








Statistics of Financial Markets

Chapter 1: Data, returns and indicators



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 (1/2)

- **ADF**: Augmented Dickey–Fuller (test)
- **AI**: Artificial Intelligence
- **BET**: Bucharest Exchange Trading (index)
- **BET-TR**: BET Total Return (index)
- **BNR**: Banca Națională a României (the National Bank of Romania)
- **BVB**: Bursa de Valori București (the Bucharest Stock Exchange)
- **CAGR**: Compound Annual Growth Rate
- **CI**: Confidence Interval
- **COVID-19**: COronaVIrus Disease 2019
- **CRSP**: Center for Research in Security Prices (University of Chicago)
- **ECB**: European Central Bank
- **EOD**: End Of Day (daily closing data)
- **EODHD**: EOD Historical Data (market-data provider)
- **FRED**: Federal Reserve Economic Data

Acronyms used in this material (2/2)

- **GARCH**: Generalised AutoRegressive Conditional Heteroskedasticity
- **GDP**: Gross Domestic Product
- **LSEG**: London Stock Exchange Group (data vendor, formerly Refinitiv)
- **MDD**: Maximum DrawDown
- **NYSE**: New York Stock Exchange
- **OHLCV**: Open, High, Low, Close, Volume
- **OMV**: Österreichische Mineralölverwaltung — Austrian Mineral Oil Administration (OMV group)
- **OUG**: Ordonanță de Urgență a Guvernului (government emergency ordinance)
- **SE**: Standard Error
- **SFE**: Statistics of Financial Markets, Exercises (Quantlet collection of the textbook)
- **SR**: Sharpe Ratio
- **US**: United States

Today's Question, Route and Learning Outcomes

- **Question**: how do we turn raw market data into numbers that let us compare investments?
  - a price level says little; a return, a risk measure and a drawdown say a lot
- **Route and outcomes**: after this chapter you can
  - choose a reliable data source and check a price series before using it
  - compute simple and log returns, and aggregate them over time and across assets
  - annualise means and volatilities with the actual frequency of the data
  - describe a return series by its moments, quantiles and extremes
  - compute and interpret CAGR, Sharpe, Sortino, maximum drawdown, Calmar and the volatility drag
  - judge whether a difference between two indicators is larger than its estimation error
- Real data throughout: BET, S&P 500, DAX, Bitcoin and five stocks of the BVB

Reading and Tools

- Textbook: Franke, Härdle & Hafner (2019), *Statistics of Financial Markets*, 5th ed., Ch. 11 (returns, definitions)
  - exercises with solutions: Borak, Härdle & López-Cabrera (2013)
- Complementary: Tsay (2010), Ch. 1 (asset returns); Campbell, Lo & MacKinlay (1997), Ch. 1
- Python Quantlets of this chapter: Quantlets/Ch_01
  - ported from the SFE Quantlets of the textbook (SFEReturns, SFEcomplogreturns, SFEportlogreturns, SFEfdescStat)
  - each chart in these slides links to the Quantlet that draws it
- Lecture notebook: open in Google Colab
- Video course: Statistics of Financial Markets

What Financial Data Are

- **Market data**: what happens on a trading venue
  - trade prices, best bid and ask quotes, traded volume
  - index levels, exchange rates, interest rates, crypto prices
- **Corporate actions**: events that change the price without changing the value of a holding
  - dividends, splits, share consolidations, bonus shares
- **Fundamental and macro data**
  - balance sheets and earnings; inflation, GDP, policy rates
- **Alternative data**
  - news, social media, satellite images, blockchain transactions
- This course works mainly with **daily market data**

Ticker Tape and Electronic Data Feeds

- From 1867, **stock tickers** printed trade prices on a paper tape, sent by telegraph
  - the first real-time price feed: the price of the last trade, minutes after it happened
- Today: electronic feeds from the exchanges, resold by data vendors
  - every quote and trade, time-stamped to the microsecond
  - daily summaries (OHLCV) are what most studies use
- The ticker survived: scrolling prices and news on building fronts; ticker symbols such as TLV, SNP, H2O on the BVB



