implemented even in the case of infrequently traded assets. I show that the spread estimator theoretically depends on the time interval between observable prices. The practical application of the standard CS measure requires continuous trading. As long as this is not true, their estimator produces an upward bias. It must be noted that my aim is not to infer any conclusion about the appropriateness of the CS measure as a liquidity proxy. My generalization suffers from all the drawbacks of the standard measure because it relays on the same theoretical assumptions. Instead, my proposal improves its practical estimation because 1) avoids making assumptions about the price during a day without trades and 2) adjusts the resulting estimator incorporating the gap between observable prices.

CS propose an estimator for the unobservable bid–ask spread that only requires high and low daily prices. They assume that the stock price follows a constant diffusion process, and the daily high price (H) is a buyer-initiated trade and the daily low price (L) is a seller-initiated trade. Then, the log high-low ratio for observable (o) and true/actual (A) prices for day t are related as following:

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

with $\alpha = ln(2 + S/2 - S)$ and S is the spread.

Additionally, the authors assume that the spread is constant over two-day periods and then the equation for the log of the high–low log ratio over the two days is

$$ln\left(\frac{H_{t,t+1}^o}{L_{t,t+1}^o}\right) = ln\left(\frac{H_{t,t+1}^A}{L_{t,t+1}^A}\right) + \alpha. \qquad (2)$$

The square of equations (1) and (2) illustrates the main idea of the paper: an observed change in price has one component due to price volatility and another to the bid–ask spread. The volatility is proportional to the data frequency, while the spread is not. Therefore, working simultaneously with two frequencies (one and two days), it is possible to identify (and estimate) the two components. If additionally the Jensen's inequality is ignored, the solution is

$$\propto = \frac{\sqrt{2\beta}-\sqrt{\beta}}{3-2\sqrt{2}} - \sqrt{\frac{\gamma}{3-2\sqrt{2}}}, \qquad (3)$$

$$\sigma = \frac{\sqrt{\beta/2}-\sqrt{\beta}}{(3-2\sqrt{2})\sqrt{k_1}} + \sqrt{\frac{\gamma}{(3-2\sqrt{2})k_1}}, \qquad (4)$$

where β is the sum for two consecutive days of the square of the high–low log ratio, and γ is the square of the high–low log ratio over the two days.

$$\beta = \sum_{j=0}^{1}\left[ln\left(\frac{H_{t+j}}{L_{t+j}}\right)\right]^2, \quad \gamma = \left[ln\left(\frac{H_{t,t+1}}{L_{t,t+1}}\right)\right]^2.$$

Therefore, to estimate this measure for one specific day, observable high and low prices in two consecutive days are required. That is, to estimate a daily series of this measure, ideally the asset must be traded on all days and at least twice a day. In the original paper of CS, the practical implementation of the estimator is for the stock market, where the assumption of observable consecutive prices is reasonable. For those days when the stock shows only one trade, the authors propose an adjustment that allows estimating high and low different prices. And for the rare cases of no trades during a day, high and low prices are assumed to be the same than those observed in the most recent prior trading day. However, authors recognize that this assumption overstates the spread because it imposes the variance component to be zero.

Recently, this liquidity measure has been translated to the corporate bond market by Schestag, Schuster, and Uhrig-Homburg (2016; SSU hereafter), among others. Theirs is a very relevant paper that compares a huge set of liquidity proxies employing both intraday and daily data. The authors conclude that the CS measure is one of the best proxies because it appropriately captures differences in cross section and time-varying patterns. Their practical estimation of the measure exactly follows that in the original paper. However, corporate bonds are not continuously traded and the number of days without trades is higher than for stocks. Therefore, it is expected that the imposition of zero volatility for days without trades would produce an upward bias in the spread higher in the corporate bond market than in the stock market.

This paper illustrates that this is the case. I find that the empirical application of the CS transaction cost proxy to the corporate bond market can be problematic, even when the sample consists of bonds with high liquidity requirements.[1] I propose a modified version of the CS measure that accounts for irregular intervals between two observable prices. Thus it can be applied when assets are infrequently traded and avoids making assumptions about the volatility for days with non-observable prices. I empirically compare the standard CS estimator and my adjusted measure and find that the former is 16% higher on average in time and across bonds. The difference between the two estimators is statistically significant for 1,661 individual bonds from a total sample of 3,526. Generally speaking, significant differences appear for bonds with an average distance between trades of, at least, 1.0948 days.

The reminder of the paper is organized as follows. Section 2 derives the modified (adjusted) CS measure. Section 3 describes the data. Section 4 includes information about the organization of the sample and the practical estimation of the two liquidity proxies. Section 5 provides the results of comparing the two estimators. Finally, Section 6 concludes the paper.

2. A generalized bid–ask estimator that accounts for non-consecutive trading days

The CS spread estimation for one day requires observable prices for two consecutive days ($n = 2$). However, corporate bonds (and other financial assets) are not continuously traded daily and thus high and low prices are not observable for all days in a sample period. I generalize this measure to be also applied when trades are not consecutive ($n > 2$), such that the bid–ask spread is only estimated for days with trades and incorporates the effect of a time gap of n - 1 days between two observable prices.

The foundation of my proposal is the same as for CS: I combine the information at a daily frequency with the information at a lower frequency; however, this other frequency changes each day and is defined by the observation of prices.

