Imperial College
Business School

Lecture 2

Measuring liquidity

1

Outline

- Explicit and implicit trading costs
- Spread-based measures:
  - Quoted, effective and realized spread
  - VWAP
- Measures based on order flow and volume data:
  - Price impact
  - Non-trading measures
- Roll's measure: based only on transaction prices
  - Derivation and assumptions
  - Possible biases and their corrections
- Liquidity's time dimension: implementation shortfall

2

Reading

Chapter 2, Foucault, Pagano and Röell, *Market Liquidity,* 2nd edition, 2024.

Explicit and implicit trading costs

Measuring trading costs

- **Explicit trading costs**:
  - platform trading fees
  - transaction taxes
  - brokerage commissions
  - clearing and settlement fees
- **Implicit trading costs**:
  - **bid-ask spread**: as seen in Session 1, it exists not only in dealer markets but also in LOB markets
  - **market impact**: price pressure from a large trade
  - **delay costs**: adverse change in price between order submission and execution

Trading costs in European equity markets (except UK)

Quarter	Delay Costs	Impact Costs	Comm. Costs	Total
2016				
Q1	26	15	8	49
Q2	17	16	8	40
Q3	20	13	8	41
Q4	19	13	7	40
2017				
Q1	17	11	7	35
Q2	12	12	7	31
Q3	15	12	7	34
Q4	16	13	7	36
2018				
Q1	16	12	5	34
Q2	20	11	5	36
Q3	24	12	5	41
Q4	27	15	5	46
2019				
Q1	18	14	5	37
Q2	16	14	5	35
Q3	19	13	5	37
Q4	18	13	5	36
2020				
Q1	24	22	5	51
Q2	22	28	4	54
Q3	24	15	4	43

Source: Virtu Global Cost Review, 2020, available at https://www.virtu.com/thinking/thought-leadership/

Measuring implicit trading costs: data requirements

- **Bid-ask spreads** require
  - bid and ask quotes
  - possibly transaction price
  - submission and execution times
- **Gap between transaction price and VWAP** and **Roll's measure** require transaction prices
- **Price impact of orders** requires transaction prices and order imbalance
- **Non-trading measures of illiquidity** require trade volumes

Bid-ask spread

Quoted spread

- a = ask price (to buy q shares)
- b = bid price (to sell q shares)

For very small q, these are the "best bid and offer" quotes (BBO)…

- Absolute spread: $S \equiv a - b$

… and this is the cost for a small "round-trip trade" at the BBO ("market touch" in a dealer market)

- Midprice: $m \equiv \dfrac{a+b}{2}$
- Relative spread: $s \equiv \dfrac{S}{m}$

(also: "percentage quoted spread")

Quoted spread (2)

- If q exceeds the smallest possible quantity, look at the *average quoted spread:*

$$s(q) \equiv \frac{\overline{a}(q) - \overline{b}(q)}{m}$$

increasing in q

- The quoted spread measures liquidity for someone placing a market order
- In LOB markets, it neglects the liquidity offered by hidden orders
- In dealer markets, it may:
  - understate liquidity, as it neglects price improvement
  - overstate it, if dealers' quotes are purely indicative

Effective spread

- The effective *half*-spread is the distance of the transaction price *p* from the mid-price *m*:

$$S_e \equiv d(p-m)$$

where *d* is direction of the trade: +1 for a buy (a trade initiated by a buyer), –1 for a sell

- It measures liquidity for a hypothetical trade, based on past data: *retrospective*, while quoted spread is *prospective*

- In relative terms: $$s_e \equiv \frac{S_e}{m}$$

Effective spread (2)

- Commonly used, and can be computed for various trade sizes
- A trader wanting to place a large order can split it in many smaller trades $\Rightarrow$ the relevant transaction price *p* is the average of the prices obtained in the various trades
- But an outside observer (econometrician), who cannot observe the *orders* that generated a set of trades:
  - must make (arbitrary) assumptions to relate trades to orders
  - must estimate trade direction *d* (Lee-Ready algorithm) – yet, trades may be reported with delay $\Rightarrow$ misaligned from quotes!

