



# Statistics of Financial Markets

## Seminar 1: Returns and Stationarity



Daniel Traian PELE  
Antoaneta Amza

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

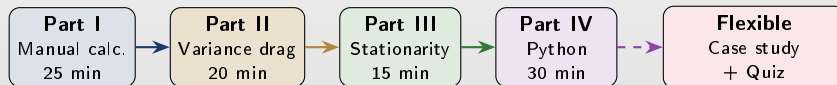
## Seminar Contents

- ▣ **Objectives** – Learning goals and seminar roadmap
- ▣ **Part I: Manual calculation** – Computing simple and log returns by hand
- ▣ **From calculation to intuition** – Why symmetric returns don't cancel out
- ▣ **Part II: Variance Drag** – The hidden cost of volatility
- ▣ **Part III: Stationarity** – Why we work with returns, not prices
- ▣ **Part IV: Python implementation** – Downloading data, computing returns, statistical tests
- ▣ **Case study: S&P 500** – Empirical analysis of real financial data
- ▣ **Homework** – Analysis of a Romanian stock (BSE)
- ▣ **Summary** – Competence checklist and key takeaways
- ▣ **Seminar planning** – Chapter 1 roadmap (4 weeks)
- ▣ **Quick quiz** – Test your understanding (15 minutes)
- ▣ **References** – Key textbooks and online resources

## Learning Objectives — Seminar 1

By the end of this seminar, you will be able to:

1. **Compute** simple and logarithmic returns from price data
2. **Compare** the temporal aggregation of both return types
3. **Explain** variance drag and its impact
4. **Argue** why we analyse returns rather than prices (stationarity)
5. **Implement** return calculations in Python



Total: 90 min lecture + case study and quiz as supplementary material

## Exercise 1: Computing simple and log returns

## Price data

| $t$   | 0   | 1   | 2  | 3   |
|-------|-----|-----|----|-----|
| $P_t$ | 100 | 110 | 99 | 120 |

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}} \text{ (simple)}$$

- ▣  $R_1 = 10/100 = +10.00\%$
- ▣  $R_2 = -11/110 = -10.00\%$
- ▣  $R_3 = 21/99 = +21.21\%$

$$r_t = \ln(P_t/P_{t-1}) \text{ (logarithmic)}$$

- ▣  $r_1 = \ln(1.10) = +9.531\%$
- ▣  $r_2 = \ln(0.90) = -10.536\%$
- ▣  $r_3 = \ln(1.2121) = +19.237\%$

## Key observation

- ▣  $R_1 + R_2 = 0\%$ , but the price dropped to 99! Simple returns are **not additive** — they compound multiplicatively:  $\prod(1 + R_t)$ .

## Exercise 1: Additivity and comparison

### Verification: log additivity

- ▣  $r_1 + r_2 + r_3 = +18.232\%$
- ▣  $\ln(P_3/P_0) = \ln(1.20) = +18.232\% \checkmark$

### Simple compounding (multiplicative)

$$\prod_{t=1}^3 (1 + R_t) = 1.10 \times 0.90 \times 1.2121 = 1.20$$

- ▣  $R_{\text{total}} = 20\%; \quad e^{r_{\text{total}}} - 1 = 20\% \checkmark$

|       | Per. 1 | Per. 2 | Per. 3 | Total  |
|-------|--------|--------|--------|--------|
| $R_t$ | +10.00 | -10.00 | +21.21 | +20.00 |
| $r_t$ | +9.53  | -10.54 | +19.24 | +18.23 |

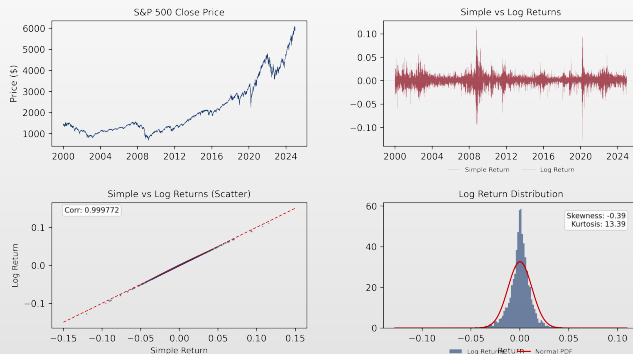
- ▣ Units: %. Both methods  $\Rightarrow$  same terminal wealth.

