








Statistics of Financial Markets

Seminar 0: Introduction



Daniel Traian PELE

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

Acronyms used in this material

- **AI**: Artificial Intelligence
- **BET**: Bucharest Exchange Trading (index)
- **BET-TR**: BET Total Return (index)
- **BNR**: Banca Națională a României (the National Bank of Romania)
- **BVB**: Bursa de Valori București (the Bucharest Stock Exchange)
- **COVID-19**: COronaVIrus Disease 2019
- **EOD**: End Of Day (daily closing data)
- **EODHD**: EOD Historical Data (market-data provider)
- **MDD**: Maximum DrawDown
- **US**: United States
- **UTC**: Coordinated Universal Time (Temps Universel Coordonné)

Today's Questions and Seminar Structure

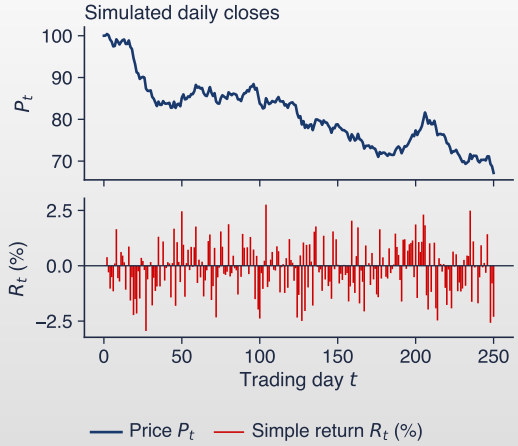
- Questions of the seminar
  - **how do we turn a price series into returns?**
  - **how risky was an asset, per year?**
  - **what was the worst loss from a peak?**
  - **how do we know that the prices are right?**
- Exercises: in the order of the Exercise Map (next slide)
  - **Part I**: setup, then the solved exercises A1–A3, B1 and B2
  - **Part II**: your turn (A4, B3, B4) and the critique of an AI answer (C2)
- Outputs to produce
  - the answers to the A exercises; the table or figure each B exercise asks for
  - the errors found in the AI answer of C2 and the corrected answer
  - the interpretation sentences each exercise asks for
- Nothing is handed in: the proposed exercises are practice for the exam and the project
- Notebook for today: open the seminar notebook in Google Colab; each task names its section
- Contact: Prof. Daniel Traian Pele, danpele@ase.ro

Exercise Map

Exercise	Question	Type	Model
A1	simple and log returns on paper	Solved	–
A2	annualising a daily mean and volatility	Solved	–
A3	drawdown of a short price path	Solved	–
B1	S&P 500, 2015–2026: returns, risk, drawdown	Solved	–
B2	data check: one suspicious EUR/RON quote	Solved	–
A4	the same calculations on new numbers	Proposed	A1–A3
B3	BET-TR, 2015–2026	Proposed	B1
B4	Bitcoin: 365 or 252 days a year?	Proposed	A2, B1
C2	find the errors in an AI answer	Proposed	A2, B1

- **[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: on paper; Part B: real data, each ending with an interpretation question; Part C: critique of an AI answer

Prerequisites for Today (1/7): Prices and Simple Returns



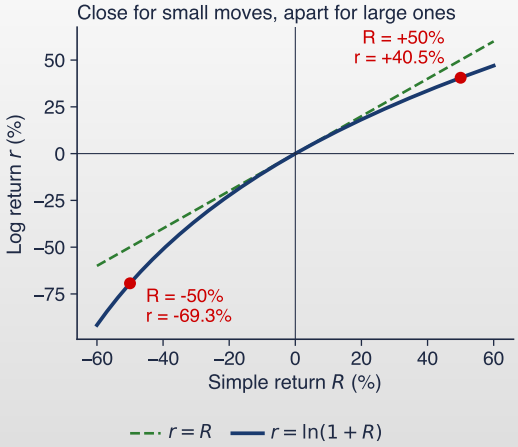
- **Price** P_t: the closing price of day t, $t = 0, 1, \ldots, T$
  - for ETFs and shares: the **adjusted close**, corrected for dividends and splits
- **Simple return** of day t

$$R_t = \frac{P_t}{P_{t-1}} - 1 = \frac{P_t - P_{t-1}}{P_{t-1}}$$

- R_t: the percentage gain of day t, relative to the close P_{t-1} of the day before
  - $R_t = 0.02$: the price rose by 2%; $R_t = -0.02$: it fell by 2%
  - $R_t \geq -1$: a price cannot fall below zero
- In the chart: a simulated price (top) and its daily returns (bottom)
  - the price wanders; the returns oscillate around 0