Realized spread

- As the effective spread measures trading costs, one may think that it also measures dealers' profits (zero-sum)
- But after a trade, dealers' quotes adjust, generally in the same direction: after filling a sell order, quotes decrease $\Rightarrow$ the shares bought by the dealer are less valuable $\Rightarrow$ his profits are less than the effective spread on the trade
- The *realized half-spread* accounts for this, by comparing *p* with a subsequent *m*:

$$S_r \equiv d_t(p_t - m_{t+\Delta}) = \underbrace{d_t(p_t - m_t)}_{\substack{effective \\ half-spread}} - \underbrace{d_t(m_{t+\Delta} - m_t)}_{\substack{midquote \\ revision}}$$

Average quoted, effective & realized spread on two days

16 March, 2020 (High volatility)

Stock	Quoted spread	Effective spread	Realized spread	Price impact
Apple	3.2	1.8	0.6	1.2
JPMorgan Chase	5.1	2.8	0.9	1.8
Pfizer	3.7	2.4	0.6	1.8
Tesla	18.0	6.5	2.3	4.1

5 August, 2020 (Low volatility)

Stock	Quoted spread	Effective spread	Realized spread	Price impact
Apple	1.0	0.5	0.1	0.4
JPMorgan Chase	1.3	0.9	0.3	0.6
Pfizer	2.6	1.2	0.6	0.5
Tesla	7.2	2.0	1.1	0.9

Another benchmark: VWAP

- What is the right benchmark for the transaction price p_t?
  - The effective spread compares it to *current* midprice m_t
  - The realized spread with the *subsequent* midprice $m_{t+\Delta}$
- Another popular benchmark is the *Volume-Weighted Average Price* (VWAP), possibly over the trading day:

$$VWAP \equiv \frac{\text{€ } volume\ of\ trading}{no.\ of\ shares\ traded} = \sum_{t \in T} w_t p_t$$

where w_t is the weight of the t^{th} trade in total volume.

Gaming VWAP

- Suppose a buy side investor – say, a mutual fund – uses VWAP to assess execution quality by a broker
- Then the broker may time the trade so as to “look good”:
  - by buying after a price decline (selling after a price rise) or, lacking this, postponing the trade to another day
  - more generally, by trickling in the order very slowly, to make average execution price as close a possible to VWAP
- To prevent such “gaming”, the investor may want to monitor the broker’s execution timing, to prevent excessive delay $\Rightarrow$ take also the time dimension of liquidity into account: more on this later …

Price impact

Price impact

- Buy orders tend to raise the mid-price (and sell orders to decrease it) $\rightarrow$ the impact of orders can be expressed as

$$\underbrace{\Delta m_t}_{\text{midquote change between t–1 and t}} = \underbrace{\lambda}_{\text{price impact coefficient}} \underbrace{x_t}_{\text{order imbalance between t–1 and t}} + \varepsilon_t$$

$\Rightarrow 1/\lambda \equiv$ market depth

- Stoll (2000) estimates λ on a sample of NYSE, AMSE and NASDAQ stocks and finds that it is positive for 98% of them, and larger for small-caps

Stoll’s price impact estimates

Journal of Finance, 2000, Table IV, p. 1497.

Price Impact Regressions

$$\Delta P_t = \lambda_0 + \lambda I_t + \lambda_2 I_{t-1} + e_t$$

ΔP_t is the change in the closing quote midpoint adjusted for the return on the S&P 500 index; I_t is the difference between the daily share volume on the ask side and on the bid side expressed as a percentage of daily volume. Regressions are run using 61 daily observations for each stock. The table summarizes the average values of the coefficients.