### Conclusion

- ▣ Log returns **add up**, simple returns **multiply**
- ▣ We use log returns in statistical modelling for temporal additivity

## Visualisation: Return analysis (S&P 500, 2000–2024)

Ch.1: Returns Analysis (S&P 500, 2000–2024)



- ▣ **Scatter:** correlation  $\approx 0.9998$  between  $R_t$  and  $r_t$  — nearly identical for small values
- ▣ **Histogram:** heavier tails than Normal ( $K = 13.4$ ,  $S = -0.39$ )

## Quick poll — Vote now!

 Question

- ☐ You invest **\$100** in an asset. It rises by  $+50\%$  in one year, then falls by  $-50\%$  the next year.
- ☐ **How much are you left with?**

**A)**  
\$100  
(0% net)

**B)**  
\$90  
( $-10\%$ )

**C)**  
\$75  
( $-25\%$ )

**D)**  
\$50  
( $-50\%$ )

## Quick poll — Answer

### Calculation

- ▣  $\$100 \times 1.50 = \$150$ , then  $\$150 \times 0.50 = \text{\textcolor{red}{\$75}}$
- ▣ You lose **25%**, not 0%!

### Why?

- ▣ The 50% gain applies to \$100, but the 50% loss applies to \$150 — a *larger base*
- ▣ This is **variance drag** in action: volatility *destroys* wealth
- ▣ Formula:  $\text{loss} \approx \sigma^2/2$  per year (we prove this in Part II)

## Guided discussion: Why don't +50% and -50% cancel out?

You just saw: \$100  $\rightarrow$  \$150  $\rightarrow$  \$75. But what about log returns?

### Simple returns (multiplicative)

$$(1.50)(0.50) = 0.75 \neq 1.00$$

- ▣  $R_1 + R_2 = +50\% + (-50\%) = 0\%$
- ▣ But terminal wealth: \$75, not \$100!
- ▣ Simple returns are **not** additive

### Log returns (additive)

$$r_{up} = \ln(1.50) = +40.55\%$$

$$r_{down} = \ln(0.50) = -69.31\%$$

- ▣ Sum:  $-28.77\% = \ln(0.75)$  ✓
- ▣ Log returns **correctly** reflect the loss

### Key message

- ▣ The asymmetry is dramatic: you need +100% to recover from -50%
- ▣ The higher the volatility, the more severe the loss

## Variance Drag: Formula and intuition

### Definition 1 (Variance drag)

- ▣ Relationship between the expected arithmetic and geometric return:

$$\mathbb{E}[r] \approx \mathbb{E}[R] - \frac{1}{2}\sigma^2$$

- ▣  $\mathbb{E}[R]$  — expected simple return
  - ▣  $\mathbb{E}[r]$  — expected log return
  - ▣  $\sigma^2$  — variance of simple returns
- 
- ▣ Variance drag grows **quadratically** with volatility: doubling  $\sigma \Rightarrow$  quadrupling the loss
  - ▣ This is a consequence of **Jensen's inequality**:  $\mathbb{E}[\ln(1 + R)] < \ln(1 + \mathbb{E}[R])$  since  $\ln$  is concave
  - ▣ **Implication**: volatility *destroys* wealth in the long run, even if the arithmetic return is positive

## Exercise 2: Two assets, same mean, different volatility

## Problem

| Asset | $\mathbb{E}[R]$ | $\sigma$ | $\mathbb{E}[r] \approx \mathbb{E}[R] - \sigma^2/2$ |
|-------|-----------------|----------|----------------------------------------------------|
| A     | 12%             | 10%      | ?                                                  |
| B     | 12%             | 30%      | ?                                                  |

## Solution

- Asset A:  $\mathbb{E}[r]_A \approx 0.12 - 0.5 \times (0.10)^2 = 0.12 - 0.005 = 11.50\%$
- Asset B:  $\mathbb{E}[r]_B \approx 0.12 - 0.5 \times (0.30)^2 = 0.12 - 0.045 = 7.50\%$

## Question for students

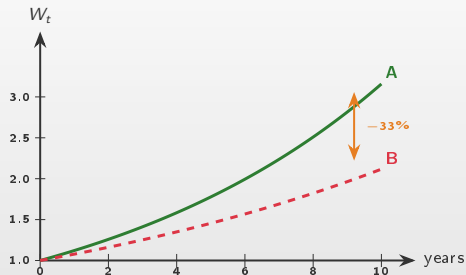
- Which asset produces more wealth in the long run?
- Compute  $W_{10}$  for both ( $W_0 = 1$ )

## Exercise 2: Solution — Long-run wealth

$$W_T = W_0 \cdot e^{T \times \mathbb{E}[r]}$$

- ▣ A:  $W_{10} = e^{10 \times 0.115} = 3.16$  (tripled)
- ▣ B:  $W_{10} = e^{10 \times 0.075} = 2.12$  (doubled)

| Asset | $\mathbb{E}[R]$ | $\sigma$ | $\mathbb{E}[r]$ | $W_{10}$ |
|-------|-----------------|----------|-----------------|----------|
| A     | 12%             | 10%      | 11.50%          | 3.16     |
| B     | 12%             | 30%      | 7.50%           | 2.12     |



