









Statistics of Financial Markets

Seminar 2: Classical distributions and stylised facts



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

- **ACF**: Autocorrelation Function
- **AI**: Artificial Intelligence
- **AIC**: Akaike Information Criterion
- **BET**: Bucharest Exchange Trading (index)
- **CDF**: Cumulative Distribution Function
- **CLT**: Central Limit Theorem
- **EODHD**: EOD Historical Data (market-data provider)
- **JB**: Jarque–Bera (test)
- **MLE**: Maximum Likelihood Estimation
- **OMV**: Österreichische Mineralölverwaltung — Austrian Mineral Oil Administration (OMV group)
- **QQ**: Quantile–Quantile (plot)
- **SE**: Standard Error

Today's Question and Seminar Structure

- Question of the seminar
  - **how far are real daily returns from the Normal distribution, and how do we measure it?**
- Exercises: in the order of the Exercise Map (next slide)
- Outputs to produce
  - the answers to the A exercises; the table or figure each B exercise asks for
  - a short argued answer to C1; 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 seminar is for practice; the solutions of the [Proposed] exercises are discussed in class
- 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

Task	Question	Type	Model
A1, A2	how likely is a large daily loss under the Normal distribution?	Solved, Proposed	A1
A3, A4	what is the price after one and ten years under lognormal prices?	Solved, Proposed	A3
A5, A6	when is a sum approximately Normal (CLT)?	Solved, Proposed	A5
A7, A8	the Jarque–Bera statistic and the moments of a Student-t	Solved, Proposed	A7
B1, B2	how non-Normal are daily returns, and how precise is the kurtosis?	Solved, Proposed	B1
B3, B4	does a Student-t fit better? how many 4-sigma days?	Solved, Proposed	B3
B5–B7	aggregational Gaussianity, volatility clustering, leverage and asymmetry	Solved, Proposed	B5
C1, C2	are Bitcoin's tails getting thinner? what is wrong in an AI answer?	Proposed	B3, B5

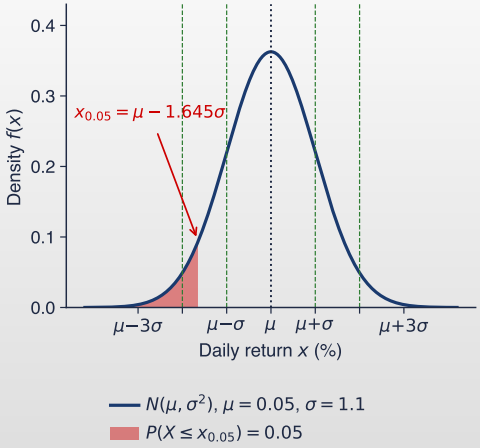
- **[Solved]**: full solution in the slides and in the notebook, a model to follow; **[Proposed]**: you solve it, following the model
- **A**: calculations on paper; **B**: real data, in the notebook; **C**: open questions and the critique of an AI answer

Data Used

Series	Source	Price used	Period
S&P 500, DAX, BET	EODHD	close	2010–2026
Bitcoin	EODHD	close, 7 days a week	2014–2026
OMV Petrom (SNP)	EODHD	adjusted close	2010–2026

- Daily log returns in %: $r_t = 100\,(\ln P_t - \ln P_{t-1})$
- Weekends and repeated holiday closes are dropped (except Bitcoin); each series on its own calendar
- In the notebook: `returns('sp500')`, `moments(r)`, `t_fit(r)`; no account or key is needed

Prerequisites for Today (1/15): the Normal Distribution



- $X \sim N(\mu, \sigma^2)$ has the density

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\Big(-\frac{(x-\mu)^2}{2\sigma^2}\Big)$$

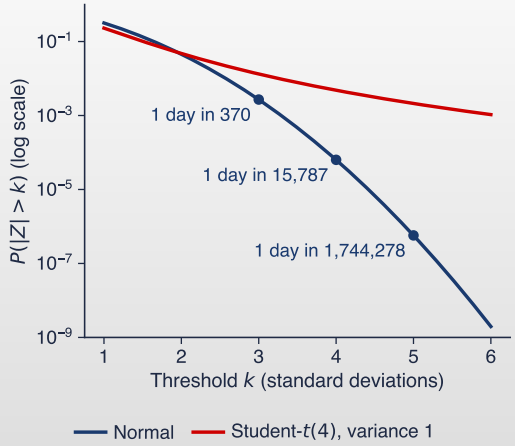
- μ: the mean (centre of the bell); σ^2: the variance (its width); $\exp(x) = e^x$
  - the area under f between two points is a probability
- Standardise: the distance from the mean in standard deviations

$$Z = \frac{X-\mu}{\sigma} \sim N(0,1), \qquad P(X \le x) = \Phi\Big(\frac{x-\mu}{\sigma}\Big)$$

- Φ: the CDF (cumulative distribution function) of $N(0,1)$
  - quantile of order α: $x_\alpha = \mu + \sigma z_\alpha$, with $z_\alpha = \Phi^{-1}(\alpha)$ and $P(X \le x_\alpha) = \alpha$
  - $z_{0.05} = -1.645$, $z_{0.025} = -1.96$, $z_{0.01} = -2.326$; red area in the chart: $\alpha = 0.05$

SFM_ch2_seminar

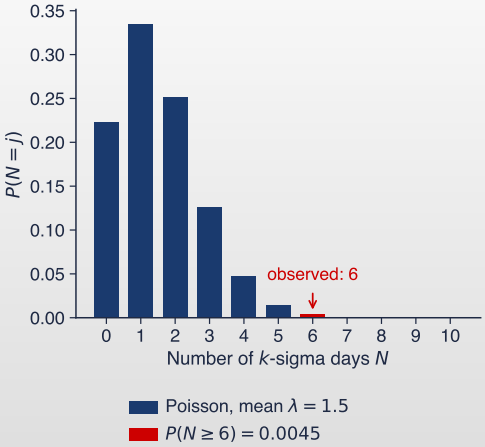
Prerequisites for Today (2/15): k-Sigma Days