	NYSE/AMSE	Nasdaq
Mean λ_0	−0.0314	0.0311
$t(\lambda_0)^*$	−10.494	16.957
mean t of individual regressions	−0.242	0.529
% positive	40.4	71.2
% positive and significant (5%)	1.3	6.7
% negative and significant (5%)	4.1	0.46
Mean λ	0.007671	0.007298
$t(\lambda)^*$	30.121	29.708
mean t	2.435	2.711
% positive	97.9	97.9
% positive and significant (5%)	63.1	71.2
% negative and significant (5%)	0.0	0.0
Mean λ_2	−0.000701	0.000168
$t(\lambda_2)^*$	−6.504	1.982
mean t	−0.064	0.240
% positive	48.7	59.7
% positive and significant (5%)	2.5	4.7
% negative and significant (5%)	4.3	2.6
Mean adj R^2	0.0929	0.1192
Number of days	61	61
Number of stocks	1,706	2,184

10% increase in imbalance ⇒ price impact of 7.5 cents

Imperial College Business School 19

19

Price impact (2)

- Order imbalance tends to be correlated with trading volume $Vol_t = |x_t|$
- So if order imbalance data are not available, one can use *trading volume* to explain the *absolute* change in prices $|\Delta m_t|$ (Hasbrouck, 2007)
- Same idea in other volume based measures, such as the “*illiquidity ratio*” by Amihud (2002):

$$I_t \equiv \frac{|\Delta m_t|}{|x_t|} \equiv \frac{|r_t|}{Vol_t}$$

or its inverse, the Amivest liquidity ratio $L_t \equiv 1/I_t$

Imperial College Business School 20

20

Other volume-based measures

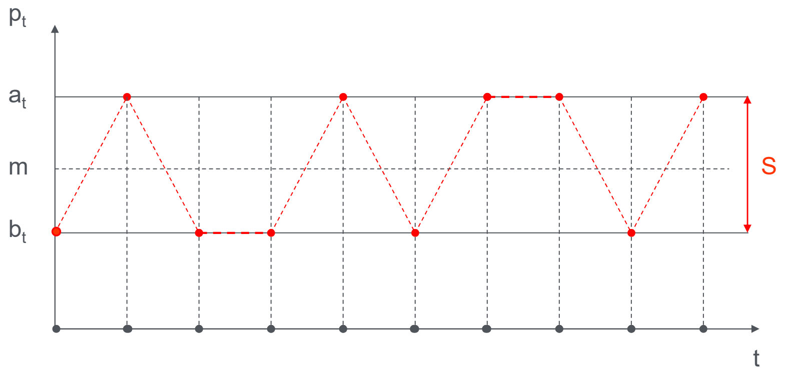
- Trading volume *per se* is a poor proxy for liquidity: *e.g.* during the 2008-09 crisis, trading volume was high, yet all spread measures of liquidity were high
- But the frequency of “no trading” may still tell us something about liquidity: if trading is too costly, people will not trade (Bekaert *et al*., 2006)
- Also, easy to measure even in emerging markets: look at frequency of “stale prices” (“zeros measure” of illiquidity)
- More sophisticated method: Lesmond, Ogden and Trcinka (1999) propose a maximum likelihood estimate of trading costs (LOT) based on observed no-trade intervals

Roll’s measure

Roll's measure

- Spread measures require data about *quotes*, and sometimes *transaction prices*
- Price impact and volume-based measures requires data about *quotes* and *orders or trades*
- What to do if you have no data about quotes and trades?
- Roll (1984) proposed an ingenious measure of the bid-ask spread based on *transaction prices alone*
- Roll's idea: sometimes orders hit the ask and sometimes they hit the bid ("bid-ask bounce") $\Rightarrow$ negative covariance in transaction-to-transaction price changes (returns) $\Rightarrow$ exploit this to estimate the bid-ask spread

Intuition: simple case with constant midquote



Deriving Roll's measure

- Suppose that the midprice follows a random walk:

$$m_t = m_{t-1} + \varepsilon_t$$

with
$$\mathrm{E}(\Delta m_t) = 0, \;\; \mathrm{E}(\varepsilon_t \varepsilon_{t-1}) = 0$$

- Transaction prices are

$$p_t = \begin{cases} a_t = m_t + S/2 & \text{if } d_t = +1 \\ b_t = m_t - S/2 & \text{if } d_t = -1 \end{cases}$$

