








Modelling Financial Markets

Seminar 10: Market Microstructure — Hands-on



Daniel Traian PELE
Drd. Antoaneta Amza, antoaneta.amza@csie.ase.ro

Bucharest University of Economic Studies
IDA Institute Digital Assets
Blockchain Research Center
AI4EFin Artificial Intelligence for Energy Finance
Romanian Academy, Institute for Economic Forecasting
MSCA Digital Finance

How Precisely Can We Measure Trading Costs?

- Question of the seminar
  - **how precisely can trading costs and liquidity be measured from public prices and volumes?**
- Today, in order: A1, A2, B1, B2, B3, B4, C1
- Deliverables
  - the answers to the A exercises
  - one figure and one table per B exercise
  - the corrected C1 code
  - one interpretation sentence per exercise
- Notebook for today: `chapter10_seminar_notebook.ipynb` (Colab)

What Do We Observe in Each Bar?

- OHLC means Open, High, Low, Close: four prices within one interval. Volume counts traded shares; traded value also depends on price.
- UTC means Coordinated Universal Time. We convert the fund time stamps to New York time before choosing the trading session; Bitcoin retains UTC.
- A tick is the minimum permitted price step. A basis point (bp; pb in Romanian) is one hundredth of a percentage point: 25 bp means 0.25 percent. The tick is in price units, whereas bp expresses a relative change.
- The standard error (SE) of an estimate: how much the estimate would vary from one sample to another.

Which Markets Do We Compare?

- US (United States), USD (US dollar). An ETF (Exchange-Traded Fund) is a fund traded on an exchange; SPY (SPDR S&P 500 ETF Trust) is the ETF on the S&P 500 index (Standard & Poor's 500).
- US stocks, by ticker (Nasdaq or NYSE): AAPL (Apple), MSFT (Microsoft), NVDA (NVIDIA), JPM (JPMorgan Chase), GME (GameStop), MSTR (Strategy, formerly MicroStrategy).
- BVB (Bursa de Valori București, the Bucharest Stock Exchange), BET (the BVB benchmark index, Bucharest Exchange Trading), BET-TR (BET Total Return, with dividends reinvested), TVBETETF (the ETF BET Patria–Tradeville, listed on BVB, tracking the BET), BNR (Banca Națională a României, the National Bank of Romania), RON (the Romanian leu).
- VIX (the volatility index of the Cboe exchange, formerly CBOE, Chicago Board Options Exchange) is computed from option quotes: it is implied volatility, not the volatility realised in our sample. Crypto: BTC (Bitcoin), ETH (Ether, the Ethereum network), SOL (Solana), XRP (the token of the XRP Ledger).

Which Romanian Companies Do the Tickers Identify?

- BVB tickers: TLV (Banca Transilvania), BRD (BRD–Groupe Société Générale), SNP (OMV Petrom), H2O (Hidroelectrica), SNG (Romgaz).
- SNN (Nuclearelectrica), TGN (Transgaz), TEL (Transelectrica), EL (Electrica), DIGI (Digi Communications). OMV stands for Österreichische Mineralölverwaltung, the Austrian OMV group.
- Compare liquidity only after converting traded values to the same currency and checking how dividends and splits are treated.

Data Used

Series	Content	Window	Units
SPY 5-minute bars	OHLC and volume; UTC time stamps converted to New York time, bars 09:30–15:55, only days with all 78 bars	12 Oct 2020 – 18 Sep 2026, 1 477 days	USD; volume in shares
Bitcoin 5-minute bars	OHLC, UTC time	31 Aug 2024 – 18 Sep 2026	USD
US, daily (7)	SPY, AAPL, MSFT, NVDA, JPM, GME, MSTR	19 Sep 2024 – 18 Sep 2026	USD; volume in shares
BVB, daily (11)	TLV, SNP, H2O, BRD, SNG, SNN, TGN, EL, DIGI, TEL and the BET ETF TVBETETF	same	lei; converted to USD at the BNR rate of the day
Crypto, daily (7)	BTC, ETH, SOL, XRP, Dogecoin, Cardano, Litecoin	same, 7 days a week	volume already in USD
VIX; BET-TR	VIX (CBOE implied-volatility index) and BET-TR (BET total-return index), daily closes	as in B4, B8	index points

- Daily series: open, high, low, close, adjusted close and volume; OHLC scaled by the ratio adjusted close / close (dividends and splits); traded value = split-adjusted close $\times$ volume; days with high = low or zero volume dropped
- BNR reference rate (RON per USD), official daily fixing; 1 bp = 0.01%

Setup: Run and Verify

```
# the data-loading functions are in the notebook's Data section
s = intraday_spy()          # SPY 5-minute bars
print(len(s), s['date'].nunique())
print(s.index[0], s.index[-1])
b = intraday_btc()          # Bitcoin, UTC
print(len(b), b['date'].nunique())
r = returns('TLV', '2024-09-19')
dv = dollar_volume('TLV', '2024-09-19')
print(len(r), dv.median() / 1e6)
d = ohlc('SPY', '2024-09-19')
print(len(d))  # daily Open, High, Low, Close bars
```

- Expected output
  - 115 206 SPY bars, 1 477 days
  - first bar 12 Oct 2020 09:30, last 18 Sep 2026 15:55 (New York time)
  - 213 754 Bitcoin bars on 749 days
  - 492 daily returns for Banca Transilvania; median traded value USD 3.46 million
- Daily SPY: 501 OHLC rows
- If a count differs, stop and check the series, the time zone and the start date

Notation Used in Every Exercise

- p_t: trade price (or bar close); m_t, q_t, c, $s = 2c$: slides "Bid, Ask and the Spread" and "Autocovariances and the Roll Model"; relative spreads in bp
- $\hat{\gamma}_0$, $\hat{\gamma}_1$: sample autocovariances of price changes, demeaned within the sample (slide "Autocovariances and the Roll Model"); T = number of price changes (78 closes give 77 changes)
- Intraday returns: log changes of the close within a day; the first bar runs from the open; $|r|$ in bp
- Amihud illiquidity ratio (ILLIQ): slide "Illiquidity: the Amihud Ratio"; day and block bootstraps: slide "Resampling: Day Bootstrap, Block Bootstrap, Monte Carlo"; every resampling uses the random seed 42
- One worked line: 78 closes, $\hat{\gamma}_1 = -1.0 \cdot 10^{-8}$ in log units gives $\hat{s} = 2\sqrt{10^{-8}} = 2 \cdot 10^{-4} = 2$ bp

Bid, Ask and the Spread

- Quotes: the **bid** b_t is the best price at which you can sell, the **ask** a_t the best price at which you can buy

$$m_t = \frac{a_t + b_t}{2}, \qquad c = \frac{a_t - b_t}{2}, \qquad s = 2c = a_t - b_t, \qquad S_t^Q = \frac{a_t - b_t}{m_t}$$

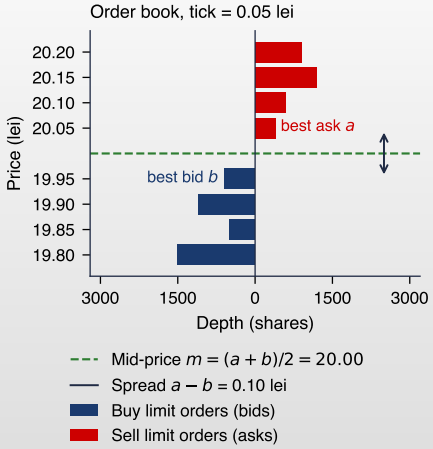
  - m_t: mid-price; c: half-spread; S_t^Q: quoted relative spread, in basis points (1 bp $= 0.01\%$)
  - a quoted spread is never below one **tick**, the smallest price step

- **Effective spread**: what a trade at price p_t actually costs

$$S_t^E = \frac{2q_t\,(p_t - m_t)}{m_t}, \qquad q_t = \begin{cases} +1 & \text{buyer-initiated trade} \\ -1 & \text{seller-initiated trade} \end{cases}$$

- The spread pays the liquidity provider for order processing, inventory risk and adverse selection (trading with better-informed counterparties)
- **Round trip**: buy at a, sell at b: the cost is $a - b$, **one** spread, half on each leg

The Limit Order Book



- **Limit order**: a price and a quantity, waiting in the book; **market order**: executes at once against the best opposite quotes
- Best bid b: the highest buy limit price; best ask a: the lowest sell limit price; depth: the shares waiting at each price
- Example (chart): $b = 19.95$, $a = 20.05$ lei, tick 0.05 lei
  - $m = 20$, $s = 0.10$ lei $= 0.5\% = 50$ bp; the spread is two ticks
  - one leg costs $c/m = 25$ bp, a round trip 50 bp
- A market buy larger than the depth at a also takes the next levels: its average price rises with the size (price impact)

Autocovariances and the Roll Model

- Sample autocovariances of the price changes $\Delta p_t = p_t - p_{t-1}$, $t = 1, \dots, T$

$$\hat{\gamma}_0 = \frac{1}{T}\sum_{t=1}^{T}\left(\Delta p_t - \overline{\Delta p}\right)^2, \qquad \hat{\gamma}_1 = \frac{1}{T-1}\sum_{t=2}^{T}\left(\Delta p_t - \overline{\Delta p}\right)\left(\Delta p_{t-1} - \overline{\Delta p}\right), \qquad \hat{\rho}_1 = \frac{\hat{\gamma}_1}{\hat{\gamma}_0}$$

  - T: the number of price changes; $\hat{\gamma}_1$ is the mean of the $T-1$ products; $\overline{\Delta p}$: the sample mean (centring)
  - γ_0, γ_1: the population values; hat: an estimate
- **Bid-ask bounce**: trades fall at random on the bid or the ask, so price changes tend to reverse
- **Roll model** Roll (1984): a random-walk efficient price m_t plus a bounce of $\pm c$

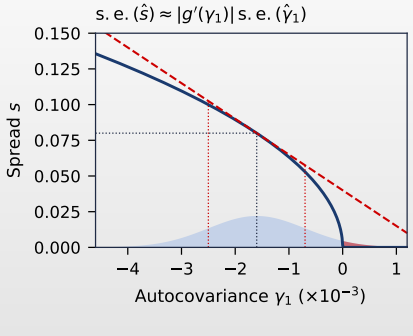
$$m_t = m_{t-1} + \varepsilon_t, \qquad p_t = m_t + c\, q_t, \qquad q_t = \pm 1 \text{ with probability } 1/2$$

  - ε_t: news, independent and identically distributed (i.i.d.), mean 0, variance σ^2; the model implies $\gamma_1 = -c^2$ (derived in A1)

$$\hat{s} = 2\sqrt{-\hat{\gamma}_1} \text{ (defined only if } \hat{\gamma}_1 < 0)$$

  - example: $\hat{\gamma}_1 = -0.0004$ lei^2: $\hat{c} = 0.02$, $\hat{s} = 0.04$ lei; if $\hat{\gamma}_1 \geq 0$, report "undefined" or truncate at 0 and say so

The Precision of the Roll Estimator



- Sampling variance of $\hat{\gamma}_1$ in the Roll model, with independent signs $q_t = \pm 1$ Harris (1990)

$$T \operatorname{Var}(\hat{\gamma}_1) \to \gamma_0^2 + \gamma_1^2$$

- **Delta method**: SE of a smooth function g of an estimate (first-order Taylor expansion)

$$\operatorname{SE}\left(g(\hat{\theta})\right) \approx |g'(\theta)| \operatorname{SE}(\hat{\theta})$$