## Key conclusion

- ▣ **High volatility destroys wealth!** Asset B has a drag of 4.5%/yr — over 10 years, it costs  $\approx 33\%$  of wealth
- ▣ *“Volatility is the cost of uncertainty.”*

## Exercises 3–5

Exercise 3: Additivity of log returns — Show that  $r_{0,T} = \ln(P_T/P_0) = \sum_{t=1}^T r_t$

- *Hint:* write  $\frac{P_T}{P_0} = \frac{P_1}{P_0} \cdot \frac{P_2}{P_1} \cdots \frac{P_T}{P_{T-1}}$  and apply  $\ln$  to the product.

Exercise 4: Variance drag — Show that  $\mathbb{E}[r] \approx \mathbb{E}[R] - \frac{1}{2}\sigma^2$

- Using the Taylor expansion:  $\ln(1 + R) \approx R - R^2/2$ , apply  $\mathbb{E}[\cdot]$  to both sides.  
□ *Hint:*  $\mathbb{E}[R^2] = \text{Var}(R) + (\mathbb{E}[R])^2 \approx \sigma^2$  for  $\mathbb{E}[R] \approx 0$ .

Exercise 5: Jensen's inequality — Derive  $\mathbb{E}[\ln(1 + R)] \leq \ln(1 + \mathbb{E}[R])$

- The function  $f(x) = \ln(x)$  is concave. State Jensen's inequality and interpret: what does this imply for the relationship between the geometric and arithmetic return?

## Exercises 3–5: Solutions

## Solution Ex. 3 — Additivity

$$\square \quad \frac{P_T}{P_0} = \prod_{t=1}^T \frac{P_t}{P_{t-1}} \Rightarrow \ln\left(\frac{P_T}{P_0}\right) = \sum_{t=1}^T \underbrace{\ln\left(\frac{P_t}{P_{t-1}}\right)}_{=r_t} \quad \square$$

## Solution Ex. 4 — Variance drag

- $r = \ln(1 + R) \approx R - \frac{R^2}{2}$ . Apply  $\mathbb{E}[\cdot]$ :
- $\mathbb{E}[r] \approx \mathbb{E}[R] - \frac{1}{2}\mathbb{E}[R^2] = \mathbb{E}[R] - \frac{1}{2}(\sigma^2 + \mu^2) \approx \mathbb{E}[R] - \frac{1}{2}\sigma^2 \quad (\text{since } \mu^2 \ll \sigma^2) \quad \square$

## Solution Ex. 5 — Jensen

- Jensen's inequality: if  $f$  is concave, then  $\mathbb{E}[f(X)] \leq f(\mathbb{E}[X])$
- With  $f = \ln$ ,  $X = 1 + R$ :  $\mathbb{E}[\ln(1 + R)] \leq \ln(1 + \mathbb{E}[R])$
- Interpretation: geometric return < arithmetic return; the gap increases with  $\sigma$  □

## Why do we work with returns, not prices?

Prices:  $P_t \sim I(1)$

- ▣  $\mathbb{E}[P_t]$ ,  $\text{Var}(P_t)$  are **not** constant
- ▣ Regression on prices  $\Rightarrow$  **spurious regression** (Granger & Newbold, 1974)

Returns:  $r_t \sim I(0)$

- ▣ Approximately stationary (mean  $\approx 0$ , variance  $\approx \text{const.}$ )
- ▣ Statistical methods are **valid**

### Definition 2 (Weak stationarity (covariance stationarity))

A process  $\{X_t\}$  is weakly stationary if,  $\forall t, h$ :

$$\mathbb{E}[X_t] = \mu, \quad \text{Var}(X_t) = \sigma^2 < \infty, \quad \text{Cov}(X_t, X_{t+h}) = \gamma(h).$$

### Consequence

- ▣  $P_t \sim I(1)$ : non-stationary (unit root)
- ▣  $r_t = \Delta \ln P_t \sim I(0)$ : stationary
- ▣ Classical methods (OLS,  $t$ -tests,  $F$ -tests) **require stationarity**; on  $I(1)$  series they produce spurious results

## Random walk and non-stationarity

Random walk:  $P_t = P_{t-1} + \varepsilon_t$ ,  $\varepsilon_t \stackrel{iid}{\sim} (0, \sigma^2)$

### Why is $P_t$ non-stationary?

- ▣ Iterating:  $P_t = P_0 + \sum_{s=1}^t \varepsilon_s$
- ▣  $\text{Var}(P_t) = t \sigma^2 \xrightarrow{t \rightarrow \infty} \infty$
- ▣ Variance **grows linearly** with  $t$