SFM_ch0_seminar

Prerequisites for Today (2/7): Log Returns



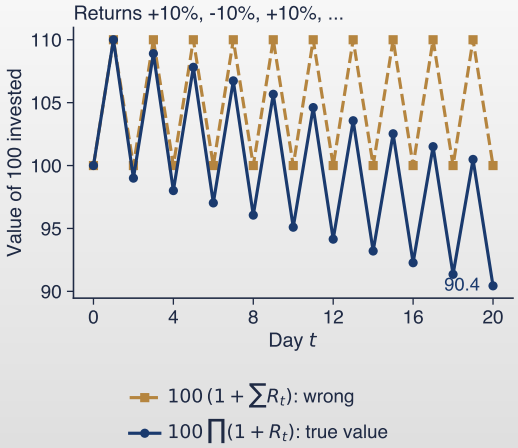
- **Log return** of day t

$$r_t = \ln \frac{P_t}{P_{t-1}} = \ln(1 + R_t), \qquad R_t = e^{r_t} - 1$$

- ln: the natural logarithm; e^x: its inverse, the exponential
  - the second formula turns a log return back into a simple return
- **Small moves**: $r_t \approx R_t$
  - $R_t = 1\%$ gives $r_t = 0.995\%$: for daily data the two are almost equal
- **Large moves**: the two differ
  - $R = -50\%$ gives $r = -69.3\%$; $R = +50\%$ gives $r = +40.5\%$
  - always $r_t \leq R_t$: the blue curve lies below the diagonal $r = R$

SFM_ch0_seminar

Prerequisites for Today (3/7): Returns over Several Days



- **Simple returns compound**

$$1 + R_{0\to T} = \prod_{t=1}^{T}(1 + R_t) = \frac{P_T}{P_0}$$

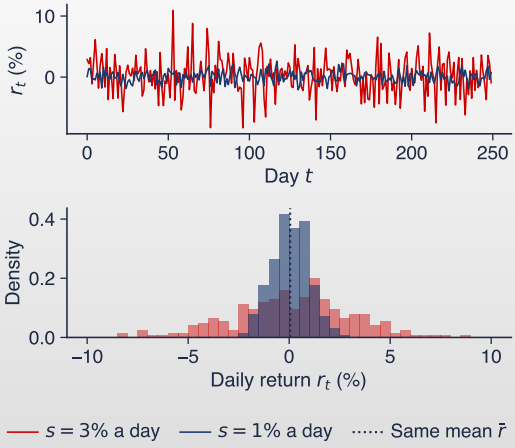
- $R_{0\to T}$: the **cumulative return** from day 0 to day T
  - $\prod_{t=1}^{T}$: the product of the factors $1 + R_t$ of the days $1, \ldots, T$
  - 100 invested on day 0 are worth $100\,P_t/P_0$ on day t
- **Log returns add up**

$$\ln \frac{P_T}{P_0} = \sum_{t=1}^{T} r_t$$

- In the chart: $+10\%$ and -10% in turn, for 20 days
  - the sum of the simple returns stays 0, yet the value falls to $100 \times (1.1 \times 0.9)^{10} = 90.4$

SFM_ch0_seminar

Prerequisites for Today (4/7): Mean and Volatility



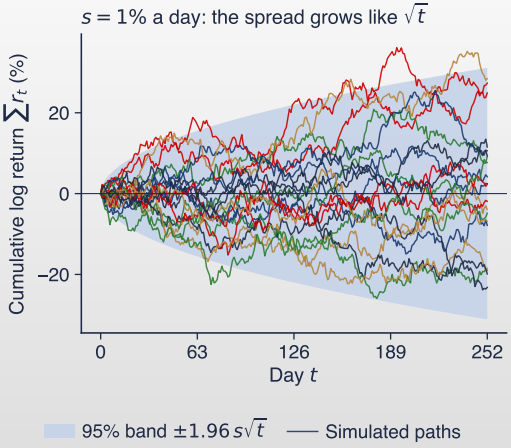
- From n daily returns $r_1, \ldots, r_n$

$$\bar{r} = \frac{1}{n}\sum_{t=1}^{n} r_t, \qquad s = \sqrt{\frac{1}{n-1}\sum_{t=1}^{n}(r_t - \bar{r})^2}$$