  - chart: $g = 2\sqrt{-\gamma_1}$ near $\gamma_1 = -0.0016$; the tangent maps ± 1 SE of $\hat{\gamma}_1$ into $\hat{s}$
  - near 0 the curve is steep and the linear approximation fails; the red tail ($\hat{\gamma}_1 > 0$) has no $\hat{s}$

- **Centring bias**: subtracting the sample mean pushes $\hat{\gamma}_1$ down, so Roll finds a spread even when $c = 0$ (B2)

$$\mathbb{E}[\hat{\gamma}_1] \approx \gamma_1 - \frac{\gamma_0 + 2\gamma_1}{T}$$

Spreads from High and Low Prices: Corwin–Schultz

- Corwin–Schultz Corwin & Schultz (2012): the high of a bar is usually a buy at the ask, the low a sell at the bid, so the log range = volatility + spread
  - H_t, L_t: high and low of bar (or day) t; $H_{t,t+1}$, $L_{t,t+1}$: high and low over both bars
  - the volatility part grows with the length of the interval, the spread part does not
- β: sum of the two squared one-bar ranges; γ: squared two-bar range

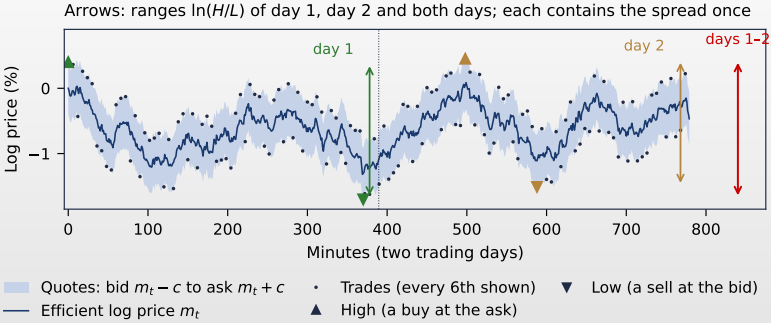
$$\beta = \sum_{j=0}^{1} \left[\ln \frac{H_{t+j}}{L_{t+j}} \right]^2, \qquad \gamma = \left[\ln \frac{H_{t,t+1}}{L_{t,t+1}} \right]^2$$

- α: the part of the log range left for the spread; $\hat{S}$: the relative spread ($\hat{S} \approx \alpha$ for small α)

$$\alpha = \frac{\sqrt{2\beta} - \sqrt{\beta}}{3 - 2\sqrt{2}} - \sqrt{\frac{\gamma}{3 - 2\sqrt{2}}}, \qquad \hat{S} = \frac{2(e^{\alpha} - 1)}{1 + e^{\alpha}}$$

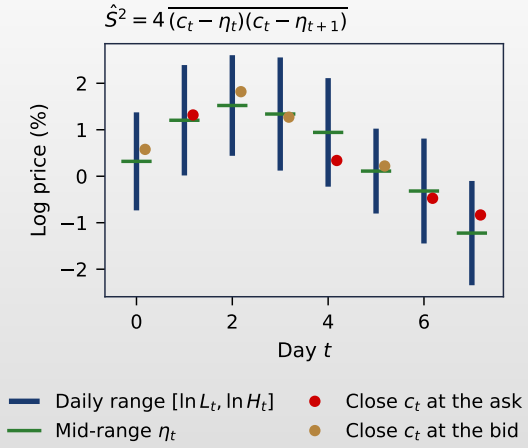
  - a negative two-bar estimate is set to 0, then the estimates are averaged over all pairs
  - example: one-day ranges 2%, two-day range 2.5%: $\beta = 0.0008$, $\gamma = 0.000625$, $\alpha = 0.0079$, $\hat{S} \approx 79$ bp; with no spread the two-day range would be $\sqrt{2} \times 2\% = 2.83\%$

Highs, Lows and the Spread on a Simulated Path



- Two simulated trading days: efficient log price m_t, quotes $m_t \pm c$ with $c = 40$ bp, trades at the bid or the ask
- Each high (triangle up) is a buy at the ask, each low a sell at the bid: every range $\ln(H/L)$ contains the full spread once
- The two-day range is longer than each one-day range only through the volatility part: comparing γ with β isolates the spread

Spreads from Closes and Ranges: Abdi–Ranaldo



- Abdi–Ranaldo Abdi & Ranaldo (2017): the log close c_t is at the bid or the ask; the mid-range proxies the log mid-price

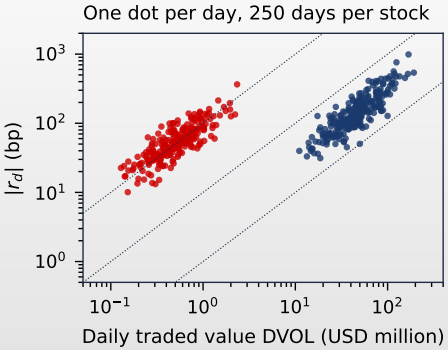
$$\eta_t = \frac{\ln H_t + \ln L_t}{2}$$

$$\hat{S}^2 = 4\,\overline{(c_t - \eta_t)(c_t - \eta_{t+1})}, \qquad \hat{S} = \sqrt{\max(\hat{S}^2, 0)}$$

  - overline: the mean over days; $c_t - \eta_t$: the close against today's mid-range, $c_t - \eta_{t+1}$: against tomorrow's
  - news moves η_{t+1} independently of the bounce in c_t: on average only the spread term $S^2/4$ remains
  - example: a mean product of $0.002 \cdot 0.001$: $\hat{S}^2 = 8 \cdot 10^{-6}$, $\hat{S} = 28$ bp

- Both range estimators mix spread and volatility: the larger the volatility relative to the spread, the less they measure the spread

Illiquidity: the Amihud Ratio



- **Amihud ratio** Amihud (2002): the absolute return per unit of traded value (DVOL, dollar volume), averaged over D days

$$\mathrm{ILLIQ} = \frac{1}{D}\sum_{d=1}^{D}\frac{|r_d|}{\mathrm{DVOL}_d}, \qquad \mathrm{DVOL}_d = P_d \times V_d$$

  - r_d: daily log return of the adjusted price; P_d: price; V_d: shares traded; DVOL in USD
  - units: bp per USD 1 million; it measures **price impact**, not the spread
  - example: $|r_d| = 100$ bp on USD 2 million gives 50 bp per USD 1 million
- Chart: the illiquid stock moves as much on USD 0.5 million as the liquid one on USD 50 million: $\mathrm{ILLIQ} = 110.3$ against 3.2
- The denominator must be a value, not a number of shares (otherwise the ratio depends on the price level)

Resampling: Day Bootstrap, Block Bootstrap, Monte Carlo

- **Day bootstrap** Efron (1979): draw D whole days with replacement, recompute the statistic $\hat{\theta}_b^*$, repeat B times; the 95% confidence interval (CI):

$$\mathrm{CI}_{95\%} = \left[q^*_{2.5\%},\ q^*_{97.5\%}\right]$$

  - q_p^*: the p quantile of $\hat{\theta}_1^*, \ldots, \hat{\theta}_B^*$; whole days keep the dependence between the bars of a day; days are treated as independent
- **Moving-block bootstrap** Künsch (1989) for daily series: resample blocks of L consecutive days (e.g. $L = 20$) and glue them; it keeps the dependence shorter than L
- **Monte Carlo reference**: simulate the model (e.g. Roll with a known c) many times, compute the statistic on each sample
  - the 2.5% and 97.5% quantiles of the simulated values: what the statistic looks like if the model is true
- **Spearman correlation**: the Pearson correlation of the ranks $R(x_i)$, $R(y_i)$ (1 = smallest)

$$r_S = \mathrm{Corr}\left(R(x_i),\ R(y_i)\right) = 1 - \frac{6\sum_i d_i^2}{n(n^2-1)}, \qquad d_i = R(x_i) - R(y_i)$$

  - the second form holds without ties; $r_S = 1$: same ordering; robust to outliers and to monotone transformations (e.g. logs)
- Annualised volatility: the s.d. of daily log returns times $\sqrt{A}$, $A = 252$ (stocks) or 365 (crypto)

Regression in Logs: Elasticity and R^2

- **Log-log regression** estimated by Ordinary Least Squares (OLS) on n days

$$\ln y_d = a + b \ln x_d + e_d, \qquad \hat{b} = \frac{\sum_d (\ln x_d - \overline{\ln x})(\ln y_d - \overline{\ln y})}{\sum_d (\ln x_d - \overline{\ln x})^2}$$

  - b: the **elasticity**: a 1% higher x goes with a b% higher y; e_d: the error
  - example: $b = 0.9$ and x up 10%: y up about 9%

- **Coefficient of determination**: the share of the variance of $\ln y$ explained by the regression

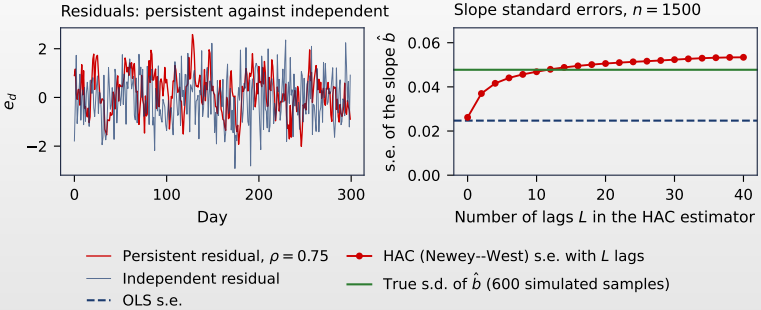
$$R^2 = 1 - \frac{\sum_d \hat{e}_d^2}{\sum_d \left(\ln y_d - \overline{\ln y}\right)^2} \in [0, 1]$$

- The usual OLS standard error assumes independent errors with constant variance

$$\mathrm{SE}_{\mathrm{OLS}}(\hat{b}) = \frac{\hat{\sigma}_e}{\sqrt{\sum_d (\ln x_d - \overline{\ln x})^2}}, \qquad \hat{\sigma}_e^2 = \frac{1}{n-2} \sum_d \hat{e}_d^2$$

  - check it with the Autocorrelation Function (ACF) of the residuals $\hat{e}_d$: a slow decay means persistent errors and a too small OLS SE

HAC Errors: Dependence Changes Uncertainty

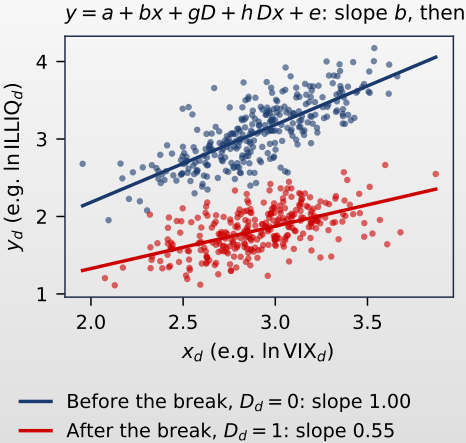


- **HAC** (heteroskedasticity and autocorrelation consistent) Newey & West (1987): with $g_d = x_d\,\hat{e}_d$ (x_d: the regressors of day d), the sandwich

$$\hat{V} = (X'X)^{-1}\,\hat{S}\,(X'X)^{-1}, \qquad \hat{S} = \sum_d g_d g_d' + \sum_{\ell=1}^{L}\left(1 - \frac{\ell}{L+1}\right)\sum_d \left(g_d g_{d-\ell}' + g_{d-\ell} g_d'\right)$$

- L: number of lags; $1 - \ell/(L+1)$: Bartlett weights; chart: persistent residuals ($\rho = 0.75$) and regressor: OLS SE 0.025, HAC ($L = 20$) 0.051, true s.d. 0.048

