








Statistics of Financial Markets

Chapter 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
- **ARCH**: AutoRegressive Conditional Heteroskedasticity
- **BET**: Bucharest Exchange Trading (index)
- **BVB**: Bursa de Valori București (the Bucharest Stock Exchange)
- **CDF**: Cumulative Distribution Function
- **CLT**: Central Limit Theorem
- **EGARCH**: Exponential GARCH
- **ES**: Expected Shortfall
- **GARCH**: Generalised AutoRegressive Conditional Heteroskedasticity
- **GJR-GARCH**: Glosten–Jagannathan–Runkle GARCH
- **JB**: Jarque–Bera (test)
- **LB**: Ljung–Box (test)
- **MLE**: Maximum Likelihood Estimation
- **OMV**: Österreichische Mineralölverwaltung — Austrian Mineral Oil Administration (OMV group)
- **OUG**: Ordonanță de Urgență a Guvernului (government emergency ordinance)
- **QQ**: Quantile–Quantile (plot)
- **SFE**: Statistics of Financial Markets, Exercises (Quantlet collection of the textbook)

Today's Question, Route and Learning Outcomes

- **Question**: is the Normal distribution a good model for daily returns, and what do real returns look like?
  - the answer decides how we measure risk, price options and test hypotheses
- **Route and outcomes**: after this chapter you can
  - compute probabilities and quantiles of the Normal and lognormal distributions
  - state the central limit theorem (CLT) and say when the Normal approximation works
  - estimate skewness and excess kurtosis and test normality with the Jarque–Bera test
  - read a QQ plot and fit a Student-t distribution by maximum likelihood
  - check six stylised facts of returns Cont (2001) and explain what each means for a model
- Real data, 2010–2026: S&P 500, DAX, BET, Bitcoin and four BVB stocks (SNP, BRD, SNN, SNG)

Reading and Tools

- Textbook: Franke, Härdle & Hafner (2019), *Statistics of Financial Markets*, 5th ed., Sec. 3.3 (Normal and lognormal) and Sec. 11.2 (stylised facts)
  - exercises with solutions: Borak, Härdle & López-Cabrera (2013)
- Complementary: Tsay (2010), Sec. 1.2 (distributional properties of returns); Cont (2001)
- Python Quantlets of this chapter: Quantlets/Ch_02
  - ported from the SFE Quantlets of the textbook (SFEDaxReturnDistribution, SFElognormal, SFEclt, SFENormalApprox1–4)
  - each chart in these slides links to the Quantlet that draws it
- Lecture notebook: open in Google Colab
- Video course: Tukey's g- and h- transformations

An Extreme Day under the Normal Distribution

- 16 March 2020: the S&P 500 fell by -12.8% (log return) in one day
- Daily mean and standard deviation 2010–2026: 0.045% and 1.084%, so the fall was $(-12.77 - 0.045)/1.084 = -11.8$ standard deviations
  - under the Normal distribution, the probability of such a day is about 2×10^{-32}
  - that is, less than once in 10^{29} years of trading
- Yet such days happen: 19 October 1987, October 2008, March 2020, April 2025
- Conclusion: the model, not the market, is wrong

Question for the Room: How Often Is a 4-Sigma Day?

- A k-**sigma day**: a daily return more than k standard deviations away from the mean
- **What do you think?** If daily returns were Normal, how often would a 4-sigma day occur?
- **What do you think?** How often does it occur in the S&P 500 since 2010?
- **Answer**
  - Normal distribution: once every 15 787 trading days, about once in 63 years
  - S&P 500: 28 days in 4 201 days, once every 150 days, about 1.7 per year

What a Distribution Is Used For

- **Risk**: the probability of a large loss
  - VaR 1% (Value at Risk): the loss exceeded with probability 1%; a quantile of the distribution (Chapter 10)
- **Pricing**: the Black–Scholes formula assumes lognormal prices
  - heavier tails make deep out-of-the-money options more valuable
- **Inference**: *t*-tests and confidence intervals assume (approximate) normality
  - heavy tails make standard errors unreliable in small samples
- A wrong distribution gives wrong numbers, even with perfect data

The Benchmark Model of This Chapter

- Price P_t, log return $r_t = \ln(P_t/P_{t-1})$, so $P_t = P_{t-1}e^{r_t} = P_0 e^{r_1+\cdots+r_t}$
  - notation: **i.i.d.** = independent and identically distributed
- **Benchmark**: r_t i.i.d. $N(\mu, \sigma^2)$
  - $N(\mu, \sigma^2)$: the Normal distribution with mean μ and variance σ^2 (daily mean and variance of the log returns)
  - prices follow a geometric random walk; P_t is lognormal Osborne (1959)
  - the first model of this kind, for arithmetic price changes: Bachelier (1900)
- This chapter tests each assumption of the benchmark in turn: normality, identical distribution and independence

Carl Friedrich Gauss and the Normal Distribution

- De Moivre (1733) found the bell curve as the limit of binomial probabilities
- Gauss (1809) used it for the errors of astronomical observations and derived least squares from it
- Laplace (1810) proved a first version of the central limit theorem
- Hence its other name: the **Gaussian** distribution
- In finance it became the default model of returns, because sums of Normal variables stay Normal



Carl Friedrich Gauss (1777–1855)

Portrait: Christian Albrecht Jensen (1840); public domain; Wikimedia Commons

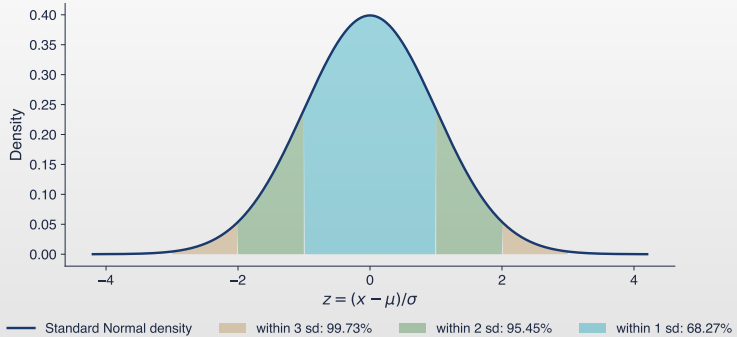
The Normal Distribution: Definition

- $X \sim N(\mu, \sigma^2)$ has the **PDF** (probability density function)
  - $f(x) = \dfrac{1}{\sigma\sqrt{2\pi}} \exp\Big(-\dfrac{(x-\mu)^2}{2\sigma^2}\Big), \quad x \in \mathbb{R}$
  - two parameters: the mean μ (location) and the variance σ^2 (spread)
- **Standard Normal distribution**: $Z = (X - \mu)/\sigma \sim N(0, 1)$
  - its **CDF** (cumulative distribution function) is $\Phi(z) = P(Z \leq z)$, with no closed form; tables or software
- Probabilities of X come from Z: $P(X \leq x) = \Phi\big((x - \mu)/\sigma\big)$
- Quantile of order α: $x_\alpha = \mu + \sigma z_\alpha$, the value with $P(X \leq x_\alpha) = \alpha$
  - $z_\alpha = \Phi^{-1}(\alpha)$: the quantile of $N(0, 1)$; $z_{0.01} = -2.33$, $z_{0.05} = -1.64$

Properties Used in Finance

- Symmetric around μ: mean = median = mode
- Skewness 0 and kurtosis 3 (defined in Section 5)
- **Closed under addition**: if $X \sim N(\mu_1, \sigma_1^2)$ and $Y \sim N(\mu_2, \sigma_2^2)$ are independent, then $X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$
  - h-day log return under the benchmark: $r_t(h) \sim N(h\mu, h\sigma^2)$
  - volatility grows with $\sqrt{h}$: the square-root-of-time rule used in Chapter 1
- Linear combinations of jointly Normal returns are Normal, so portfolio returns stay Normal
- Fully described by two numbers: once μ and σ are known, every probability is known