- $\bar{r}$: the **sample mean**, the average daily return
  - $r_t - \bar{r}$: the deviation of day t from the mean
- s: the sample standard deviation, called **volatility**
  - the typical distance of a daily return from the mean
  - s^2: the sample variance; $n-1$ because $\bar{r}$ is estimated from the same data
- In the chart: two simulated assets with the same mean
  - $s = 1\%$ (blue) and $s = 3\%$ (red) a day: the same centre, a three times wider spread

SFM_ch0_seminar

Prerequisites for Today (5/7): Annualisation



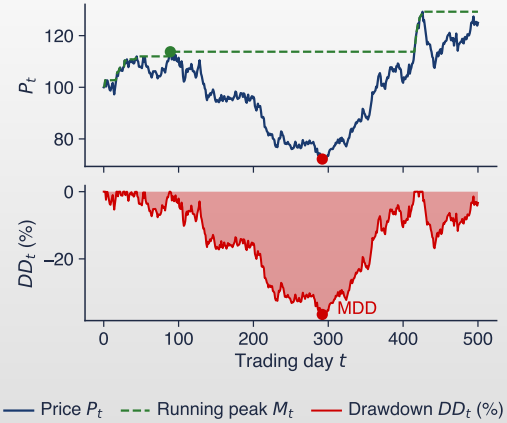
- With A returns per year

$$\bar{r}_{\text{year}} = A\,\bar{r}, \qquad s_{\text{year}} = \sqrt{A}\,s$$

- A: the number of observations per year of the series
  - $A \approx 252$ for exchanges (weekdays without holidays), $A = 365$ for Bitcoin, $A = 12$ for monthly data
- **Why** $\sqrt{A}$: for independent days, variances add up
  - $\mathrm{Var}(r_1 + \cdots + r_A) = A\,s^2$, so the standard deviation is $\sqrt{A}\,s$
  - the mean grows like A, the volatility only like $\sqrt{A}$
- In the chart: 20 simulated paths of $\sum r_t$, with $s = 1\%$ a day
  - the 95% band $\pm 1.96\, s\sqrt{t}$ widens like $\sqrt{t}$: after 252 days, $\pm 1.96 \times 15.9\% = \pm 31.1\%$

SFM_ch0_seminar

Prerequisites for Today (6/7): Drawdown



- **Running peak** and **drawdown** on day t

$$M_t = \max_{s \le t} P_s, \qquad DD_t = \frac{P_t}{M_t} - 1 \le 0$$

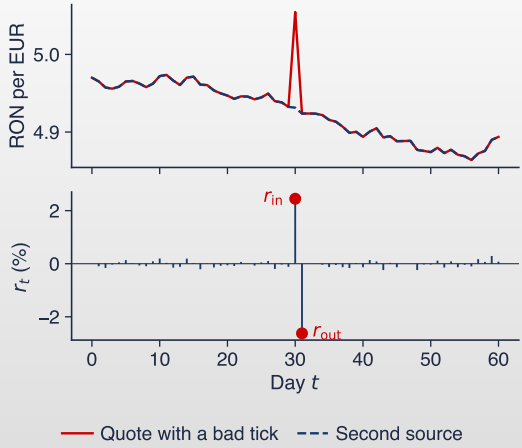
- M_t: the highest price over the days $s = 0, \ldots, t$ (green dashed line)
  - DD_t: the loss from that peak; $DD_t = 0$ on every new maximum
  - the red area is the **underwater** curve: how deep and how long the investor stays below the peak
- **Maximum drawdown**

$$\text{MDD} = \min_t DD_t$$

- the worst loss from a peak to a later trough (from the green to the red point)
  - it is reported as a negative percentage, for example -25%

SFM_ch0_seminar

Prerequisites for Today (7/7): Data Check



- **Bad tick**: an isolated wrong quote in an otherwise correct series
  - before computing statistics, look at the data for prices that cannot be true
- **Its signature**: a jump into the day that reverses out of it

$$r_{\mathbf{in}} = \ln \frac{P_t}{P_{t-\mathbf{1}}}, \qquad r_{\mathbf{out}} = \ln \frac{P_{t+\mathbf{1}}}{P_t} \approx -r_{\mathbf{in}}$$