- A **k-sigma day**: a return more than k standard deviations from the mean
$$|r_t - \bar{r}| > k\,s$$
- r_t: the log return of day t; $\bar{r}$, s: the sample mean and standard deviation
- Under the Normal distribution: probability and **expected number** in n days
$$P(|Z| > k) = 2\big(1 - \Phi(k)\big), \qquad E[N] = n\,P(|Z| > k)$$
- N: the number of k-sigma days among the n days
  - $P(|Z| > 4) = 6.3 \times 10^{-5}$: about one day in 15 787, once in 63 years of trading (252 days a year)
  - a heavy-tailed $t(4)$ with the same variance makes 4-sigma days about 30 times more frequent

SFM_ch2_seminar

Prerequisites for Today (3/15): the Poisson Count and Its p-value



- Rare events, each day an independent trial: N is approximately **Poisson**

$$P(N = j) = e^{-\lambda}\,\frac{\lambda^j}{j!}, \qquad j = 0, 1, 2, \ldots$$

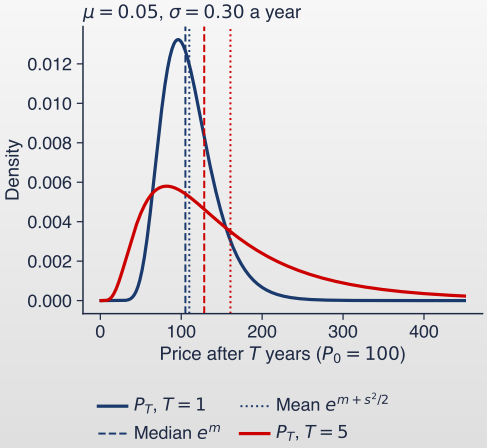
- $\lambda = n\,P(|Z| > k)$: the expected number under the Normal model
  - $j! = 1 \cdot 2 \cdots j$; mean and variance of N are both λ
- **Poisson p-value** of an observed count $N_{\mathbf{obs}}$

$$p = P(N \geq N_{\mathbf{obs}}) = 1 - \sum_{j=0}^{N_{\mathbf{obs}}-1} e^{-\lambda}\,\frac{\lambda^j}{j!}$$

- the probability of at least the observed count if the Normal model were true
  - a tiny p: the model, not luck, explains the count; in the chart, $\lambda = 1.5$ and 6 observed days give $p = 0.0045$

SFM_ch2_seminar

Prerequisites for Today (4/15): Lognormal Prices



- Y is **lognormal**, $Y \sim LN(m, s^2)$, if $\ln Y \sim N(m, s^2)$

$$\text{median} = e^m, \qquad E[Y] = e^{m+s^2/2}$$

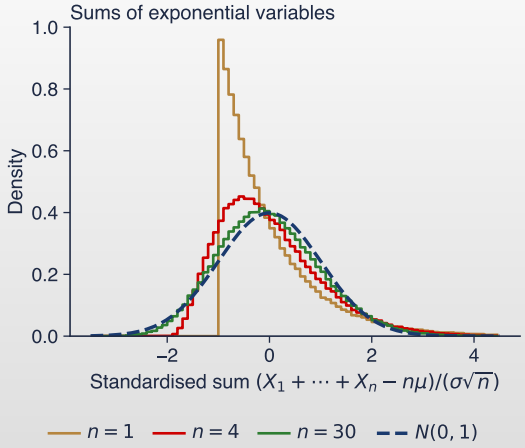
- m, s^2: the mean and the variance of $\ln Y$, not of Y
  - $P(Y \le y) = \Phi\big((\ln y - m)/s\big)$, because $Y \le y$ exactly when $\ln Y \le \ln y$
  - the mean exceeds the median: the long right tail pulls it up
- Yearly log returns i.i.d. $N(\mu, \sigma^2)$; price after T years

$$P_T = P_0\, e^{r_1 + \cdots + r_T}, \qquad \ln \frac{P_T}{P_0} \sim N(T\mu,\ T\sigma^2)$$

- i.i.d.: independent and identically distributed; P_0: today's price; T: the horizon in years
  - so $m = \ln P_0 + T\mu$ and $s = \sigma\sqrt{T}$: the spread grows with the horizon (red curve, $T = 5$)

SFM_ch2_seminar

Prerequisites for Today (5/15): the Central Limit Theorem



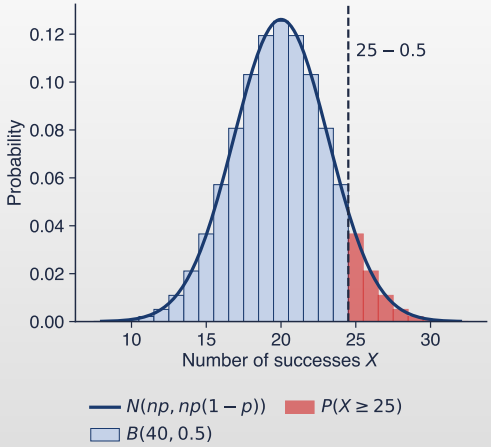
- **CLT** (central limit theorem): for $X_1, \ldots, X_n$ i.i.d. with mean μ and finite variance σ^2

$$\frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} \approx N(0, 1)$$

- equivalently, the sum is approximately $N(n\mu,\ n\sigma^2)$
  - the approximation improves as n grows
- Three conditions
  - independence; identical distribution; finite variance
  - with infinite variance (for example Student-t with $\nu \leq 2$) the limit is not Normal (Chapter 3)
- In the chart: sums of very skewed exponential variables
  - $n = 1$ is far from the bell; with $n = 30$ the sum is already close to $N(0, 1)$