A Break in the Slope and the Wald Test



- **Dummy variable** $D_d = 1$ from a date fixed in advance, 0 before

$$y_d = a + b\,x_d + g\,D_d + h\,D_d x_d + e_d$$

  - slope b before the break, $b + h$ after it; g: shift of the intercept
- **Wald test** of H_0: $h = 0$ (no change in the slope)

$$W = \left(\frac{\hat{h}}{\mathrm{SE}(\hat{h})} \right)^2 \sim \chi^2_1 \text{ under } H_0$$

  - reject at 5% if $W > 3.84$; $p = P(\chi^2_1 > W)$; use the HAC SE when the residuals are persistent
  - not rejecting does not prove stability: it may only mean low power
- A linear **trend** t (years since the first day) as an extra regressor absorbs a slow drift in y

Bayes, Gaussian Projections and Poisson Likelihood

- The extension packs (after the lecture) use three tools; the models themselves (Glosten–Milgrom, Kyle, Almgren–Chriss and probability of informed trading, PIN) come in Chapter 10
- **Bayes' rule**: update the probability of a value V_H after observing a buy

$$P(V_H \mid \text{buy}) = \frac{P(\text{buy} \mid V_H)\, P(V_H)}{P(\text{buy} \mid V_H)\, P(V_H) + P(\text{buy} \mid V_L)\, P(V_L)}$$

- **Projection theorem**: for jointly Normal (v, y), the best forecast of v given y is linear

$$\mathbb{E}[v \mid y] = \mathbb{E}[v] + \frac{\operatorname{Cov}(v,y)}{\operatorname{Var}(y)}\,(y - \mathbb{E}[y]), \qquad \operatorname{Var}(v \mid y) = \operatorname{Var}(v) - \frac{\operatorname{Cov}(v,y)^2}{\operatorname{Var}(y)}$$

- **Poisson likelihood** of a count k with mean λ, and log-sum-exp for mixtures

$$\ln P(k) = k \ln \lambda - \lambda - \ln k!, \qquad \ln \sum_j e^{z_j} = z_{\max} + \ln \sum_j e^{z_j - z_{\max}}$$

  - subtracting $z_{\max}$ avoids overflow when the z_j are large (thousands of trades a day)

Exercise Map

Exercise	Question	Type	Worked example
A1	what do bid-ask bounces do to price changes?	Solved	–
A2	how noisy is one Roll estimate?	Proposed	A1
B1	is the intraday U-shape systematic?	Solved	–
B2	which frequency gives a credible spread?	Solved	–
B3	do BVB stocks look as liquid as US stocks?	Proposed	B2
B4	how strongly does illiquidity move with the VIX?	Solved	–
C1	what is wrong in an artificial intelligence (AI) answer?	Proposed	A1, B2
Extension packs	A3–A10 (information models, execution, probability of informed trading (PIN)), B5–B8, C2, C3 (optional, after the seminar)	both	named on each task

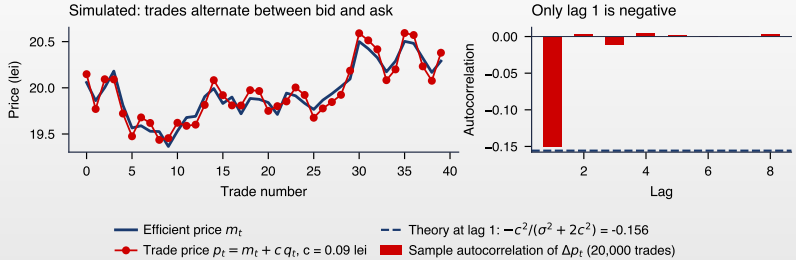
- **[Solved]**: full solution in the slides and the notebook, a model to copy; **[Proposed]**: you solve it, the solution is discussed in class
- Part A: derivations on paper; Part B: estimation and inference on real data, each ending with an interpretation question; Part C: open questions and AI (Artificial Intelligence) critique

A1 [Solved]: The Roll Model — Task

- **Questions**
  - what does the bid-ask bounce do to the moments of price changes?
  - how do we recover the spread from them?
- **Data and assumptions**
  - The Roll model of slide "Autocovariances and the Roll Model" Roll (1984): $p_t = m_t + c\, q_t$, with m_t a random walk with i.i.d. news ε_t (mean 0, variance σ^2)
  - Signs $q_t = \pm 1$, i.i.d. with probability 1/2, independent of ε
  - Numbers: a BVB stock at about 20 lei, daily $\mathrm{Var}(\Delta p) = 0.0520$, $\mathrm{Cov}(\Delta p_t, \Delta p_{t-1}) = -0.0081$ (lei^2)
- **Tasks**
  1. Write Δp_t in terms of ε_t, q_t and q_{t-1}, then derive $\mathrm{Var}(\Delta p_t)$ and $\mathrm{Cov}(\Delta p_t, \Delta p_{t-1})$.
  2. From the two numbers, compute c, s in lei and in % of the price, σ^2, the bounce share $2c^2/\mathrm{Var}(\Delta p)$ and the lag-1 autocorrelation ρ_1.
  3. Extension: let the signs be a stationary Markov chain with $\mathrm{Corr}(q_t, q_{t-k}) = \rho^k$, $\mathrm{Var}(q_t) = 1$, still independent of ε; expand the four covariance terms, and find the limit of $2\sqrt{-\gamma_1}$ for $\rho = 0.3$.
- **Report**: the two moments and a table: c, s (lei), s (%), σ^2, bounce share, ρ_1, and the extension limit
- **Interpretation**: does persistent order flow make the Roll spread too high or too low?

Back Exercise list

A1 [Solved]: Picture the Bounce Before the Algebra



- Simulated with the A1 numbers ($c = 0.09$ lei, $\sigma^2 = 0.0358$), 20 000 trades: the trade price jumps between $m_t + c$ and $m_t - c$ around the efficient price
- Sample lag-1 autocorrelation of Δp_t: -0.150, close to the theoretical value derived in the solution; no autocorrelation at longer lags; notebook: section A1

MFM_ch10_seminar

A1 [Solved]: Moments and Numbers

- Step 1. $\Delta p_t = \varepsilon_t + c(q_t - q_{t-1})$: the efficient innovation plus the change of side
- Step 2. Independence: $\mathrm{Var}(\Delta p_t) = \sigma^2 + c^2(1+1) = \sigma^2 + 2c^2$
- Step 3. $\mathrm{Cov}(\Delta p_t, \Delta p_{t-1}) = c^2\mathrm{Cov}(q_t - q_{t-1}, q_{t-1} - q_{t-2})$; the only common term is q_{t-1}, with a minus sign: $-c^2$
- Step 4. $c = \sqrt{0.0081}$; $\sigma^2 = 0.0520 - 2c^2$; $\rho_1 = -0.0081/0.0520$
- Interpretation: about a third of the daily variance is bid-ask noise, so volatility from trade prices overstates the efficient volatility

Quantity	Value
c	0.090 lei
$s = 2c$	0.18 lei
$s/20$	0.9%
$2c^2$	0.0162
σ^2	0.0358
bounce share	31.2%
ρ_1	-0.156

A1 [Solved]: Correlated Trade Signs

- Step 5. With $\mathrm{Cov}(q_t, q_{t-k}) = \rho^k$, expand the four terms of $\mathrm{Cov}(q_t - q_{t-1}, q_{t-1} - q_{t-2})$
- $\mathrm{Cov}(q_t, q_{t-1}) - \mathrm{Cov}(q_t, q_{t-2}) - \mathrm{Var}(q_{t-1}) + \mathrm{Cov}(q_{t-1}, q_{t-2}) = \rho - \rho^2 - 1 + \rho = -(1-\rho)^2$
- Step 6. $\mathrm{Cov}(\Delta p_t, \Delta p_{t-1}) = -c^2(1-\rho)^2 = -0.0081 \cdot 0.49 = -0.0040$
- Step 7. The half-spread estimator converges to $c(1-\rho) = 0.063$ lei instead of 0.090; the spread estimator to 0.126 instead of 0.18 lei: both 30% too low
- Interpretation: persistent order flow (orders split into several trades) makes reversals rarer, hides part of the bounce and biases the Roll spread **downwards**

A2 [Proposed]: Precision of the Roll Estimator — Task

- **Worked example to follow**: A1
- **Question**: how much does the sample size T matter for one Roll estimate, holding the population moments fixed?
- **Data and assumptions**
  - Population moments of A1: $\gamma_0 = 0.0520$, $\gamma_1 = -0.0081$, so $c = 0.090$, $s = 0.18$ lei; the Roll estimator $\hat{s}$ (slide "Autocovariances and the Roll Model") Roll (1984); its precision: Harris (1990)
  - T = number of price changes from $T + 1$ prices; the binary-sign variance of $\hat{\gamma}_1$ and the delta method: slide "The Precision of the Roll Estimator"; Φ: distribution function of the standard Normal distribution
- **Tasks**
  1. Derive $g'(\gamma_1)$ for $g(\gamma_1) = 2\sqrt{-\gamma_1}$ and write $\mathrm{SE}(\hat{s})$ by the delta method.
  2. Fill the table below for $T = 78$ and $T = 250$ price changes.
  3. Approximate $P(\hat{\gamma}_1 > 0)$ with the Normal distribution centred at the population covariance, using the standard error from step 2; explain the direction of the demeaning effect.
- **Report**: the formula for $\mathrm{SE}(\hat{s})$, the completed table and one sentence on the direction of the centring effect
- **Interpretation**: how often would a true Roll market show a positive covariance in one sample of 78 changes?

T	$\mathrm{SE}(\hat{\gamma}_1)$	$\mathrm{SE}(\hat{s})$ (lei)	$\mathrm{SE}(\hat{s})/s$	$P(\hat{\gamma}_1 > 0)$
78	...	...	...	...
250	...	...	...	...

Back Exercise list

B1 [Solved]: The Intraday U-Shape of SPY — Task

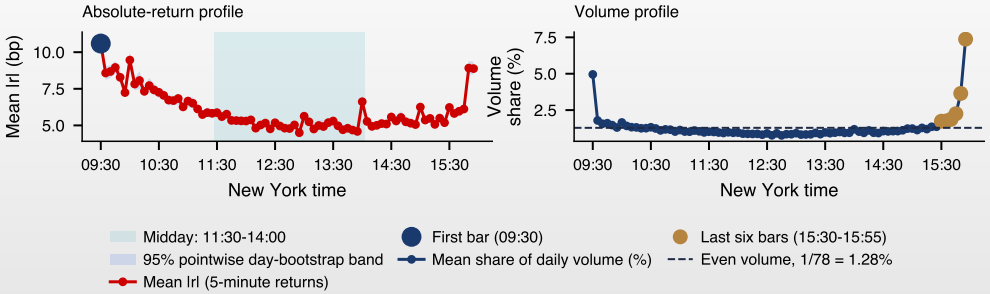
- **Questions**
  - are volatility and volume systematically higher at the open and the close?
  - is the pattern driven by a few extreme days?
- **Data and assumptions**
  - SPY 5-minute bars, 1 477 full days (slide "Data Used"); $|r|$ in bp; bars are labelled by their start time
  - One row per day: $|r|$ of the first bar (09:30), the mean $|r|$ of the 31 midday bars starting 11:30 to 14:00 inclusive, and the share of the day's volume in the last six bars
- **Tasks**
  1. Compute the ratio of the cross-day means, $\overline{|r|}_{\mathbf{open}}/\overline{|r|}_{\mathbf{mid}}$, and the mean last-30-minute volume share.
  2. Build 95% day-bootstrap intervals for both statistics (1 000 resamples of complete rows, seed 42).
  3. Plot mean $|r|$ and the mean volume share by bar with pointwise day-bootstrap bands.
- **Report**: the two estimates with intervals and the two profiles
- **Interpretation**
  - why resample days rather than bars?
  - what does the pattern imply for a trader who must buy 1 million shares today?