The 68–95–99.7 Rule



- Within 1, 2 and 3 standard deviations: 68.27%, 95.45% and 99.73% of the probability
- Outside 3 standard deviations: one day in 370; the tails of the Normal distribution fall off like $e^{-z^2/2}$, very fast

SFM_ch2_normal_lognormal

Frequency of Large Moves under the Normal Distribution

k	$P(\lvert Z\rvert > k)$	once every ... trading days	once every ... years
1	0.3173	3	0.013
2	0.0455	22	0.087
3	2.7×10^{-3}	370	1.5
4	6.3×10^{-5}	15 787	63
5	5.7×10^{-7}	1 744 278	6 922
6	2.0×10^{-9}	506 797 346	2 011 101

- $P(|Z| > k)$: the probability that a standard Normal variable lies more than k standard deviations from 0
- Years computed with 252 trading days a year
- Each extra standard deviation makes a move 7 to 291 times rarer
- A 6-sigma day should not happen in the whole history of stock markets

SFM_ch2_normal_lognormal

Worked Example: Large Losses of the S&P 500

- S&P 500, 2010–2026: $n = 4\,201$ daily log returns, mean 0.045%, standard deviation 1.08%
- A loss larger than 3%: $z = (-3 - 0.045)/1.08 = -2.81$
  - Normal distribution: $\Phi(-2.81) = 0.25\%$, so 10.5 days expected
  - observed: 49 days
- A loss larger than 5%: $z = -4.65$
  - Normal distribution: 6.9×10^{-3} days expected; observed: 8 days
- BET: 13 days below -5% against 5.1×10^{-3} expected
- The Normal distribution underestimates the frequency of large losses by orders of magnitude

Recap: The Normal Distribution

- Two parameters, symmetric, closed under addition, volatility scales with $\sqrt{h}$
- Its tails are thin: a 4-sigma day once every 63 years
- Real returns have many more large moves than it allows

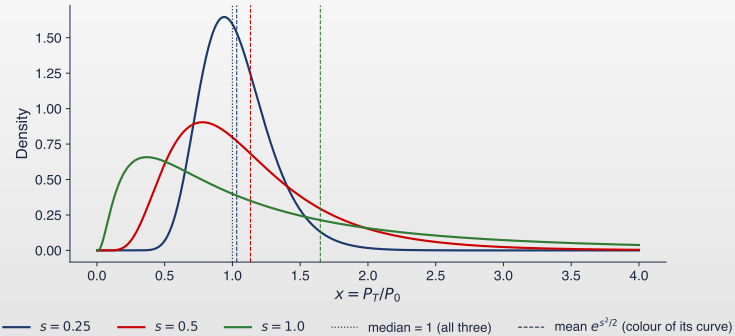
From Normal Log Returns to Lognormal Prices

- A positive variable Y is **lognormal**, $Y \sim LN(m, s^2)$, if $\ln Y \sim N(m, s^2)$
  - m, s^2: the mean and the variance of $\ln Y$, not of Y
- Density: $f(y) = \dfrac{1}{ys\sqrt{2\pi}} \exp\Big(-\dfrac{(\ln y - m)^2}{2s^2}\Big), \quad y > 0$
- Under the benchmark, $\ln(P_T/P_0) = r_1 + \cdots + r_T \sim N(T\mu, T\sigma^2)$
  - T: the horizon in days; so $P_T/P_0 \sim LN(T\mu, T\sigma^2)$ Osborne (1959)
- Why lognormal and not Normal prices?
  - prices cannot be negative; a Normal price could be
  - the simple return $R = e^r - 1 > -1$: you cannot lose more than you invested
- The Black–Scholes model is built on it

Mean, Median and Mode

- If $Y \sim LN(m, s^2)$:
- median e^m; mode e^{m-s^2}; mean $E[Y] = e^{m+s^2/2}$
  - mode: the most likely value; median: the value exceeded with probability 1/2
  - always mode $<$ median $<$ mean: a long right tail (positive skewness)
- Variance: $\mathrm{Var}(Y) = (e^{s^2} - 1)\, e^{2m+s^2}$
- The mean comes from $E[e^X] = e^{m+s^2/2}$ for $X \sim N(m, s^2)$ (Seminar 2 derives it)
  - the extra $s^2/2$ is the volatility drag of Chapter 1, seen from the other side
- Which one matters? The median is the outcome of a typical investor; the mean is pulled up by a few very good paths

Lognormal Densities (SFElognormal)



- $m = 0$: the median is 1 for all three curves; the dashed lines mark the means
- $s = 0.25$: mean 1.03, mode 0.94, almost symmetric; $s = 1$: mean 1.65, mode 0.37, skewness 6.18
- Long horizons or high volatility: the gap between mean and median widens

SFM_ch2_normal_lognormal

Worked Example: 100 Lei in the BET

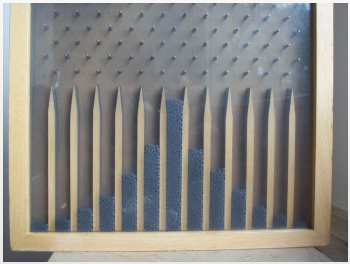
- BET, 2010–2026: annual mean log return $\mu = 11.6\%$ and volatility $\sigma = 16.9\%$ (daily moments $\times$ 250.6 and $\sqrt{250.6}$)
- After 1 year, $P_{\mathbf{1}}/P_{\mathbf{0}} \sim LN(\mu, \sigma^{\mathbf{2}})$
  - median $100e^{\mu} = 112.3$; mean $100e^{\mu+\sigma^{\mathbf{2}}/\mathbf{2}} = 113.9$
  - probability of a loss $\Phi(-\mu/\sigma) = 24.8\%$; 5% quantile $100e^{\mu-\mathbf{1.645}\sigma} = 84.9$
  - $\Phi(-\mu/\sigma) = P(r < 0)$: the probability of a negative log return; $-1.645 = z_{\mathbf{0.05}}$
- After 10 years, $P_{\mathbf{10}}/P_{\mathbf{0}} \sim LN(10\mu, 10\sigma^{\mathbf{2}})$
  - median 317.7; mean 366.8; probability of a loss 1.5%; 5% quantile 131.6
- The probability of a loss falls with the horizon; after 10 years even the 5% quantile is above 100, but only if the historical mean persists
- Caveat: these numbers inherit every weakness of the Normal model of log returns

Recap: The Lognormal Distribution

- Normal log returns $\Rightarrow$ lognormal prices: positive, right-skewed
- Median e^m below mean $e^{m+s^2/2}$: the gap is the volatility drag
- Report the median and a low quantile, not only the mean

The Galton Board

- Francis Galton (1889): balls fall through rows of pins and bounce left or right with probability 1/2
- The final position of a ball is a **sum** of many small independent shocks
- The heaps form a bell curve: binomial probabilities close to the Normal density
- The same argument was used for returns: a daily return is the sum of many small price changes during the day



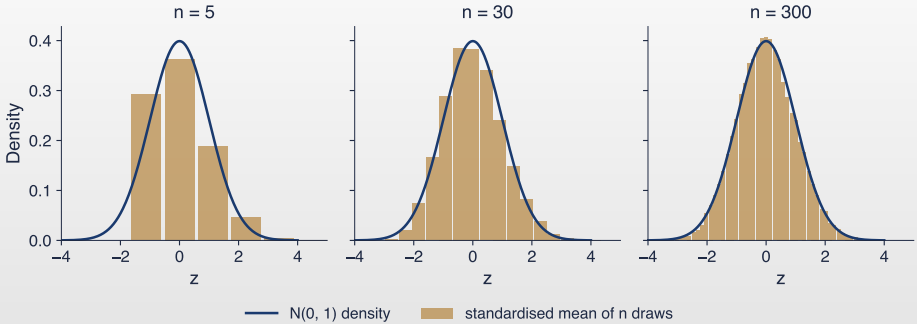
A Galton board after the balls have fallen