SFM_ch2_seminar

Prerequisites for Today (6/15): the Binomial and Its Normal Approximation



- $X \sim B(n, p)$: the number of successes in n independent trials with probability p

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \qquad k = 0, 1, \ldots, n$$

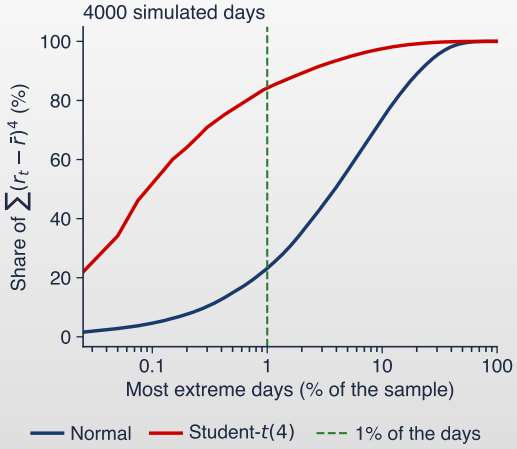
- $\binom{n}{k} = \frac{n!}{k!\,(n-k)!}$: the number of ways to choose the k successes
  - mean $E[X] = np$, variance $\mathrm{Var}(X) = np(1-p)$
  - X is a sum of n independent 0/1 variables, so the CLT applies
- Normal approximation with the **continuity correction**

$$P(X \geq k) \approx 1 - \Phi\Big(\frac{k - 0.5 - np}{\sqrt{np(1-p)}}\Big)$$

- the integer k covers the interval $[k - 0.5, k + 0.5]$ (dashed line in the chart)
  - good when $np(1-p)$ is large (rule of thumb: above 10)

SFM_ch2_seminar

Prerequisites for Today (7/15): Skewness and Excess Kurtosis



- Central sample moments and the two shape measures

$$m_k = \frac{1}{n}\sum_{t=1}^{n}(r_t - \bar{r})^k, \qquad S = \frac{m_3}{m_2^{3/2}}, \qquad K = \frac{m_4}{m_2^2} - 3$$

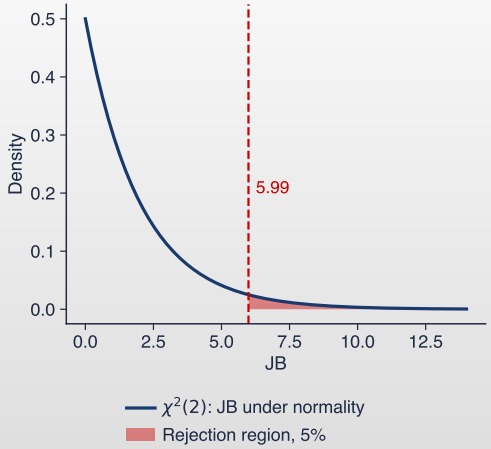
- m_2: the variance; $S < 0$: a longer left tail; $K > 0$: heavier tails than the Normal distribution
  - for the Normal distribution $S = K = 0$; the kurtosis itself is $m_4/m_2^2 = K + 3$
- Standard errors under normality and the z statistics

$$\mathrm{SE}(S) = \sqrt{6/n}, \qquad \mathrm{SE}(K) = \sqrt{24/n}$$

- $z_S = S/\mathrm{SE}(S)$, $z_K = K/\mathrm{SE}(K)$: each approximately $N(0,1)$ under normality
- In the chart: the share of $\sum(r_t - \bar{r})^4$ due to the most extreme days
  - heavy tails: 1% of the days give most of m_4, so K is fragile

SFM_ch2_seminar

Prerequisites for Today (8/15): the Jarque–Bera Test



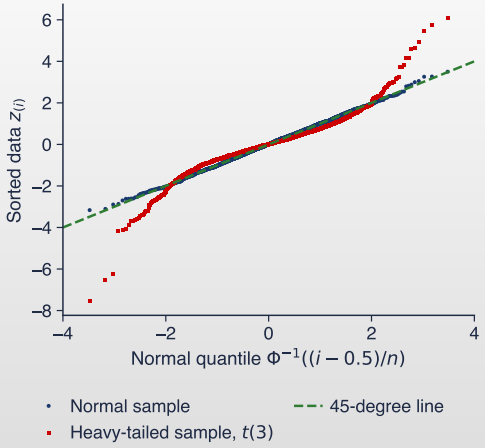
- **Jarque–Bera** Jarque & Bera (1987): one statistic for both shape measures

$$\text{JB} = \frac{n}{6}\left(S^2 + \frac{K^2}{4}\right) = z_S^2 + z_K^2$$

- zero for a Normal shape, large when S or K is far from 0
  - H_0: the data come from a Normal distribution
- Reference distribution: under H_0, JB $\approx \chi^2(2)$
  - $\chi^2(2)$: the chi-squared distribution with 2 degrees of freedom, the sum of two squared $N(0,1)$
  - reject normality at 5% if JB > 5.99 (red region); p-value $= P(\chi^2(2) > \text{JB}) = e^{-\text{JB}/2}$
- With thousands of days, even small departures give a huge JB