- **Comparison with a second source**, for example the BNR rate
  - percentage gap of day t: $g_t = 100\,(P_t / P_t^{\mathbf{BNR}} - 1)$
  - $P_t^{\mathbf{BNR}}$: the reference rate of the same day
- **Effect**: one wrong quote adds two large terms $(r_t - \bar{r})^{\mathbf{2}}$ to s
  - the volatility then describes the error, not the market

SFM_ch0_seminar

Data Used Today

Series	Price used	Unit	Calendar	Exercise
S&P 500	index, close	USD	US trading days	B1, C2
BET-TR	index with dividends, close	RON	Romanian trading days	B3
Bitcoin	BTC/USD, close	USD	every day, 24:00 UTC	B4, C2
EUR/RON, official	BNR reference rate	RON per EUR	Romanian working days	B2
EUR/RON, market	series from EODHD, close	RON per EUR	weekdays	B2

- Daily data from EODHD (EOD Historical Data), saved in the course repository; sample from 2 January 2015 to 18 September 2026
- BNR: the National Bank of Romania; its reference rate is fixed once a day, around 13:00 Bucharest time
- BET-TR: the BET index of the BVB (Bucharest Stock Exchange) with dividends reinvested

Setup (1/2): Open the Notebook and Save a Copy

- **Question**: can you run Python on the course data in the browser?
- **Inputs**: Google Colab: Google's free notebook service; Python runs on Google's servers, a Google account is enough
- **Tasks**
  1. Open the seminar notebook in Colab.
  2. Choose *File* → *Save a copy in Drive*, or run the Drive-save cell under the banner, so that your changes are kept.
  3. Run the cells of the section *Setup* in order with Shift + Enter; they import the libraries and define the data loader.
  4. If a cell fails, choose *Runtime* → *Restart session* and run again from the top.
- **Report**: nothing; you are ready when the check on the next slide prints the expected numbers

Setup (2/2): Load the Data

```
# load_close(name, start): one daily series
# from the course data, with the right
# price column and calendar
p = load_close('sp500', start='2015-01-02')
print(len(p), p.index[0].date(),
      p.index[-1].date())
print(p.head(3).round(2))
```

- `SERIES` in the notebook lists every available name

- **Expected output**
  - 2 944 prices, from 2 January 2015 to 18 September 2026
  - first three closes: 2058.20, 2020.58, 2002.61
- If your count or first value differs, check the name and the start date before going on

A1 [Solved]: Simple and Log Returns — Task

- **Question**: can two daily returns simply be added to get the two-day return?
- **Inputs**: prices $P_0 = 100$, $P_1 = 110$, $P_2 = 99$
- **Tasks**
  1. Compute the simple returns R_1, R_2 and the log returns r_1, r_2.
  2. Add the two simple returns, and then add the two log returns.
  3. Compute the two-day returns $P_2/P_0 - 1$ and $\ln(P_2/P_0)$, and compare them with the sums of step 2.
- **Report**: a table with the columns day, price, simple return, log return, and one sentence on which returns add up
- Notebook: section A1

A1 [Solved]: Solution

t	P_t	R_t	r_t
0	100	–	–
1	110	+10.00%	+9.53%
2	99	−10.00%	−10.54%
sum		0.00%	−1.01%
$0 \to 2$		−1.00%	−1.01%

- **Step by step**
  - $R_1 = 110/100 - 1$, $r_1 = \ln 1.1$
  - $R_2 = 99/110 - 1$, $r_2 = \ln 0.9$
- **Why**
  - log returns add up over time: $(\ln P_1 - \ln P_0) + (\ln P_2 - \ln P_1) = \ln P_2 - \ln P_0 = \ln(P_2/P_0)$, because the intermediate term $\ln P_1$ cancels
  - simple returns compound: $1.1 \times 0.9 - 1 = -1\%$
- **Interpretation**
  - adding simple returns reports 0% for a path that lost 1%: add log returns, compound simple returns