Following equation (1), the square of the high–low ratio for day t is:

$$\left[ln\left(\frac{H_t^O}{L_t^O}\right)\right]^2 = \left[ln\left(\frac{H_t^A}{L_t^A}\right)\right]^2 + 2\alpha ln\left(\frac{H_t^A}{L_t^A}\right) + \alpha^2. \qquad (5)$$

Assuming that it is possible to have trades that are not on consecutive days, with n - 1 days between the current trade and the previous one, I define the sum of (5) over n single days in the period as follows:

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

Taking the expectations in (6):

$$E\left\{\sum_{j=0}^{n-1}\left[ln\left(\frac{H_{t+j}^O}{L_{t+j}^O}\right)\right]^2\right\} = \sum_{j=0}^{n-1} E\left\{\left[ln\left(\frac{H_{t+j}^A}{L_{t+j}^A}\right)\right]^2\right\} + 2\alpha\sum_{j=0}^{n-1} E\left[ln\left(\frac{H_{t+j}^A}{L_{t+j}^A}\right)\right] + n\alpha^2. \qquad (7)$$

Parkinson (1980) and Garman and Klass (1980) show that, under the constant diffusion assumption, an estimator for the variance of the price change in a time interval can be obtained using only the distance between the maximum and the minimum prices observed in the interval. The advantage of this estimator is that we do not need to measure the continuous price sample path during the interval and, however, it is more efficient than the standard estimator that uses closing prices, for example. Using the derivation of the probability distribution for the distance between the maximum and the minimum prices in Feller (1951), Parkinson (1980) shows that the moments of the distance are proportional to the variance. In particular, the two first moments are

$$E\left\{\left[ln\left(\frac{H_t^A}{L_t^A}\right)\right]^2\right\} = k_1\sigma^2 \text{ and } E\left[ln\left(\frac{H_t^A}{L_t^A}\right)\right] = k_2\sigma, \qquad (8)$$

with $k_1 = 4ln(2)$, $k_2 = \sqrt{8/\pi}$, and σ the volatility of each single day.

Introducing equations (8) into (7), we obtain

$$E\left\{\sum_{j=0}^{n-1}\left[ln\left(\frac{H_{t+j}^o}{L_{t+j}^o}\right)\right]^2\right\} = nk_1\sigma^2 + 2\alpha nk_2\sigma + n\alpha^2. \qquad (9)$$

The left-hand side of (9) is denoted β^n and thus the expression is

$$nk_1\sigma_{HL}^2 + 2\alpha nk_2\sigma_{HL} + n\alpha^2 - \beta^n = 0. \qquad (10)$$

On the other hand, the square of the high–low ratio from the n-day period is

$$\left[ln\left(\frac{H_{t,t+n-1}^o}{L_{t,t+n-1}^o}\right)\right]^2 = \left[ln\left(\frac{H_{t,t+n-1}^A}{L_{t,t+n-1}^A}\right)\right]^2 + 2\alpha ln\left(\frac{H_{t,t+n-1}^A}{L_{t,t+n-1}^A}\right) + \alpha^2. \qquad (11)$$

Again, taking the expectations in (11), incorporating the volatility proportionalities given in (8), and denoting the left-hand side as γ^n, we obtain

$$nk_1\sigma^2 + 2\alpha\sqrt{n}k_2\sigma + \alpha^2 - \gamma^n = 0. \qquad (12)$$

Using the equations (10) and (12) we can solve for the two unknowns α and σ. Moreover, if we ignore Jensen's inequality, and thus $k_2\sigma = \sqrt{k_1}\sigma$, the closed-form solutions are:

$$\alpha^* = \frac{\sqrt{n\beta^n} - \sqrt{\beta^n}}{n+1-2\sqrt{n}} - \sqrt{\frac{\gamma^n}{n+1-2\sqrt{n}}} \qquad (13)$$

$$\sigma^* = \frac{\sqrt{\beta^n/n} - \sqrt{\beta^n}}{(n+1-2\sqrt{n})\sqrt{k_1}} + \sqrt{\frac{\gamma^n}{(n+1-2\sqrt{n})k_1}}. \qquad (14)$$

Equations (13) and (14) match the standard CS estimators (equations (3) and (4)) if the asset is daily traded continuously and n =2 for all the days during the asset trading life. But my estimators can also be used for the general case of n - 1 days between trades. In the practical application, the n-day period is defined by the availability of prices. Therefore, on the one hand, my proposal does not impose to estimate the spread and volatility all the days. If there is no trade one day, the estimators are not computed. In contrast to the standard case in which the volatility estimator would be imposed to be zero and then the spread proxy would be upward biased. On the other hand, for days with observable prices, if $n > 2$, both the spread and the volatility adjusted estimators are lower than the standard one because they recognize the distance between trades. The two effects produce a positive difference between the standard CS spread estimator and the adjusted one. Next I empirically evaluate the magnitude and significance of such a difference.

3. Data

The initial sample consists of intraday U.S. corporate bond transaction data from the Trade Reporting and Compliance Engine (TRACE) and the corresponding bond characteristics from the Fixed Income Securities Database (FISD), all provided by Mergent, from July 2002 to December 2014. I apply the filters described by Dick-Nielsen (2009, 2014) to clean duplicates, corrections, and reversals and a median filter with five standard deviations computed with a previous window of 90 natural days is used to eliminate extreme outliers and erroneous reports.[2]

From the intraday frequency data, I construct a daily sample. For each bond and day, I record the high, low, and last prices, the number of trades within the day, and the cumulative trading volume at the end of the day. Then, some requirements are applied. I include only bonds with FISD information. I delete trades on holidays or days outside of the bond's life, defined by the offering and maturity dates available in the FISD. Additionally, bonds in default are only included up to three months before the default date and after the reinstated date, if any. This provides a final sample of 89,654 bonds. I complete the daily information with the rating associated with each bond and day using the historical rating changes for each of the four rating types available in the FISD: Standard & Poor's, Moody's, Fitch, and DP Information. Descriptive statistics of this complete and clean database are provided in panel A of Table 1.