SFM_ch2_seminar

Prerequisites for Today (9/15): the QQ Plot



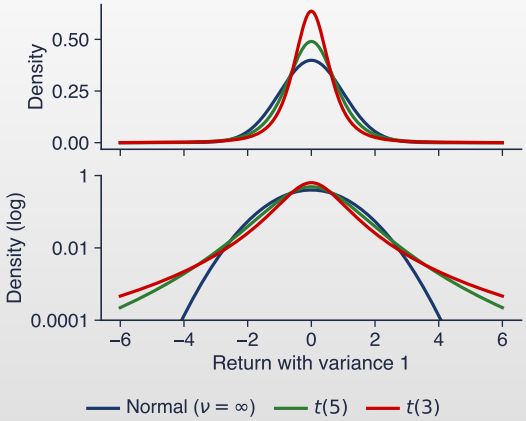
- **QQ plot**: the sorted data against the quantiles of a model

$$\left(F^{-1}\left(\frac{i-0.5}{n}\right),\ z_{(i)}\right), \qquad i = 1, \ldots, n$$

- $z_{(i)}$: the i-th smallest standardised return, $z_t = (r_t - \bar{r})/s$
  - F^{-1}: the quantile function of the model; for $N(0,1)$, $F^{-1} = \Phi^{-1}$
  - $(i-0.5)/n$: the share of the data below the i-th point
- How to read it
  - points on the 45-degree line: the model fits (blue)
  - an S-shape, low on the left and high on the right: heavier tails than the model (red)

SFM_ch2_seminar

Prerequisites for Today (10/15): the Student-t Distribution



- **Student-$t(\nu)$** Student (1908): a symmetric density with power-law tails

$$f(x) \propto \left(1 + \frac{x^2}{\nu}\right)^{-(\nu+1)/2}$$

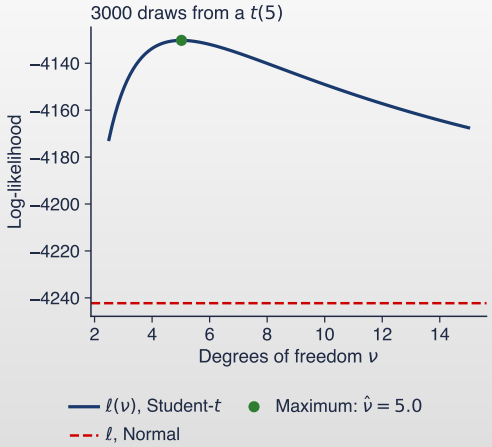
- $\nu > 0$: the degrees of freedom; smaller ν, heavier tails; $\nu \to \infty$: the Normal distribution
  - $\propto$: proportional to (a constant makes the area equal to 1)
- Moments of $T \sim t(\nu)$; they are infinite below the bounds

$$\text{Var}(T) = \frac{\nu}{\nu - 2} \ (\nu > 2), \qquad K = \frac{6}{\nu - 4} \ (\nu > 4)$$

- For returns: $r = m + s\,T$, with location m and scale s
  - standard deviation $s\sqrt{\nu/(\nu-2)}$; the chart rescales each t to variance 1
  - a taller centre and heavier tails than the Normal (log scale, bottom)

SFM_ch2_seminar

Prerequisites for Today (11/15): Maximum Likelihood and AIC



- **MLE** (maximum likelihood estimation): the parameters that make the data most likely

$$\ell(\theta) = \sum_{t=1}^{n} \ln f(r_t; \theta), \qquad \hat{\theta} = \arg\max_{\theta} \ell(\theta)$$

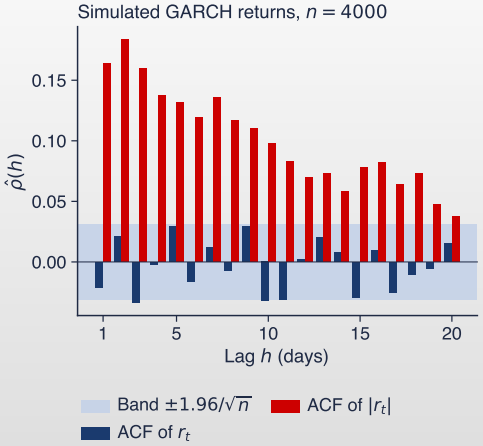
- θ: the parameters, for the t: (ν, m, s); f: the density of the model
  - ℓ: the log-likelihood; a hat marks an estimate, for example $\hat{\nu}$
- **AIC** (Akaike information criterion): fit penalised by the number of parameters

$$\text{AIC} = 2k - 2\ell(\hat{\theta})$$

- k: the number of parameters (2 for the Normal, 3 for the t); the lower AIC is preferred
  - in the chart: $\ell(\nu)$ peaks near the true $\nu = 5$, far above the Normal ℓ

SFM_ch2_seminar

Prerequisites for Today (12/15): the ACF and the Ljung–Box Test



- **ACF** (autocorrelation function): the correlation of returns h days apart

$$\hat{\rho}(h) = \frac{\sum_{t=h+1}^{n}(r_t - \bar{r})(r_{t-h} - \bar{r})}{\sum_{t=1}^{n}(r_t - \bar{r})^2}$$

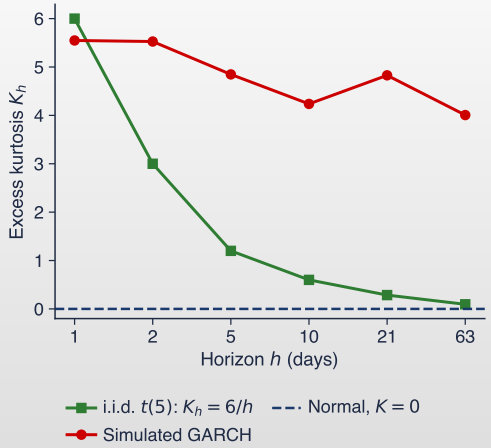
- h: the lag; 95% band under independence: $\pm 1.96/\sqrt{n}$
  - the same formula with $|r_t|$ in place of r_t measures the dependence of the size of moves
- **Ljung–Box** Ljung & Box (1978): the first m lags tested together

$$Q(m) = n(n+2)\sum_{h=1}^{m}\frac{\hat{\rho}(h)^2}{n-h} \approx \chi^2(m) \text{ under } H_0$$

