Assessment of Bid-Ask Spread

Tornike Kikacheishvili

November 2025

1 Introduction

This document presents an assessment of bid-ask spread on high-frequency data. The assignment has provided three price series for the same underlying instrument from three different sources. The goal is to analyse these price series and provide an assessment as to which source is better for a broker to use. The analysis include metrics Time Weighted Relative Quoted Spread, Corwin and Schultz method, Roll model and Roll model with CSS extension. All the computation is implemented in Python code.

2 Data

Data consist of timestamp, bid and ask only, from three different sources. Which are given from 01/01/2025 to 03/02/2025 with in microseconds. First provider (ID 0) has 20,533,743 prices, second provider (ID 1) has 22,582,671 prices and third provider (ID 2) has around 10,329,249 prices. Also, Provider ID 2 contains negative bid-ask spread and due to the small amount of cases (380), it is removed during the analysis. During the assignment any other information such as type of underlying instrument, traded prices, trade volume or tick size was not provided. Therefore all the metrics are chosen based on the given limited dataset.

3 Metrics

3.1 Time Weighted Relative Quoted Spread

This metrics as it is explained in reference[1], quoted spread is the difference between the best bid and the best ask prices available on the market at a certain point in time. In order to obtain a relative measure, the quoted spread is divided by the mid-price, which is the average the of best bid and ask prices. Order of bid and ask change frequently in microseconds, therefore simple average would not be enough accurate to assess spread especially with different amount of prices for each dataset, time-weighted average of the intraday relative quoted spread is used to see difference between each dataset.Equation (1) represents the time-weighted average of the daily relative quoted spread (TWS).

$$TWS = \frac{1}{T} \sum_{t=1}^{Q} \frac{Ask\ Price_t - Bid\ Price_t}{(Ask\ Price_t + Bid\ Price_t)/2} * (T_{t+1} - T_t) \tag{1}$$

T - the total time with an available spread measured in seconds
Q - the number of updates in the best bid or ask orders
T_t - the timestamp of t-th spread update measured in seconds.

TWS is a daily liquidity measure calculated with all intraday updates of the best bid and ask orders.

Table 1: Time Weighted Spread Summary Statistics

	Value (%)	bps (* 10000)	Days
Provider ID_0	0.0127	1.27	29
Provider ID_1	0.9410	94.10	29
Provider ID_2	0.0547	5.47	29

The Time Weighted Relative Quoted Spread (TWS) is a direct measure of liquidity based on quoted spreads from order book data. Lower values indicate tighter spreads and higher liquidity. Higher values suggest illiquidity, thin order books, or volatile conditions.

This metrics is more precise than indirect estimators like Corwin-Schultz, which is discussed below, for quoted liquidity it requires granular tick data and can be noisy in low-volume periods.

From broker side to choose the best source for gain profits from wide spread is Provider ID 2, but nowadays market is very competitive and I would choose Provider ID 0, it is cost effective and liquid for traders.

3.2 Corwin and Schultz

As it is explained reference [2] the estimator is based on two principles: first, high prices are almost always matched against the offer, and low prices are almost always matched against the bid. The ratio of high-to-low prices reflects fundamental volatility as well as the bid-ask spread. Second, the component of the high-to-low price ratio that is due to volatility increases proportionately with the time elapsed between two observations.

$$S_t = \frac{2\left(e^{a_t} - 1\right)}{1 + e^{a_t}} \tag{2}$$

where

$$\alpha_t = \frac{\sqrt{2\beta_t} - \sqrt{\beta_t}}{3 - 2\sqrt{2}} - \sqrt{\frac{\gamma_t}{3 - 2\sqrt{2}}} \tag{3}$$

$$\beta_t = \mathrm{E}\left[\sum_{j=0}^{1}\left[\log\left(\frac{H_{t-j}}{L_{t-j}}\right)\right]^2\right] \tag{4}$$

$$\gamma_t = \left[\log\left(\frac{H_{t-1,t}}{L_{t-1,t}}\right)\right]^2 \tag{5}$$

and $H_{t-1,t}$ is the high price over 2 bars ($t-1$ and t), whereas $L_{t-1,t}$ is the low price over 2 bars ($t-1$ and t). Because $\alpha_t < 0 \Rightarrow S_t < 0$, the authors recommend setting negative alphas to 0.

Table 2: Corwin and Schultz Summary Statistics

	CS (%)
Provider ID_0	0.1334
Provider ID_1	0.0000
Provider ID_2	0.1439

Lower CS values indicate tighter spreads (higher liquidity), while higher values suggest wider spreads (lower liquidity or higher trading costs). Values near 0 are common for highly liquid assets like large-cap stocks. Provider ID 0 and 1 indicates a moderately tight spreads, meanwhile Provider ID 2 has a zero spread, which is unusual and likely an artifact of the data, which may cause by overnight returns or high-low variation (e.g., single-price days in illiquid assets).

3.3 Roll Model

Reference [3] explains Roll model, which is used for measure bid-ask spread and next chapter shows extension of this model for using to enhance bid-ask spread analysis.