The aim of this paper is to analyze the magnitude of the possible bias that the standard CS liquidity proxy produces for corporate bonds' daily data. To make comparisons, I apply the same sample selection criterion as SSU do: The period is from October 1, 2004, to September 30, 2012, and the bonds must show at least one year of active trading and must be traded on at least 75% of the trading days. This selection produces a sample of 3,526 bonds.[3] Panel B of Table 1 contains the descriptive statistics of these selected bonds.

Comparing panel B of Table 1 with the results in Table 1 of SSU, we can see the similitude of the samples. My selected bonds have lower dispersion in maturities, are slightly better rated, and show a slightly higher number of trades per bond.[4] Comparing panels A and B and looking at the median (Q0.5), for example, we find selected bonds to be more active, as expected. They show an average number of trades of 7.58 per day (2.56 in the entire sample) and a total number of trades of 6,677 (61 in the entire sample) and are traded on 91.31% of the days (17.52% in the entire sample). Additionally, the offering amount is considerably larger for these more active bonds. Other differential characteristics in the selected sample regarding the full universe are the following: Bonds have longer maturities, larger coupons, worse rating classifications, and a slightly higher Treasury spread and more than the 50% of the bonds are issued by industrial firms.

4. Transaction cost proxies and practical issues

For each bond, I estimate the daily series of the bid–ask spread, $S = \frac{2(e^{\alpha}-1)}{1+e^{\alpha}}$, using both the standard CS estimator in equation (3) and the adjusted estimator in equation (13).

In the standard case,

$$\hat{\alpha} = \frac{\sqrt{2\hat{\beta}} - \sqrt{\hat{\beta}}}{3-2\sqrt{2}} - \sqrt{\frac{\hat{\gamma}}{3-2\sqrt{2}}},$$

$$\hat{\beta} = \left[log\left(\frac{H_{t-1}}{L_{t-1}}\right)\right]^2 + \left[log\left(\frac{H_t}{L_t}\right)\right]^2,$$

$$\hat{\gamma} = \left[log\left(\frac{H_{t-1,t}}{L_{t-1,t}}\right)\right]^2,$$

where H_t (L_t) is the highest (lowest) price on day t and $H_{t-1,t}$ ($L_{t-1,t}$) is the highest (lowest) price in the period from t - 1 to t. When the bond does not trade at all during one day, I adopt the practical solution suggested by CS and use the most recent high and low prices. The estimator is computed for all days in the bond's life.

In the adjusted case, the estimator is computed only for those days with at least one trade:

$$\widehat{\alpha^*} = \frac{\sqrt{n\,\widehat{\beta^n}} - \sqrt{\widehat{\beta^n}}}{n+1-2\sqrt{n}} - \sqrt{\frac{\widehat{\gamma^n}}{n+1-2\sqrt{n}}},$$

$$\widehat{\beta^n} = \left[log\left(\frac{H_{t-n+1}}{L_{t-n+1}}\right)\right]^2 + \left[log\left(\frac{H_t}{L_t}\right)\right]^2,$$

$$\widehat{\gamma^n} = \left[log\left(\frac{H_{t-n+1,t}}{L_{t-n+1,t}}\right)\right]^2,$$

where n - 1 is the number of days between the current and the previous observable prices and $H_{t-n+1,t}$ ($L_{t-n+1,t}$) is the highest (lowest) price in the period from t - n + 1 to t.

For both estimators, I follow the adjustment adopted by CS for those days in which the high and low prices are equal; I assume the previous observable high and low prices if the price today is within the previous price range. In other cases, the previous high or low prices are corrected with the excess of today's price and the previous price range. If the spread estimator is negative for a day, it is set to zero.

To analyze and compare the two proxies, we split the sample of bonds into four subsamples regarding precisely the number of days without trade (observable prices). Specifically, for each bond, I compute the average number of days between trades. The minimum is one for bonds that are traded all days during their life and the maximum is 1.3361. I use the quartiles of the distribution of this variable to split the sample so that the four subsamples contain the same number of bonds (882 in subsamples 1 and 3 and 881 in subsamples 2 and 4) but where the average number of days between trades is below 1.0252 for subsample 1, ranges from 1.0252 to 1.0948 for subsample 2, ranges from 1.0948 to 1.2015 for subsample 3, and is above 1.2015 for subsample 4.

The four panels in Table 2 contain descriptive statistics for the subsamples. For example, bonds in subsample 1 have a very low value for the number of days between trades, 1.01, on average, with also a very low standard deviation. Looking at the median values, we see they are traded 17.57 times per day, on average, with a trading volume of 5,577,183 USD per day, and show a total of 15,618 trades during their life. Comparing among panels, the activity indicators decrease from panel A to panel D. The number of days with a trade, the average number of trades per day, the total number of trades in the bond's life, and the trading volume decrease monotonically from subsample 1 to subsample 4. Additionally, relations between bond characteristics and liquidity patterns are observed; more active bonds have lower maturity, a much higher offering amount, lower coupons, and better rating classifications. Finally, the most active bonds are issued in higher proportions by financial institutions.