Back Exercise list

B1 [Solved]: Resampling Days, Step by Step

- Step 1. Point estimates: $\overline{|r|}_{\text{open}} = 10.6$ bp, $\overline{|r|}_{\text{mid}} = 5.1$ bp, ratio 2.06; last 30 minutes 18.6% of daily volume against $6/78 = 7.7\%$ if even
- Step 2. One resample: draw 1 477 day-rows with replacement (here 933 distinct days appear), keep each row whole, recompute
  - First resample: 10.7 bp and 5.0 bp, ratio 2.13
- Step 3. Repeat 1 000 times; the 2.5% and 97.5% percentiles: ratio $[1.98, 2.15]$, volume share $[18.3, 19.0]\%$
- Why whole days: the bars of one day share its volatility and news, so resampling days keeps that dependence; days are still treated as independent, although volatility persists

B1 [Solved]: Volatility and Volume Profiles



- Both intervals are narrow and exclude the even benchmarks (ratio 1, 7.7%): a systematic pattern, not a few extreme days; the bands are pointwise, not a simultaneous band for the whole curve
- Interpretation: a large buyer finds most volume at the open and the close; the open is also the most volatile, so the best time also depends on spreads, impact, urgency and the risk of waiting

MFM_ch10_seminar

B2 [Solved]: Spread Estimators at Two Frequencies — Task (1/2)

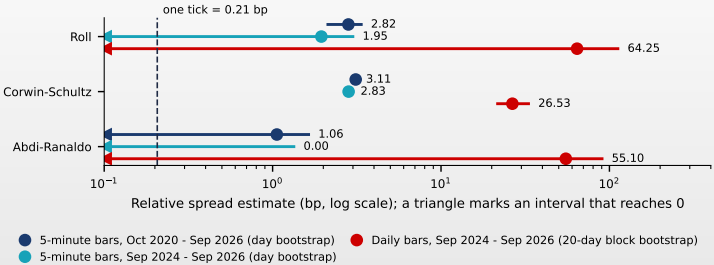
- **Question**: which sampling frequency gives a credible bid-ask spread for SPY, the most traded ETF?
- **Data and assumptions**
  - Intraday: the 1 477 days of SPY 5-minute bars; per day 78 closes give 77 log returns
  - Daily: SPY daily bars, 19 Sep 2024 – 18 Sep 2026 (501 days); one tick = USD 0.01. Daily Roll uses one demeaned lag-1 autocovariance across the whole window
- Recipes (log prices; all spreads relative)
  - Roll: for intraday bars, per day the demeaned lag-1 autocovariance of the 77 returns, averaged over days, then $2\sqrt{-\tilde{\gamma}_1}$, set to 0 if $\tilde{\gamma}_1 \geq 0$ Roll (1984)
  - Corwin–Schultz: the two-bar estimate for each pair of consecutive bars (negatives set to 0), averaged; daily bars add the overnight adjustment Corwin & Schultz (2012)
  - Abdi–Ranaldo: the mean of $4(c_t - \eta_t)(c_t - \eta_{t+1})$, then the square root, 0 if negative Abdi & Ranaldo (2017)

Back Exercise list

B2 [Solved]: Spread Estimators at Two Frequencies — Task (2/2)

- **Tasks**
  1. Compute the three estimators from the 5-minute bars with 95% day-bootstrap intervals (resample precomputed daily moment rows, not prices).
  2. Repeat on the same two-year window as the daily bars, so that only the frequency differs.
  3. Compute the three estimators from daily bars with a moving-block bootstrap (blocks of 20 days, 1 000 resamples), and report how often each is truncated at 0.
  4. Compare every estimate with one tick relative to the price, a floor for the quoted (not the effective) spread.
  5. Simulate the Roll model with SPY's bar variance, 78 prices a day, 1 477 days per sample, 200 samples, for a one-tick spread (c = half a tick) and $c = 0$; compute the share of days with $\hat{\gamma}_1 > 0$ and the pooled estimate Harris (1990).
  6. Correct the pooled 5-minute Roll estimate for the demeaning bias of slide "The Precision of the Roll Estimator", with $T = 77$.
- **Report**: an interval plot of the estimates, the truncation shares, and the Monte Carlo results next to the observed ones
- **Interpretation**
  - which frequency gives a credible spread?
  - what do the daily estimators measure?
  - is the share of positive covariances evidence against Roll?

B2 [Solved]: Estimates and Intervals



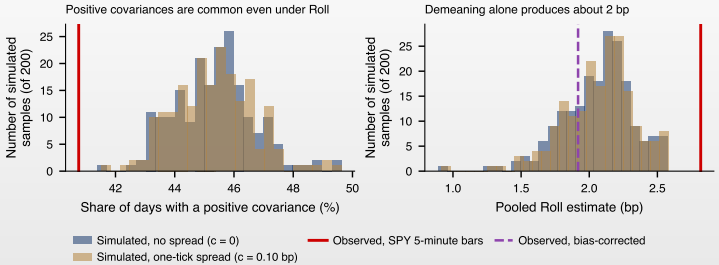
- 5-minute (2020–2026): Roll 2.82 [2.09, 3.43], Corwin–Schultz 3.11 [3.03, 3.20], Abdi–Ranaldo 1.06 [0.00, 1.67] bp; same two years (496 days): 1.95, 2.83, 0.00 bp
- Daily: Roll 64.3 [0.0, 114.4], Corwin–Schultz 26.5 [21.3, 33.7], Abdi–Ranaldo 55.1 [0.0, 92.2] bp; truncated at 0 in 9.7% (Roll) and 3.4% (Abdi–Ranaldo) of resamples; 5-minute Abdi–Ranaldo in 11.2%
- One tick is 0.21 bp: daily estimates are two orders of magnitude above it, because the daily SD (Standard Deviation) is 103.4 bp; even 5-minute Corwin–Schultz is 15 ticks

MFM_ch10_seminar

A One-Tick Spread and the Sign of the Autocovariance

- **What do you think**: if the SPY spread is exactly one tick, on what share of days is the lag-1 autocovariance of 5-minute returns positive, although the Roll model predicts $\gamma_1 = -c^2 < 0$?
- **Answer**: almost half: 45.3% of days in the simulation, the same as without a spread ($c = 0$)
- The signal is small: $c^2 \approx 0.01$ bp^2, against a bar variance of $9.2^2 \approx 85$ bp^2
- The standard deviation of $\hat{\gamma}_1$ on a day with 77 returns is about $85/\sqrt{77} \approx 9.6$ bp^2: the daily sign is close to random
- Demeaning shifts $\hat{\gamma}_1$ by about $-\gamma_0/77 \approx -1.1$ bp^2, so the share stays below 50%: $\Phi(-1.1/9.6) \approx 0.45$ (Φ: the distribution function of the standard Normal distribution)

B2 [Solved]: Monte Carlo under the Roll Model



- One tick: $c = 0.10$ bp against a bar SD of 9.2 bp; share of positive days 45.3% [43.0, 48.1], with $c = 0$ 45.3%: indistinguishable; observed 40.8%, below both simulated reference intervals
- Pooled Roll: 2.07 bp (one tick), 2.06 bp ($c = 0$), observed 2.82 bp; bias correction (the bias formula solved for γ_1, with $T = 77$ and the day means $\bar{\gamma}_0$, $\bar{\gamma}_1$) $\hat{\gamma}_1^c = (\bar{\gamma}_1 + \bar{\gamma}_0/77)/(1 - 2/77)$ gives 1.92 bp
- Interpretation: only intraday bars give credible magnitudes; positive covariances are expected under Roll, so they do not reject it; after the approximate demeaning correction, negative autocovariance remains that the one-tick Roll benchmark does not explain; the correction does not separate the bounce, heteroskedasticity, sign dependence or discreteness. These are simulated reference intervals, not bootstrap CIs

MFM_ch10_seminar

B3 [Proposed]: Spread Estimates Across 25 Assets — Task

- **Worked example to follow**: B2
- **Questions**
  - do daily spread proxies rank BVB stocks as liquid as US stocks?
  - does the Amihud ratio agree?
- **Data and assumptions**
  - The 25 assets listed on the Data Used slide: 7 US, 10 BVB stocks plus the BET ETF (11), 7 crypto; daily bars, 19 Sep 2024 – 18 Sep 2026
  - Estimators as in B2 (daily recipes); volatility annualised with 252 days (stocks) or 365 (crypto); Amihud in bp per USD 1 million, crypto volume already in USD, BVB converted at the BNR rate
- **Tasks**
  1. Compute Corwin–Schultz and Abdi–Ranaldo for each asset, and the median of each estimator in each group.
  2. Count, per group, the assets for which Abdi–Ranaldo is set to 0.
  3. Compute the Spearman correlations of Corwin–Schultz with volatility, of Amihud with the median traded value, and of Corwin–Schultz with Amihud, and draw the three scatter plots.
- **Report**: a table with rows US, BVB, crypto and columns Corwin–Schultz median, Abdi–Ranaldo median, Abdi–Ranaldo zeros; the three correlations and plots
- **Interpretation**
  - which measure would you trust to compare markets?
  - what would you need to validate it?

Back Exercise list

B4 [Solved]: Illiquidity and the VIX — Task (1/2)

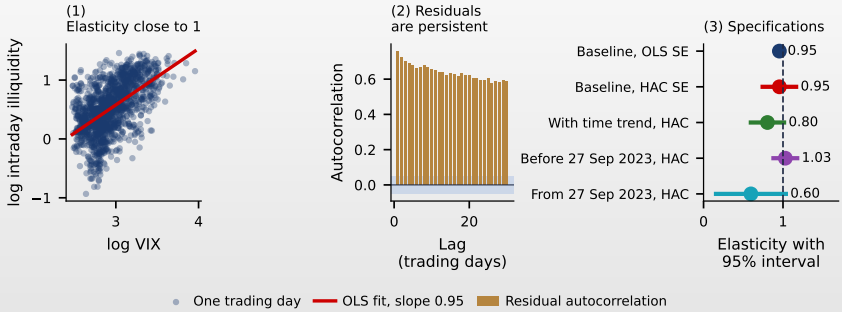
- **Questions**
  - by how much does SPY's intraday illiquidity rise when implied volatility rises?
  - how precise is that elasticity?
- **Data and assumptions**
  - $\mathrm{ILLIQ}_d = \frac{1}{K_d} \sum_k 10^4 |r_{d,k}| / (\mathrm{DVOL}_{d,k} / 10^8)$: bp per USD 100 million, from the 5-minute bars of day d, $k = 1, \ldots, K_d$ ($r_{d,k}$: log return, $\mathrm{DVOL}_{d,k}$ = close $\times$ volume; 10^4 converts to bp, 10^8 to USD 100 million); K_d: the number of bars with nonzero value, usually 78
  - Aligned on trading days with the close of the VIX index: 1 477 days, 2020-10-12 – 2026-09-18; mean ILLIQ 1.83

Back Exercise list

B4 [Solved]: Illiquidity and the VIX — Task (2/2)

- **Tasks**
  1. Estimate $\ln \mathrm{ILLIQ}_d = a + b \ln \mathrm{VIX}_d + e_d$ by OLS (the log-log regression of slide "Regression in Logs: Elasticity and R^2") and plot the fit.
  2. Plot the residual autocorrelations; report OLS and HAC standard errors (slide "HAC Errors: Dependence Changes Uncertainty", 20 lags) and the 95% HAC interval Newey & West (1987).
  3. Extension: add a linear trend t (in years from the first day) and re-estimate b.
  4. Extension: with $D_d = 1$ from 2023-09-27 (the middle of the sample, fixed before estimation), add $gD_d + hD_d \ln \mathrm{VIX}_d$ and test $h = 0$ with the HAC Wald test of slide "A Break in the Slope and the Wald Test" at 5%.
- **Report**: a table of specifications ($\hat{b}$, SE, interval) and the coefficient plot
- **Interpretation**
  - why do the two standard errors differ?
  - is the elasticity stable?
  - is the relation causal?