- H_0: no autocorrelation up to lag m; reject at 5% if $Q(10) > \chi^2_{0.95}(10) = 18.31$
  - in the chart: r_t almost uncorrelated, $|r_t|$ clearly autocorrelated: **volatility clustering**

SFM_ch2_seminar

Prerequisites for Today (13/15): Aggregational Gaussianity



- h-**day returns**: sums of h daily log returns, on non-overlapping blocks

$$r_j^{(h)} = \sum_{t=(j-\mathbf{1})h+\mathbf{1}}^{jh} r_t, \qquad j = 1, \ldots, \lfloor n/h \rfloor$$

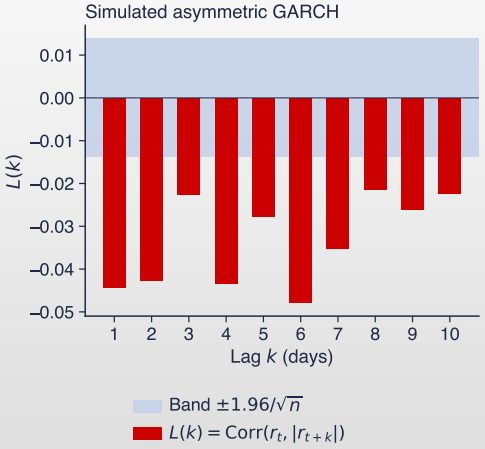
- $\lfloor n/h \rfloor$: the number of complete blocks; fewer data as h grows
- For i.i.d. daily returns, shape measures shrink with h

$$S_h = \frac{S}{\sqrt{h}}, \qquad K_h = \frac{K}{h}$$

- **Aggregational Gaussianity**: longer-horizon returns are closer to the Normal distribution
  - volatility clustering slows the decline (red curve in the chart)
  - with few blocks, $\mathrm{SE}(K_h) = \sqrt{24/\lfloor n/h \rfloor}$ is large and tests lose power

SFM_ch2_seminar

Prerequisites for Today (14/15): Leverage and Gain/Loss Asymmetry



- **Leverage effect**: a fall today is followed by larger moves

$$L(k) = \mathrm{Corr}\big(r_t,\ |r_{t+k}|\big) < 0$$

- k: the number of days ahead; a negative $L(k)$ outside the band $\pm 1.96/\sqrt{n}$
  - for a firm, a fall raises its debt relative to its equity, hence its risk
- **Gain/loss asymmetry**: large losses bigger and more frequent than large gains
  - compare $|q_{\mathbf{0.01}}|$ with $q_{\mathbf{0.99}}$; q_p: the sample quantile of order p
- **Binomial test**: among $N = D + U$ extreme days, D falls and U rises

$$H_{\mathbf{0}}:\ D \sim B(N, 0.5), \qquad p = P\big(|D - N/2| \geq |D_{\mathbf{obs}} - N/2|\big)$$

- A small p: falls and rises are not equally likely

SFM_ch2_seminar

Prerequisites for Today (15/15): Stylised Facts, Bootstrap and Rank Correlation

- **Stylised facts** Cont (2001): statistical properties shared by most assets and periods
  - heavy tails; no linear autocorrelation; volatility clustering; aggregational Gaussianity; leverage; gain/loss asymmetry
- **Bootstrap**: the precision of a statistic $\hat{\theta}$ (for example S or K) without a formula
  - draw n days with replacement from the sample and recompute $\hat{\theta}_b^*$; repeat for $b = 1, \ldots, B$
  - 95% interval: the 2.5% and 97.5% percentiles of $\hat{\theta}_{\mathbf{1}}^*, \ldots, \hat{\theta}_B^*$; their standard deviation is the bootstrap SE
  - unlike $\sqrt{24/n}$, it does not assume Normal data
- **Spearman rank correlation** of n pairs (x_i, y_i), for example (year, $\hat{\nu}$)

$$\hat{\rho}_S = 1 - \frac{6 \sum_{i=\mathbf{1}}^{n} d_i^{\mathbf{2}}}{n(n^{\mathbf{2}} - 1)}$$

- d_i: the difference between the rank of x_i and the rank of y_i (rank 1 = the smallest value)
  - $\hat{\rho}_S \in [-1, 1]$; $+1$: a monotone increasing link; its p-value tests $H_{\mathbf{0}}$: no monotone link
  - with few points (a few years), only a very strong trend is significant

A1: A Large Daily Loss under the Normal Distribution [Solved]

Task

- Daily log returns of an index are $N(\mu, \sigma^2)$ with $\mu = 0.05\%$ and $\sigma = 1.1\%$; there are 252 trading days a year.
- 1. Compute the probability of a daily loss larger than 2% and the expected number of such days per year.
- 2. Compute the 1% quantile of the daily return.
- 3. Compute the expected number of 4-sigma days in 10 years.
- Report: a probability, a number of days, a quantile and an expected count.

Solution

- 1. $z = (-2 - 0.05)/1.1 = -1.86$; $P = \Phi(-1.86) = 3.12\%$; $252 \times 3.12\% = 7.9$ days a year
- 2. $x_{0.01} = 0.05 + 1.1 \times (-2.326) = -2.51\%$
- 3. $P(|Z| > 4) = 6.3 \times 10^{-5}$; in 2 520 days: 0.16 days, so most decades would see none
- Lecture 2 shows that real indices have one such day every few months.