Hence:
$$p_t = m_t + \frac{S}{2} d_t \;\; \Rightarrow \;\; \Delta p_t = \frac{S}{2} \Delta d_t + \varepsilon_t$$

Deriving Roll's measure (2)

Suppose that trades are
1) balanced,
2) serially uncorrelated, and
3) uncorrelated with "news":

$$1)\ \mathrm{E}(d_t) = 0, \;\; 2)\ \mathrm{E}(d_t d_{t-s}) = 0, \;\; 3)\ \mathrm{E}(d_t \varepsilon_t) = \mathrm{E}(d_t \varepsilon_{t+s}) = 0.$$

$$\Rightarrow \;\; \mathrm{cov}(\Delta p_{t+1}, \Delta p_t) = \mathrm{E}(\Delta p_{t+1} \Delta p_t) - \underbrace{\mathrm{E}(\Delta p_{t+1}) \mathrm{E}(\Delta p_t)}_{=0}$$

$$= \mathrm{E}\left[\left(\frac{S}{2} \Delta d_{t+1} + \varepsilon_{t+1}\right)\left(\frac{S}{2} \Delta d_t + \varepsilon_t\right)\right]$$

$$= \frac{S^2}{4} \mathrm{E}(\Delta d_{t+1} \cdot \Delta d_t)$$

Deriving Roll's measure (3)

$$\operatorname{cov}(\Delta p_{t+1}, \Delta p_t) = \frac{S^2}{4} \mathrm{E}(\Delta d_{t+1} \cdot \Delta d_t)$$

$$= \frac{S^2}{4} \mathrm{E}\left[(d_{t+1} - d_t)(d_t - d_{t-1})\right]$$

$$= \frac{S^2}{4} \mathrm{E}\left[d_{t+1}d_t - d_{t+1}d_{t-1} - d_t d_t + d_t d_{t-1}\right]$$

$$= \frac{S^2}{4} \underbrace{\left[-\mathrm{E}(d_t^2)\right]}_{=-\frac{1}{2}(+1)^2 - \frac{1}{2}(-1)^2 = -1} \quad = -\frac{S^2}{4}$$

$$\boxed{\Rightarrow \quad S_R = 2\sqrt{-\operatorname{cov}(\Delta p_{t+1}, \Delta p_t)}}$$

Roll's measure and its limits

- Scaling by the price, one can estimate the relative spread:

$$\boxed{s_R = 2\sqrt{-\operatorname{cov}(r_{t+1}, r_t)}}$$

- Roll's measure is simple and easy to estimate, but it yields a biased estimate of *S* if any of its stringent assumptions fails:
  - balanced order flow
  - random and serially independent trade direction
  - no informational content in the order flow
  - constant expected return

a) Unbalanced order flow

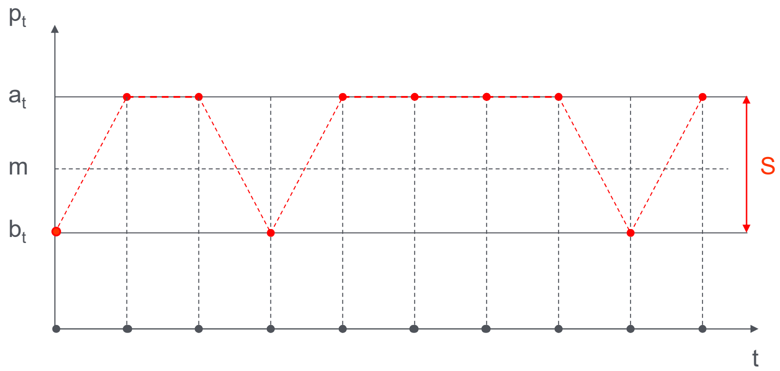
- If the probability of a buy order is η and that of a sell is $1-\eta$, with $\eta \neq ½$, then:

$$\mathrm{cov}(\Delta p_{t+1}, \Delta p_t) = \frac{S^2}{4} \overbrace{\mathrm{E}(\Delta d_{t+1} \cdot \Delta d_t)}^{-4\eta(1-\eta)} \quad \Rightarrow S_a = \sqrt{-\frac{\mathrm{cov}(\Delta p_{t+1}, \Delta p_t)}{\eta(1-\eta)}}$$

