Journal of Empirical Finance 48 (2018) 36–57

Contents lists available at ScienceDirect

Journal of Empirical Finance

journal homepage: www.elsevier.com/locate/jempfin

Bid–ask spread estimator from high and low daily prices: Practical implementation for corporate bonds

Belén Nieto

Financial Economics and Accounting, University of Alicante, Ctra. San Vicente sn, E-03690 Alicante, Spain

ARTICLE INFO

Keywords:
Transaction cost
Corporate bonds
Liquidity

1. Introduction

Liquidity has become an important driver of asset prices. Specially, corporate bond returns have great exposure to liquidity risk. For example, during the 2008–2009 crisis, 30% of the variability in credit spreads could be explained by liquidity. The vast numbers of papers measuring liquidity and evaluating its effect on market prices are therefore not surprising.[1] Most of the liquidity proxies designed and previously estimated in the stock market have been directly translated to this market. However, the structure of the stock and bond markets is clearly different. Schestag et al. (2016; SSU hereafter) evaluate the appropriateness of a large set of daily liquidity proxies by comparing them with measures based on intraday data. These authors conclude that Corwin and Schultz's (2012; hereafter CS) transaction cost measure appropriately captures cross-sectional differences and time-varying patterns.

CS propose daily estimators for the two unobservable components of price changes, the bid–ask spread and volatility, which employ only high and low daily prices. The idea is to combine the information regarding a single day with a two-day period. Independently of the theoretical assumptions of their model, estimation of the CS measure presents practical concerns that are not addressed by SSU and which can be particularly relevant to corporate bonds. First, having a daily spread estimator requires that the asset be traded on all days and at least twice a day. This empirical requirement is hard to hold for corporate bonds, even if the most active ones are selected. Second, under the model, the variance in a two-day interval is twice the variance for one day in the interval. Again, this assumption is especially problematic in this market, since it is characterized by high volatility precisely in moments of low liquidity. Third, volatility and spread estimates depend on each other and can only be computed numerically with a high time cost, unless Jensen's inequality is ignored. This paper evaluates the effects that these practical issues can have on the accuracy of volatility and the spread estimates for corporate bonds.

With regard to the first point above, in the original paper of CS, the estimation is applied to the stock market, where the assumption of observable consecutive prices is reasonable. In their sample, on average, there is no trade in 4.11% of all working days. For these days without trades, the authors suggest the assumption that the high and low prices are the same as those observed the most recent prior trading day. This assumption imposes zero volatility, so that the final estimators understate the volatility and overstate the spread. In the sample of bonds used in this paper, the problem of infrequent trading affects 16% of trading days. To avoid the unrealistic imposition of zero volatility, I propose an adjusted version of the CS spread and volatility estimators that account for

E-mail address: belen.nieto@ua.es.

[1] Bao et al. (2011), Bongaerts et al. (2017), De Jong and Driessen (2012), Edwards et al. (2007), Feldhütter (2012), and Friewald et al. (2012) are some examples.

https://doi.org/10.1016/j.jempfin.2018.06.003
Received 5 May 2017; Received in revised form 8 June 2018; Accepted 11 June 2018
Available online 19 June 2018
0927-5398/

irregular intervals between two observable prices. Therefore, the adjusted estimators can be obtained without assumptions about prices on days without trades. This is the first contribution of this paper.

With respect to the second point above, as CS recognize, in periods of high volatility, the variance in a two-day interval can be more than twice the variance for a single day. In these cases, the resulting spread estimate is negative. The authors propose setting these negative daily values to zero since this approach produces more accurate estimates than other alternatives. In the sample of CS, negative spreads are obtained in 29.26% of the days, on average, across stocks. This percentage is similar in my sample of bonds. However, my sample includes the crisis period, which is characterized by highly volatile financial markets. Specifically, the volatility of daily bond returns is 1.86%, on average, between April 2007 and December 2009 and 1.04% during the remainder of the sample period. Therefore, spreads can be estimated as especially low or negative and the imposition of a zero spread against just omitting negative observations will reduce the average spread estimate. I compare the accuracy of the two approaches.

Finally, regarding the third point, I evaluate the differences in estimates between ignoring Jensen's inequality, which is computationally easy, and not ignoring it, which suffers from the high time cost of numerical computation.

With all of this in mind, this paper's second contribution is a detailed empirical analysis that allows for a clear practical implementation of the measure based on high and low daily prices in the case of corporate bonds (but also for other assets with non-continuous trading and non-constant volatility). I analyze the importance of the effect produced by the assumptions imposed in the standard estimation by evaluating the accuracy of the different estimators that relax these practical impositions. To do so, I compare the different CS-based volatility estimators with the realized volatility and the different CS-based spread estimators with three transaction cost benchmarks based on high-frequency data.

The first result is that the standard high–low volatility and spread estimators are significantly downward biased. On average, the CS volatility estimator is around 25% lower than the realized volatility in both time series and cross-sectional analyses. The CS spread estimator is 11% to 27% lower than the intraday proxy, depending on the benchmark, in the time series and even higher biases are obtained in the cross section. More importantly, both estimators show a problematic decreasing trend from the most active (liquid) bonds to less active bonds.

The volatility analysis shows that the negative bias is generally observed for all the bonds but monotonically increases in absolute value with the volatility level (percentage of days without trades). Jensen's inequality produces slightly larger estimates than the standard approach does and thus contributes to reducing the bias. However, the differences between the two are unremarkable. The use of the generalized version that accounts for non-continuous trading proposed here significantly reduces this bias and allows one to obtain volatility estimates that increase with effective volatility.

Regarding the spread, the standard CS proxy appears to be upward biased in the subsample that contains the 25% of the most liquid bonds but downward biased for the remaining bonds. The upward bias is eliminated when the expression corrects for the number of days between consecutive trades. The downward bias is considerably reduced when negative spreads are discarded instead of set equal to zero. The combination of the two adjustments reproduces reasonably well the time series and the cross-sectional distribution of the intraday proxies for the 50% of bonds in the central position of the sample in terms of trading frequency. As in the volatility estimation, if, in addition, Jensen's inequality is taken into account, the accuracy of estimates slightly improves. For the whole sample of bonds and, on average, for the three intraday proxies, the bias created by the standard methodology is −0.19 in the time series and −0.17 in the cross-section, in contrast to 0.05 and −0.03, respectively, obtained with the estimator that accounts for non-continuous trading and Jensen's inequality and discards negative spreads.

The remainder of the paper is organized as follows. Section 2 summarizes the CS proposal and its standard estimation in practice. Section 3 presents the adjusted estimator that accounts for infrequent trading and discusses alternatives that do not impose unrealistic assumptions. Section 4 describes the data. Section 5 compares the different estimation proposals with the standard one, while Section 6 evaluates their accuracy through their comparison with some benchmark proxies. Section 7 concludes the paper.

2. CS spread estimator

2.1. Volatility and bid–ask spread measures

CS propose an estimator based on the assumptions 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:

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

where $\alpha = ln\left[(2+S)/(2-S)\right]$ and S is the spread.

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

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

The square of Eqs. (1) and (2) illustrates the main idea of the paper: the high–low price ratio has one component due to price volatility and another due to the bid–ask spread. The volatility is proportional to the data frequency, whereas the spread is