5. Comparison between the standard and the adjusted CS measures

As explained before, the daily series of the standard and adjusted bid–ask spread estimators are obtained for each bond. In this section, their comparison is carried out. First, I compare their cross-sectional distributions within each subsample of bonds, and at a daily frequency. Second, I compare the time series distributions at individual level. Additionally I evaluate the time series patterns of the two estimators at aggregate level and at a monthly frequency in this case.

5.1. Cross-sectional results

For each bond, I compute the time series mean of the standard and the adjusted proxies. Then, I define the relative difference between the two means as the difference between the standard and the adjusted measures divided by the standard one. Descriptive statistics (the mean, the standard deviation, and the three quartiles) of the cross-sectional distribution of these three series are computed within each subsample of bonds. These are displayed in panels A to D of Table 3.

Rows 1-3 in each panel of Table 3 provide cross-sectional descriptive statistics for the mean of the standard measure, the mean of the adjusted measure, and their relative difference, respectively. Row 4 displays the value of the statistics to test the null hypothesis that descriptive statistics are equal for the two measures and row 5 provides the corresponded p-value. In all cases I use nonparametric tests which are shortly described below. To complete the perspective I also compute standard parametric tests of equal means (t-test) or variances (F-test) and p-values are provided in brackets in row 6.

For the comparison of the two means I use the Wilcoxon rank sum test with the null that the two samples come from identical continuous distributions with the same mathematical expectation. For the comparison of standard deviations I apply the Ansari-Bradley test that is also based on the rank sums but makes inference about the equally of the two dispersion parameters. This test requires that the two samples have equal medians. If this is not the case (on the basis of the conclusion from the previous test), I first subtract the medians. In both cases, the test statistics are normally distributed under the null and the alternative hypothesis is two-sided. The comparison between the three quartiles is done by a Pearson's chi-squared test. The null is that the frequency distribution in the observed samples is consistent with a theoretical distribution. In the case of two samples the statistic is:

$$\sum_{i=1}^{2} \frac{\left(O_i^{<q} - E_i^{<q}\right)^2}{E_i^{<q}} + \sum_{i=1}^{2} \frac{\left(O_i^{\geq q} - E_i^{\geq q}\right)^2}{E_i^{\geq q}},$$

where O_i is the observed frequency for sample i and E_i is the expected theoretical frequency, for values lower than q and higher or equal than q. The expected frequency is estimated with the values of the two samples simultaneously and q indicates the quantile: 0.25, 0.5 or 0.75. Under the null, the difference between the observed and the expected frequencies for the two samples is zero and the statistic has a Chi-squared distribution with one degree of freedom. Finally, the p-value in the last column of Table 3 refers to the Kolmogorov-Smirnov nonparametric test to compare the complete distributions of values in the two samples.

Starting with panel A of Table 3, the first column indicates that the mean value across all the bonds in subsample 1 of the mean of the standard CS proxy is 0.89%. The corresponding value for the adjusted CS proxy is slightly lower (0.87%), confirming an upward difference for the standard measure of 2.51%, on average. The second column indicates that the standard proxy shows higher cross-sectional dispersion. The columns for the three quartiles again show that the values for the standard measure are higher than for the corresponding adjusted measure, such that upward difference occurs in the complete distribution. Although the magnitude of the relative difference is higher in the left tail (for bonds with smaller transaction costs).

In panels B to D, we find the same pattern than in panel A: The mean of the standard CS measure is upward biased in relation to the adjusted measure for the full cross-sectional distribution and the magnitude of the bias decreases with the level of transaction costs.[5] And comparing the numbers in the different panels of Table 3, we can see that the bias increases as the trading frequency decreases. On the basis of the median, the bias is 1.85%, 8.71%, 21.11%, and 31.56% from the most to the less traded bonds, respectively.

The results of the equality tests depend on the panel. In panel A, all the tests that compare the five statistics and the complete distribution indicate that differences between the standard and the adjusted estimators are not significant for this subsample which contains the most liquid bonds. In the second subsample (panel B), I find that means are significantly different but the null of equal quartiles cannot be rejected at 5% level. Consistently, the test to compare distributions cannot reject the null at the standard significance level. Results in panels C and D are clearly different. In both panels and for all different tests we can conclude that the cross-sectional distributions of the standard and the adjusted estimators are significantly different. Therefore, the assumption of continuous trading and the use of the standard CS expression produce the bid–ask spread component to be overestimated, even when we work with a restrictive subsample of highly traded bonds. And the longer the gap between available data, the greater the difference between the standard CS proxy and my adjusted proposal. This difference is considerably large and statistically significant for bonds with an average number of days between trades higher than 1.0948.

Finally, it is striking that the bid–ask spread estimator (both the standard and the adjusted one) will be higher for the most active bonds than the others. It is not in the interest of this paper to investigate the reasons, but this result could be related to trade sizes (e.g., splitting strategies). SSU find that transaction costs are significantly higher for smaller trades. As seen in Table 2, the number of trades per day is related to the number of days with at least one trade. A higher number of trades could be associated with smaller sizes. To shed light on this point, I estimate an alternative illiquidity proxy that analyzes the price impact in terms of volume: the Amihud (2002) ratio. The cross-sectional statistics for each subsample of bonds show that, in this case, the mean, Q0.25, Q0.5, and Q0.75 are monotonously increasing from the most to the less active bonds. This result points out the importance of the two dimensions, price and quantity, of the liquidity function.