Photo: Klaus-Dieter Keller (2010); public domain; Wikimedia Commons

The Central Limit Theorem

- **CLT** (Lindeberg–Lévy): $X_1, \ldots, X_n$ i.i.d. with mean μ and finite variance σ^2
  - $\dfrac{\sqrt{n}\,(\bar{X}_n - \mu)}{\sigma} \xrightarrow{d} N(0,1)$ as $n \to \infty$
  - $\bar{X}_n$: the mean of the n variables; $\xrightarrow{d}$: convergence in distribution, the CDF of the left side approaches Φ
  - equivalently, the sum $X_1 + \cdots + X_n$ is approximately $N(n\mu, n\sigma^2)$
- Three conditions
  - independence; identical distribution; finite variance
- How fast? The error falls like $1/\sqrt{n}$ and grows with the skewness of X (Berry–Esseen bound)
- It explains why the Normal distribution is everywhere, and why sample means have Normal confidence intervals

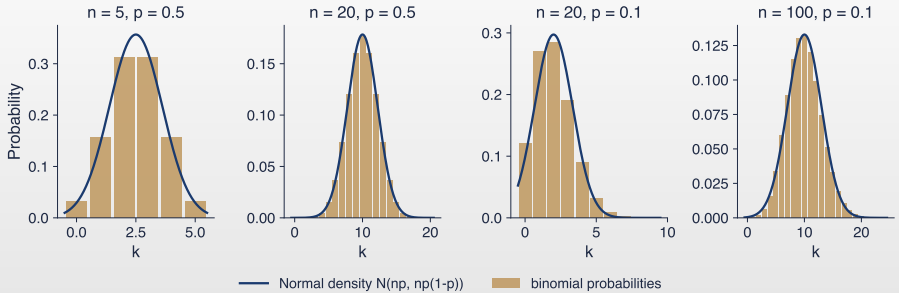
Illustration of the CLT (SFEclt)



- Standardised mean of n Bernoulli(0.2) draws (a skewed variable), 20 000 simulations
- Share above 1.96: 5.8% for $n = 5$, 2.7% for $n = 30$, 2.8% for $n = 300$; Normal distribution: 2.5%
- Skewness of the mean, $(1 - 2p)/\sqrt{np(1-p)}$ with $p = 0.2$: 0.67, 0.27, 0.09; it falls like $1/\sqrt{n}$

SFM_ch2_clt_normal_approx

The Normal Approximation of the Binomial (SFENormalApprox)



- $B(n, p)$: the number of successes in n independent trials with success probability p; $B(n, p) \approx N\big(np, np(1-p)\big)$
- Continuity correction: $P(X \leq k) \approx \Phi\big((k + 0.5 - np)/\sqrt{np(1-p)}\big)$; the $+0.5$ spreads each integer k over $[k - 0.5, k + 0.5]$
- Largest error of the CDF: 0.001 for $n = 20, p = 0.5$; 0.037 for $n = 20, p = 0.1$; 0.017 for $n = 100, p = 0.1$
- Rule of thumb: the approximation is good when $np(1-p)$ is large (at least about 9)

SFM_ch2_clt_normal_approx

Question for the Room: Does the CLT Make Daily Returns Normal?

- A daily log return is the sum of thousands of intraday log returns
- **What do you think?** Should daily returns then be close to the Normal distribution?
- **Answer**: not necessarily; each condition of the CLT can fail
  - dependence: calm and turbulent days come in clusters (Section 8)
  - changing distribution: volatility differs from day to day
  - infinite variance: Mandelbrot (1963) proposed it for cotton prices (Chapter 3, α-stable distributions)
- The data decide: the next sections measure how far returns are from the Normal distribution

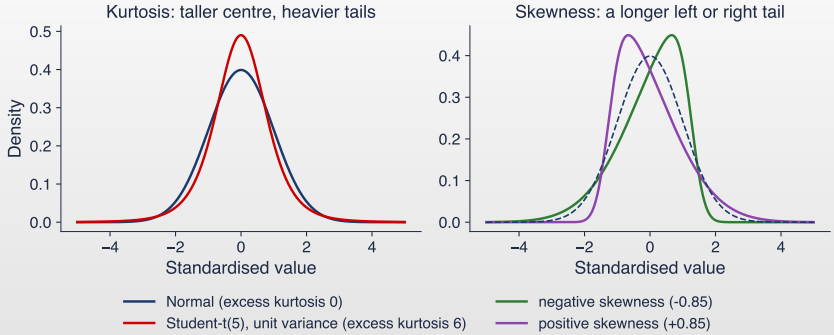
Recap: The Central Limit Theorem

- Sums of many i.i.d. terms with finite variance are approximately Normal
- Convergence is slower for skewed variables; $np(1-p) \geq 9$ for the binomial
- Returns may violate independence, identical distribution or finite variance

Four Moments

- Mean $\mu = E[X]$; variance $\sigma^2 = E[(X-\mu)^2]$; $E[\cdot]$ is the expected value
- **Skewness**: $\gamma_1 = E[(X-\mu)^3]/\sigma^3$
  - the third power keeps the sign of each deviation; $\gamma_1 < 0$: a longer left tail, large losses more frequent than large gains
- **Kurtosis**: $\gamma_2 = E[(X-\mu)^4]/\sigma^4$; Normal distribution: $\gamma_2 = 3$
  - **excess kurtosis** $\gamma_2 - 3$: positive means heavy tails and a tall, narrow centre (**leptokurtic**)
- Both are free of units: they describe the shape, not the location or the scale
- Fourth powers make the kurtosis very sensitive to a few extreme days

Skewness and Kurtosis: Graphical Illustration



- Left: same mean and variance; the Student-$t(5)$ curve (Section 7) has excess kurtosis 6: more mass in the centre and in the tails, less in the shoulders
- Right: skew-Normal densities with skewness ± 0.85; the dashed curve is the Normal distribution

SFM_ch2_moments_jarque_bera

Sample Moments

- Central sample moments: $m_k = \frac{1}{n}\sum_{t=1}^{n}(r_t - \bar{r})^k$, the average k-th power of the deviations from the mean (m_2: variance)
- Sample skewness $S = m_3/m_2^{3/2}$; sample excess kurtosis $K = m_4/m_2^2 - 3$
- If the data are i.i.d. Normal: $S \approx N(0, 6/n)$ and $K \approx N(0, 24/n)$
  - with $n = 4\,201$: standard errors $\sqrt{6/n} = 0.038$ and $\sqrt{24/n} = 0.076$
- These standard errors hold only under normality; for heavy-tailed data they are far too small (Seminar 2, B1)

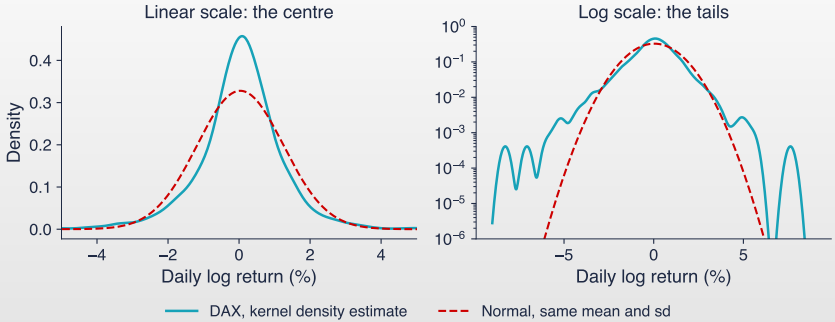
Daily Log Returns: Moments, 2010–2026

Series	from	n	Mean %	S.d. %	Skew.	Exc. kurt.	Min %
S&P 500	2010	4 201	0.045	1.08	−0.61	13.5	−12.8
DAX	2010	4 244	0.034	1.22	−0.45	7.4	−13.1
BET	2010	4 186	0.046	1.07	−0.98	17.9	−11.9
Bitcoin	2014	4 384	0.118	3.49	−0.70	12.0	−46.5
OMV Petrom (SNP)	2010	3 815	0.076	1.62	−0.54	8.9	−15.8
BRD	2010	3 771	0.044	1.66	−0.38	7.3	−18.2
Nuclearelectrica (SNN)	2013	2 911	0.099	1.52	0.27	5.1	−9.4
Romgaz (SNG)	2013	2 909	0.090	1.40	−0.34	5.3	−11.0