Back Exercise list

A2: The Same Questions for Bitcoin [Proposed]

Solution discussed in class

Task

- Daily log returns of Bitcoin are $N(0.12\%, 3.6\%^2)$; Bitcoin trades 365 days a year. Model: A1.
- 1. Compute the probability of a daily loss larger than 10% and the expected number of such days per year.
- 2. Compute the 1% quantile of the daily return.
- 3. Compute the expected number of 4-sigma days in 10 years.
- Report: a probability, a number of days, a quantile and an expected count.

Back Exercise list

A3: Lognormal Prices after One and Ten Years [Solved]

Task

- A stock costs 100 lei; its yearly log returns are i.i.d. $N(0.07, 0.18^2)$.
- 1. Compute the median and the mean of the price after one year.
- 2. Compute the probability that the price after one year is below 100 lei.
- 3. Compute the 5% quantile of the price after one year.
- 4. Repeat 1–3 for ten years.
- Report: eight numbers and one sentence on the horizon.

Solution

- 1. Median $100e^{\mathbf{0.07}} = 107.25$; mean $100e^{\mathbf{0.07}+\mathbf{0.0162}} = 109.00$
- 2. $P = \Phi(-0.07/0.18) = \Phi(-0.389) = 34.9\%$
- 3. $100e^{\mathbf{0.07}+\mathbf{0.18}\times(\mathbf{-1.645})} = 79.77$
- 4. $T = 10$: $10\mu = 0.7$, $\sqrt{10}\sigma = 0.569$; median 201.38, mean 236.79, $P(\text{loss}) = 10.9\%$, 5% quantile 78.96
- A loss becomes less likely with time, but the 5% quantile does not improve: 79.77 after one year, 78.96 after ten.

Back Exercise list

A4: Derivation of the Mean $e^{m+s^2/2}$ [Proposed]

Solution discussed in class

Task

- Let $X \sim N(m, s^2)$ and $Y = e^X$. Model: A3; exercise of the type in Borak, Härdle & López-Cabrera (2013).
- 1. Write $E[e^X] = \int e^x f(x)\,dx$ and complete the square in the exponent to show that $E[e^X] = e^{m+s^2/2}$.
- 2. Use the same idea for $E[e^{2X}]$ and derive $\mathrm{Var}(Y) = (e^{s^2} - 1)e^{2m+s^2}$.
- 3. For $m = 0.07$, $s = 0.18$, compute the mean and the median of Y and their gap.
- Report: the two derivations and one sentence linking the gap to the volatility drag of Chapter 1.

Back Exercise list

A5: Up Days in a Year (Normal Approximation) [Solved]

Task

- An index rises on a given day with probability $p = 0.53$, independently across days; a year has 252 trading days.
- 1. Give the distribution of the number X of up days, its mean and its standard deviation.
- 2. Approximate $P(X \geq 140)$ with the Normal distribution and the continuity correction.
- 3. Compare with the exact binomial value from the notebook.
- Report: two moments, two probabilities, one sentence on the quality of the approximation.

Solution

- 1. $X \sim B(252, 0.53)$; $E[X] = 133.56$, $\text{sd} = \sqrt{252 \times 0.53 \times 0.47} = 7.92$
- 2. $z = (139.5 - 133.56)/7.92 = 0.750$; $P \approx 1 - \Phi(0.750) = 22.67\%$
- 3. Exact: 22.69%
- $np(1-p) \approx 63$ is large, so the CLT works very well here.

A6: Is a Monthly Return Normal? [Proposed]

Solution discussed in class

Task

- Daily log returns are i.i.d. with mean 0.04% and standard deviation 1.2%. Model: A5.
- 1. Give the approximate distribution of the 21-day log return and its two parameters.
- 2. Compute the probability of a 21-day loss larger than 10%, and of a 5-day loss larger than 10%.
- 3. Say whether the CLT still applies if the daily returns are Student-t with $\nu = 3$.
- 4. Say whether it applies with $\nu = 1.5$.
- Report: two distributions, two probabilities, two answers with a reason each.

Back Exercise list

A7: Jarque–Bera on Paper [Solved]

Task

- A series of $n = 4000$ daily returns has sample skewness $S = -0.6$ and excess kurtosis $K = 12$.
- 1. Compute the standard errors of S and K under normality and the two z statistics.
- 2. Compute JB and the share of it that comes from the kurtosis.
- 3. Decide at the 5% level.
- Report: two standard errors, two z values, JB, the decision.

Solution

- 1. $\sqrt{6/4000} = 0.0387$, $\sqrt{24/4000} = 0.0775$; $z_S = -15.5$, $z_K = 154.9$
- 2. $\mathrm{JB} = \frac{4000}{6}(0.36 + 36) = 240 + 24\,000 = 24\,240$, almost all from K
- 3. $\mathrm{JB} > 5.99$: reject normality
- JB is the sum of the two squared z statistics: $z_S^2 + z_K^2$.

Back Exercise list

A8: Moments of the Student-t [Proposed]

Solution discussed in class

Task

- $T \sim t(\nu)$; a model for daily returns is $r = m + s\,T$. Model: A7.
- 1. Compute the variance and the excess kurtosis of $t(5)$ and $t(10)$.
- 2. Find ν such that the excess kurtosis is 3.
- 3. Find the scale s that gives a daily standard deviation of 1.2% for $\nu = 5$ and for $\nu = 6$.
- 4. Compare $P(|X| > 4)$ for a unit-variance $t(5)$ (value in the notebook) with the Normal value.
- Report: four moments, one ν, two scales and one ratio.