B4 [Solved]: Fit and Residuals



- (1) $\hat{b} = 0.95$, $R^2 = 0.30$: a 10% higher VIX goes with about $0.95 \times 10\%$ higher illiquidity; (2) residual lag-1 autocorrelation is 0.76 and decays slowly: OLS SE 0.038, HAC SE 0.121 (3.2 times larger), HAC CI $[0.72, 1.19]$

MFM_ch10_seminar

B4 [Solved]: Specifications and the Break Test

Specification	$\hat{b}$	SE	95% interval
Baseline, OLS SE	0.95	0.038	[0.88, 1.03]
Baseline, HAC SE	0.95	0.121	[0.72, 1.19]
With trend (HAC)	0.80	0.120	[0.57, 1.04]
Before 2023-09-27 (HAC)	1.03	0.090	[0.85, 1.21]
From 2023-09-27 (HAC)	0.60	0.238	[0.13, 1.06]

- Trend -0.062 a year (SE 0.021): illiquidity falls over time, the VIX link remains; break: $\hat{h} = -0.43$, Wald $(\hat{h}/\mathrm{SE})^2 = (-0.433/0.254)^2 = 2.90$, $p = 0.089$: not rejected at 5%
- Interpretation: persistent residuals make OLS errors far too small; the elasticity is close to 1; non-rejection at 5% does not establish stability; not causal, since both respond to news

C1 [Proposed]: Critique an AI Answer — the Answer

- **Prompt**: "I have daily closes `P` (lei) and volumes `V` (number of shares, positive every day) for a BVB stock. Estimate the Roll spread and the Amihud illiquidity. What does a round trip cost me?"
- **Answer, text**: Roll: $\mathrm{Cov}(\Delta p_t, \Delta p_{t-1}) = -c^2$ and $s = 2c$; if the covariance is positive, take its absolute value, the formula still gives the spread; Amihud: the average of $|r_d|$ divided by the volume of day d, per 1 million; result: $s = 0.05$ lei on a 20-lei stock, 0.25% of the price, so a round trip costs $2s$, 0.50%
- **Answer, code**

```
import numpy as np
dp = P.diff().dropna().values
g = np.cov(dp[1:], dp[:-1])[0, 1]
s_roll = 2 * np.sqrt(abs(g))
s_pct = 100 * s_roll / P.mean()
r = np.log(P).diff().dropna()
illiq = 1e6 * (r.abs() / V.loc[r.index]).mean()
```

C1 [Proposed]: Critique an AI Answer — Task

- **Worked example to follow**: A1, B2
- **Questions**
  - which parts of the AI answer are wrong?
  - how would you detect each error before using the numbers?
- **Data and assumptions**
  - A conceptual audit: no data file is needed; `P` and `V` are hypothetical
  - Check every formula and every unit in the answer against the notation introduced today
- **Tasks**
  1. The answer contains three errors; identify each one in the text and, where applicable, in the code.
  2. For each error, state how you would detect it: a test, a published number or a formula from today's primer.
  3. Correct each error: give the formula, the code line and the corrected number where the information permits one; state the units of your Amihud ratio.
  4. Log the errors you actually found in `AI_ERRORS.md`: request, answer, error, how found, correction.
- **Report**: the three errors with detection and correction
- **Interpretation**: which error would change an investment decision?

Team Project: Choose One

- C2: is illiquidity priced on the BVB? C3: the implied price of the speed race for SPY; or: Roll, Corwin–Schultz and Abdi–Ranaldo for three assets not covered here (documenting the data you use)
- Deliverables: a reproducible Quantlet and a report of at most two pages
- Assessment criteria: question, data and conventions, method, uncertainty (intervals, robustness), limitations
- Declare AI use in `AI_USE.md`; log every AI error you actually found in `AI_ERRORS.md`

Extension Packs: What Each One Needs

Pack	Exercises	Prerequisites
1. Information models	A3 [Solved], A4, A5 [Solved], A6	lecture section *Information and Prices*, Bayes' rule, projections
2. Optimal execution	A7 [Solved], A8	lecture section *Price Impact and Optimal Execution*
3. Informed trading (PIN)	A9 [Solved], A10	Poisson likelihood, log-sum-exp
4. Illiquidity across markets	B5	B2, B3
5. Volume and price impact	B6 [Solved]	B1, instrumental-variable (IV) estimation
6. The Bitcoin clock	B7	B1
7. Price discovery	B8 [Solved]	cointegration, Vector Error Correction Model (VECM)
8. Projects	C2, C3	B4, B5; the lecture case study

- Each pack has its own notebook section; do them in order within a pack; C2 and C3 are possible team projects

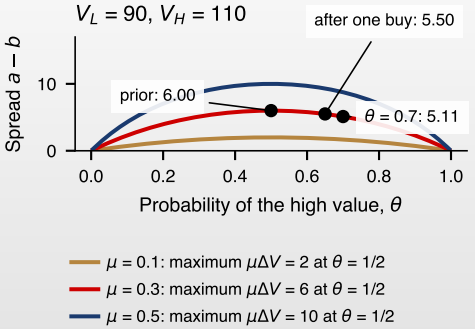
A3 [Solved]: Glosten–Milgrom for Any Prior — Task

- **Questions**
  - how wide is the spread when the market maker learns from order flow?
  - when is it widest?
- **Data and assumptions**
  - $V \in \{V_L, V_H\}$, $\Delta V = V_H - V_L$; $\mathcal{F}_t$ = the trades observed so far, $\theta = P(V = V_H \mid \mathcal{F}_t)$ Glosten & Milgrom (1985)
  - Likelihoods: $P(\text{buy} \mid V_H) = (1+\mu)/2$, $P(\text{buy} \mid V_L) = (1-\mu)/2$ (informed share μ, uninformed buy or sell with probability 1/2)
  - Pre-trade quotes for the next trade: $a = \mathbb{E}[V \mid \mathcal{F}_t, \text{buy}]$, $b = \mathbb{E}[V \mid \mathcal{F}_t, \text{sell}]$; $V_L = 90$, $V_H = 110$, $\mu = 0.3$
- **Tasks**
  1. At $\theta = 0.5$, compute $P(V_H \mid \text{buy})$, the ask, the bid and the spread.
  2. After one buy, the posterior becomes the new θ: compute the next-trade quotes and spread.
  3. Derive the spread for any $0 < \theta < 1$ and show that it is largest at $\theta = 1/2$; check it at $\theta = 0.7$.
  4. Extension: show that $\mathbb{E}[\text{next trade price} \mid \mathcal{F}_t] = \mathbb{E}[V \mid \mathcal{F}_t]$ and that beliefs are a martingale.
- **Report**: the quotes, the spread formula and the martingale argument
- **Interpretation**: why is the spread largest when the market maker is most uncertain?

Back Exercise list

A3 [Solved]: Quotes and the General Spread

- Step 1 ($\theta = 0.5$). $P(V_H \mid \text{buy}) = 0.65 \cdot 0.5/(0.65 \cdot 0.5 + 0.35 \cdot 0.5) = 0.6500$: $a = 90 + 20 \cdot 0.6500 = 103.00$, $b = 97.00$, spread $6.00 = \mu \Delta V$
- Step 2. After a buy $\theta = 0.6500$: $a = 105.50$, $b = 100.00$, spread 5.50
- Step 3. With $x = 2\theta - 1$: $P(V_H \mid \text{buy}) = \theta(1+\mu)/(1+\mu x)$, $P(V_H \mid \text{sell}) = \theta(1-\mu)/(1-\mu x)$
- Step 4. $a - b = \mu \Delta V (1 - x^2)/(1 - \mu^2 x^2) \le \mu \Delta V$, equality only at $x = 0$; at $\theta = 0.7$: 5.11
- Step 5. $P(\text{buy})a + P(\text{sell})b = \mathbb{E}[V \mid \mathcal{F}_t]$ (total expectation); iterated expectations give the martingale
- Interpretation: at $\theta = 1/2$ one trade moves beliefs most, so adverse selection costs most



MFM_ch10_seminar

A4 [Proposed]: How Fast Does the Market Maker Learn? — Task

- **Worked example to follow**: A3
- **Questions**
  - how many consecutive buys does it take before the bid exceeds 108?
  - how does this number depend on μ?
- **Data and assumptions**
  - A3 with $\theta_{\mathbf{0}} = 1/2$, $V_L = 90$, $V_H = 110$; true value V_H; a **conditioned scenario** in which every trade is a buy (the expected learning time is not asked for)
  - Timeline: after k buys the belief is θ_k; the bid quoted then conditions on a hypothetical next sell
- **Tasks**
  1. Show that each buy multiplies the odds $\theta/(1-\theta)$ by $(1+\mu)/(1-\mu)$ and each sell divides them by the same factor.
  2. Show that the bid after k buys is $V_L + \Delta V\, \theta_{k-\mathbf{1}}$.
  3. Show that bid > 108 requires odds above 9, derive $k_{\min} = \lfloor \ln 9 / \ln \frac{\mathbf{1}+\mu}{\mathbf{1}-\mu} \rfloor + 2$, ($\lfloor \cdot \rfloor$: integer part), and evaluate it for $\mu = 0.1, 0.3, 0.5$, checking the boundary case.
  4. Compare the conditional scenario with simulated paths in which uninformed traders also sell.
- **Report**: the three values of $k_{\min}$ and a plot of the bid against k
- **Interpretation**: how does the speed of price discovery depend on the share of informed traders?

Back Exercise list

A5 [Solved]: Deriving the Kyle Equilibrium — Task

- **Questions**
  - how aggressively does an insider trade?
  - how much does the market maker move the price per share?
- **Data and assumptions**
  - Timeline: (1) the insider sees $v \sim N(p_0, \sigma_v^2)$ and sends x; (2) noise traders send $u \sim N(0, \sigma_u^2)$, independent; (3) the market maker sees only $y = x + u$ and sets $p = \mathbb{E}[v \mid y]$ Kyle (1985)
  - Units: v, p in USD; x, u, y in shares; β in shares per USD; λ in USD per share. Numbers: $\sigma_v = 2$, $\sigma_u = 10\,000$ and $20\,000$
- **Tasks**
  1. Insider best response: given $p = p_0 + \lambda y$, maximise the expected profit and write β in $x = \beta(v - p_0)$.
  2. Market-maker pricing: given $x = \beta(v - p_0)$, derive λ from the projection theorem.
  3. Solve the two equations; compute the expected profit, $\mathrm{Var}(x)$ and $\mathrm{Var}(v \mid y)$.
  4. Fill a two-column table ($\sigma_u = 10\,000$ and $20\,000$): β, λ, profit, price change for $y = 15\,000$.
- **Report**: the equilibrium formulas and the table
- **Interpretation**: why does more noise trading help the insider?