5.2. Time series results

In this final section, I focus on the time series behavior of the standard and the adjusted CS estimators. In this case I compare the time series distribution of the two transaction costs estimators for each individual bond. As before, specifically I compare the means, the standard deviations, the three quartiles, and the complete distributions and test their equality with the nonparametric tests described in Section 5.1. Table 4 provides, for each descriptive statistic and for each subsample, the number of bonds for which the null is rejected and thus significant differences between the two measures are obtained.

As expected, the two estimators are very similar in the subsample of the most traded bonds and the null of equal means, standard deviations, and complete distributions cannot be rejected for any bond in this group. The only significant difference is observed for the first quartile for 162 bonds. The number of rejections increases from subsample 1 to subsample 4. In subsample 2, I find that around 10% of the bonds show differences in the mean and the standard deviation between the standard and the adjusted measures. The proportion is lower if we compare the median or the complete distribution. Consistently with the cross-sectional analysis, the number of significant differences are much higher in subsamples 3 and 4. Around 80% (95%) of the bonds in subsample 3 (4) show relevant differences in descriptive statistics of the distribution between the standard and the adjusted CS measures.

Finally, I compute aggregate measures as the average across the whole sample of 3,526 bonds, and at monthly frequency, by averaging daily series within each month. Figure 1 illustrates the evolution of the two aggregate transaction cost estimators. Again, the upward difference between the standard CS measure and the adjusted one is confirmed. On average, the standard measure produces a bid–ask spread estimator 0.153%[6] higher than the adjusted measure does. Both series move relatively closely in normal times, but the distance is considerably accentuated during the crisis period. The estimation of the spread is overstated up to 1% in October 2008, when the frequency of trading achieves the minimum in the sample period.

5. Conclusion

This paper shows that the standard empirical application of the bid–ask spread estimator proposed by CS can produce important biases when the asset does not trade all days in the sample period. This is the general case for corporate bonds. However, recent papers have exactly translated this estimator (originally developed for stocks) to the corporate bond market. Moreover, SSU conclude that this transaction cost estimator is a good liquidity proxy on the basis of both cross-sectional and time series comparisons among a huge set of liquidity measures.

My paper aims to compute the magnitude of the bias that this practice produces. To do so, I modify the original bid–ask spread estimator to account for non-consecutive daily observable prices. The derivation of my proposed adjusted measure follows the original idea of CS but is generalized to an *n*-day period between two observable prices. This avoids assuming that high and low prices in days without trades are those previously observed. Thus, the volatility component estimate is not biased toward zero and, therefore, the spread is not upward biased. Additionally, the adjusted estimator incorporates the effect of the time distance between observable prices.

I estimate the daily series of the standard and adjusted versions of the CS measure for a sample of highly traded bonds, that is, traded on at least 75% of the trading days. This sample selection implies that the maximum for the average distance between trades is 1.3361 days. I evaluate the difference between the standard and adjusted measures in four subsamples that result from splitting the sample of bonds using the quartiles of the distribution of the average number of days between trades. On the median, the relative difference is 1.85%, 8.71%, 21.11%, and 31.56% from the most to the less traded bonds, respectively. I compare the mean, the standard deviation, and the three quartiles of the cross-sectional distribution for the two estimators within each subsample and the null of equal statistics is consistently rejected in the two last subsamples (bonds with an average number of days between trades of 1.09 at least). The same conclusion is obtained when descriptive statistics of time series distribution of the two estimators are compared at individual level. Finally, I average the liquidity measures across all bonds and evaluate the time series evolution of the upward difference. The standard proxy produces a transaction cost estimator only 0.05% higher than the adjusted one during stability periods. However, in stressed periods, when the frequency of trading is reduced, the bias can be very large. In my sample period, the standard estimation of the spread is overstated up to 1% in October 2008.

Acknowledgments

This work was supported by the Generalitat Valenciana (PROMETEOII/2013/015), the Spanish Department of Economy and Competitiveness (ECO2015-67035-P and ECO2014-58434-P), and the Bank of Spain (Excellence in Education and Research Grant 2016). I would like to thank David Abad and Gonzalo Rubio for their comments always appropriate and helpful.

References

Amihud, Yakov, 2002. Illiquidity and stock returns: Cross-section and time-series effects, Journal of Financial Markets 5, 31–56. DOI: 10.1016/S1386-4181(01)00024-6.

Corwin, Shane A., Schultz, Paul, 2012. A simple way to estimate bid–ask spreads from daily high and low prices, Journal of Finance 67, 719–760. DOI: 10.1111/j.1540-6261.2012.01729.x.

Dick-Nielsen, Jens, 2009. Liquidity biases in TRACE, Journal of Fixed Income 19(2), 43–55. DOI: 10.3905/jfi.2009.19.2.043.

________, 2014. How to clean enhanced TRACE data. SSRN working paper, available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2337908.

Feller, William, 1951. The Asymptotic Distribution of the Range of Sums of Independent Random Variables. Annals of Mathematical Statistics 22 (3), 427–432.

Garman, Mark B., Klass, Michael J., 1980. On the estimation of security price volatilities from historical data, Journal of Business 53, 67–78.

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

Rossi, Marco, 2014. Realized volatility, liquidity, and corporate yield spreads. SSRN working paper, available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1571437.