- $\eta\,(1-\eta)$ is maximal (and equal to ¼) for $\eta = ½$ $\Rightarrow$ Roll's measure
- So with $\eta \neq ½$, Roll's measure S_R *underestimates* the spread:

$$S_a = \frac{2\sqrt{-\mathrm{cov}(\Delta p_{t+1}, \Delta p_t)}}{2\sqrt{\eta(1-\eta)}} = \frac{S_R}{2\sqrt{\eta(1-\eta)}} \quad \Rightarrow S_R = \underbrace{2\sqrt{\eta(1-\eta)}}_{<1}\, S_a$$

Intuition: when buys are more frequent than sells (η>1/2)



b) Autocorrelated orders

- If the trade direction d_t is autocorrelated, with prob($d_t = d_{t-1}$) = $\delta \neq$ ½, then:

$$\mathrm{cov}(\Delta p_{t+1}, \Delta p_t) = \frac{S^2}{4} \overbrace{\mathrm{E}(\Delta d_{t+1} \cdot \Delta d_t)}^{-4(1-\delta)^2}$$

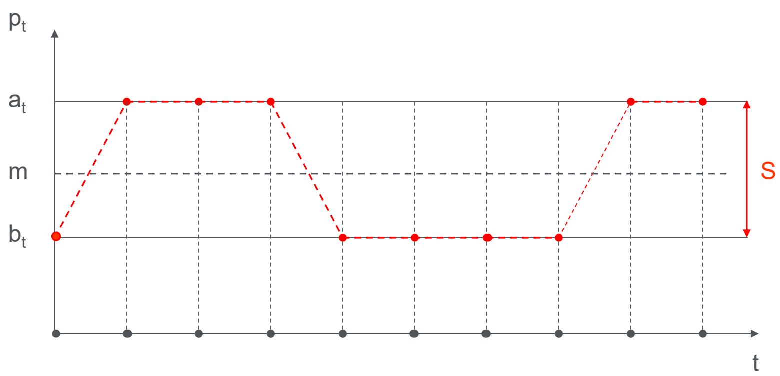
this is −1 with no autocorrelation

$$\Rightarrow S_b = \frac{1}{1-\delta}\sqrt{-\mathrm{cov}(\Delta p_{t+1}, \Delta p_t)} = \frac{2\sqrt{-\mathrm{cov}(\Delta p_{t+1}, \Delta p_t)}}{2(1-\delta)} = \frac{S_R}{2(1-\delta)}$$

- Choi et al. (1988) estimate $\delta = 0.7 \Rightarrow$ Roll's measure S_R *underestimates* the spread at 0.6 of its true value:

$$S_b = \frac{1}{0.3}\sqrt{-\mathrm{cov}(\Delta p_{t+1}, \Delta p_t)} = \frac{S_R}{2 \cdot 0.3} = \frac{S_R}{0.6} \Rightarrow S_R = 0.6 \cdot S_b$$

Intuition: when orders are positively autocorrelated (δ>1/2)



c) Informative order flow

- As we shall see in the next class, if the order flow is informative (i.e., if d_t and ε_t are correlated), orders have a *direct impact on the mid-price m_t*:
  - after a buy order, liquidity suppliers revise *both* ask and bid prices upwards $\rightarrow m_t$ increases
  - after a sell, they revise *both* ask and bid prices downwards $\rightarrow m_t$ decreases
- Hence, orders do *not just* induce bid-ask bounce around a *given* midquote $\Rightarrow$ yet another reason why the usual Roll measure may *underestimate* the bid-ask spread

d) Varying expected return

- With a time-varying expected return in returns:

$$m_t = m_{t-1} + \overline{r}_t + \varepsilon_t \qquad \Rightarrow \; \Delta p_t = \overline{r}_t + \frac{S}{2}\Delta d_t + \varepsilon_t$$

$$\Rightarrow \; \operatorname{cov}(\Delta p_{t+1}, \Delta p_t) = \operatorname{cov}(\overline{r}_{t+1}, \overline{r}_t) - \frac{S^2}{4}$$