SFM_ch0_seminar

A2 [Solved]: Annualisation — Task

- **Question**: what do a daily mean and a daily volatility mean per year?
- **Inputs**: a stock index with daily mean log return 0.05% and daily volatility 1.2%; Bitcoin with daily volatility 3%
- **Tasks**
  1. Annualise the mean and the volatility of the index with $A = 252$.
  2. Annualise the volatility of Bitcoin with $A = 365$, and then with $A = 252$.
  3. Explain which value of A is right for Bitcoin.
- **Report**: three annual numbers and one sentence on the choice of A
- Notebook: section A2

Back Exercise list

A2 [Solved]: Solution

- **Index**, $A = 252$
  - mean: $252 \times 0.05\% = 12.6\%$ a year
  - volatility: $\sqrt{252} \times 1.2\% = 15.9 \times 1.2\% = 19.0\%$ a year
- **Bitcoin**
  - $A = 365$: $\sqrt{365} \times 3\% = 57.3\%$ a year
  - $A = 252$: $\sqrt{252} \times 3\% = 47.6\%$ a year, too low
- **Why the square root**
  - the variances of independent days add up: Prerequisites for Today (5/7)
- **Interpretation**
  - A must be the number of observations per year of the series: Bitcoin trades every day, so $A = 365$

SFM_ch0_seminar

A3 [Solved]: Drawdown on Paper — Task

- **Question**: what was the worst loss from a previous peak?
- **Inputs**: prices on days 0–5: 100, 120, 90, 110, 130, 104
- **Tasks**
  1. Compute the running peak M_t for each day.
  2. Compute the drawdown DD_t for each day (formula: Prerequisites for Today (6/7)).
  3. Find the maximum drawdown and the days of its peak and of its trough.
- **Report**: a table with t, P_t, M_t, DD_t, the maximum drawdown, and whether the price recovered by day 5
- Notebook: section A3

A3 [Solved]: Solution

t	P_t	M_t	DD_t
0	100	100	0.0%
1	120	120	0.0%
2	90	120	−25.0%
3	110	120	−8.3%
4	130	130	0.0%
5	104	130	−20.0%

- **Step by step**
  - day 2: $90/120 - 1 = -25\%$; day 3: $110/120 - 1$
  - day 4: a new peak, so $M_4 = 130$ and $DD_4 = 0$
  - day 5: $104/130 - 1 = -20\%$, measured from the new peak
- **Maximum drawdown**: -25.0%, from the peak of day 1 (120) to the trough of day 2 (90)
- **Interpretation**
  - the price recovered on day 4, but on day 5 the investor is again 20% below the peak

SFM_ch0_seminar

B1 [Solved]: The S&P 500, 2015–2026 — Task

- **Question**: what return and what risk did the S&P 500 show from 2015 to 2026?
- **Inputs**: S&P 500 closing prices, 2 January 2015 – 18 September 2026 (`load_close('sp500')`)
- **Tasks**
  1. Compute the daily simple and log returns, and check the first two step by step from the first three prices.
  2. Compute the number of returns per year A, the annualised mean log return and the annualised volatility.
  3. Compute the cumulative return $P_T/P_0 - 1$, and check it against $\exp(\sum_t r_t) - 1$.
  4. Plot the value of 100 invested and the drawdown, and find the maximum drawdown with its dates.
- **Report**: A, the annual mean and volatility, the cumulative return, the maximum drawdown with its peak and trough dates, and one sentence of interpretation
- Notebook: section B1

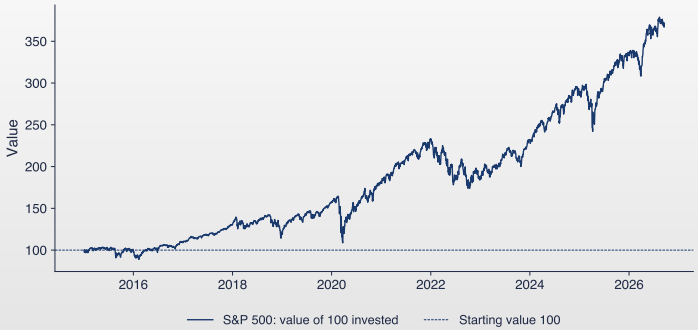
Back Exercise list

B1 [Solved]: Solution (1/3) — Step by Step

- **First returns**, from $P_0 = 2058.20$, $P_1 = 2020.58$, $P_2 = 2002.61$
  - $R_1 = P_1/P_0 - 1 = -1.83\%$, $r_1 = -1.84\%$
  - $R_2 = -0.89\%$, $r_2 = -0.89\%$
- **Annualisation**
  - 2 943 returns in 11.7 years: $A = 251.5$ returns a year
  - daily mean 0.045%, daily volatility 1.12%
  - annual mean 11.2%, annual volatility 17.7%
- **Cumulative return**
  - $P_T/P_0 - 1 = 7650.50 \,/\, 2058.20 - 1 = 271.7\%$; $\exp(\sum_t r_t) - 1$ gives the same 271.7%