- Every series has excess kurtosis between 5.1 (SNN) and 17.9 (BET): far from 0
- Every series except SNN has negative skewness
- Worst days: S&P 500 on 16 March 2020, BET on 19 December 2018 (OUG 114, a tax on bank assets), Bitcoin on 12 March 2020

SFM_ch2_moments_jarque_bera

DAX Returns vs the Normal Density (SFEDaxReturnDistribution)



- Kernel density estimate (a smoothed histogram) against the Normal density with the same mean and standard deviation
- Centre: peak 0.46 vs 0.33; 76.8% of days within one standard deviation, vs 68.3% for the Normal distribution
- Tails (log scale): at -6% the estimated density is 439 times the Normal one

SFM_ch2_moments_jarque_bera

Data Checks before Tail Estimation

- Banca Transilvania (TLV), 27–31 May 2016: the close falls from 2.83 to 2.09 lei, the ex-date of a bonus share issue
- The adjusted close in our data: 3.6014, then 3.0995, then 3.8060
  - log returns -15.0% and $+20.5\%$ on two consecutive days: the adjustment was applied one day late
- Excess kurtosis of TLV, 2010–2026: 17.5 with these two days, 13.5 without them
- Two wrong days out of thousands change the kurtosis by a third: this is why TLV is not in the tables of this chapter
- Rule from Chapter 1: list the largest moves and check each one, especially pairs of opposite jumps

Recap: Moments

- Skewness measures asymmetry, excess kurtosis measures tail weight relative to the Normal distribution
- Daily returns: mostly negative skewness and excess kurtosis between 5 and 18
- Kurtosis is driven by a few days: check them before trusting it

The Jarque–Bera Test

- H_0: the data are Normal, so skewness $= 0$ and excess kurtosis $= 0$
- **JB** statistic Jarque & Bera (1980), Jarque & Bera (1987): $\text{JB} = \frac{n}{6}\left(S^2 + \frac{K^2}{4}\right)$
  - the sum of two squared standardised statistics: $(S/\sqrt{6/n})^2 + (K/\sqrt{24/n})^2$
- Under H_0 and for large n: $\text{JB} \sim \chi^2(2)$; reject at 5% if $\text{JB} > 5.99$
  - $\chi^2(2)$: the chi-squared distribution with 2 degrees of freedom, the distribution of a sum of two squared independent $N(0,1)$ variables; 5.99 is its 95% quantile
  - $\text{JB} = 0$ for a perfect Normal shape; large values mean skewness or excess kurtosis far from 0
- Simple, uses only two moments, and is the standard test reported in empirical finance

Worked Example: Jarque–Bera for the S&P 500

- $n = 4\,201$, $S = -0.606$, $K = 13.49$
- $\text{JB} = 700.17 \times (0.3673 + 45.497) = 32\,113$
- Critical value 5.99: normality is rejected; the p-value is zero to any number of decimals
- The kurtosis term dominates: almost all of JB comes from $K^2/4$
- JB for the other series: all rejected at any usual level
  - indices and Bitcoin: BET 56 762, DAX 9 909, Bitcoin 26 482
  - BVB stocks: SNP 12 832, BRD 8 370, SNN 3 174, SNG 3 415

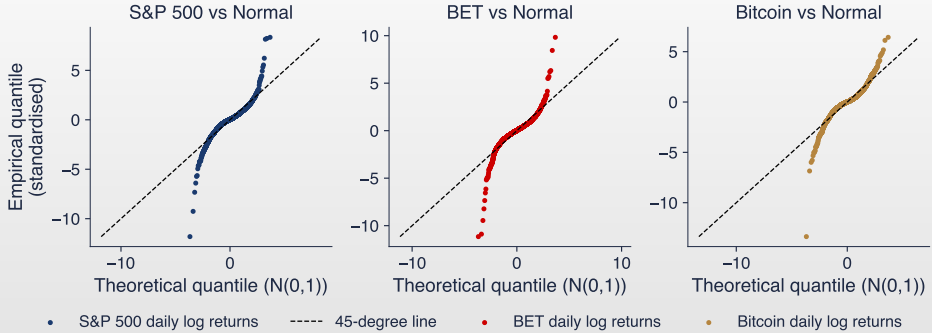
Limitations of the Jarque–Bera Test

- With thousands of observations it rejects even tiny departures
  - report the size of S and K, not only the p-value
- It assumes i.i.d. data; with volatility clustering its standard errors are too small
  - S&P 500: Normal-theory 95% interval for K: $[13.34, 13.64]$; bootstrap interval: $[6.7, 20.6]$ (Seminar 2, B1)
- A rejection says the data are not Normal; it does not say which distribution they follow
- Other tests (Kolmogorov–Smirnov, Anderson–Darling) compare the whole CDF: Chapter 6

QQ Plots

- A **QQ plot** (quantile–quantile plot) compares the empirical quantiles with those of a model Wilk & Gnanadesikan (1968)
  - sort the data: $r_{(1)} \leq r_{(2)} \leq \cdots \leq r_{(n)}$
  - plot $r_{(i)}$ against the model quantile $F^{-1}\big((i-0.5)/n\big)$
  - $r_{(i)}$: the i-th smallest return; F^{-1}: the quantile function of the model; $(i-0.5)/n$: the share of the data below $r_{(i)}$
- Reading
  - points on the 45-degree line: the model fits
  - an S-shape, low end below and high end above the line: heavier tails than the model
  - only one end bent away: skewness
- Unlike JB, it shows **where** the model fails: centre, shoulders or tails

QQ Plots against the Normal Distribution



- Standardised returns against $N(0,1)$ quantiles: all three bend away from the line at both ends
- S&P 500: lowest standardised return -11.8 where the Normal quantile is -3.7; BET: -11.2 vs -3.7
- The lower tail bends more than the upper one: negative skewness

SFM_ch2_qq_plots

Recap: Testing Normality

- JB combines skewness and kurtosis; it rejects normality for every series
- Report the size of the departure; Normal-theory standard errors are too small
- QQ plots show where the model fails: here, in both tails

"Student": William Sealy Gosset

- Chemist and brewer at Guinness, Dublin, working with small samples of barley and hops
- The brewery did not allow its staff to publish under their own names: he signed "Student" Student (1908)
- He derived the distribution of $(\bar{X} - \mu)/(s/\sqrt{n})$ for Normal data: the t distribution
- In finance it is used for a different reason: its tails are heavier than those of the Normal distribution Praetz (1972), Blattberg & Gonedes (1974)



William Sealy Gosset (1876–1937)