News ticker on the Reuters building, New York

Photo: bgilliard (2005); CC BY-SA 2.0; Wikimedia Commons

Data Frequency

- **Tick data**: every trade and quote
  - millions of records per day; noise from the bid–ask bounce
- **Intraday bars**: 1, 5 or 60 minutes
  - used for realised volatility (Chapter 8)
- **Daily data**: one row per trading day
  - the standard of this course and of most academic studies
  - exchanges trade about 250 days a year; crypto markets trade 365 days a year
- **Weekly, monthly, yearly data**
  - fewer observations, closer to the horizon of long-term investors
- Rule: annualise with the **actual** number of observations per year of each series

Where the Data Come From

Source	What it provides	Access
Exchanges (BVB, Deutsche Börse, NYSE)	official prices, volumes, index levels	website, paid feeds
Data vendors (Bloomberg, LSEG, EODHD)	prices from many venues, adjusted prices, fundamentals	subscription
Central banks (BNR, ECB, Federal Reserve)	reference exchange rates, interest rates	free
Statistical databases (FRED, Eurostat)	macro series, yields	free
Academic databases (CRSP, Kenneth French library)	cleaned returns, including delisted firms; factor returns	university licence or free
Crypto exchanges and aggregators	prices 24/7, on-chain data	free or subscription

- A vendor collects and cleans; the **primary source** of a number (exchange, central bank) is the reference when two sources disagree

The Data of This Course

- Daily market data from **EODHD** (EOD Historical Data), saved once in the course repository
  - 91 series: indices, ETFs, stocks (BVB, US, Europe), crypto, exchange rates, bond yields
  - columns: date, open, high, low, close, adjusted close, volume
- EUR/RON: the official **BNR** reference rate
  - the EUR/RON series from EODHD has erroneous quotes (Data Errors: Two EUR/RON Series)
- Conventions
  - indices, exchange rates, crypto: *close*; stocks and ETFs: *adjusted close*
  - weekend quotes and repeated holiday closes are dropped, except for crypto
  - each series is analysed on its own calendar
- Comparison period of the chapter: 1 January 2015 to 18 September 2026

Daily Bars: OHLCV

- **OHLCV**: open O_t, high H_t, low L_t, close C_t, volume V_t of day t
- By construction $L_t \le O_t, C_t \le H_t$; the range $H_t - L_t$ measures the intraday spread of prices
- Banca Transilvania (TLV), the last five trading days in our data:

Date	Open	High	Low	Close	Volume
14 Sep 2026	34.52	34.86	34.28	34.54	346 819
15 Sep 2026	34.40	34.68	33.10	34.00	676 653
16 Sep 2026	34.00	34.10	33.20	33.40	853 078
17 Sep 2026	33.40	33.40	33.40	33.40	1 084 380
18 Sep 2026	34.22	35.80	34.22	35.50	1 884 988

- Check: on 17 September 2026 open = high = low = close with 1 084 380 shares traded; a flat bar with a large volume is suspicious

Corporate Actions and the Adjusted Close

- The traded close P_t jumps on the day of a corporate action; the value of a holding does not
- **Adjustment factor** F on the event day
  - cash dividend D: $F = 1 - D/P_{t-1}$
  - split m-for-n (m new shares for n old): $F = n/m$; a consolidation is a split with $m < n$
- **Adjusted close**: every earlier price is multiplied by the factors of all later events, $P_t^{\text{adj}} = P_t \prod_{k:\, t_k > t} F_{t_k}$
  - t_k: the day of event k; the product runs over all events after day t, so the last prices stay unchanged
  - returns computed from adjusted prices include the dividend and have no artificial jumps
- With dividends, the simple return is $R_t = (P_t + D_t)/P_{t-1} - 1$, with D_t the dividend paid on day t Franke, Härdle & Hafner (2019)

Worked Example: Adjusted Prices

- A share trades at 100; on day 5 it pays a cash dividend of 2; on day 10 it splits 2-for-1
- Adjustment factors
  - dividend: $F_5 = 1 - 2/100 = 0.98$; split: $F_{10} = 1/2 = 0.50$
- Adjusted prices
  - before day 5: $P_t^{\text{adj}} = P_t \times 0.49$, so 100 becomes 49
  - between days 5 and 9: $P_t^{\text{adj}} = P_t \times 0.50$; from day 10: $P_t^{\text{adj}} = P_t$
- The adjusted series changes every time a new event happens: always keep the download date
- Indices: a **price index** ignores dividends; a **total return index** reinvests them