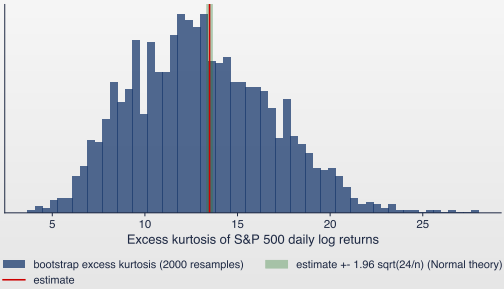
Back Exercise list

B1: How Non-Normal Is the S&P 500? [Solved]

- **Question**: how precisely can we measure the distance of the daily log returns of the S&P 500 from the Normal distribution?
- **Inputs**: S&P 500 closes, 2010–2026, daily log returns in %
- **Tasks**
  1. Compute the mean, standard deviation, skewness, excess kurtosis and the worst day with its date.
  2. Compute the Normal-theory standard errors of S and K, and the JB statistic.
  3. Bootstrap S and K with 2000 resamples of the days and take the 2.5% and 97.5% percentiles.
  4. Interpretation: why is the bootstrap interval of K so much wider than the Normal-theory one?
- **Report**: a table of moments, JB, two intervals for K, one sentence
- Notebook: section B1

Back Exercise list

B1: Solution [Solved]



- $n = 4\,201$; mean 0.045%, s.d. 1.08%; worst day -12.8% on 16 March 2020
- $S = -0.61$ (SE 0.038), $K = 13.49$ (SE 0.076); JB $= 32\,113$: normality rejected
- Normal theory: $K \in [13.34, 13.64]$; bootstrap: $[6.7, 20.6]$; for S, bootstrap $[-1.54, 0.29]$
- Interpretation: $\sqrt{24/n}$ assumes Normal data; with heavy tails K depends on a few days, and whether they are resampled moves it a lot; the bootstrap s.d. of K is 50 times larger
- The skewness interval contains 0: the sign of S is not certain

SFM_ch2_seminar

B2: Three More Series [Proposed]

- **Question**: which of the BET, Bitcoin and OMV Petrom is farthest from the Normal distribution, with and without its most extreme day?
- **Inputs**: daily log returns, 2010–2026 (Bitcoin from 2014), each on its own calendar; model: B1
- **Tasks**
  1. Compute n, mean, standard deviation, S, K and JB for each series.
  2. Remove the single largest absolute return of each series and recompute K.
  3. Interpretation: what does the change in K tell you about kurtosis as a summary of risk?
- **Report**: one table and two sentences
- Notebook: section B2

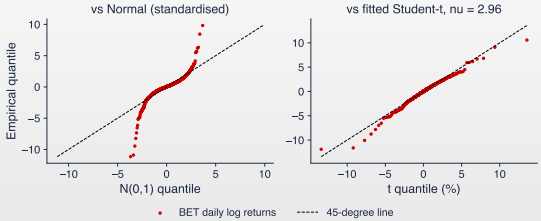
Back Exercise list

B3: Normal or Student-t for the BET? [Solved]

- **Question**: does a Student-t describe the daily returns of the BET better than the Normal distribution?
- **Inputs**: BET closes, 2010–2026, daily log returns in %
- **Tasks**
  1. Draw the QQ plot of the standardised returns against $N(0, 1)$.
  2. Fit a Student-t by maximum likelihood and draw the QQ plot against the fitted t.
  3. Compare the log-likelihoods and the AIC of the two models.
  4. Interpretation: where does the Normal distribution fail?
  5. Interpretation: where does the fitted t fail?
- **Report**: $\hat{\nu}$, $\hat{m}$, $\hat{s}$, two AIC values, two QQ plots, one sentence
- Notebook: section B3

Back Exercise list

B3: Solution [Solved]



- $\hat{\nu} = 2.96$, $\hat{m} = 0.078\%$, $\hat{s} = 0.628\%$; implied s.d. $\hat{s}\sqrt{\hat{\nu}/(\hat{\nu}-2)} = 1.10\%$
- $\ell_t = -5496$ vs $\ell_N = -6224$; $\text{AIC}_N - \text{AIC}_t = 1454$: the t wins clearly
- Lowest return -11.9%; matching quantile: Normal distribution -3.9%, fitted t -13.4%
- Interpretation: the Normal distribution misses both tails (S-shape); the t fits the centre and most of the tails, but its extreme quantiles go slightly beyond the data

SFM_ch2_seminar

B4: How Many 4-Sigma Days? [Proposed]

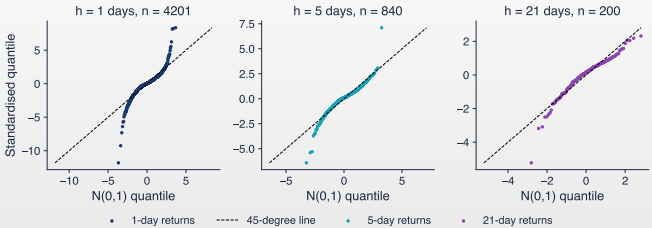
- **Question**: which model predicts the number of 4-sigma days of the S&P 500, DAX, BET and Bitcoin?
- **Inputs**: daily log returns, 2010–2026 (Bitcoin from 2014); model: B3
- **Tasks**
  1. Count the days with $|r_t - \bar{r}| > 4s$ in each series.
  2. Compute the expected count under the Normal distribution and the Poisson probability of seeing at least the observed count.
  3. Compute the expected count under the fitted Student-t (same threshold $\bar{r} \pm 4s$).
  4. Interpretation: are the tails of the Student-t too heavy, too light or about right for each series?
- **Report**: one table and one sentence per series
- Notebook: section B4

Back Exercise list

B5: Does the S&P 500 Become Normal at Longer Horizons? [Solved]

- **Question**: are 5-day and 21-day returns of the S&P 500 closer to the Normal distribution than daily returns?
- **Inputs**: S&P 500 daily log returns, 2010–2026; h-day returns are sums over non-overlapping blocks of h days
- **Tasks**
  1. Build the 5-day and 21-day log returns.
  2. Compute n, S, K, the standard error $\sqrt{24/n}$ and JB with its p-value at $h = 1, 5, 21$.
  3. Draw the three QQ plots against $N(0, 1)$.
  4. Interpretation: is the monthly distribution Normal?
  5. Interpretation: why does the test have less power at $h = 21$?
- **Report**: a table with three rows, three QQ plots, two sentences
- Notebook: section B5