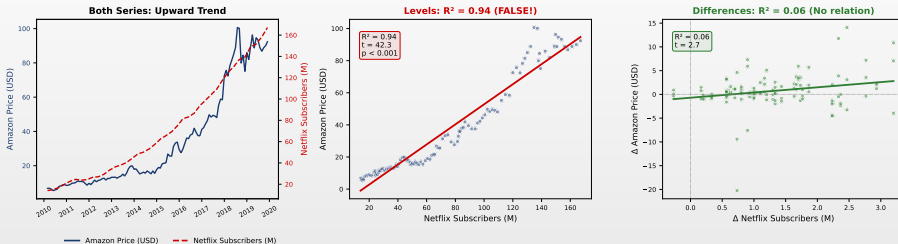
### But $r_t = \Delta \ln P_t$ is stationary

- ▣  $\mathbb{E}[r_t] = \mu$  (constant)
- ▣  $\text{Var}(r_t) = \sigma^2$  (constant)
- ▣  $\text{Cov}(r_t, r_{t+h}) = \gamma(h)$  — depends only on  $h$

### Strict vs. weak

- ▣ **Strict**: the joint distribution  $(X_{t_1}, \dots, X_{t_k})$  is invariant to time shifts
- ▣ **Weak**: only the first two moments are invariant (sufficient for OLS, ACF)
- ▣ Strict  $\Rightarrow$  weak (if  $\mathbb{E}[X_t^2] < \infty$ ), but **not** conversely

## Spurious regression: Amazon Price vs. Netflix Subscribers (2010–2020)



### In levels (price vs. subscribers)

|               |                  |
|---------------|------------------|
| $R^2$         | 0.94             |
| $\hat{\beta}$ | 0.63             |
| t-stat        | 42.3             |
| Conclusion    | <b>Spurious!</b> |

### In differences ( $\Delta P_t$ vs. $\Delta S_t$ )

|               |                        |
|---------------|------------------------|
| $R^2$         | 0.06                   |
| $\hat{\beta}$ | 1.10                   |
| t-stat        | 2.7                    |
| Conclusion    | <b>No relationship</b> |

■ **Lesson:** A stock price and a subscriber count — **completely different** variables, but both trending  $\Rightarrow$

Statistics of Financial Markets.  $R^2 = 0.94$  (Granger & Newbold, 1974). In differences:  $R^2 = 0.06$ .

## Python implementation: Download and compute

### Download data + returns + descriptive statistics

```
import yfinance as yf
import numpy as np
import matplotlib.pyplot as plt
from statsmodels.tsa.stattools import adfuller
from scipy.stats import jarque_bera

ticker = "AAPL"
data = yf.download(ticker, start="2020-01-01", end="2024-01-01")
data['log_ret'] = np.log(data['Close'] / data['Close'].shift(1))
data['simple_ret'] = data['Close'].pct_change()
r = data['log_ret'].dropna()

print(f"Daily_mean: {r.mean():.6f}")
print(f"Annualised_vol: {r.std()*np.sqrt(252):.4f}")
print(f"Skewness: {r.skew():.4f}")
print(f"Kurtosis: {r.kurtosis()+3:.4f}") # scipy: excess
```

## Python implementation: Statistical tests

### ADF test (stationarity) and Jarque-Bera (normality)

```
# ADF test on log-prices vs. returns
adf_price = adfuller(np.log(data['Close'].dropna()))
adf_ret    = adfuller(r)
print(f"ADF_log-price: stat={adf_price[0]:.3f}, p={adf_price[1]:.4f}")
print(f"ADF_returns: stat={adf_ret[0]:.3f}, p={adf_ret[1]:.4f}")

# Jarque-Bera test
jb_stat, jb_p = jarque_bera(r)
print(f"\nJarque-Bera: JB={jb_stat:.1f}, p-value={jb_p:.2e}")

# ACF on r_t and r_t^2
from statsmodels.tsa.stattools import acf
acf_r    = acf(r, nlags=30)
acf_r2   = acf(r**2, nlags=30)
```

- ▣ `adfuller` returns: (statistic, p-value, lags, nobs, critical values)
- ▣  $p < 0.05 \Rightarrow$  reject  $H_0$  of unit root

## AI Exercise (I): Prompt Engineering

### Task

- Write a prompt for an LLM (ChatGPT, Claude, etc.) that **fully** solves Exercise 1 from this seminar
- The prompt should include: input data, required formulas, desired output format

### Example of a good prompt

- *"I have closing prices  $P_0 = 100$ ,  $P_1 = 110$ ,  $P_2 = 99$ ,  $P_3 = 120$ . Compute simple returns  $R_t$  and log returns  $r_t$  for each period. Verify the additivity of log returns. Present the results in a table."*

### Reflection questions

- Is the AI answer mathematically correct? Verify with manual calculation
- What happens if you change the wording? Do you get a different result?
- Did the AI explain **why** log returns are additive?