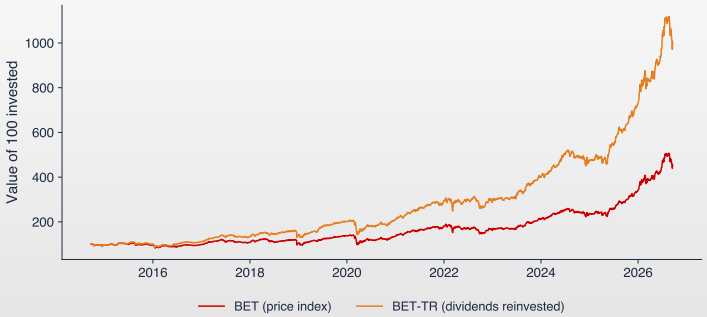
Real Data: Banca Transilvania, Close vs Adjusted Close



- On 16 August 2022 the close jumps from 2.12 to 21.20 RON: a share consolidation, 10 old shares into one
- Log return from the close: +230% in one day; from the adjusted close: 0.0%
- Since 2015 the daily returns of the two series differ by more than 2 percentage points on 24 days (dividends, bonus shares)

SFM_ch1_data_quality

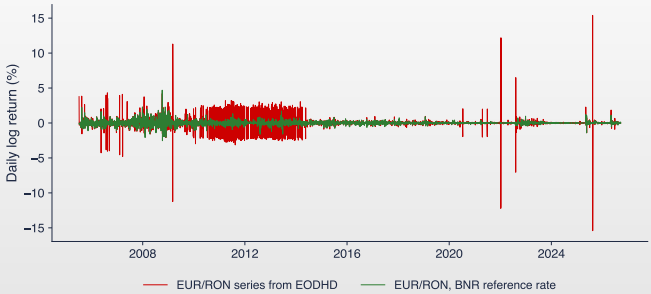
The Effect of Dividends: BET vs BET-TR



- 100 invested on 23 September 2014: 455 in the BET (price index), 1 005 in the BET-TR (dividends reinvested)
- Growth rate: 13.5% vs 21.2% a year; dividends add 7.8 percentage points a year
- Compare like with like: the DAX is a total return index, the S&P 500 and the BET are price indices

SFM_ch1_data_quality

Data Errors: Two EUR/RON Series



- Same exchange rate, 5 154 common days: the EODHD series moves by more than 1 percentage point more than the BNR rate on 412 days
- Annualised volatility: 13.4% (EODHD) vs 4.5% (BNR); largest absolute EODHD move 15.4% on 14 August 2025, the reversal of a wrong quote of the previous day
- Fix: use the primary source of the number, the BNR reference rate

SFM_ch1_data_quality

Three Biases That Overstate Performance

- **Survivorship bias**: only firms or funds that still exist are in the sample
  - the losers have disappeared, so the average return is too high Brown et al. (1992)
  - example: today's members of the BET, followed back to 2000
- **Look-ahead bias**: using information not yet available at the decision date
  - annual reports published in March used for decisions in January
- **Data snooping**: trying many rules on the same data and reporting the best one
  - hundreds of published return predictors are likely false discoveries Harvey, Liu & Zhu (2016)
- Each bias inflates returns; none shows up in the summary statistics

A Data Checklist

- Source: the primary source of the number and the column used (close or adjusted close)
- Calendar: weekends, holidays, missing days; same dates for all assets in a joint analysis
- Jumps: list the 10 largest absolute returns and check each against the news
- Flat bars, zero volume, repeated prices: holidays filled with the last price or stale quotes
- Units and currency: RON, EUR, USD; index points; prices in cents
- Keep the raw file and the download date; never edit data manually

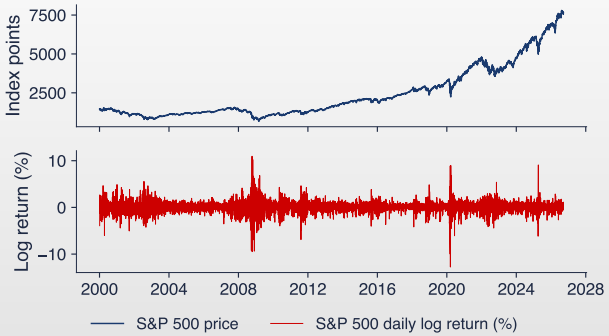
Recap: Data

- Prices come from exchanges, vendors and central banks; the primary source of a number is the reference
- Use adjusted prices for stocks, total return indices when dividends matter
- Real data contain errors: check jumps, flat bars and a second source
- Survivorship, look-ahead and data snooping inflate results silently