Schestag, Raphael, Schuster, Philipp, Uhrig-Homburg, Marliese, 2016. Measuring liquidity in bond markets, Review of Financial Studies 29 (5), 1170–1219. DOI:10.1093/rfs/hhv132.

Table 1. Descriptive statistics of the TRACE database and selected bonds

Panel A: All TRACE data after filters for duplicates, cancelations, corrections, reversals, and outliers. Only market days and dates during the bond's life and until three months up to the default are included. The sample period is from July 1, 2002, to December 31, 2014.

Number of bonds:	89,654				
Industry distribution:	*Industrial*	*Financial*	*Utilities*	*Government*	*Miscellaneous*
	16.46%	56.09%	3.36%	23.95%	0.15%

	Mean	*St. Dev.*	*Q0.25*	*Q0.5*	*Q0.75*
Time to maturity at issuance	9.769	14.964	3.016	5.247	10.047
Offering amount (mn. USD)	178.074	446.565	3.63	20	150
Coupon	3.643	2.972	0.95	3.54	5.65
Treasury spread at issuance	162.949	126.494	84	125	197
Rating (average in the bond's life)	5.397	4.220	1.25	5	7.833
Days with trade (%)	40.215	47.292	5.349	17.518	60.252
Average number of trades per day	4.113	11.555	2	2.560	3.778
Total number of trades per bond	971.399	4,285.111	11	61	352

Panel B: Selected bonds with liquidity requirements: one year between the first and last transaction dates and trades on at least 75% of trading days. The sample period is from October 1, 2004, to September 30, 2012.

Number of bonds:	3,526				
Industry distribution:	*Industrial*	*Financial*	*Utilities*	*Government*	*Miscellaneous*
	50.77%	40.36%	3.37%	5.39%	0.11%

	Mean	*St. Dev.*	*Q0.25*	*Q0.5*	*Q0.75*
Time to maturity at issuance	10.713	8.619	5.027	9.826	10.055
Offering amount (mn. USD)	1,068.557	967.832	500	750	1,250
Coupon	5.539	2.093	4.350	5.625	6.875
Treasury spread at issuance	176.132	135.745	85	134	215
Rating (average in the bond's life)	7.610	4.076	4.988	6.750	9.891
Days with trade (%)	90.020	7.885	83.154	91.314	97.478
Average number of trades per day	10.893	9.869	5.234	7.583	12.452
Total number of trades per bond	10,107.984	11,878.617	3,355	6,677	12,259

This table reports the descriptive statistics of bond characteristics for the overall TRACE data set (panel A) and the sample selected (panel B). The Treasury spread is the difference between the bond yield and the yield on Treasury securities with the same maturity. The Treasury yields for all possible maturities are obtained by interpolation. Days with trade refer to the percentage of days with at least one trade in the bond's trading life span.

Table 2. Descriptive statistics of the subsamples of bonds, based on frequency of trading

	Panel A: Subsample 1, 882 bonds # Days between trades (average) <1.0252					*Panel B: Subsample 2, 881 bonds #Days between trades (average) in [1.0252, 1.0948)*				
Industry distribution:	*Industrial*	*Financial*	*Utilities*	*Govern.*	*Misc.*	*Industrial*	*Financial*	*Utilities*	*Govern.*	*Misc.*
	39.34%	52.15%	0.34%	8.16%	0.00%	52.21%	39.50%	2.95%	5.22%	0.11%
	Mean	*St. Dev.*	*Q0.25*	*Q0.5*	*Q0.75*	*Mean*	*St. Dev.*	*Q0.25*	*Q0.5*	*Q0.75*
Time to maturity at issuance	8.899	6.344	5.022	8.516	10.030	9.993	7.952	5.022	9.205	10.041
Offering amount (millions USD)	1,771.502	1,209.557	1,000	1,500	2,250	1,083.670	868.215	501.500	875	1,250
Coupon	5.156	1.684	4.125	5.3	6.125	5.466	2.151	4.250	5.500	6.804
Treasury spread at issuance	172.048	122.116	90	135	210	184.016	150.689	83	135	225
Rating (average in the bond's life)	6.125	3.457	4.333	5.712	7.623	7.395	4.083	4.667	6.456	9.500
Days with trade (%)	98.991	0.715	98.499	99.079	99.531	94.760	1.810	93.220	94.926	96.432
Average number of trades per day	21.300	13.601	12.353	17.565	24.994	9.668	4.772	6.826	8.641	10.946
Total number of trades per bond	19958.4	18210.7	7959	15618.5	25574	9311.1	6866.3	4380.3	7428	12714
Average number of days between trades	1.010	0.007	1.005	1.009	1.015	1.055	0.020	1.036	1.053	1.072
Maximum number of trades in a day	218.832	248.895	84	153	266	146.583	204.181	54	95	157
Average trading volume per day (000 USD)	7,254.8	6,251.8	3,458.1	5,577.2	8,921.6	3,816.6	3,073.9	1,972.1	3,091.5	4,707.4
Max. trading volume in a day (000 USD)	111,962.5	83,727.3	55,596.5	91,910.8	140,409.1	72,928.3	64,323.4	36,881.2	56,763.1	86,696.9
	Panel C: Subsample 2, 882 bonds #Days between trades (average) in [1.0948, 1.2015)					*Panel D: Subsample 1, 881 bonds # Days between trades (average) ≥1.0252*				
Industry distribution:	*Industrial*	*Financial*	*Utilities*	*Govern.*	*Misc.*	*Industrial*	*Financial*	*Utilities*	*Govern.*	*Misc.*
	55.33%	35.49%	4.88%	4.08%	0.23%	56.19%	34.28%	5.33%	4.09%	0.11%
	Mean	*St. Dev.*	*Q0.25*	*Q0.5*	*Q0.75*	*Mean*	*St. Dev.*	*Q0.25*	*Q0.5*	*Q0.75*
Time to maturity at issuance	12.108	9.558	5.060	10	12.011	11.853	9.775	5.062	9.995	11.937
Offering mount (millions USD)	808.520	678.782	450	600	1,000	610.032	546.459	300	500	750
Coupon	5.708	2.184	4.400	5.750	7.125	5.826	2.244	4.7	6	7.2
Treasury spread at issuance	171.981	128.104	87	135	212.5	177.045	143.335	83	127	220
Rating (average in the bond's life)	8.314	4.109	5.667	7.891	10.583	8.657	4.153	5.667	8.278	11.613
Days with trade (%)	87.455	2.369	85.509	87.596	89.501	78.867	2.310	76.864	78.808	80.696
Average number of trades per day	6.970	3.925	4.946	5.988	7.644	5.629	3.675	3.901	4.675	6.076
Total number of trades per bond	6,492.8	5,378.0	2,662	5,200.5	8,721	4,662.6	4,052.9	1,914.8	3,552	6,349.5
Average number of days between trades	1.144	0.031	1.117	1.141	1.168	1.268	0.037	1.237	1.268	1.300
Maximum number of trades in a day	102.644	108.004	41	72	120	81.402	93.440	30	55	96
Average trading volume per day (000 USD)	2,861.4	2,346.0	1,481.9	2,311.0	3,545.3	2,144.3	1,746.7	984.4	1,825.8	2,769.5
Max. trading volume in a day (000 USD)	57,489.8	75,633.8	26,674.3	43,926.8	69,898.2	42,582.8	36,197.6	18,469.1	34,192.9	55,140.1