SFM_ch0_seminar

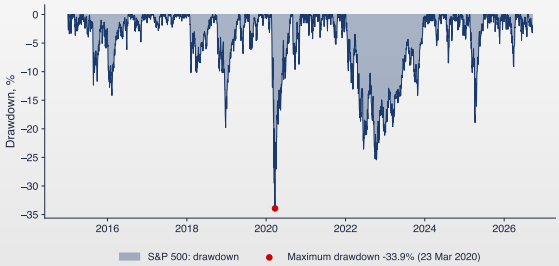
B1 [Solved]: Solution (2/3) — Value of 100 Invested



- 100 invested on 2 January 2015 were worth 372 on 18 September 2026 (price index, without dividends)

SFM_ch0_seminar

B1 [Solved]: Solution (3/3) — Drawdown and Interpretation



- **Maximum drawdown**: −33.9%, from the peak of 19 February 2020 to 23 March 2020 (COVID-19)
- Second deepest episode: 2022, when interest rates rose
- **Interpretation**
  - a good average year (11.2%) hides a fall of a third in one month
  - report the volatility and the drawdown next to the mean

SFM_ch0_seminar

B2 [Solved]: One Suspicious Quote — Task

- **Question**: is the largest EUR/RON move in the market series a real price?
- **Inputs**: the EUR/RON series from EODHD (`eurron_eodhd`) and the BNR reference rate (`eurron`), 2015–2026
- **Tasks**
  1. For each day, compute the gap between the EODHD quote and the BNR rate of the same day, in percent.
  2. Find the day with the largest absolute gap, and print the quotes of the day before and the day after.
  3. Compute the log returns into and out of that day.
  4. Plot both series in the month around that day.
- **Report**: the date, the quote, the BNR rate, the two log returns, and your decision: real price or bad tick
- Notebook: section B2

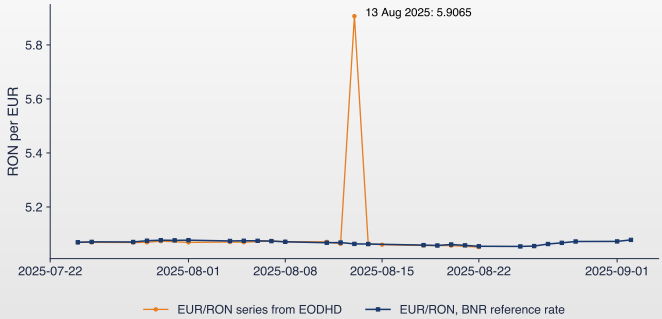
Back Exercise list

B2 [Solved]: Solution (1/2) — Step by Step

- **Largest gap**: 13 August 2025
  - EODHD quote 5.9065 against the BNR rate 5.0633: a gap of +16.7%
  - day before (12 August 2025): 5.0636; day after (14 August 2025): 5.0629
- **Two log returns**
  - into the day: +15.4%; out of the day: −15.4%
  - for comparison, the largest daily move of the BNR rate since 2015 is 1.4%
- **Decision**: a bad tick
  - a jump of 15% that fully reverses the next day, absent from the official rate
  - kept in the data, it would add two of the largest daily moves of the leu in a decade
- In the course we use the BNR rate for EUR/RON

SFM_ch0_seminar

B2 [Solved]: Solution (2/2) — Chart and Interpretation



- One quote far from every other quote and from the official rate: always look at the data before computing statistics
- Interpretation: a volatility computed with this tick would describe an error, not the market

SFM_ch0_seminar

A4 [Proposed]: New Numbers — Task

- **Question**: can you repeat A1–A3 without the solution in front of you?
- **Inputs**: a share at $P_0 = 50$, $P_1 = 55$, $P_2 = 44$; a fund at 200, 180, 220, 165, 198, 230 on days 0–5; daily mean 0.03% and daily volatility 1.5%
- **Tasks**
  1. Model: A1, A2 and A3 [Solved].
  2. Compute the two simple and the two log returns of the share, their sums and the two-day returns.
  3. Compute the drawdown of the fund for each day and its maximum drawdown.
  4. Annualise the daily mean and volatility with $A = 252$.
- **Report**: the two tables, the maximum drawdown, the two annual numbers, and one sentence on which return sum matches the two-day return
- Notebook: section A4