## AI Exercise (II): Code Generation and Verification

### Task

- Ask an LLM: *“Write Python code that downloads AAPL data for 2020–2024, computes log returns, tests stationarity with the ADF test, and displays a price vs. returns chart.”*
- Run the generated code — does it work *as-is*?

### Verification checklist

- Are the imports correct? (yfinance, numpy, statsmodels)
- Return formula: `np.log(P/P.shift(1))` or `pct_change()`?
  - ▶ Caution: `pct_change()` = **simple** returns, not log!
- ADF test: correct interpretation of the *p*-value?
- Does the plot have titles, axis labels, legend?

### Pedagogical objective

- AI is a **tool**, not a substitute — generated code must **always be verified**
- Common error: confusion between simple and log returns

## Discussion based on charts

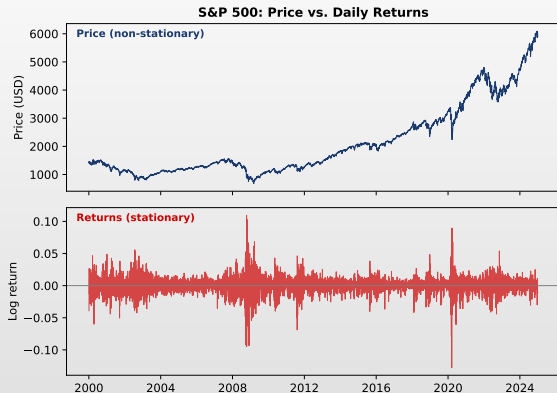
### Questions for students — analysing the generated charts

1. Which series appears **stationary**? Which does not? Why?
2. Where do you observe **volatility clustering**?
  - ▶ Hint: look at March 2020 (COVID-19).
3. Are large returns followed by large returns (of any sign)?
4. Are there visibly “calm” periods alternating with “turbulent” ones?

### Visually observable stylised facts

- **Clustering**: volatile periods come in groups
- **Fat tails**: a few extreme returns (far from the mean)
- **Mean**  $\approx 0$ : returns fluctuate around zero
- These are the **stylised facts** of financial returns (Cont, 2001)

## Case study: S&P 500 — Price vs. Returns



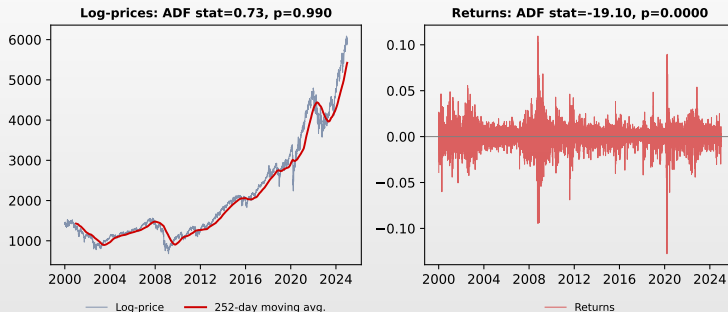
### Data used

- S&P 500 Index
- Period: 2000–2024
- Frequency: daily

### What do we observe?

- Price: upward trend, 2008 and 2020 crises visible
- Returns: oscillate around zero
- Volatility clustering evident during crisis periods

## Case study: S&amp;P 500 — ADF Test



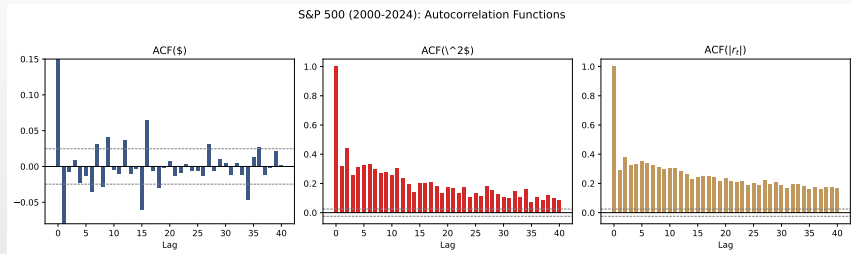
## Log-prices

- $H_0$  is **not** rejected ( $p = 0.98$ )
- Non-stationary series,  $I(1)$
- Unit root confirmed

## Log-returns

- $H_0$  rejected at 1% ( $p \approx 0$ )
- Stationary series,  $I(0)$
- Statistical methods are valid