Photo: unknown author (1908); public domain; Wikimedia Commons

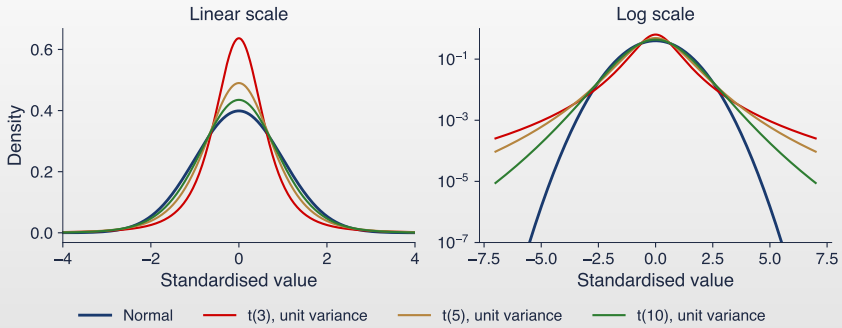
The Student-t Distribution: Definition

- $T = Z/\sqrt{V/\nu}$ with $Z \sim N(0,1)$ independent of $V \sim \chi^2(\nu)$: $T \sim t(\nu)$
  - $\nu > 0$: the **degrees of freedom**, which control the tails
  - $\chi^2(\nu)$: the distribution of a sum of ν squared independent $N(0,1)$ variables; Γ: the gamma function, $\Gamma(k) = (k-1)!$ for integer k
- Density: $f(x) = \dfrac{\Gamma\big((\nu+1)/2\big)}{\sqrt{\nu\pi}\,\Gamma(\nu/2)}\Big(1 + \dfrac{x^2}{\nu}\Big)^{-(\nu+1)/2}$
- Location–scale version for returns: $r = m + s\,T$, with location m, scale s and ν
- As $\nu \to \infty$, $t(\nu) \to N(0,1)$: the Normal distribution is a limiting case

Moments and Tails

- Mean 0 if $\nu > 1$; variance $\nu/(\nu-2)$ if $\nu > 2$; excess kurtosis $6/(\nu-4)$ if $\nu > 4$
- Moments of order ν and higher are infinite
  - $\nu \leq 4$: infinite kurtosis; $\nu \leq 2$: infinite variance
- **Power-law tails**: $P(|T| > x) \approx c\, x^{-\nu}$ for large x
  - $c > 0$: a constant; doubling x divides the tail probability by about 2^{ν}
  - the Normal tail falls like $e^{-x^2/2}$, much faster than any power
  - ν is also the **tail index**, estimated directly in Chapter 5
- To compare with the Normal distribution at the same variance, use $T\sqrt{(\nu-2)/\nu}$, which has variance 1

Student-t Densities with Unit Variance



- Smaller ν: a taller centre and heavier tails; on the log scale the Normal tails drop like a parabola
- $P(|X| > 4)$ at unit variance: $t(3)$ 0.62%, $t(5)$ 0.36%, $t(10)$ 0.12%, Normal distribution 0.0063%
- With $\nu = 5$, a 4-sigma day is 56 times more likely than under the Normal distribution

SFM_ch2_student_t_fit

Fitting a Student-t by Maximum Likelihood

- **MLE** (maximum likelihood estimation): choose (ν, m, s) that maximise the log-likelihood
  - $\ell(\nu, m, s) = \sum_{t=1}^{n} \ln f\big((r_t - m)/s; \nu\big) - n \ln s$
  - $f(\cdot; \nu)$: the $t(\nu)$ density; $-n \ln s$ corrects for rescaling the returns by s
  - no closed form: a numerical optimiser (`scipy.stats.t.fit` in the notebook)
- **AIC** (Akaike information criterion) Akaike (1974): $\text{AIC} = 2k - 2\ell$
  - k = number of parameters (2 for the Normal distribution, 3 for the t); the lower AIC wins
  - $\Delta\text{AIC} = \text{AIC}_{\text{Normal}} - \text{AIC}_t > 0$ favours the t
- Early evidence for stock prices: Praetz (1972), Blattberg & Gonedes (1974)

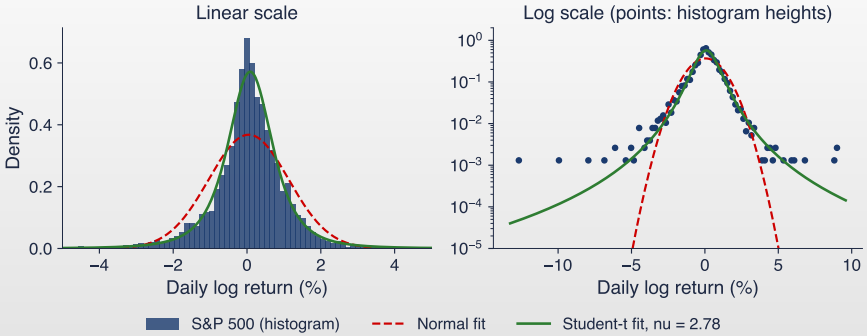
Fitted Student-t: Degrees of Freedom, 2010–2026

Series	$\hat{\nu}$	ΔAIC
S&P 500	2.78	1213
DAX	3.31	739
BET	2.96	1454
Bitcoin	2.23	1335
OMV Petrom (SNP)	3.79	679
BRD	4.67	478
Nuclearelectrica (SNN)	3.08	541
Romgaz (SNG)	3.40	482

- All $\hat{\nu}$ between 2.23 (Bitcoin) and 4.67 (BRD)
- $\hat{\nu} < 4$ for 7 of the 8 series: the fitted model has infinite kurtosis
- ΔAIC is at least 478: the t beats the Normal distribution everywhere

SFM_ch2_student_t_fit

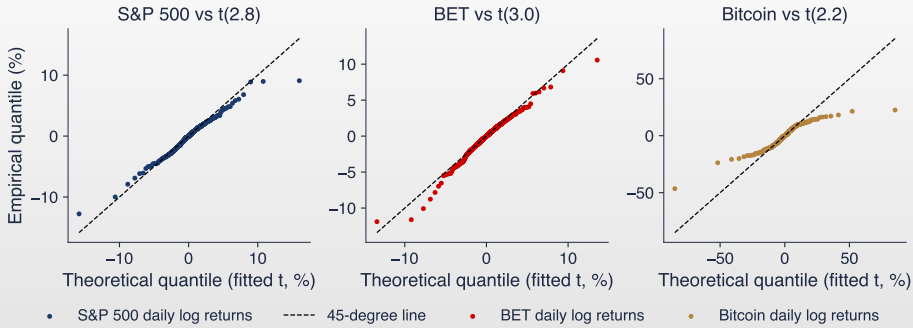
S&P 500: Normal vs Student-t Fit



- Linear scale: the t ($\hat{\nu} = 2.78$) follows the tall centre that the Normal distribution misses
- Log scale: the Normal density collapses beyond $\pm 4\%$; the t follows the observed tails

SFM_ch2_student_t_fit

QQ Plots against the Fitted Student-t



- The centre and the moderate tails now lie on the line
- Extremes: the fitted t expects lower minima than observed: S&P 500 -15.8% vs -12.8%; Bitcoin -85.0% vs -46.5%
- With $\hat{\nu}$ close to 2 the model overshoots the most extreme quantiles

SFM_ch2_qq_plots

Limitations of the Student-t Distribution

- **Symmetry**: the t cannot reproduce the negative skewness
  - skewed t distributions exist; they are beyond this chapter
- **Independence**: an i.i.d. t model has no volatility clustering
  - GARCH models with t errors combine both Bollerslev (1987) (Chapter 9)
- **Single tail index**: one ν for both tails and for the centre
  - extreme value theory models each tail separately (Chapter 5)
- Still the simplest heavy-tailed model, and a large improvement on the Normal distribution

Recap: Student-t

- One extra parameter ν controls the tails; power-law tails, the Normal distribution as $\nu \to \infty$
- Fitted $\hat{\nu}$ between 2 and 5 for daily returns; AIC strongly prefers the t
- It misses skewness and volatility clustering

Definition of a Stylised Fact

- A **stylised fact**: a statistical property shared by many assets, markets and periods Cont (2001)
- Qualitative, not exact: "heavy tails", not "$\nu = 3.2$"
- Rama Cont lists eleven; we check six on our data
- A good model of returns must reproduce them; the benchmark i.i.d. Normal model fails most of them
- Earlier evidence: Mandelbrot (1963), Fama (1965)