Panels A to D report the descriptive statistics of security characteristics and activity indicators for the corporate bonds in each subsample. The sample selection criteria consist of bonds that are active during at least one year and traded on at least 75% of trading days. The subsamples are created by splitting the sample on the basis of the quartiles for the average number of days between trades, as indicated in each panel. The Treasury spread is the difference between the bond yield at issuance and the yield on Treasury securities with the same maturity. The Treasury yields for all possible maturities are obtained by interpolation. Days with trade refer to the percentage of days with at least one trade in the bond's trading life.

Table 3. Cross-sectional distribution of standard and adjusted transaction cost estimators

Panel A: <1.0252 *# Bonds: 882*	*Mean*	*St. Dev.*	*Q0.25*	*Q0.5*	*Q0.75*	*Complete Distribution*
Standard CS	0.8898	0.6665	0.3485	0.7485	1.2577	
Adjusted CS	0.8730	0.6598	0.3417	0.7283	1.2393	
Difference (%)	2.51	2.54	3.28	1.85	1.03	
Equality Stat	0.587	0.043	0.148	0.145	0.305	
P-value	(0.557)	(0.966)	(0.700)	(0.703)	(0.581)	(0.999)
Parametric p-value	[0.60]	[0.76]				
Panel B: [1.0252, 1.0948) *# Bonds: 881*	*Mean*	*St. Dev.*	*Q0.25*	*Q0.5*	*Q0.75*	*Complete Distribution*
Standard CS	0.7144	0.5733	0.2954	0.5554	0.9760	
Adjusted CS	0.6532	0.5340	0.2623	0.5036	0.8952	
Difference (%)	9.89	6.09	13.36	8.71	5.53	
Equality Stat	2.357	-1.801	2.546	2.472	2.725	
P-value	(0.018)	(0.072)	(0.111)	(0.116)	(0.099)	(0.070)
Parametric p-value	[0.02]	[0.04]				
Panel C: [1.0948, 1.2015) *# Bonds: 882*	*Mean*	*St. Dev.*	*Q0.25*	*Q0.5*	*Q0.75*	*Complete Distribution*
Standard CS	0.7508	0.7505	0.3456	0.5811	0.9902	
Adjusted CS	0.5964	0.5345	0.2642	0.4660	0.7911	
Difference (%)	21.13	9.26	26.58	21.11	15.31	
Equality Stat	6.143	-4.292	29.633	25.479	20.828	
P-value	(8.09e-10)	(1.77e-5)	(5.22e-8)	(4.47e-7)	5.02e-6)	(1.45e-6)
Parametric p-value	[7.14e-7]	[1.94e-23]				
Panel D: >=1.2015 *# Bonds: 881*	*Mean*	*St. Dev.*	*Q0.25*	*Q0.5*	*Q0.75*	*Complete Distribution*
Standard CS	0.7203	0.4815	0.3587	0.5810	0.9685	
Adjusted CS	0.5147	0.3787	0.2361	0.4099	0.7255	
Difference (%)	30.65	10.77	37.49	31.56	24.79	
Equality Stat	10.187	-6.297	96.985	57.392	57.644	
P-value	(2.28e-24)	(3.04e-10)	(0)	(3.57e-14)	(3.14e-14)	(5.54e-20)
Parametric p-value	[8.98e-23]	[1.39e-12]				