## Case study: ACF — Returns vs. Squared Returns



$$\text{ACF}(r_t) \approx 0$$

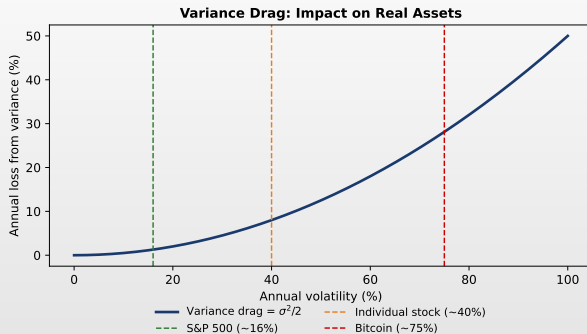
- ▣ Returns are **not** autocorrelated
- ▣ Consistent with the Efficient Market Hypothesis (EMH)
- ▣ Cannot predict  $r_{t+1}$  from  $r_t$

$$\text{ACF}(r_t^2) \neq 0 \text{ — slow decay}$$

- ▣ **Volatility clustering**: volatile periods persist
- ▣ Stylised fact #1 (Cont, 2001)
- ▣ Motivation for GARCH models:  $\sigma_t^2$  depends on  $r_{t-1}^2$

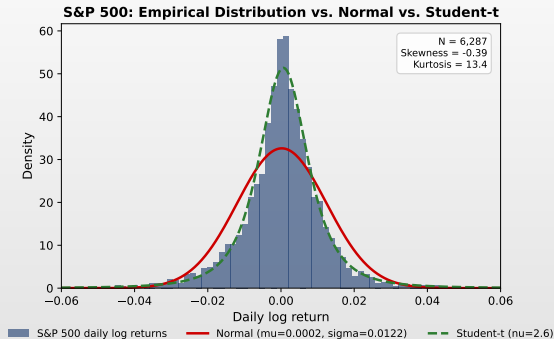
⚠ Returns are uncorrelated but **not** independent — dependence through higher-order moments

## Case study: Variance Drag — Impact on real assets



- ▣ **S&P 500** ( $\sigma \approx 16\%$ ): variance drag  $\approx 1.3\%/yr$  — negligible
- ▣ **Individual stocks** ( $\sigma \approx 30\text{--}40\%$ ): drag  $4\text{--}8\%/yr$  — significant
- ▣ **Bitcoin** ( $\sigma \approx 75\%$ ): drag  $\approx 28\%/yr$  — devastating in the long run
- ▣ **Conclusion:** diversification reduces  $\sigma$  and hence reduces variance drag

## Case study: S&P 500 — Empirical distribution vs. Normal



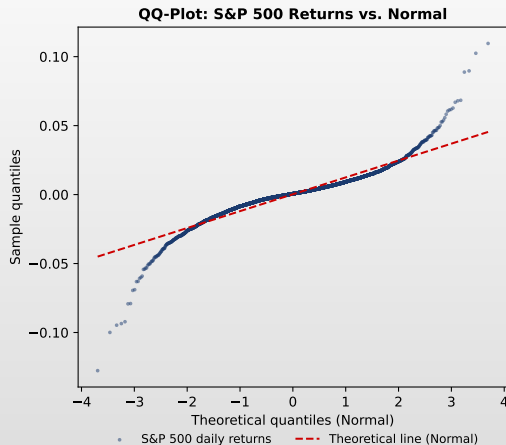
### Real data: S&P 500 (2000–2024)

- $N = 6\,287$  daily log returns
- Empirical histogram vs. Normal curve vs. Student- $t$

### What do we observe?

- Central peak **sharper** than Normal
- Tails **heavier** — crashes far more frequent
- $S = -0.39$ ,  $K = 13.4$  (Normal:  $S = 0$ ,  $K = 3$ )
- Student- $t$  with  $\nu \approx 3$  fits better

## Case study: QQ-Plot — Visual normality check



### How to read a QQ-plot?

- Sample quantiles vs. Normal distribution quantiles
- If data are Normal  $\Rightarrow$  points fall on the 45° line

### What do we observe?

- **S-shape:** tails deviate from the line
  - ▶ Left tail: losses more extreme than Normal predicts
  - ▶ Right tail: gains more extreme
- The Jarque-Bera test formally confirms:
  - ▶  $JB = \frac{n}{6} (S^2 + \frac{\kappa^2}{4}) \sim \chi^2_2$

## Case study: Jarque-Bera Test — Formal rejection of normality

## Definition 3 (Jarque-Bera statistic)

- ▣  $H_0: r_t \sim \mathcal{N}(\mu, \sigma^2)$
- ▣  $JB = \frac{n}{6} \left( S^2 + \frac{(K-3)^2}{4} \right) \xrightarrow{d} \chi^2_2$
- ▣  $S$  = skewness,  $K$  = kurtosis,  $n$  = sample size

