The Bias of Simple Bid-Ask Spread Estimators

Scott H. Irwin
UIUC

Pedro Tremacoldi-Rossi
Columbia

IAQF & Thalesians Seminar Series
September 2023

Measuring market illiquidity

You observe a lot of (daily) price data p_t and want to estimate bid-ask spreads S

On one hand, data is noisy — not great for measurement

- trading not continuous, real-world frictions...

Measuring market illiquidity

You observe a lot of (daily) price data p_t and want to estimate bid-ask spreads S

On one hand, data is noisy — not great for measurement

- trading not continuous, real-world frictions...

On the other, noise in data is exactly what we want to measure!

- $S_t = 2\,(p_t - \mathcal{P}_t)$: want to estimate S, don't observe efficient price $\mathcal{P}_t$
- 1 equation, 2 unknowns

Measuring market illiquidity

You observe a lot of (daily) price data p_t and want to estimate bid-ask spreads S

On one hand, data is noisy — not great for measurement

- trading not continuous, real-world frictions...

On the other, noise in data is exactly what we want to measure!

- $S_t = 2\left(p_t - \mathcal{P}_t\right)$: want to estimate S, don't observe efficient price $\mathcal{P}_t$
- 1 equation, 2 unknowns

Solution:

1. Fix S and augment t: $\{p_t\left(S, \mathcal{P}_t\right)\}_{t \in \mathbb{Z}_+}$ (get more equations)
2. Come up with theory linking $\mathcal{P}_t$ over t, back out S

From identification challenge to implementation challenge

But assumption of constant S becomes increasingly unrealistic when t grows

- S varies more than 100% in 20% of day-over-day
- settle on "daily" estimates computed using t and $t+1$, then average

Basic technology used in all simple bid-ask spread estimators

- daily p_t, compute statistic over $\{t, t+1\}$
- estimate via method of moments

From identification challenge to implementation challenge

But assumption of constant S becomes increasingly unrealistic when t grows

- S varies more than 100% in 20% of day-over-day
- settle on "daily" estimates computed using t and $t+1$, then average

Basic technology used in all simple bid-ask spread estimators

- daily p_t, compute statistic over $\{t, t+1\}$
- estimate via method of moments

Structural tension: theory for $\mathcal{P}_t$ is asymptotic, data for estimation uses $n(t) = 2$

Roll (1984) estimator

Use closing price c_t for p_t:

$$\widehat{S}^{Roll} = 2\sqrt{-Cov\left(\Delta c_t, \Delta c_{t+1}\right)}$$

1. Because $\widehat{S}^{Roll}$ is nonlinear, averages are biased due to Jensen's inequality
   - $\mathbb{E}\left[\widehat{S}^{Roll}\right] - S = -S\left[\frac{\mu_4/\sigma^4 + 4}{8\left(n(t) - 1\right)}\right]$
2. Empirical implementation uses $\widehat{Cov}$, not population Cov: $\widehat{S^*}^{Roll}$
   - Harris (1990): $\widehat{Cov}$ is biased in "small" samples (eg 2-200+ days of data!)
   - $\mathbb{E}\left[\widehat{Cov}\right] - Cov = -\sigma^2/n(t)$

Roll (1984) estimator

Use closing price c_t for p_t:

$$\widehat{S}^{Roll} = 2\sqrt{-Cov\left(\Delta c_t, \Delta c_{t+1}\right)}$$

1. Because $\widehat{S}^{Roll}$ is nonlinear, averages are biased due to Jensen's inequality
   - $\mathbb{E}\left[\widehat{S}^{Roll}\right] - S = -S\left[\frac{\mu_4/\sigma^4 + 4}{8\left(n(t) - 1\right)}\right]$
2. Empirical implementation uses $\widehat{Cov}$, not population Cov: $\widehat{S^*}^{Roll}$
   - Harris (1990): $\widehat{Cov}$ is biased in "small" samples (eg 2-200+ days of data!)
   - $\mathbb{E}\left[\widehat{Cov}\right] - Cov = -\sigma^2/n(t)$