Rama Cont

Photo: Renate Schmid (2012); CC BY-SA 2.0 DE; Wikimedia Commons

Six Stylised Facts

- **1. Heavy tails**: large moves are far more frequent than under the Normal distribution
- **2. Aggregational Gaussianity**: returns over longer horizons are closer to the Normal distribution
- **3. Absence of linear autocorrelation**: past returns barely predict future returns
- **4. Volatility clustering**: large moves follow large moves, of either sign
- **5. Leverage effect**: falls raise future volatility more than rises
- **6. Gain/loss asymmetry**: large losses are larger and more frequent than large gains

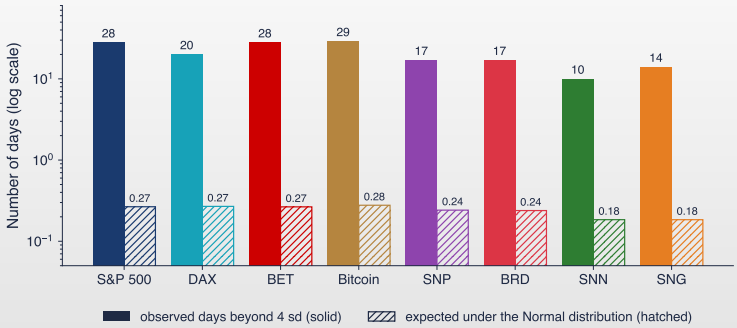
Fact 1: Heavy Tails, Number of k-Sigma Days

Series	n	$> 3\sigma$	$> 4\sigma$	$> 5\sigma$	$> 6\sigma$	p-value (Poisson, 4σ)
S&P 500	4 201	62	28	12	8	2×10^{-46}
DAX	4 244	56	20	5	4	1×10^{-30}
BET	4 186	70	28	16	11	2×10^{-46}
Bitcoin	4 384	80	29	10	4	6×10^{-48}
Normal distribution, $n = 4\,201$		11.34	0.27	2.4×10^{-3}	8.3×10^{-6}	

- $> k\sigma$: number of days with $|r_t - \bar{r}| > k\,s$; last row: number expected under the Normal distribution
- p-value: probability of at least that many 4-sigma days if the Normal count is Poisson with mean $n \times 6.3 \times 10^{-5}$
- S&P 500: 105 times more 4-sigma days than the Normal distribution allows

SFM_ch2_sigma_days

Fact 1: 4-Sigma Days in Every Market



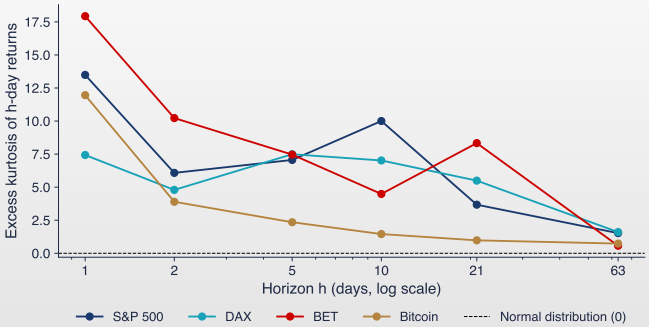
- Observed days beyond 4 standard deviations (log scale) against the Normal expectation of 0.18–0.28 days
- Indices and Bitcoin: one 4-sigma day every 150–212 days; BVB stocks: 10–17 days in total
- Under a fitted Student-t, the expected numbers are close to the observed ones: S&P 500 34.9, BET 28.1 (Seminar 2, B4)

SFM_ch2_sigma_days

Fact 2: Aggregational Gaussianity

- h-day log return: $r_t(h) = r_t + r_{t-1} + \cdots + r_{t-h+1}$, taken on non-overlapping blocks
  - if the CLT applies, the distribution of $r_t(h)$ approaches the Normal distribution as h grows
- The stylised fact: the shape changes with the horizon; it is not the same at all time scales Cont (2001)
- Price to pay: fewer observations
  - S&P 500, 17 years: 4 201 daily, 200 monthly and 66 quarterly returns
  - estimates of kurtosis at long horizons are very noisy

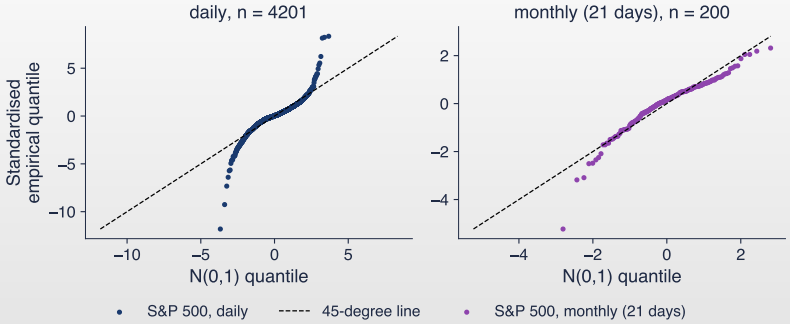
Fact 2: Kurtosis Falls with the Horizon



- Excess kurtosis, daily → monthly → quarterly: S&P 500 13.5 → 3.7 → 1.5; BET 17.9 → 8.3 → 0.6; Bitcoin 12.0 → 1.0 → 0.7
- The fall is slow and not monotone: one crash month (March 2020) dominates a sample of 200 months
- Quarterly returns: JB no longer rejects normality for the BET (p-value 0.07) and Bitcoin (p-value 0.22)

SFM_ch2_cont_stylised_facts

Fact 2: S&P 500, Daily vs Monthly QQ Plots



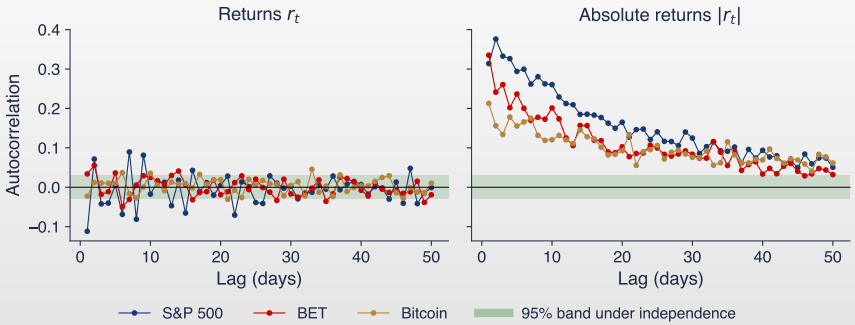
- Monthly returns lie much closer to the line than daily returns
- But monthly returns are not Normal: skewness -1.18, excess kurtosis 3.7, $n = 200$; the lowest point is March 2020
- Why so slow? Volatility clustering makes daily returns dependent, which slows the CLT

SFM_ch2_cont_stylised_facts

Fact 3: Absence of Linear Autocorrelation

- **ACF** (autocorrelation function): $\rho(h) = \text{Corr}(r_t, r_{t-h})$, $h = 1, 2, \ldots$
  - sample version $\hat{\rho}(h) = \sum_t (r_t - \bar{r})(r_{t-h} - \bar{r}) / \sum_t (r_t - \bar{r})^2$
  - h: the lag in days; $\rho(h) \in [-1, 1]$; for i.i.d. data, $\hat{\rho}(h) \approx N(0, 1/n)$: 95% band $\pm 1.96/\sqrt{n}$
- **LB** (Ljung–Box) test Ljung & Box (1978) of H_0: $\rho(1) = \cdots = \rho(m) = 0$
  - $Q(m) = n(n+2) \sum_{h=1}^{m} \hat{\rho}(h)^2/(n-h) \sim \chi^2(m)$; for $m = 10$, reject if $Q > 18.3$
  - m: the number of lags tested together; Q adds up the squared autocorrelations, so large values mean some $\rho(h) \neq 0$
- Why absent? If returns were predictable, traders would exploit it until the predictability disappears (efficient markets, Chapter 7)