Computation on S&P 500 (2000–2024,  $n = 6287$ )

$$\square JB = \frac{6287}{6} ((-0.39)^2 + (13.4 - 3)^2 / 4) = 1047.8 \times 27.19 = 28\,488$$

| Asset   | $n$   | $S$   | $K$  | JB     | $p$ -value   |
|---------|-------|-------|------|--------|--------------|
| S&P 500 | 6 287 | −0.39 | 13.4 | 28 488 | $< 10^{-16}$ |
| AAPL    | 2 516 | −0.31 | 8.6  | 3 168  | $< 10^{-16}$ |
| Bitcoin | 3 287 | −0.15 | 7.4  | 2 675  | $< 10^{-16}$ |

- ▣ Critical value:  $\chi^2_{2,0.01} = 9.21$
- ▣ All  $JB \gg 9.21 \Rightarrow$  reject  $H_0$  at **any** significance level

## Case study: Probability of extreme values

How likely are crashes under the normality assumption?

- Under  $\mathcal{N}(\mu, \sigma^2)$ , a  $5\sigma$  event has prob.  $\approx 2.87 \times 10^{-7}$
- Black Monday** (19 Oct. 1987): S&P 500  $-22.6\%$  in one day ( $> 20\sigma$ )

| Magnitude        | Normal prob.              | Observed frequency           |
|------------------|---------------------------|------------------------------|
| Move $> 3\sigma$ | 0.27% (1 in 370 days)     | $\sim 1$ in 50–100 days      |
| Move $> 4\sigma$ | 0.006% (1 in 16 000 days) | $\sim 1$ in 500–1000 days    |
| Move $> 5\sigma$ | 1 in 3.5 mil. days        | $\sim 1$ in 5000–10 000 days |

### Conclusion

- Returns are **leptokurtic**
- The Normal distribution dramatically underestimates tail risk

### Practical implication

- VaR under normality **underestimates** extreme losses
- Empirically  $\approx$  Student- $t$  with  $\nu \in [4, 8]$

## Case study: Empirical moments — Cross-asset comparison

- Sample statistics for daily log returns, S&P 500 (2000–2024), others (2015–2024):

| Asset   | Mean    | Std   | Skewness | Kurtosis | Min    | Max    |
|---------|---------|-------|----------|----------|--------|--------|
| S&P 500 | +0.040% | 1.00% | −0.39    | 13.4     | −12.8% | +9.4%  |
| AAPL    | +0.058% | 1.80% | −0.31    | 8.6      | −13.0% | +11.9% |
| Bitcoin | +0.195% | 3.90% | −0.15    | 7.4      | −46.5% | +22.5% |
| EUR/USD | +0.000% | 0.45% | +0.08    | 5.2      | −2.7%  | +3.1%  |

- All assets: excess kurtosis  $\gg 0 \Rightarrow$  fat tails confirmed
- Stocks and indices: negative skewness  $\Rightarrow$  crash risk dominates
- Bitcoin: highest volatility ( $\sigma = 3.9\%$  daily  $\approx 62\%$  annualised)
- EUR/USD: mean  $\approx 0$ , nearly symmetric, but heavy tails

## Homework: Seminar 1

### Analysis of a Romanian stock (BSE)

- Choose a stock listed on the Bucharest Stock Exchange (e.g.: TLV, SNP, BRD, FP, SNG). Period: min. 3 years.

### Computational requirements

1. Daily log returns (chart)
2. Descriptive statistics:  $\hat{\mu}$ ,  $\hat{\sigma}$ ,  $S$ ,  $K$
3. Annualised volatility ( $\times \sqrt{252}$ )
4. **ADF test** on log-prices and on returns (report statistic,  $p$ -value, conclusion)
5. **Jarque-Bera test**: compute  $JB$ , report  $p$ -value, interpret

### Interpretation requirements

- Are the returns stationary? (ADF)
- Are they normally distributed? (JB)
- Compare  $\sigma$  with S&P 500 ( $\approx 16\%$  annualised)
- Compute variance drag — is it significant?

### Deadline

- Next seminar
- Python code + charts + 1 page interpretation

## What you should know after Seminar 1

### Competence checklist

- ✓ **Simple vs. log difference:** simple are multiplicative, log are temporally additive
- ✓ **Return compounding:**  $1 + R_{\text{total}} = \prod (1 + R_t)$  vs.  $r_{\text{total}} = \sum r_t$
- ✓ **Variance drag:**  $\mathbb{E}[r] \approx \mathbb{E}[R] - \sigma^2/2$ ; volatility destroys wealth
- ✓ **Stationarity rationale:**  $P_t \sim I(1)$ ,  $r_t \sim I(0)$ ; statistical methods require stationarity
- ✓ **Python implementation:** data download, compute  $r_t$ , price vs. returns charts