We call 1. **moment bias** and 2. **small sample bias**

This paper

Moment and **small sample bias** are well-known for Roll

Since then, many estimators for S developed using Roll's framework (c_t)

- no tangible performance improvement (eg French and Roll (1986), Thompson and Waller (1987), Lesmond et al. (1999), Hasbrouck (2004))

Progress with Corwin and Schultz (2012): high-low estimator (+1,200 citations)

- daily high and low instead of close prices

This paper

Moment and **small sample bias** are well-known for Roll

Since then, many estimators for S developed using Roll's framework (c_t)

- no tangible performance improvement (eg French and Roll (1986), Thompson and Waller (1987), Lesmond et al. (1999), Hasbrouck (2004))

Progress with Corwin and Schultz (2012): high-low estimator (+1,200 citations)

- daily high and low instead of close prices

This paper: The same bias sources also affect the high-low estimator

- + *any* moment-based estimator using the range (eg Abdi and Ranaldo (2017))
- abstract from why do we even want to estimate effective spreads (eg Hagströmer (2021) critique)

Outline

Exact bias expression for the high-low estimator

- bias depends on underlying true spread (mainly) and volatility
- introduce non-classical measurement error
- cross-sectional regressions will be contaminated
- *all* model violations combined account for only 5% of performance issues

Bias bounds for empirical use

- bound 80% of US stocks (misses very illiquid stocks)
- correlate 98% with actual bias

Lessons for future low-frequency spread estimators

- work on small sample properties
- possibly abandon Roll framework altogether

Building blocks

True (efficient) log price evolves as GBM

Can't observe true price, real-world data is noisy

Observed prices p_t track true prices $\mathcal{P}_t$ according to

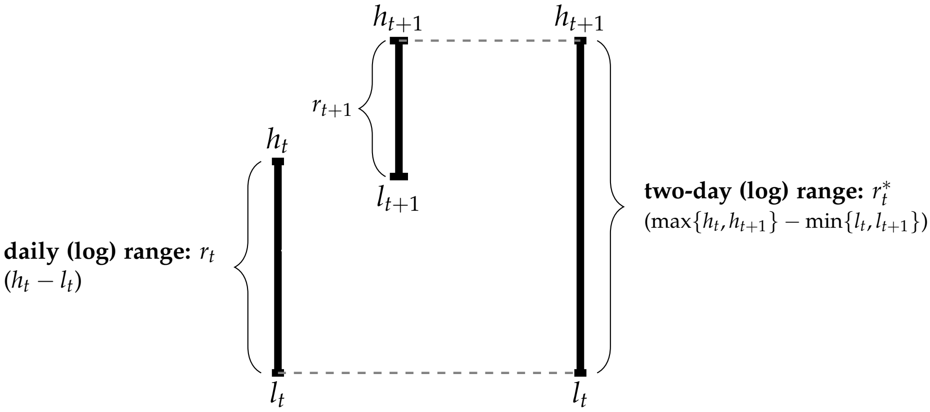
$$p_t = \mathcal{P}_t + \frac{S}{2} Q_t$$

- S: total effective spread
- Q_t: order flow indicator ($+1$ for buyer-initiated, -1 for seller-initiated)
- assume: $Q_t \perp\!\!\!\perp P_t$

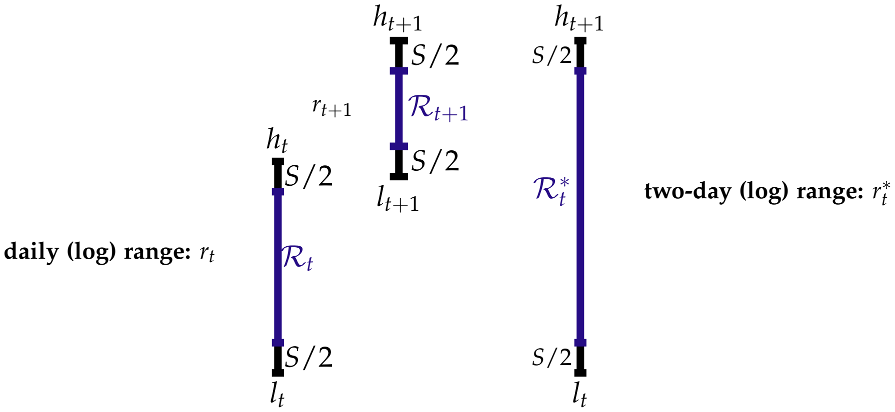
Sometimes p_t overstates the true price by $S/2$; sometimes understates it by $S/2$