Facts 3 and 4: ACF of Returns and of Absolute Returns



- Left: returns; most autocorrelations are inside or close to the band; S&P 500 $\hat{\rho}(1) = -0.11$, BET 0.03, Bitcoin -0.02
- Right: absolute returns; all positive, slowly decaying: S&P 500 0.31 at lag 1, 0.26 at lag 10, 0.05 at lag 50

SFM_ch2_cont_stylised_facts

Fact 3: Weak Autocorrelation of Returns

- Ljung–Box $Q(10)$ for returns: S&P 500 199, BET 46, Bitcoin 20; critical value 18.3
  - formally, some autocorrelation is significant
- Two reasons not to over-read this
  - the values are economically small: $|\hat{\rho}(h)| \leq 0.11$, so at most 1.2% of the variance is explained
  - the band $\pm 1.96/\sqrt{n}$ assumes i.i.d. returns; with volatility clustering it is too narrow, so "significant" is found too often
- The S&P 500 value at lag 1 comes mostly from March 2020, when large falls and rebounds alternated: without 20 February–30 April 2020, $\hat{\rho}(1) = -0.04$
- Robust tests of predictability (variance ratios): Chapter 7

Fact 4: Volatility Clustering

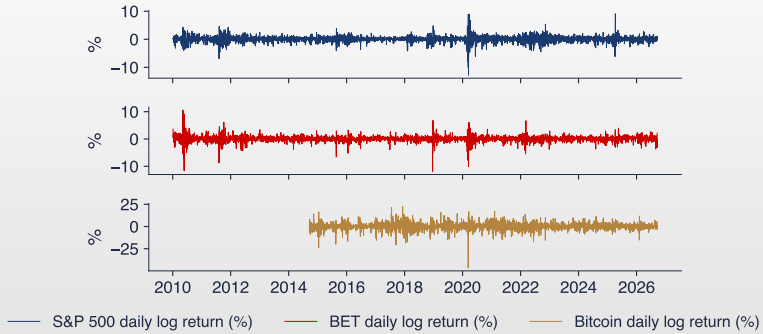
- "[L]arge changes tend to be followed by large changes—of either sign—and small changes tend to be followed by small changes" Mandelbrot (1963)
- The sign of tomorrow's return is unpredictable; its size is not
- Evidence: ACF of $|r_t|$ positive for 50 lags
  - Ljung–Box $Q(10)$ of $|r_t|$: S&P 500 3859, BET 2120, Bitcoin 1086
- So returns are uncorrelated but **not independent**
- Models: ARCH and GARCH (Chapter 9)



Benoit Mandelbrot (1924–2010)

Photo: Rama (2007); CC BY-SA 2.0 fr; Wikimedia Commons

Fact 4: Calm and Turbulent Periods



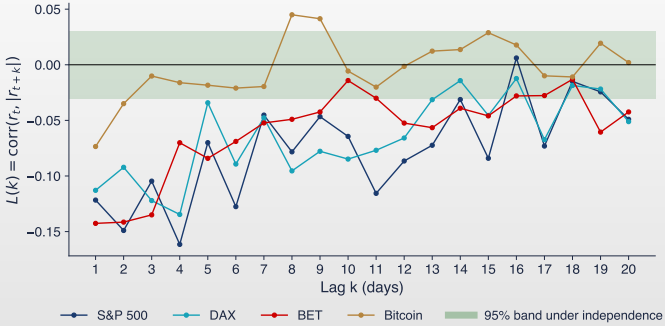
- Turbulent periods: 2010–2011 (euro area debt crisis), December 2018 for the BET, March 2020 everywhere, April 2025
- Calm years in between: the volatility of a single day depends on the period it falls in

SFM_ch2_cont_stylised_facts

Fact 5: The Leverage Effect

- Measure: $L(k) = \text{Corr}(r_t, |r_{t+k}|)$, $k = 1, 2, \ldots$
  - $L(k) < 0$: a fall today is followed by larger moves than a rise of the same size
- Explanation by financial leverage Christie (1982)
  - when the share price falls, the debt-to-equity ratio rises, so equity becomes riskier
- A second explanation, volatility feedback: higher expected volatility raises required returns and lowers prices
- Models with an asymmetric response (EGARCH, GJR-GARCH): Chapter 9

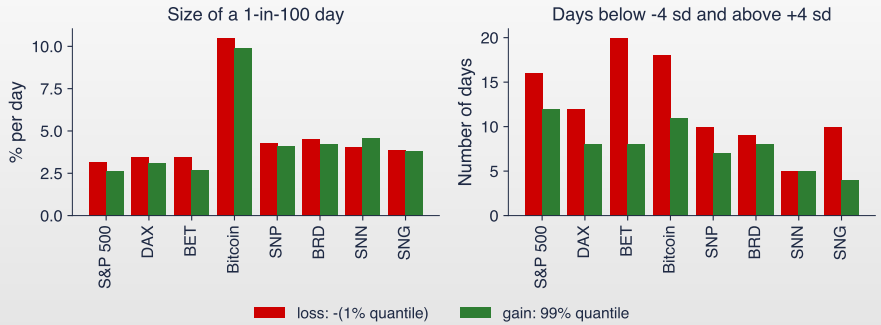
Fact 5: Leverage Correlations



- Stock indices: $L(k) < 0$ for most of the first 20 lags; mean of $L(1..5)$: S&P 500 -0.12, DAX -0.10, BET -0.11
- Bitcoin: $L(1) = -0.07$, then inside the band: no firm has debt behind Bitcoin, so the leverage mechanism is absent

SFM_ch2_cont_stylised_facts

Fact 6: Gain/Loss Asymmetry



- Left: a 1-in-100 loss is larger than a 1-in-100 gain for the indices: BET 3.45% vs 2.66%, S&P 500 3.15% vs 2.61%
- Right: days below -4σ vs above $+4\sigma$: BET 20 vs 8, S&P 500 16 vs 12, Bitcoin 18 vs 11
- Exception: SNN, with positive skewness and a larger right quantile

SFM_ch2_sigma_days

The Six Stylised Facts in Eight Series

Series	Exc. kurt.	Exc. kurt., 21 days	$\hat{\rho}_r(1)$	$\hat{\rho}_{\lvert r\rvert}(10)$	$\bar{L}(1..5)$	$\lvert q_{0.01}\rvert / q_{0.99}$
S&P 500	13.5	3.7	−0.11	0.26	−0.12	1.21
DAX	7.4	5.5	0.01	0.16	−0.10	1.12
BET	17.9	8.3	0.03	0.20	−0.11	1.30
Bitcoin	12.0	1.0	−0.02	0.12	−0.03	1.07
OMV Petrom (SNP)	8.9	3.3	−0.02	0.13	−0.03	1.04
BRD	7.3	0.4	0.03	0.11	−0.05	1.07
Nuclearelectrica (SNN)	5.1	2.3	0.09	0.08	−0.03	0.88
Romgaz (SNG)	5.3	1.4	0.07	0.08	−0.04	1.02

- Columns: excess kurtosis of daily and 21-day returns; ACF of returns at lag 1 and of $|r_t|$ at lag 10; mean $L(k)$ over lags 1–5; 1% over 99% quantile
- Fact 1 holds for every series, fact 2 with noise; fact 3 only approximately, $|\hat{\rho}_r(1)|$ up to 0.11 (S&P 500; SNN 0.09); fact 4 everywhere, weakest for SNN and SNG (0.08)
- Fact 5: clear for the indices (−0.12 to −0.10), equally weak for Bitcoin, SNP, SNN and SNG (−0.04 to −0.03); fact 6 fails for SNN (0.88), close to 1 for SNP and SNG
- A comparison of the return distributions of cryptocurrencies and traditional assets: Pele et al. (2023)