- So the correct spread estimator becomes:

$$S_d = 2\sqrt{-\operatorname{cov}(\Delta p_{t+1}, \Delta p_t) + \operatorname{cov}(\overline{r}_{t+1}, \overline{r}_t)}$$

- If expected returns are positively autocorrelated at high frequencies (“momentum”), we have another reason why the usual Roll measure S_R *underestimates* the spread

Empirical performance of Roll's measure

- One or more of these reasons may explain why empirically Roll's measure underestimates the spread compared to the quoted and effective spread
- They also explain why sometimes it cannot even be computed: at high frequency, the covariance of price changes is often positive → root of a negative number!
- True for roughly half of the stocks in Roll's (1984) study
- As we shall see, Roll's measure captures only *one* of the *three* components of the spread: that arising from order-processing costs, *i.e.* the real resource costs of trading

Time dimension of liquidity

- All the liquidity measures seen so far do not account for its time dimension
- Trickling in an order slowly to minimize price impact may imply partial execution → opportunity cost in terms of forgone returns
- We may want a measure of trading costs that gives some weight not only to execution costs but also to this opportunity cost → "implementation shortfall"
- Idea: benchmark the actual portfolio's performance against a hypothetical "paper portfolio" for which trade occurs *costlessly* <u>*and*</u> *instantaneously* at mid-prices

Implementation shortfall

- Benchmark = "paper portfolio" acquired at midquote m_0 when the (signed) order q was sent to the broker (at time 0):

$$R_p = q(m_t - m_0)$$

- At time t the client reviews the broker's performance relative to that of the paper portfolio, R_p

- If the broker executed a fraction κ of the order at an average price $\overline{p}$, the *actual* gain on his position is

$$R_a = \kappa q(m_t - \overline{p}).$$

Implementation shortfall (2)

- The implementation shortfall (IS) is

$$\begin{aligned} IS \equiv R_p - R_a &= q(m_t - m_0) - \kappa q(m_t - \overline{p}) \\ &= \underbrace{\kappa q(\overline{p} - m_0)}_{execution\ cost} + \underbrace{(1-\kappa) q(m_t - m_0)}_{opportunity\ cost} \end{aligned}$$

- The 1st term can itself be broken down into a price pressure component at execution time τ and a delay cost component:

$$\underbrace{\kappa q(\overline{p} - m_0)}_{execution\ cost} = \underbrace{\kappa q(\overline{p} - m_\tau)}_{price\ pressure} + \underbrace{\kappa q(m_\tau - m_0)}_{delay\ \cos t}$$

Implementation shortfall (3)

- It is generally computed by averaging over many trades:

$$E(IS) = \kappa E\left[q(\bar{p} - m_0)\right] + (1-\kappa)E\left[q(m_t - m_0)\right]$$

- 2nd term can be positive (q correlated with Δm) because of the price pressure produced by the order flow
- 1st term can often be lowered by more patient trading (lower κ) $\Rightarrow$ tradeoff with 2nd term: patient = slow!
- Tradeoff depends on market resilience: *e.g.*, how fast the LOB (limit order book) is replenished after large orders

Key takeaways

- Some trading costs are explicit, some implicit
- Implicit trading costs include (i) the bid-ask spread, (ii) the price impact of orders and (iii) delay costs
- Depending on the available data, one can use different measures of implicit trading costs:
  - bid-ask spreads require quotes, sometimes also transaction prices
  - price impact measures require quotes and order flow data
  - Amihud's illiquidity ratio requires returns (prices) and volume data
  - non-trading and Roll's measure only require transaction price data
- The implementation shortfall also accounts for the time dimension of liquidity: partial execution and delay costs