- if we can sign Q_t, then we can precisely sign $p_t - \mathcal{P}_t$

For every two consecutive trading days t and $t+1$



For every two consecutive trading days t and $t+1$



Range history

Use of range in finance to estimate volatility dates to Parkinson (1980) and Garman and Klass (1980)

- The $\mathcal{R}_t$ part of the range
- "Revived" by Alizadeh et al. (2002) and more recently by Christensen and Podolskij (2007), Martens and van Dijk (2007)

Based on *asymptotic* results from Feller (1951)

- neglected literature from hydrology: range is biased in finite data!
- Anis and Lloyd (1953), Şen (1977) derive the expected range for small samples
- for $n(t) = 1$, bias is approximately -0.798σ

High-low spread estimator

1. Daily high is buyer-initiated, daily sell is seller-initiated
   - stylized fact: holds for 95% of US stock-days
2. Volatility is proportional to trading horizon
3. + usual: spread constant over two consecutive days, GBM

Key relationship: $r_t = h_t - l_t = (\mathcal{H}_t + S/2) - (\mathcal{L}_t - S/2) = \mathcal{R}_t + S$

- with $\mathbb{E}[\mathcal{R}_t] = 1.596\sigma$
- $r_t + r_{t+1}$ has 2 times the spread and volatility over 2 days: $\mathcal{R}_t + \mathcal{R}_{t+1} + 2S$
- r_t^* has 1 spread and volatility over 2 days: $\mathcal{R}_t^* + S$

Corwin and Schultz (2012) formula

Rewrite the high-low estimator as:

$$\widehat{S}^{HL} = \left(1 + \sqrt{2}\right) \left(\sqrt{\mathbb{E}\left[r_t^2 + r_{t+1}^2\right]} - r_t^* \right)$$

Corwin and Schultz (2012) formula

Rewrite the high-low estimator as:

$$\widehat{S}^{HL} = \left(1+\sqrt{2}\right)\left(\sqrt{\mathbb{E}\left[r_t^2 + r_{t+1}^2\right]} - r_t^*\right) = S + \left(1+\sqrt{2}\right)\left(\sqrt{2}\mathbb{E}\left[\mathcal{R}_t\right] - \mathcal{R}_t^*\right)$$

- *if* we could use $\mathbb{E}\left[r_t^2 + r_{t+1}^2\right]$, $\widehat{S}^{HL}$ is unbiased iff $\sqrt{2}\mathbb{E}\left[\mathcal{R}_t\right] = \mathcal{R}_t^*$
- which is the case asymptotically proposition 1 (I)

But in practice we can only estimate $\widehat{S^*}^{HL}$ which uses $r_t + r_{t+1}$:

- $\widehat{S^*}^{HL} = \omega S + \left(1+\sqrt{2}\right)\left(\phi\mathcal{R}_t^{min} - \mathcal{R}_t^*\right)$
- where $\phi \equiv \sqrt{1+\kappa^2}$, $\kappa \equiv \max\{r_t, r_{t+1}\} / \min\{r_t, r_{t+1}\}$, $\omega \equiv \left(\phi + \sqrt{2}\phi - 1 - \sqrt{2}\right)$, and $\mathcal{R}_t^{min} = \min\{\mathcal{R}_t, \mathcal{R}_{t+1}\}$

Empirical estimator

So when is the estimator actually used in practice unbiased?

Empirical estimator

So when is the estimator actually used in practice unbiased?

- virtually never.
- (only when $r_t = r_{t+1}$) proposition 1 (II)

But the result above relies on the asymptotic distribution

- similar to first source of bias in Roll measure

Empirical estimator

So when is the estimator actually used in practice unbiased?