Back Exercise list

A5 [Solved]: Equilibrium and Numbers

- Step 1. $\mathbb{E}[(v - p_0 - \lambda x - \lambda u)x \mid v] = (v - p_0)x - \lambda x^2$; first-order condition: $x = (v - p_0)/(2\lambda)$, $\beta = 1/(2\lambda)$
- Step 2. (v, y) jointly Normal: $\lambda = \mathrm{Cov}(v, y)/\mathrm{Var}(y) = \beta\sigma_v^2/(\beta^2\sigma_v^2 + \sigma_u^2)$
- Step 3. With $2\lambda\beta = 1$: $\beta^2\sigma_v^2 = \sigma_u^2$, so $\beta = \sigma_u/\sigma_v$, $\lambda = \sigma_v/(2\sigma_u)$
- Step 4. Profit $\beta\sigma_v^2 - \lambda\beta^2\sigma_v^2 = \sigma_v\sigma_u/2$; $\mathrm{Var}(x) = \sigma_u^2$; $\mathrm{Var}(v \mid y) = \sigma_v^2/2$ in both economies
- Interpretation: λ measures informativeness per share; more noise makes the flow less informative, the insider trades more and earns more, and half the information still enters the price

	$\sigma_u = 10^4$	$\sigma_u = 2 \cdot 10^4$
β	5 000	10 000
λ	0.0001	0.00005
profit (USD)	10 000	20 000
Δp ($y = 15\,000$)	1.50	0.75

A6 [Proposed]: Kyle's Lambda as a Regression Slope — Task

- **Worked example to follow**: A5
- **Questions**
  - what does an Amihud-type ratio estimate in a Kyle economy?
  - does the choice of volume denominator matter?
- **Data and assumptions**
  - A5 with $p_0 = 50$ USD, $\sigma_v = 2$, $\sigma_u = 10\,000$; simulation (seed 42): 200 000 independent auctions, $v = p_0 + \sigma_v z_1$, $u = \sigma_u z_2$, $x = \beta(v - p_0)$, $y = x + u$, $p = p_0 + \lambda y$, with z_1, z_2 independent standard Normal draws
  - $|x| + |u|$ is a constructed gross-order proxy (the two orders counted separately), not exchange volume
- **Tasks**
  1. Show that the population OLS slopes of $p - p_0$ and of $v - p_0$ on y both equal λ, and give $\mathrm{Var}(y)$.
  2. Show that $|r|/(p_0|y|)$, with $r = (p - p_0)/p_0$, equals λ/p_0^2, and convert it to bp per USD 1 million (factor $10^4 \times 10^6$).
  3. Replace $|y|$ by $|x| + |u|$, derive the direction of the bias and measure it by simulation; repeat with $\sigma_u = 20\,000$.
- **Report**: the slopes, the ratio for both denominators and both noise levels
- **Interpretation**
  - what does the Amihud ratio measure in a Kyle world?
  - why is it biased with gross volume?

Back Exercise list

A7 [Solved]: Almgren–Chriss Execution — Task

- **Questions**
  - how should a risk-averse seller split 1 million shares over five days?
  - what does the risk reduction cost?
- **Data and assumptions**
  - Parameters Almgren & Chriss (2001): $X = 10^6$ shares, $T = 5$ days, $N = 5$, $\tau = 1$; $\sigma = 0.95$ USD per day$^{1/2}$; temporary impact $\eta = 2.5 \cdot 10^{-6}$, permanent $\gamma = 2.5 \cdot 10^{-7}$ (USD per share per share); risk aversion $\lambda = 2 \cdot 10^{-6}$ per USD
  - x_k = holdings after trade k ($x_0 = X$, $x_5 = 0$), $n_k = x_{k-1} - x_k$; $\mathbb{E}[C] = \frac{1}{2}\gamma X^2 + \frac{\tilde{\eta}}{\tau}\sum_{k=1}^{5} n_k^2$, $\mathrm{Var}[C] = \sigma^2 \tau \sum_{k=1}^{5} x_k^2$, $\tilde{\eta} = \eta - \gamma\tau/2$
- **Tasks**
  1. Compute $\tilde{\eta}$, $\tilde{\kappa}^2 = \lambda\sigma^2/\tilde{\eta}$ and κ from $\cosh(\kappa\tau) = 1 + \tilde{\kappa}^2\tau^2/2$.
  2. Fill a table of k, $x_k = X\sinh(\kappa(5-k))/\sinh(5\kappa)$ and n_k.
  3. Compute $\mathbb{E}[C]$ and $\mathrm{sd}[C]$ and compare with selling evenly ($\lambda = 0$).
  4. Extension: minimise $\int_0^T (\eta\dot{x}^2 + \lambda\sigma^2 x^2)\,dt$ with $x(0) = X$, $x(T) = 0$ by Euler–Lagrange and compare κ_c with κ.
- **Report**: the schedule table, the cost–risk comparison and the holdings plot
- **Interpretation**: what does the trader buy with the extra expected cost?

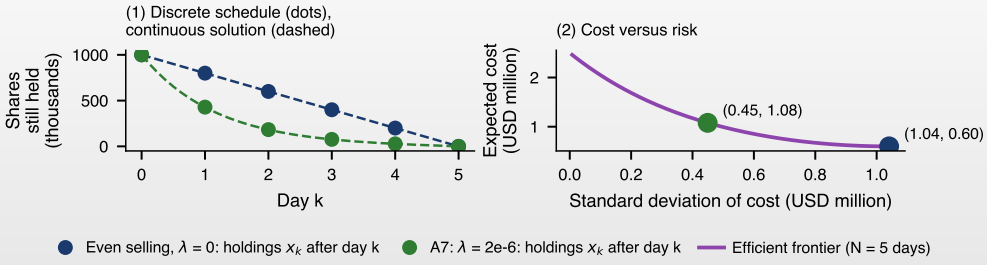
Back Exercise list

A7 [Solved]: Schedule, Cost and Risk

- Step 1. $\tilde{\eta} = 2.5 \cdot 10^{-6} - 1.25 \cdot 10^{-7} = 2.375 \cdot 10^{-6}$; $\tilde{\kappa}^2 = 0.76$; $\cosh \kappa = 1.38$, $\kappa = 0.8463$, $\sinh(5\kappa) = 34.40$
- Step 2. The table: 57% is sold on day 1
- Step 3. $\mathbb{E}[C] = 1.078$, $\mathrm{sd}[C] = 0.449$ million USD; evenly: 0.600 and 1.041
- Interpretation: front-loading pays 0.48 million USD more expected impact to cut the SD of cost by more than half

k	x_k	n_k
0	1 000 000	–
1	428 599	571 401
2	182 933	245 666
3	76 296	106 637
4	27 643	48 652
5	0	27 643

A7 [Solved]: Euler–Lagrange and the Frontier



- $F = \eta \dot{x}^2 + \lambda \sigma^2 x^2$: $\frac{d}{dt}\partial F/\partial \dot{x} = \partial F/\partial x \Rightarrow \ddot{x} = \kappa_c^2 x$, $\kappa_c^2 = \lambda\sigma^2/\eta = 0.722$; with $x(0) = X$, $x(T) = 0$: $x(t) = X \sinh(\kappa_c(T-t))/\sinh(\kappa_c T)$, $x(1) = 427\,151$
- $\kappa_c = 0.8497$ against $\kappa = 0.8463$: close here because discretisation and $\tilde{\eta}$ partly offset ($\kappa\tau \approx 0.85$ is not small); dots: the discrete schedule, dashed: the continuous one

MFM_ch10_seminar

A8 [Proposed]: A More Risk-Averse Seller — Task

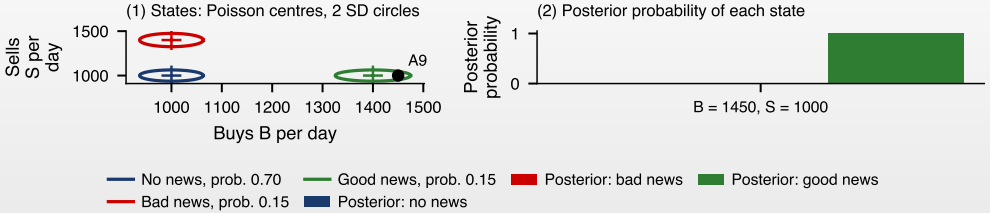
- **Worked example to follow**: A7
- **Question**: how does the schedule change when risk aversion is ten times higher?
- **Data and assumptions**
  - A7 parameters: $X = 10^6$, $T = N = 5$, $\sigma = 0.95$, $\eta = 2.5 \cdot 10^{-6}$, $\gamma = 2.5 \cdot 10^{-7}$; only λ changes, to $2 \cdot 10^{-5}$
- **Tasks**
  1. Compute $\tilde{\kappa}^2$, κ, κ_c and the holdings x_k.
  2. Compute $\mathbb{E}[C]$ and $\mathrm{sd}[C]$ and the ratios new/A7, with their direction.
  3. Overlay the holdings of A7 and A8 on one chart.
  4. Optional: explain why κ and κ_c now differ much more.
- **Report**: one comparison table and the chart
- **Interpretation**: when would a trader choose this schedule?

A9 [Solved]: The PIN Likelihood, Step by Step — Task

- **Questions**
  - how do we evaluate the likelihood of one day of buys and sells without numerical overflow?
  - what does it say about that day?
- **Data and assumptions**
  - Easley–Kiefer–O'Hara–Paperman (EKOP) Easley et al. (1996), three states: no news, prob. $1-\alpha$, $(B, S) \sim$ Poisson(ε_b), Poisson(ε_s); bad news, $\alpha\delta$, sells rate $\mu + \varepsilon_s$; good news, $\alpha(1-\delta)$, buys rate $\mu + \varepsilon_b$
  - $\alpha = 0.3$, $\delta = 0.5$, $\mu = 400$, $\varepsilon_b = \varepsilon_s = 1\,000$; one day with $B = 1\,450$, $S = 1\,000$
- **Tasks**
  1. Write the likelihood as a three-component mixture and compute $\mathrm{PIN} = \alpha\mu/(\alpha\mu + \varepsilon_b + \varepsilon_s)$.
  2. Explain why evaluating the terms directly fails in double precision.
  3. Write each weighted log component, combine them as $m + \ln \sum e^{\ell_j - m}$ with $m = \max_j \ell_j$, and normalise the same components into posterior probabilities.
- **Report**: PIN, $\ln L$, the three posterior probabilities
- **Interpretation**: is the posterior probability of good news the same thing as PIN?

Back Exercise list

A9 [Solved]: Stable Log Components and Posteriors



- PIN $= 120/2\,120 = 0.0566$; directly, $e^{-1000} \to 0$ and $1000^{1450} \to \infty$: $0 \cdot \infty =$ NaN Lin & Ke (2011). Factor $K = e^{-\varepsilon_b - \varepsilon_s}(\mu+\varepsilon_b)^B(\mu+\varepsilon_s)^S/(B!S!)$, $x_b = \varepsilon_b/(\mu+\varepsilon_b)$, x_s likewise; $\ln B!$ by log-gamma
- Components: $\ell_1 = \ln(1-\alpha) + B\ln x_b + S\ln x_s$, $\ell_2 = \ln \alpha\delta - \mu + B\ln x_b$, $\ell_3 = \ln \alpha(1-\delta) - \mu + S\ln x_s$; $\ln L = \ln K + m + \ln\sum e^{\ell_j - m} = -11.71$; posterior $e^{\ell_j - m}/\sum e^{\ell_i - m}$: good news ≈ 1
- Interpretation: 450 extra buys against a Poisson SD of about 32 classify the day almost surely; PIN $= 0.0566$ is the unconditional expected share of informed trades, not this posterior

MFM_ch10_seminar

A10 [Proposed]: Reading Days with the PIN Model — Task

- **Worked example to follow**: A9
- **Questions**
  - how confidently can known PIN parameters classify a day?
  - what changes when the parameters are unknown?
- **Data and assumptions**
  - The A9 parameters; two observed days (a) $B = 1\,000$, $S = 1\,420$; (b) $B = S = 1\,000$; (c) a hypothetical ambiguous day, $B = 1\,195$, $S = 1\,000$
- **Tasks**
  1. For each day, compute the stable log-likelihood and the three posterior probabilities.
  2. Compute PIN when μ doubles to 800.
  3. Conceptual: explain why estimating $(\alpha, \delta, \mu, \varepsilon_b, \varepsilon_s)$ from many days remains hard for very liquid stocks.
- **Report**: a table of days by posterior probabilities, and the new PIN
- **Interpretation**: can an order imbalance always be read as private information?