Back Exercise list

B3 [Proposed]: The BET-TR, 2015–2026 — Task

- **Question**: how did the Romanian market with dividends compare with the S&P 500?
- **Inputs**: BET-TR closing prices from 2 January 2015 (`load_close('bettr')`) to 18 September 2026
- **Tasks**
  1. Model: B1 [Solved]; reuse its code and change only the series name.
  2. Compute A, the annualised mean log return and the annualised volatility.
  3. Compute the cumulative return and the maximum drawdown with its dates.
  4. Compare each number with the S&P 500 result of B1.
- **Report**: a two-column table (S&P 500, BET-TR) and one sentence: why is the comparison not fully fair?
- Notebook: section B3

Back Exercise list

B4 [Proposed]: Bitcoin: 365 or 252 Days? — Task

- **Question**: how large is the error if Bitcoin is annualised like a stock index?
- **Inputs**: Bitcoin closing prices from 2 January 2015 (`load_close('btc')`) to 18 September 2026
- **Tasks**
  1. Model: A2 and B1 [Solved].
  2. Compute the number of returns per year A from the data.
  3. Compute the annualised volatility with $A = 365$, and then with $A = 252$.
  4. Compute the cumulative return and the maximum drawdown with its dates.
- **Report**: A, the two volatilities and their ratio, the cumulative return, the maximum drawdown, and one sentence on which volatility is right
- Notebook: section B4

Back Exercise list

C2 [Proposed]: Critique an AI Answer — the Answer

- **Prompt** sent to an AI assistant: "From the course data, compute the annualised volatility and the cumulative return of the S&P 500 and of Bitcoin since 2 January 2015, from daily simple returns."

```
# read_market(symbol): the daily table of the course data
for name, sym in [('S&P␣500', 'GSPC.INDX'), ('Bitcoin', 'BTC-USD.CC')]:
    p = read_market(sym)['close'].loc['2015-01-02':]   # closing prices
    R = p.pct_change().dropna()          # daily simple returns
    vol = R.std() * 252                  # annualised volatility
    cum = R.sum()                        # cumulative return
    print(name, round(100 * vol, 1), round(100 * cum, 1))
```

- **Conclusion of the AI**: "Volatility: S&P 500 280.3%, Bitcoin 873.3%. Cumulative return: S&P 500 149.6%, Bitcoin 814.0%. Bitcoin earned only about five times as much as the S&P 500."

Back Exercise list

C2 [Proposed]: Critique an AI Answer — Task

- **Question**: can you trust code that runs without an error message?
- **Inputs**: the code and the conclusion on the previous slide; the definitions of the prerequisites
- **Tasks**
  1. Model: A1, A2 and B1 [Solved].
  2. Find the three errors planted in the code.
  3. For each error, say how you would detect it: a check on the output, the prerequisites, or a solved exercise.
  4. Write the corrected code, and report the corrected volatility and cumulative return of both assets.
  5. Say whether the conclusion "Bitcoin earned only about five times as much" still holds.
- **Report**: a table AI versus corrected, and one line per error in the format: request, answer, error, how found, correction
- Notebook: section C2

Practice at Home

- **Finish the proposed exercises**: A4, B3, B4 and C2
- **Try one more series**
  - repeat B1 for the DAX (`dax`) or for gold (`gold`), and compare the annual volatilities
- **Take the Chapter 0 quiz** on the course website: 20 questions, each answer explained

Recap and Preview of the Chapter 0 Lecture

- **Today**
  - simple returns compound, log returns add up
  - annualise the mean with A and the volatility with $\sqrt{A}$; $A = 365$ for Bitcoin
  - the maximum drawdown is the worst loss from a peak
  - look at the data first: one bad tick can dominate a statistic
  - AI code can run and still be wrong
- **In the lecture**
  - How is the course organised and graded?
  - What do the 2026 markets look like in the data?
  - How did exchanges, crashes and models develop?
  - Why are daily returns not Normal?