- virtually never.
- (only when $r_t = r_{t+1}$) proposition 1 (II)

But the result above relies on the asymptotic distribution

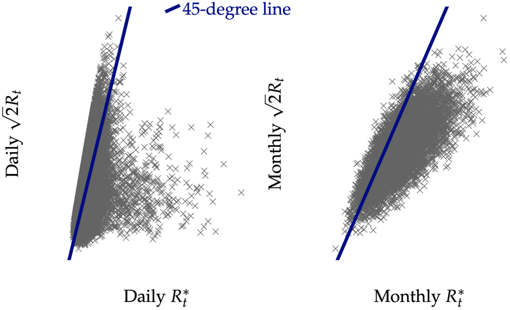
- similar to first source of bias in Roll measure

Even if consecutive ranges were equal, $\widehat{S^*}^{HL}$ would still be biased in small samples proposition 2

$$\mathbb{E}\left[\widehat{S^*}^{HL}\right] - S = \left(1+\sqrt{2}\right)\left[\underbrace{r_t^{min}\left(\phi - \sqrt{2}\right)}_{=0 \text{ if } r_t = r_{t+1}} + \underbrace{\left(\sqrt{2}\mathcal{R}_t^{min} - \mathcal{R}_t^{*}\right)}_{?}\right]$$

210,000 trading days (10,000 monthly averages)

Unbiased if and only if $\sqrt{2}\mathbb{E}\left[\mathcal{R}_t\right] = \mathbb{E}\left[\mathcal{R}_t^*\right]$:



- Even under simulated ideal conditions almost never happens

Moment bias and small sample bias

$$\mathbb{E}\left[\widehat{S^*}^{HL}\right] - S = \left(1+\sqrt{2}\right)\left[\underbrace{r_t^{min}\left(\phi-\sqrt{2}\right)}_{\textbf{Moment Bias}} + \underbrace{\left(\sqrt{2}\mathcal{R}_t^{min} - \mathcal{R}_t^*\right)}_{\textbf{Small Sample Bias}}\right]$$

- these are exactly the same sources of bias as in the Roll estimator

Moment bias and small sample bias

$$\mathbb{E}\left[\widehat{S^*}^{HL}\right] - S = \left(1+\sqrt{2}\right)\left[\underbrace{r_t^{min}\left(\phi - \sqrt{2}\right)}_{\textbf{Moment Bias}} + \underbrace{\left(\sqrt{2}\mathcal{R}_t^{min} - \mathcal{R}_t^*\right)}_{\textbf{Small Sample Bias}}\right]$$

- these are exactly the same sources of bias as in the Roll estimator
- **moment bias** only depends on observable data
- **small sample bias** is approximately

$$\sqrt{2}\mathbb{E}\left[\mathcal{R}_t\right] - \mathbb{E}\left[\mathcal{R}_t^*\right] = \sigma\sqrt{2/\pi}\left(1 - \sum_{k=1}^{2}\frac{1}{\sqrt{k}}\right) \approx -0.08\sigma$$

What determines the bias?

Proposition 3. *The bias in* $\widehat{S^*}^{HL}$ *is decreasing in* S*. Given sufficiently high values of volatility relative to the the latent spread* ($\sigma/S \approx 4$)*, the bias increases in* σ*.* proof

Consequence of estimation bias previously documented:

- marginally correlates with spread changes in FX and ETFs (Karnaukh et al. (2015), Marshall et al. (2018))
- no correlation with commodity trading costs (Marshall et al. (2011))

What determines the bias?

Proposition 3. *The bias in* $\widehat{S^*}^{HL}$ *is decreasing in* S*. Given sufficiently high values of volatility relative to the the latent spread* ($\sigma/S \approx 4$)*, the bias increases in* σ*.* proof

Consequence of estimation bias previously documented:

- marginally correlates with spread changes in FX and ETFs (Karnaukh et al. (2015), Marshall et al. (2018))
- no correlation with commodity trading costs (Marshall et al. (2011))

If you sort US stocks into monthly liquidity deciles using $\widehat{S^*}^{HL}$...