### Cognitive level achieved

- |                                                                  |                                                         |
|------------------------------------------------------------------|---------------------------------------------------------|
| <input type="checkbox"/> Operational calculation (manual + code) | <input type="checkbox"/> Mathematics–finance connection |
| <input type="checkbox"/> Economic interpretation                 | <input type="checkbox"/> Beginning of critical thinking |

## Seminar structure — Chapter 1 (4 weeks)

| Sem. | Topic                                      | Content                                                                   | Homework                             |
|------|--------------------------------------------|---------------------------------------------------------------------------|--------------------------------------|
| 1    | Returns and stationarity<br>(this seminar) | Compute $R_t$ , $r_t$ ; variance drag;<br>$I(1)$ vs $I(0)$ ; Python intro | Analyse BSE<br>stock: returns        |
| 2    | Volatility                                 | CC, rolling vol, $\sqrt{T}$ ;<br>EWMA; 20 vs 60 day comparison            | 30-day volatility<br>for BSE stock   |
| 3    | Stylised facts                             | Skew, kurtosis, QQ-plot;<br>$ACF( r_t )$ ; normality                      | Jarque-Bera test;<br>normality check |
| 4    | Performance and risk                       | Sharpe, MDD, Sortino;<br>complete mini-project                            | Mini-project:<br>full analysis       |

## Quick quiz — Seminar 1 (15 minutes)

## Answer on paper (no notes)

1. Write the **3 conditions** for weak stationarity (covariance stationarity).
2. Why is  $ACF(r_t) \approx 0$  but  $ACF(r_t^2) \neq 0$ ? What model does this motivate?
3. An asset has  $\mathbb{E}[R] = 10\%$ ,  $\sigma = 40\%$ . Compute the variance drag and wealth after 20 years ( $W_0 = 1$ ). Interpret.
4. An  $R^2 = 0.92$  between the price of stock A and the price of stock B necessarily indicates a real relationship? Argue.
5. For a sample with  $n = 1000$ ,  $S = -0.5$ ,  $K = 8$ , compute the Jarque-Bera statistic and conclude.

## Quick quiz — Answers

## Answers

1.  $\mathbb{E}[X_t] = \mu$ ;  $\text{Var}(X_t) = \sigma^2 < \infty$ ;  $\text{Cov}(X_t, X_{t+h}) = \gamma(h)$  — the first two moments are constant, autocovariance depends only on the lag  $h$ .
2. Returns are uncorrelated (EMH) but **not independent**: their magnitude is autocorrelated.  $\text{ACF}(r_t^2)$  captures **volatility clustering**  $\Rightarrow$  GARCH motivation.
3.  $\text{Drag} = \sigma^2/2 = 0.40^2/2 = 8\%/yr$ .  $\mathbb{E}[r] = 10\% - 8\% = 2\%/yr$ .  $W_{20} = e^{20 \times 0.02} = 1.49$ . The 10% arithmetic return “melts” to just 2% geometric — extreme volatility destroys wealth.
4. **No**. Both prices can be  $I(1)$  with common trends (spurious regression, Granger & Newbold 1974). Must test on returns ( $R^2$  drops dramatically) or test for cointegration.
5.  $JB = \frac{1000}{6}((-0.5)^2 + (8 - 3)^2/4) = 166.7 \times (0.25 + 6.25) = 1083.3$ .  $\chi_{2,0.01}^2 = 9.21$ .  $1083.3 \gg 9.21 \Rightarrow$  reject  $H_0$  of normality.

## References and resources

### Key references

- ▣ Franke, J., Härdle, W.K., Hafner, C.M. (2019). *Statistics of Financial Markets*. 4th Ed., Springer.
- ▣ Cont, R. (2001). Empirical Properties of Asset Returns. *Quantitative Finance*, 1(2), 223–236.
- ▣ Dickey, D.A., Fuller, W.A. (1979). Distribution of the Estimators for AR Time Series with a Unit Root. *JASA*, 74, 427–431.
- ▣ Granger, C.W.J., Newbold, P. (1974). Spurious Regressions in Econometrics. *J. of Econometrics*, 2(2), 111–120.
- ▣ Mandelbrot, B. (1963). The Variation of Certain Speculative Prices. *J. of Business*, 36(4), 394–419.

### Online resources

- ▣ **GitHub:** [https://github.com/QuantLet/Stat\\_fin\\_markets](https://github.com/QuantLet/Stat_fin_markets) — complete Python code
- ▣ **Quantlet:** <https://quantlet.com>     **Quantinar:** <https://quantinar.com>

Thank you! Questions? Seminar materials available at:

<https://danpele.github.io/SFM/>  Quantlet  Quantinar