Back Exercise list

B5 [Proposed]: Amihud Illiquidity on Three Markets — Task

- **Worked example to follow**: B2, B3
- **Questions**
  - how many times less liquid is the BVB than the US by the Amihud ratio?
  - how precise is that multiple?
  - how stable is it?
- **Data and assumptions**
  - The 25 assets of B3 (BVB includes the BET ETF); fixed group membership
  - Aggregation: average the daily ratios within each asset over 19 Sep 2024 – 18 Sep 2026, then take the group median
  - Years: 19 Sep 2024 – 18 Sep 2025 and 19 Sep 2025 – 18 Sep 2026
- **Tasks**
  1. Compute the three group medians and the ratio BVB median / US median.
  2. Build a 95% interval by resampling assets with replacement independently within each group (1 000 resamples), recomputing the median ratio each time.
  3. Compute the Spearman correlation of the asset rankings in the two years.
- **Report**: the medians, the ratio with its interval, the rank correlation and log-scale plots
- **Interpretation**
  - what does the interval condition on?
  - is illiquidity a stable asset characteristic?

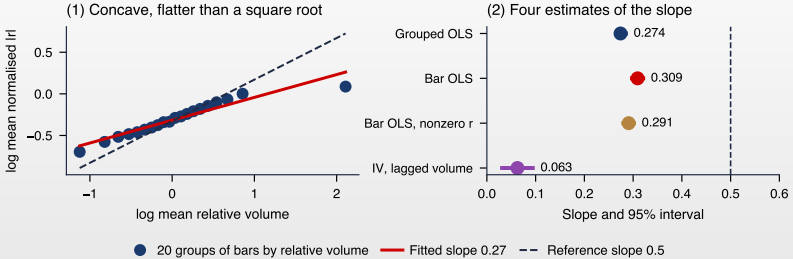
Back Exercise list

B6 [Solved]: Volume and Price Moves in SPY Bars — Task

- **Questions**
  - how do absolute returns scale with relative volume in 5-minute bars?
  - can that slope be read as a price-impact law?
- **Data and assumptions**
  - SPY 5-minute bars with positive volume: 115 052 bars on 1 477 days (115 206 bars minus 154 without volume)
  - Relative volume v = bar volume / median volume at the same time of day; $z = |r|/\sigma_{\mathbf{day}}$, $\sigma_{\mathbf{day}}$ the SD of that day's 5-minute returns
- **Tasks**
  1. Sort bars into 20 equal-size groups by v (bin edges fixed on the full sample), regress $\ln \bar{z}$ on $\ln \bar{v}$ ($\bar{z}$, $\bar{v}$: group means), and give a 95% day-bootstrap interval (300 resamples, same edges).
  2. At bar level, estimate $\ln(|r| + 10^{-\mathbf{6}}) - \ln \sigma_{\mathbf{day}} = a + b \ln v + e$ with SEs clustered by day, test $b = 0.5$, and repeat without the bars with $r = 0$.
  3. Extension: instrument $\ln v$ by the log relative volume of the same bar on the previous trading day; report the first stage and the slope estimated by two-stage least squares (2SLS), with clustered SE.
- **Report**: the grouped fit, the four slope estimates with intervals and the first stage
- **Interpretation**
  - is the relation concave?
  - is it the square-root law Tóth et al. (2011)?
  - is the exclusion restriction plausible?

Back Exercise list

B6 [Solved]: Fitted Relation and Slope Estimates



- Grouped slope 0.274 [0.262, 0.289]; bar-level 0.309 (SE 0.008), $t = (0.5 - 0.309)/0.008 \approx 24$: 0.5 rejected; 0.8% of bars have $r = 0$, without them 0.291 (SE 0.008)
- First stage slope 0.339 ($t = 32.8$): relevant; 2SLS slope 0.063 (SE 0.018); relevance does not establish exclusion: yesterday's volume at that time reflects persistent activity and scheduled news (lagged correlation of normalised $\ln|r|$ 0.043)
- Interpretation: concave but flatter than a square root; a bar mixes buyers and sellers, while the law concerns one trader's net metaorder, so slope 0.5 is not a direct test of it

MFM_ch10_seminar

B7 [Proposed]: The Bitcoin Clock — Task

- **Worked example to follow**: B1
- **Question**: does a market open 24/7 still keep the clock of traditional markets?
- **Data and assumptions**
  - Bitcoin 5-minute bars, UTC, 749 days (535 weekdays, 214 weekend days); returns spanning missing bars are discarded; coverage is incomplete: only 106 days have all 288 bars (minimum 26); all days are kept
  - Weighting: first the mean $|r|$ of each day in a window (hours $[h_1, h_2)$), then the mean over days; ratios are ratios of these day means
- **Tasks**
  1. On weekdays, compute the mean $|r|$ in 13:00–17:00 UTC and in 03:00–11:00 UTC and their ratio, with a 95% day-bootstrap interval.
  2. Compute the weekday / weekend ratio of mean $|r|$ with a 95% day-bootstrap interval (1 000 resamples, seed 42; sample whole days from the pooled sample, so weekday/weekend counts may vary).
  3. Plot the two hourly profiles with day-bootstrap bands.
- **Report**: the two ratios with intervals and the profiles
- **Interpretation**: what spread, depth and volume evidence would an execution decision need beyond volatility?

Back Exercise list

B8 [Solved]: Price Discovery: the BET ETF and BET-TR — Task (1/2)

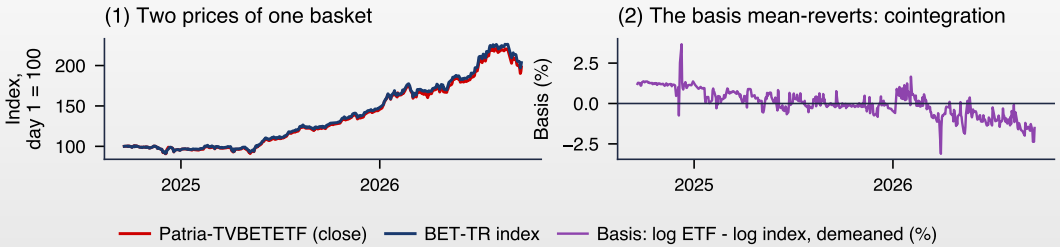
- **Questions**
  - does the BET ETF or the BET-TR index lead price discovery?
  - how sure can we be with daily closes?
- **Data and assumptions**
  - Log closes $y_t = (\ln P_t^{\mathrm{ETF}}, \ln P_t^{\mathrm{idx}})'$ of Patria-TVBETETF (days with trades) and BET-TR, common days 2024-09-19 – 2026-09-18, 493 VECM observations; the ETF reinvests dividends; BET-TR is computed, not traded
  - VECM: $\Delta y_t = a_0 + \alpha z_{t-1} + \sum_{i=1}^{p} \Gamma_i \Delta y_{t-i} + e_t$, $z_{t-1} = \ln P_{t-1}^{\mathrm{ETF}} - \ln P_{t-1}^{\mathrm{idx}}$, $\mathrm{Var}(e_t) = \Omega$; unrestricted constant; p lags of the *differences*, chosen by the Bayesian Information Criterion (BIC) over 1–10 Johansen (1991)

Back Exercise list

B8 [Solved]: Price Discovery: the BET ETF and BET-TR — Task (2/2)

- **Tasks**
  1. Plot the rebased prices and the basis; run the Johansen trace test (rank one does not by itself establish $\beta = (1, -1)'$, which is imposed).
  2. Estimate the VECM and test $\alpha_{\mathrm{ETF}} = 0$ and $\alpha_{\mathrm{idx}} = 0$ at 5%.
  3. With $\alpha_\perp = (-\alpha_{\mathrm{idx}}, \alpha_{\mathrm{ETF}})'$ compute the component share $\alpha_\perp / \sum \alpha_\perp$ Gonzalo & Granger (1995) and the information share $([\alpha_\perp' F]_j)^2 / \alpha_\perp' \Omega \alpha_\perp$ for both Cholesky orderings, F the lower-triangular factor with $FF' = \Omega$ and $[\cdot]_j$ the j-th element Hasbrouck (1995).
  4. Bootstrap: resample moving blocks of 20 regression rows $(\Delta y_t, 1, z_{t-\mathbf{1}}, \Delta y_{t-\mathbf{1}}, \ldots)$, 1 000 times; then re-estimate α on each half.
- **Report**: the test, α with asymptotic and bootstrap intervals, the ordering bounds with endpoint intervals, the two halves
- **Interpretation**
  - which price leads?
  - why are the bounds wide?
  - why can the component share leave $[0, 1]$?

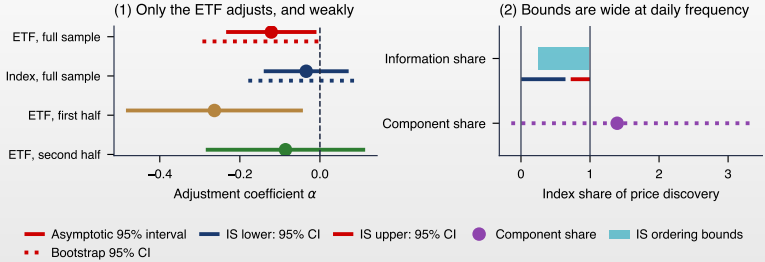
B8 [Solved]: Prices, Test and Adjustment



- BIC: 1 lag; trace 49.6 for $r = 0$ (5% critical value 15.49), 0.10 for $r \leq 1$ (3.84): one cointegrating vector
- $\alpha_{\mathrm{ETF}} = -0.12$ ($t = -2.11$, rejected at 5%), $\alpha_{\mathrm{idx}} = -0.03$ ($t = -0.63$); block bootstrap: ETF $[-0.29, 0.01]$, index $[-0.18, 0.09]$: the asymptotic test rejects, the bootstrap interval touches 0, so the evidence is weak

MFM_ch10_seminar

B8 [Solved]: Shares, Bounds and Halves



- Index information share within $[0.25, 0.98]$ (ordering bounds, not a CI); bootstrap CI of the lower bound $[0.00, 0.64]$, of the upper $[0.72, 1.00]$; component share 1.39 $[-0.14, 3.32]$
- Halves: ETF $\alpha = -0.26$ ($t = -2.34$), then -0.09 ($t = -0.85$); the index is never significant ($t = -0.77$, 0.47)
- Interpretation: the ETF appears to adjust toward the index; residual correlation 0.93 makes the ordering decisive, hence wide bounds; both α negative push the component share outside $[0, 1]$

MFM_ch10_seminar

C2: Panel, Sample and Reproducible Inputs

- Data for C2 (task follows): daily observations from 19 September 2016 to 18 September 2026; only complete months enter every stage: the illiquidity index and the autoregressive AR(1) model, October 2016–August 2026; the return regressions, November 2016–August 2026. This long window is used only here.
- US stocks: AAPL, MSFT, NVDA, JPM, GME, MSTR. Other cryptoassets: ETH, SOL, XRP, DOGE, ADA, LTC.
- The panel is unbalanced: an asset enters a monthly cross-sectional mean only when at least ten valid days are available. Missing pre-listing observations are not imputed.
- Load returns and traded value with `returns(key, start)` and `dollar_volume(key, start)`, using `start="2016-09-19"`; align dates before forming ratios.