SFM_ch2_cont_stylised_facts

Implications of the Stylised Facts for Models

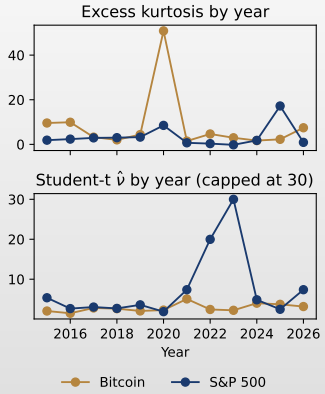
Model	1	2	3	4	5	6
i.i.d. Normal	–	trivially	✓	–	–	–
i.i.d. Student-t	✓	✓	✓	–	–	–
GARCH with t errors (Ch. 9)	✓	✓	✓	✓	–	–
EGARCH, GJR-GARCH (Ch. 9)	✓	✓	✓	✓	✓	(✓)

- Columns: the six facts in the order of this section; (✓): only for returns over several days, unless the errors are skewed
- Risk measures (VaR 1%, ES 2.5%, Chapter 10) inherit the quality of the model: a Normal VaR is too low in crises
- A related summary of tail risk, information entropy: Pele, Lazar & Dufour (2017)

Recap: Stylised Facts

- Heavy tails: 4-sigma days every few months instead of once in decades
- Returns are nearly uncorrelated, but their size is strongly autocorrelated
- Longer horizons are closer to the Normal distribution, slowly
- Stock indices: falls raise volatility, and losses are larger than gains

An Open Research Question: Evolution of Bitcoin's Tails



- As a market matures (more traders, futures, ETFs), its tails might thin out
- Bitcoin, 2015–2026: yearly $\hat{\nu}$ between 1.5 and 5.1; rank correlation with the year 0.63 (p-value 0.03)
- Why it is open: 12 noisy yearly estimates, one year (2020, excess kurtosis 51) dominated by one day
- AI tools can speed up such a study; they do not replace checking it Wang et al. (2023)

SFM_ch2_tails_by_year

Possible Uses of AI Tools

- **Literature**: find studies of tail indices of crypto assets and summarise their methods
- **Code**: draft rolling-window estimates of ν and of the excess kurtosis
- **Robustness**: propose other windows, other tail measures (Chapter 5), other crypto assets
- Example prompt
  - `Write a Python function that fits a Student-t distribution to daily Bitcoin log returns on rolling 365-day windows and returns the degrees of freedom with a bootstrap 95% interval.`

What to Check

- Data: the same price column and calendar for every year; errors among the largest moves
- Estimation: did the optimiser converge? $\hat{\nu}$ near 2 or above 30 needs a second look
- Inference: one $\hat{\nu}$ per year is noisy; report intervals, not only point estimates
- Volatility clustering: thinner yearly tails may only mean calmer years (Chapter 9)
- References: every cited paper must exist; check the DOI

Project Idea

- **Question**: do the tails of crypto assets thin out as their markets mature?
  - data: Bitcoin, Ethereum and Solana from the course data; S&P 500 as a benchmark
- Steps
  - estimate $\hat{\nu}$ and the excess kurtosis on rolling one-year windows, with bootstrap intervals
  - test for a trend; repeat on returns divided by a rolling volatility
  - compare the count of 4-sigma days per year with the S&P 500
- Deliverable: one table, one chart, and a paragraph on what the data can and cannot show
- Declare any AI use, and list the errors of the AI that you corrected

Key Formulas

Quantity	Formula
Normal PDF	$f(x) = (\sigma\sqrt{2\pi})^{-1} \exp\big(-(x-\mu)^2/(2\sigma^2)\big)$
Lognormal moments	median e^m, mean $e^{m+s^2/2}$
CLT	$\sqrt{n}(\bar{X}_n - \mu)/\sigma \xrightarrow{d} N(0,1)$
Skewness, excess kurtosis	$S = m_3/m_2^{3/2}, \quad K = m_4/m_2^2 - 3$
Jarque–Bera	$\mathrm{JB} = \frac{n}{6}(S^2 + K^2/4) \sim \chi^2(2)$
Student-$t(\nu)$	variance $\nu/(\nu-2)$, excess kurtosis $6/(\nu-4)$
AIC	$2k - 2\ell$
ACF, Ljung–Box	$\hat{\rho}(h), \quad Q(m) = n(n+2)\sum_{h=1}^{m} \hat{\rho}(h)^2/(n-h)$
Leverage	$L(k) = \mathrm{Corr}(r_t, \lvert r_{t+k}\rvert)$

Self-Assessment

- **Question**: daily returns are $N(0.05\%, 1\%^2)$; how many standard deviations is a -3% day?
  - **Answer**: $z = (-3 - 0.05)/1 = -3.05$
- **Question**: $n = 2500$, $S = -0.5$, $K = 6$; what is JB?
  - **Answer**: $\frac{2500}{6}(0.25 + 9) = 3854$; normality is rejected
- **Question**: an annual log return is $N(5\%, 20\%^2)$; what are the median and the mean of 100 invested after one year?
  - **Answer**: median $100e^{0.05} = 105.1$, mean $100e^{0.05+0.02} = 107.3$
- Next: Chapter 3, α-stable distributions

References (1/2)

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6), 716–723.
- Bachelier, L. (1900). Théorie de la spéculation. *Annales scientifiques de l'École Normale Supérieure*, 17, 21–86.
- Blattberg, R.C., Gonedes, N.J. (1974). A comparison of the stable and Student distributions as statistical models for stock prices. *Journal of Business*, 47(2), 244–280.
- Bollerslev, T. (1987). A conditionally heteroskedastic time series model for speculative prices and rates of return. *Review of Economics and Statistics*, 69(3), 542–547.
- Borak, S., Härdle, W.K., López-Cabrera, B. (2013). *Statistics of Financial Markets: Exercises and Solutions*, 2nd ed. Springer.
- Christie, A.A. (1982). The stochastic behavior of common stock variances: value, leverage and interest rate effects. *Journal of Financial Economics*, 10(4), 407–432.
- Cont, R. (2001). Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*, 1(2), 223–236.
- Fama, E.F. (1965). The behavior of stock-market prices. *Journal of Business*, 38(1), 34–105.
- Franke, J., Härdle, W.K., Hafner, C.M. (2019). *Statistics of Financial Markets: An Introduction*, 5th ed. Springer.
- Jarque, C.M., Bera, A.K. (1980). Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics Letters*, 6(3), 255–259.
- Jarque, C.M., Bera, A.K. (1987). A test for normality of observations and regression residuals. *International Statistical Review*, 55(2), 163–172.

References (2/2)

- Ljung, G.M., Box, G.E.P. (1978). On a measure of lack of fit in time series models. *Biometrika*, 65(2), 297–303.
- Mandelbrot, B. (1963). The variation of certain speculative prices. *Journal of Business*, 36(4), 394–419.
- Osborne, M.F.M. (1959). Brownian motion in the stock market. *Operations Research*, 7(2), 145–173.
- Pele, D.T., Lazar, E., Dufour, A. (2017). Information entropy and measures of market risk. *Entropy*, 19(5), 226.
- Pele, D.T., Wesselhöfft, N., Härdle, W.K., Kolossiatis, M., Yatracos, Y.G. (2023). Are cryptos becoming alternative assets? *European Journal of Finance*, 29(10), 1064–1105.
- Praetz, P.D. (1972). The distribution of share price changes. *Journal of Business*, 45(1), 49–55.
- Student (1908). The probable error of a mean. *Biometrika*, 6(1), 1–25.
- Tsay, R.S. (2010). *Analysis of Financial Time Series*, 3rd ed. Wiley.
- Wang, H., Fu, T., Du, Y., Gao, W., et al. (2023). Scientific discovery in the age of artificial intelligence. *Nature*, 620, 47–60.
- Wilk, M.B., Gnanadesikan, R. (1968). Probability plotting methods for the analysis of data. *Biometrika*, 55(1), 1–17.