...**only 22% of stocks will be in the correct group**

- other estimators will do just as poorly (or worse)

Practical considerations

Previous work suggested via data or simulations that bias worsens with small spreads and large volatility

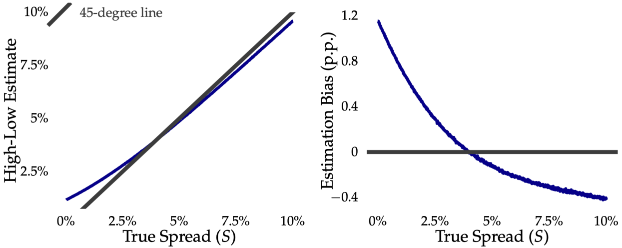
Jahan-Parvar and Zikes (2023) conjecture bias is induced by *ad-hoc* adjustments and driven by volatility

- but *ad-hoc* adjustments and model violations account for 5% of misbehavior
- volatility only starts to matter when it becomes too large relative to S
- signal-to-noise ratio < 0.25

Say you want to regress returns on liquidity risk (eg Acharya and Pedersen (2005), Pontiff and Singla (2019)) or to control for liquidity (McLean and Pontiff (2016))

- Because bias decreases in S, estimator suffers from non-classical measurement error
- attenuation bias may even flip the sign of estimated effect

Relationship between bias and spread size



For small spreads (below 1%), the high-low estimator suffers from an upward bias. As the spread becomes larger, the bias decreases. For reference, the median effective spread in US stocks during 2003-2015 is 0.3%.

Notes: This figure computes 10,000 monthly averages of the high-low spread estimator (with zeros adjustment) for each 0.1-percentage-point increment in the spread between 0.1% and 10% from simulated data. The data generating process is described in detail in the text and is designed to maintain all model assumptions, including constant true spreads in each sample. Bias is defined as the difference between the estimated spread and the true spread.

Estimating bias empirically

Because only **small sample bias** is not observable

$$S = \widehat{S^*}^{HL} - \left(1+\sqrt{2}\right)\left[r_t^{min}\left(\phi - \sqrt{2}\right) + \left(\sqrt{2}\mathcal{R}_t^{min} - \mathcal{R}_t^*\right)\right],$$

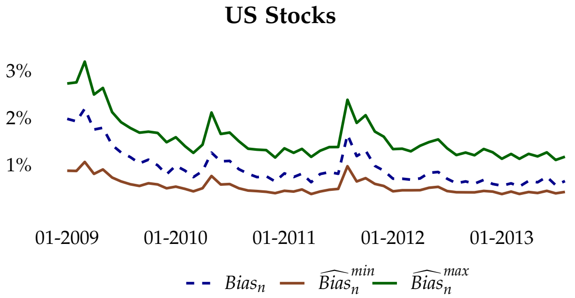
estimating the bias empirically comes down to bounding $\sqrt{2}\mathcal{R}_t^{min} - \mathcal{R}_t^*$

- sign **small sample bias**: method correctly signs 90% of cases
- use bias expression to back out $\left[\widehat{Bias_n}^{min}, \widehat{Bias_n}^{max}\right]$

These bounds work best when $\widehat{S^*}^{HL}$ is most biased (by construction)

Empirical bias bounds

- bias bounds track actual estimation bias almost perfectly: 98% correlation
- midpoint can be used in cross-sectional regs to soak up measurement error



Next steps for bid-ask spread estimators

New estimators are introduced and compared to previous ones in horse-races

- if new estimator correlates better with effective spreads in US stocks than eg Roll, it's a better mousetrap

This approach is useful but misses structural problem with technology used

- issues not only persist, but are the same despite alternative theory for $\mathcal{P}_t$
- eventually studies in other markets cast doubt on new measure being better mousetrap

Appendix

Proof 1, Proof Back

Since $\mathbb{E}\left[R_t^*\right] = \sqrt{2}\mathbb{E}\left[R_t\right]$, unbiasedness of $\widetilde{S}^{HL}$ immediately follows. For $\widehat{S}^{HL}$ to be unbiased, it must be that $\left(1+\sqrt{2}\right)\left(\phi R_t^{min} - R_t^*\right) = \left(1-\omega\right)S$. Expanding both sides yields

$$\mathbb{E}\left[\phi\right]\mathbb{E}\left[R_t\right] - \sqrt{2}\mathbb{E}\left[R_t\right] = \sqrt{2}S - \mathbb{E}\left[\phi\right]S$$

which holds if and only if $\phi = \sqrt{2}$, which in turn is implied by $r_t = r_{t+1}$, $\forall t$. Note that $\phi = \sqrt{2}$ results in $\omega = 1$. This completes the proof.