Uncertainty After Correcting Persistence

- In C2, $\hat{\phi}_c$ denotes corrected AR(1) persistence (formula: Milestone 3 of the task); v_t^c is the residual computed with that coefficient and a re-estimated intercept.
- Let $\hat{d}$ be the coefficient of v_t^c in the augmented return regression. Let s_b be the HAC standard error of its lagged-illiquidity coefficient, conditional on the estimated correction.
- For the AR(1) fit, s_ϕ is the standard error of $\hat{\phi}$ and T counts the usable transitions. The delta-method approximation: $\widehat{\mathrm{Var}}(\hat{\phi}_c) = (1 + 3/T + 9/T^2)^2 s_\phi^2, \qquad s_{b,c} = \sqrt{s_b^2 + \hat{d}^2 \widehat{\mathrm{Var}}(\hat{\phi}_c)}.$
- The implementation caps corrected persistence at 0.9999. The displayed derivative is the uncapped interior approximation, so it is not valid when the cap binds. Amihud & Hurvich (2004)
- Joint inference: a bootstrap over overlapping 12-month blocks of the triples (L_t, L_{t-1}, r_t); each replicate re-estimates the AR(1), the capped correction and the final regression. Report the 95% percentile interval, the share of replicates where the cap binds and the uncapped coefficient.

C2 [Proposed]: Does Illiquidity Carry a Premium on the BVB? — Task (1/2)

- **Worked example to follow**: B4, B5
- **Questions**: Amihud (2002) predicts that illiquidity is priced Amihud (2002)
  - does expected illiquidity raise future returns on the BVB?
  - does an unexpected rise in illiquidity lower current returns?
- **Data and assumptions**
  - Markets: BVB (return BET-TR; illiquidity from the ten blue chips), US (SPY; six US stocks), crypto (Bitcoin; six other coins); monthly log returns in %, 2016–2026
  - Index: mean daily Amihud ratio per asset-month ($\geq$ 10 days); cross-sectional mean of its log, L_t; shock: residual of the autoregressive model AR(1) $L_t = \phi_{\mathbf{0}} + \phi L_{t-\mathbf{1}} + v_t$, with $\hat{\phi}$ the estimated persistence and T the number of months

C2 [Proposed]: Does Illiquidity Carry a Premium on the BVB? — Task (2/2)

- **Tasks**
  1. Milestone 1: build the index and its shock for each market.
  2. Milestone 2: estimate $r_t = a + b_1 L_{t-1} + b_2 \,\mathrm{shock}_t + e_t$ with HAC SEs (6 lags); r_t: the market's monthly return; Amihud predicts $b_1 > 0$ (expected illiquidity is priced) and $b_2 < 0$ (an illiquidity shock lowers prices).
  3. Milestone 3: correct b_1 for the Stambaugh bias Stambaugh (1999) (the small-sample bias of $\hat{b}_1$ when L_t is persistent and its shocks are correlated with returns) with the reduced-bias estimator: $\hat{\phi}_c = \hat{\phi} + (1 + 3\hat{\phi})/T + 3(1 + 3\hat{\phi})/T^2$, residuals v_t^c from $\hat{\phi}_c$, regress r_t on L_{t-1} and v_t^c, SE including the uncertainty of $\hat{\phi}_c$ Amihud & Hurvich (2004).
  4. Milestone 4 (sorts): each January rank the blue chips with more than 150 trading days by the previous year's Amihud; the membership of the illiquid and liquid halves is set once a year and equal weights are restored every month (the mean of the members' monthly simple total returns); test the mean difference with HAC SEs.
- **Report**: a table per market, the reference charts and a discussion of power (few stocks, ten years) Pele, Mazurencu-Marinescu & Nijkamp (2013)
- **Interpretation**: which of Amihud's two predictions survives on each market?

C3 [Proposed]: The Price of the Speed Race in SPY — Task (1/2)

- **Worked example to follow**: B1, B4
- **Questions**: if the London latency-arbitrage tax applied to SPY
  - how large would the prize be?
  - how much would it move on stress days?
- **Data and assumptions**
  - Lecture case study Aquilina, Budish & O'Neill (2022): tax 0.419 bp (Table X); Table XV: col. 2 $\Pi_t = 0.4213 \cdot 10^{-4} V_t$; col. 6 $\Pi_t = 10^{-4}(0.3354\, V_t + 0.0066\, \sigma_t \bar{V})$; Table XIV scenarios 0.20–0.74 bp; SPY full days 2021–2025
  - Units: Π_t, V_t in USD; V_t = sum of close $\times$ volume over the 78 bars; $\sigma_t = 100\sqrt{252 \sum r^2}$ in %; $\bar{V}$ the SPY yearly mean, not the London sample mean
  - Outputs are implications of transferred coefficients, not measured US race profits

Back Exercise list

C3 [Proposed]: The Price of the Speed Race in SPY — Task (2/2)

- **Tasks**
  1. Daily inputs: compute V_t and σ_t; report yearly means and the yearly correlation of V_t and σ_t with 95% intervals from an i.i.d. day bootstrap and a 20-day moving-block bootstrap; compare with the daily volume–volatility correlation of 0.82 in the London sample (Section VI.A, discussion of Table XV) Aquilina, Budish & O'Neill (2022).
  2. Annual extrapolation: compute each year's prize $\sum_t \Pi_t$ under col. 2, col. 6 and the two scenarios.
  3. Show that under col. 6 the annual tax is $0.3354 + 0.0066\,\bar{\sigma}$ bp and find the $\bar{\sigma}$ above which col. 6 exceeds col. 2.
  4. Stress days: plot the tax at mean traded value, $0.3354 + 0.0066\,\sigma_t$, and the realised ratio $\Pi_t / V_t = 0.3354 + 0.0066\,\sigma_t \bar{V} / V_t$; work 7 April 2025 completely.
- **Report**: the annual table, the derivation and threshold, the daily chart and the worked day
- **Interpretation**
  - which assumptions carry a London 2015 tax over to SPY today?
  - can the data at hand sign their effect?

Acronyms used in this material (1/2)

- **AAPL**: Apple Inc. (Nasdaq ticker)
- **ACF**: AutoCorrelation Function
- **AI**: Artificial Intelligence
- **AR**: AutoRegressive (model)
- **BET**: Bucharest Exchange Trading (index)
- **BET-TR**: BET Total Return (index)
- **BIC**: Bayesian Information Criterion
- **BNR**: Banca Națională a României — *the National Bank of Romania*
- **BVB**: Bursa de Valori București — *the Bucharest Stock Exchange*
- **CBOE**: Chicago Board Options Exchange
- **CI**: Confidence Interval
- **DVOL**: Dollar Volume — traded value in USD
- **EKOP**: Easley–Kiefer–O'Hara–Paperman (model of informed trading)
- **ETF**: Exchange-Traded Fund
- **GME**: GameStop Corp. (NYSE ticker)
- **HAC**: Heteroskedasticity and Autocorrelation Consistent (standard errors)

Acronyms used in this material (2/2)

- **JPM**: JPMorgan Chase & Co. (NYSE ticker)
- **MA**: Moving Average
- **MSFT**: Microsoft Corporation (Nasdaq ticker)
- **MSTR**: Strategy Inc., formerly MicroStrategy (Nasdaq ticker)
- **NVDA**: NVIDIA Corporation (Nasdaq ticker)
- **NVIDIA**: NVIDIA Corporation (company name, US chip designer)
- **OHLC**: Open, High, Low, Close
- **OLS**: Ordinary Least Squares
- **OMV**: Österreichische Mineralölverwaltung — *Austrian Mineral Oil Administration (OMV group)*
- **PIN**: Probability of INformed trading
- **SD**: Standard Deviation
- **SE**: Standard Error
- **US**: United States
- **UTC**: Coordinated Universal Time (Temps Universel Coordonné)
- **VECM**: Vector Error-Correction Model
- **VIX**: CBOE Volatility IndeX
- **XRP**: XRP (the native token of the XRP Ledger)

References (1/2)

- Abdi, F., Ranaldo, A. (2017). A simple estimation of bid-ask spreads from daily close, high, and low prices. *Review of Financial Studies*, 30(12), 4437–4480.
- Almgren, R., Chriss, N. (2001). Optimal execution of portfolio transactions. *Journal of Risk*, 3(2), 5–39.
- Amihud, Y. (2002). Illiquidity and stock returns: cross-section and time-series effects. *Journal of Financial Markets*, 5(1), 31–56.
- Amihud, Y., Hurvich, C.M. (2004). Predictive regressions: a reduced-bias estimation method. *Journal of Financial and Quantitative Analysis*, 39(4), 813–841.
- Aquilina, M., Budish, E., O'Neill, P. (2022). Quantifying the high-frequency trading "arms race". *Quarterly Journal of Economics*, 137(1), 493–564.
- Brauneis, A., Mestel, R., Riordan, R., Theissen, E. (2021). How to measure the liquidity of cryptocurrency markets? *Journal of Banking & Finance*, 124, 106041.
- Corwin, S.A., Schultz, P. (2012). A simple way to estimate bid-ask spreads from daily high and low prices. *Journal of Finance*, 67(2), 719–760.
- Easley, D., Kiefer, N.M., O'Hara, M., Paperman, J.B. (1996). Liquidity, information, and infrequently traded stocks. *Journal of Finance*, 51(4), 1405–1436.
- Efron, B. (1979). Bootstrap methods: another look at the jackknife. *Annals of Statistics*, 7(1), 1–26.

References (2/2)

- Glosten, L.R., Milgrom, P.R. (1985). Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. *Journal of Financial Economics*, 14(1), 71–100.
- Gonzalo, J., Granger, C.W.J. (1995). Estimation of common long-memory components in cointegrated systems. *Journal of Business & Economic Statistics*, 13(1), 27–35.
- Harris, L. (1990). Statistical properties of the Roll serial covariance bid/ask spread estimator. *Journal of Finance*, 45(2), 579–590.
- Hasbrouck, J. (1995). One security, many markets: determining the contributions to price discovery. *Journal of Finance*, 50(4), 1175–1199.
- Hasbrouck, J. (2007). Empirical Market Microstructure. Oxford University Press.
- Johansen, S. (1991). Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica*, 59(6), 1551–1580.
- Kyle, A.S. (1985). Continuous auctions and insider trading. *Econometrica*, 53(6), 1315–1335.
- Künsch, H.R. (1989). The jackknife and the bootstrap for general stationary observations. *Annals of Statistics*, 17(3), 1217–1241.
- Lin, H.-W.W., Ke, W.-C. (2011). A computing bias in estimating the probability of informed trading. *Journal of Financial Markets*, 14(4), 625–640.
- Newey, W.K., West, K.D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3), 703–708.
- Pele, D.T., Mazurencu-Marinescu, M., Nijkamp, P. (2013). Herding behaviour, bubbles and log periodic power laws in illiquid stock markets: a case study on the Bucharest Stock Exchange. Social Science Research Network (SSRN) Working Paper 2305945.
- Roll, R. (1984). A simple implicit measure of the effective bid-ask spread in an efficient market. *Journal of Finance*, 39(4), 1127–1139.
- Stambaugh, R.F. (1999). Predictive regressions. *Journal of Financial Economics*, 54(3), 375–421.
- Tóth, B., Lempérière, Y., Deremble, C., de Lataillade, J., Kockelkoren, J., Bouchaud, J.-P. (2011). Anomalous price impact and the critical nature of liquidity in financial markets. *Physical Review X*, 1(2), 021006.