The order processing model was developed under the following set of assumptions.

1. *Security prices at any given time, fully incorporate all relevant available information. What this means is that the price that would be observed in the absence of market imperfections accurately reflects the intrinsic value of the asset.*

2. *Buy or sell orders are equally probable at any instant in time, on an unconditional basis, and order flows are serially independent.*

3. *The underlying price of the asset is independent of the order flow in the market. This implicitly rules out any adverse information effects.*

4. *The true value of the security is bracketed by the bid-ask spread and the spread is assumed to be constant and symmetric, at least for the time period for which the analysis is conducted.*

Mathematically, the model may be stated as follows.

$$P_t = P_t^* + \frac{s}{2} Q_t \tag{6}$$

and,

$$P_t^* = P_{t-1}^* + \mu + \epsilon_t, \tag{7}$$

P_t - the observed price at time t;
P_t^* - the intrinsic value of the asset at time t, which would have been observed in the absence of the spread;
Q_t - the indicator of the transaction type;
$Q_t = 1$ if the observed price is at the ask, and -1 if it is at the bid;
$\{\epsilon_t\}$ - are assumed to be serially uncorrelated, and have a zero mean and constant variance σ^2;
μ is assumed to be constant over time, as is the spread s.

With these assumptions, the covariance of the observed price changes equals minus the square of half the bid-ask spread, i.e., $\mathrm{Cov}(\Delta P_t, \Delta P_{t-1}) = -s^2/4$, and the estimates of the covariance can be transformed to get the spreads.

Table 3: Roll's Measure Summary Statistics

	(%)
Provider ID_0	Nan
Provider ID_1	3.1467
Provider ID_2	1.7468

The underlying idea is that asset prices bounce between bid and ask prices, and greater spread is associated with higher negative covariance. Therefore, privider ID 0 has no detectable bounce, so spread undefined. Most probably because hyper-liquid data source, meanwhile previous two measures also proved, that ID 0 source is liquid. Provide ID 1 has wide spread, signaling low liquidity, and again third measure shows the same result. It is risky for large trades. Priovider ID 2 has moderate spread, again. acceptable liquidity but with some friction.

3.4 Roll Model with extension of CSS

There are multiple extensions of Roll model, but in case of limit access of data, most relevant is to use CSS extension. As it it is explained in reference [3] *the CSS generalize the model by partially relaxing the second assumption made above. In model, although the unconditional probability of a bid or an ask transaction is the same at a particular point in time, conditional on a bid, the probability of a consecutive bid is greater than a half, and likewise for the ask. In other words, transaction types are positively serially correlated. if the conditional probability of a bid following a bid or that of an ask following an ask, is greater than half, the formula for the spread converges to that predicted by the Roll model, if we consider time intervals with an adequate number of transactions.*

when δ, the conditional probability of a bid following a bid, or of an ask following an ask, is not equal to 1/2. The covariance, in this case, depends on the length of the measurement interval. More specifically, it depends on 'N', where 'N' is the number of transactions that have occurred between t-1 and t. For the purpose of this analysis, we will assume that transactions occur at fixed intervals in time,i.e., N is fixed for a given measurement interval. How covariances are calculated, I highly suggest to check reference [3].

Synthetic mid-prices and resample data with 1 min is applied for Roll extension model, which is presented in table 4, where N is number of transaction assumed form 1 to 5, which means transaction happens every minute.

Roll model with extension improves Roll's measure results, where probability of bid (ask) following bid (ask) slightly higher than 1/2 for each data.

Table 4: Roll Model with CSS Extension Results

Provider ID	δ	**Roll Model (CSS)**
N=1		
0	0.5645	−0.005 549 89
1	0.5649	−0.043 470 72
2	0.5640	−0.008 888 08
N=2		
0	0.5645	−0.007 074 71
1	0.5649	−0.055 489 17
2	0.5640	−0.011 310 44
N=3		
0	0.5645	−0.007 284 96
1	0.5649	−0.057 156 73
2	0.5640	−0.011 641 74
N=4		
0	0.5645	−0.007 312 31
1	0.5649	−0.057 375 01
2	0.5640	−0.011 684 51
N=5		
0	0.5645	−0.024 759 38
1	0.5649	−0.194 141 05
2	0.5640	−0.039 597 47

3.5 Conclusion

This assignment evaluates liquidity across three data providers using a different methods framework encompassing quoted (TWQS), range-based (Corwin-Schultz), and covariance-based (Roll's) spread estimators solely and with CSS extension. Provider ID 0 emerges as unequivocally the most liquid and negligible trading frictions, suitable for broker in high competitive market. Provider ID 2 demonstrates moderate liquidity, while Provider ID 1 exhibits illiquidity.

References

[1] Su, E., & Tokmakcioglu, K. (2020). A comparison of bid-ask spread proxies and determinants of bond bid-ask spread.

[2] Lopez de Prado, M. (2018). Advances in Financial Machine Learning.

[3] Parameswaran, S. K., & Basu, S. (2020). Some Analytical Results for Models of the Bid-Ask Spread.