Proof 2, Proof ▸ Back

Start with

$$\left[\left(\phi+\sqrt{2}\phi-1-\sqrt{2}\right)S+\left(1+\sqrt{2}\right)\left(\phi R_t^{min}-R_t^*\right)\right]-S$$

which yields the following after straightforward algebra and by noting that $r_t = R_t + S$:

$$-2r_t^{min}+2R_t^{min}-\sqrt{2}S+\phi r_t^{min}+\sqrt{2}\phi r_t^{min}-R_t^*\left(1+\sqrt{2}\right)$$

Adding and subtracting $\sqrt{2}r_t^{min}$ from the above and further algebra finally gives

$$\left(1+\sqrt{2}\right)\left[r_t^{min}\left(\phi-\sqrt{2}\right)+\left(\sqrt{2}R_t^{min}-R_t^*\right)\right]. \tag{1}$$

We know must show that the expression in (1), which measures the daily estimation error in the high-low proxy, is equivalent in expectation to the estimator's bias. First, consider the implementable high-low spread formula and its expected value:

$$\widehat{S}^{HL}=\left(1+\sqrt{2}\right)\left(\phi r_t^{min}-r_t^*\right)$$

and then

$$\mathbb{E}\left[\widehat{S}^{HL}\right]-S=\left(1+\sqrt{2}\right)\left(\mathbb{E}\left[\phi r_t^{min}\right]-\mathbb{E}\left[R_t^*\right]\right)-S\left(2+\sqrt{2}\right)$$

which gives the estimator's bias and is identical to the expectation of (1).

Proof 3, Proof

Back

WLOG, assume that $R_t < R_{t+1}$. Therefore, $\partial\phi/\partial S < 0$ and thus

$$\frac{\partial\left(\widehat{S}^{HL} - S\right)}{\partial S} = \left(1+\sqrt{2}\right)\left(R_t^{min}\frac{\partial\phi}{\partial S} + \phi + S\frac{\partial\phi}{\partial S} - \sqrt{2}\right) =$$

$$-\frac{\left(1+\sqrt{2}\right)\left(R_t^{min}\sqrt{\dfrac{2\left(R_t^{max}+S\right)^2}{\left(R_t^{min}+S\right)^2}+2} + S\sqrt{\dfrac{2\left(R_t^{max}+S\right)^2}{\left(R_t^{min}+S\right)^2}+2} - R_t^{max} - R_t^{min} - 2S\right)}{\left(R_t^{min}+S\right)\phi} < 0$$

since $R_t^{min} = R_t < R_{t+1} = R_t^{max}$ and $S > 0$. For the volatility result, recall $\phi r_t^{min} \equiv \sqrt{\widehat{\beta}}$. Thus, the expected value of the bias is:

$$\left(1+\sqrt{2}\right)\left(\mathbb{E}\left[\sqrt{\left(r_t^{max}\right)^2 + \left(r_t^{min}\right)^2}\right] - \sqrt{2}\mathbb{E}\left[r_t^{min}\right]\right) \leq \left(1+\sqrt{2}\right)\left[\sqrt{8\ln 2\sigma^2 + 4S\sqrt{\frac{8}{\pi}}\sigma + 2S^2} - \sqrt{2}\sqrt{\frac{8}{\pi}}\sigma - \sqrt{2}S\right]$$

by Jensen's inequality. The derivative of the above wrt σ is positive iff $\dfrac{\sigma}{S} > 4.20 - \varepsilon$, where $\varepsilon \geq 0$ is the error induced by the concavity of β. This completes the proof.