B5: Solution [Solved]



h	n	S	K	$\sqrt{24/n}$	JB (p-value)
1	4201	−0.61	13.49	0.08	32 113 (≈ 0)
5	840	−0.62	7.06	0.17	1 797 (≈ 0)
21	200	−1.18	3.68	0.35	159 (2.9×10^{-35})

- Interpretation: K falls from 13.49 to 3.68, but monthly returns are still not Normal (skewness −1.18, March 2020); with only 200 months, the standard error of K is 0.35, so smaller departures would go undetected

SFM_ch2_seminar

B6: Are BET Returns Independent? [Proposed]

- **Question**: are the daily returns of the BET independent, not only uncorrelated?
- **Inputs**: BET daily log returns, 2010–2026; model: B5
- **Tasks**
  1. Compute the ACF of r_t and of $|r_t|$ at lags 1–10 and the band $\pm 1.96/\sqrt{n}$.
  2. Compute the Ljung–Box statistic $Q(10)$ for r_t and for $|r_t|$ and compare with $\chi^2_{0.95}(10)$.
  3. Interpretation: can returns be uncorrelated but not independent?
- **Report**: two ACF columns, two Q values, one sentence
- Notebook: section B6

B7: Is Bitcoin Like a Stock Index? [Proposed]

- **Question**: do the S&P 500 and Bitcoin both show the leverage effect and the gain/loss asymmetry?
- **Inputs**: daily log returns, S&P 500 2010–2026, Bitcoin 2014–2026; model: B5
- **Tasks**
  1. Compute $L(k)$ (Prerequisites for Today (14/15)) for $k = 1, \ldots, 5$ and their mean, with the band $\pm 1.96/\sqrt{n}$.
  2. Count the days below $\bar{r} - 4s$ and above $\bar{r} + 4s$; test whether falls and rises are equally likely (binomial test, $p = 0.5$).
  3. Compare $|q_{0.01}|$ with $q_{0.99}$ and examine the skewness.
  4. Interpretation: which stylised facts of stock indices does Bitcoin share?
- **Report**: one table and two sentences
- Notebook: section B7

C1: Are Bitcoin's Tails Getting Thinner? [Proposed]

- **Question**: has the tail of Bitcoin's daily returns become thinner since 2015, as its market matured?
- **Inputs**: Bitcoin and S&P 500 daily log returns, 2015–2026; models: B3, B5
- **Tasks**
  1. Fit a Student-t and compute the excess kurtosis for each calendar year and each series.
  2. Compute the Spearman rank correlation between the year and $\hat{\nu}$.
  3. Fit one t on 2015–2020 and one on 2021–2026 and compare $\hat{\nu}$.
  4. Propose one way to make the answer more reliable (intervals, other windows, other crypto assets).
  5. Interpretation: what can 12 yearly estimates tell us about a trend?
- **Report**: a table, the rank correlation with its p-value, a short plan for a project
- Notebook: section C1

Back Exercise list

C2: Audit an AI Answer [Proposed]

- A student asked an AI assistant: "Describe the distribution of daily S&P 500 returns in 2010–2026." The answer:
- (a) The excess kurtosis is 13.5; since the Normal distribution has kurtosis 0, the kurtosis of the S&P 500 is 13.5.
- (b) JB = 32 113 with p = 0.000, which proves that the returns follow a Student-t distribution.
- (c) A Normal 4-sigma day happens once every 15 787 days; the 28 such days in 4 201 days are just bad luck.
- (d) The lag-1 autocorrelation is only −0.11, so returns are independent and volatility cannot be predicted.
- (e) With yearly log returns N(7%, 18%^2), the expected value of 100 invested after one year is 100 exp(0.07) = 107.25.
- **Tasks**
  1. For each statement, say whether it is correct; if not, give the correct statement and the correct number from the notebook (section C2).
- **Report**: a list of five verdicts with one line of justification each.

Points to Remember

- The Normal distribution makes 4-sigma days almost impossible; real markets have them every few months
- Lognormal prices: the median is below the mean by the volatility drag
- Skewness and kurtosis are fragile: report bootstrap intervals and the extremes
- The Student-t fits much better than the Normal distribution, but not perfectly
- Returns are nearly uncorrelated but not independent: their size is predictable
- An AI answer is a draft: check every definition, model and number

After the Seminar

- Lecture 2 develops each topic of today: the Normal and lognormal models, the CLT, moments and tests, the Student-t and six stylised facts
- C1 can grow into a team project: rolling windows, bootstrap intervals, more crypto assets
- Reading: Franke, Härdle & Hafner (2019), Sec. 3.3 and Sec. 11.2; exercises with solutions: Borak, Härdle & López-Cabrera (2013)

References

- Borak, S., Härdle, W.K., López-Cabrera, B. (2013). *Statistics of Financial Markets: Exercises and Solutions*, 2nd ed. Springer.
- Cont, R. (2001). Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*, 1(2), 223–236.
- Franke, J., Härdle, W.K., Hafner, C.M. (2019). *Statistics of Financial Markets: An Introduction*, 5th ed. Springer.
- Jarque, C.M., Bera, A.K. (1987). A test for normality of observations and regression residuals. *International Statistical Review*, 55(2), 163–172.
- Ljung, G.M., Box, G.E.P. (1978). On a measure of lack of fit in time series models. *Biometrika*, 65(2), 297–303.
- Student (1908). The probable error of a mean. *Biometrika*, 6(1), 1–25.