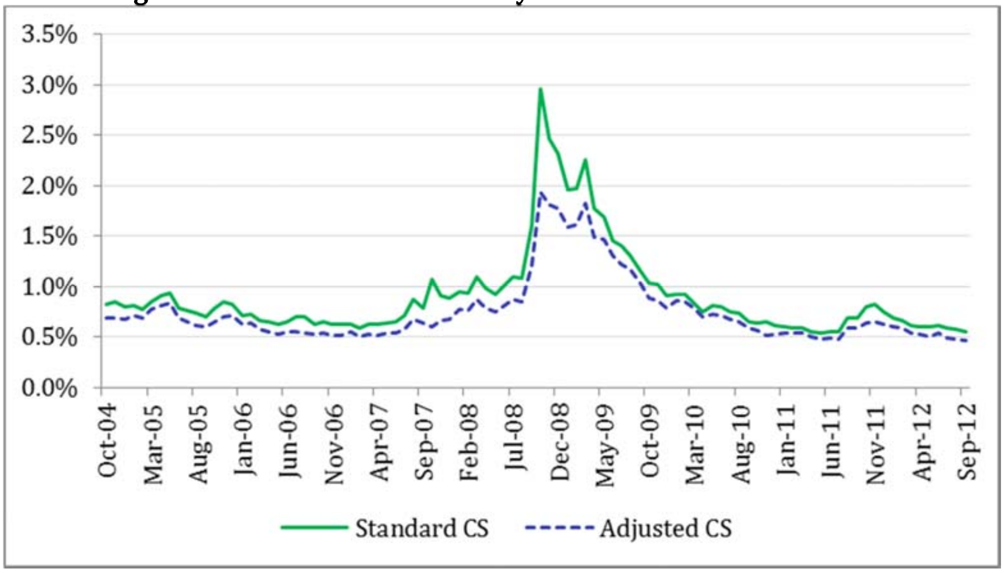
This table reports cross-sectional descriptive statistics for the mean of two transaction cost estimators, in the first and second rows, and for the mean of their relative difference. *Standard CS* refers to the Corwin and Schultz bid–ask spread and *adjusted CS* refers to an adjusted version that accounts for infrequent trading. The two proxies are computed daily using high and low prices from intraday TRACE data for the period between October 1, 2004, and September 30, 2012. The sample selection criterion consists of bonds that are active during at least one year and traded on at least 75% of trading days. The sample is divided into four subsamples on the basis of the quartiles for the average number of days between trades, as indicated in panels A to D. *Equality Stat* refers to the value of the statistic that tests the null of equal descriptive statistic for the two estimators. Wilcoxon rank sum test is used for the means, Ansari-Bradley test is used for standard deviations, the comparison between the three quartiles is done by a Pearson's chi-squared test, and finally the comparison of the complete distributions is done by the Kolmogorov-Smirnov test. Below the correspondent p-values are reported. The last row contains the p-values associated with the standard parametric t-statistic for the comparison of means and the F-statistic for the comparison of variances.

Table 4. Time series distribution of standard and adjusted transaction cost estimators. Number of significant differences

	# Bonds	*Mean*	*St. Dev.*	*Q0.25*	*Q0.5*	*Q0.75*	*Complete Distribution*
Subsample 1: <1.0252	882	0	0	162	2	0	0
Subsample 2: [1.0252, 1.0948)	881	92	93	569	49	80	15
Subsample 3: [1.0948, 1.2015)	882	710	715	808	636	612	526
Subsample 4: >=1.2015	881	859	868	868	849	789	853

Descriptive statistics for the time series distribution of the daily standard and adjusted CS estimators are compared for each individual bond. The comparison employs different nonparametric tests depending on the statistic which are indicated in the notes of Table 3. This table reports the number of bonds for which the null of equal descriptive statistic is rejected within each subsample. The subsamples are created by splitting the sample on the basis of the quartiles for the average number of days between trades, as indicated in the first column.

Figure 1. Time series of monthly transaction cost estimators



This figure displays the monthly time series of two transaction cost estimators: the Corwin and Schultz bid–ask spread (standard CS) and an adjusted version that accounts for infrequent trading (adjusted CS). The two proxies are computed daily using high and low prices from intraday TRACE data for the period between October 1, 2004, and September 30, 2012. Monthly values are averaged over all days within the month.

Footnotes

[1] Following SSU's paper, I select bonds that trade at least 75% of the days during their life span. This implies that the average distance between two trades in my sample of bonds has a maximum of 1.3361 days.

[2] After some robustness checks, I find that this filter cleans better than alternative filters that use the mean instead of the median, a different number of standard deviations, a different window length, or the filter proposed by Rossi (2014).

[3] My sample is similar but not identical to that used by SSU, probably because the filters are in a different order and my outliers filter is slightly less restrictive.

[4] Note that SSU's bond characteristics and credit ratings come from Thomson Reuters and Bloomberg, while I work directly with the FISD information. I use the average value between four rating agencies while SSU use three and I provide the offering amount instead of the average outstanding amount over the life of the bond.

[5] To compare the values for the standard liquidity proxy with those provided by SSU, I compute the cross-sectional statistics for the entire sample of bonds. The results for the mean, standard deviation, Q0.25, Q0.5, and Q0.75 are 0.77, 0.63, 0.34, 0.61, and 1.04, respectively, while the numbers of SSU are 0.94, 0.95, 0.31, 0.63, and 1.25, respectively. However, SSU work with monthly series by averaging daily measures, while my statistics are computed directly from daily series. Additionally, differences could be due to the fact that the samples of bonds are not entirely identical and the bond prices are from different sources (TRACE in my case, Bloomberg in theirs).

[6] This value is just the difference and not the relative difference.