Returns as the Object of Analysis

- A price has a unit and a scale; a return does not
  - a 1-leu move is large for a 2-leu share and small for a 500-leu share
- Returns of different assets, currencies and periods can be compared directly
- Prices trend and wander; returns fluctuate around a stable level
  - **weakly stationary** series: constant mean, constant variance, and a covariance that depends only on the lag
  - most statistical methods of this course assume (at least) weak stationarity
- The investor cares about the relative change of wealth, which is exactly a return

S&P 500: Price Level vs Daily Returns



- Top: the price trends upward with long swings; its mean and spread change over time
- Bottom: the returns fluctuate around zero, in calm and turbulent periods (2008, 2020)

SFM_ch1_prices_returns

A Formal Check: the ADF Test (1/2)

- **ADF test** (augmented Dickey–Fuller) Dickey & Fuller (1979): is there a unit root (a random walk) in the series?
  - H_0: unit root, the series is non-stationary; H_1: the series is stationary
- The test regresses the daily change of the series on its previous level

$$\Delta y_t = \alpha + \beta t + \gamma\, y_{t-1} + \sum_{j=1}^{p} \delta_j\, \Delta y_{t-j} + \varepsilon_t$$

- Notation
  - y_t: the tested series (log price or log return); $\Delta y_t = y_t - y_{t-1}$: its change from day $t-1$ to day t
  - α: constant; βt: linear trend; δ_j: coefficients of the p lagged changes, which absorb short-term autocorrelation; ε_t: error
  - $\gamma = 0$: the series does not return to a level (unit root); $\gamma < 0$: it is pulled back (stationary)
- Reject H_0 when the t-statistic of $\hat{\gamma}$ is below the critical value (Dickey–Fuller tables, not Student)

A Formal Check: the ADF Test (2/2)

- S&P 500, daily data since 2000
  - log price (with trend): statistic -2.08, 5% critical value -3.41, p-value 0.56: we do not reject the unit root
  - log returns: statistic -14.3, 5% critical value -2.86, p-value < 0.001: we reject it
- Reading
  - p-value: the probability, under H_0, of a statistic at least as negative as the observed one; below 0.05 we reject H_0
  - prices behave like a random walk; returns are stationary, so we model returns, not prices
- Unit roots and random walks are the subject of Chapter 7

SFM_ch1_prices_returns

The Simple Return

- **Simple (net) return** from $t-1$ to t Franke, Härdle & Hafner (2019), Tsay (2010)
  - $R_t = \dfrac{P_t - P_{t-1}}{P_{t-1}} = \dfrac{P_t}{P_{t-1}} - 1$
  - **gross return**: $1 + R_t = P_t / P_{t-1}$
- Bounded below: $R_t \geq -1$ (limited liability: you cannot lose more than you invested)
- Example: $P_{t-1} = 100$, $P_t = 110$: $R_t = 10\%$
- It is what an investor sees on the account statement: "the fund gained 10%"
- Notation: from here on P_t is the close of an index or the adjusted close of a stock

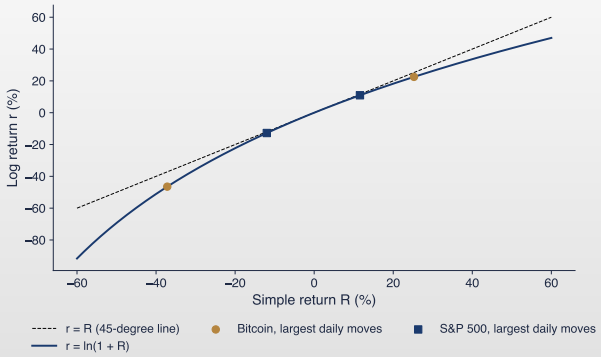
The Log Return

- **Log return** (continuously compounded return)
  - $r_t = \ln \dfrac{P_t}{P_{t-1}} = \ln P_t - \ln P_{t-1} = \ln(1 + R_t)$
  - back to the simple return: $R_t = e^{r_t} - 1$
- Name: e^r is the growth of 1 leu at rate r compounded continuously over the period
- Unbounded: $r_t \in (-\infty, +\infty)$; a total loss ($P_t = 0$) is $r_t = -\infty$
- Example: $P_{t-1} = 100$, $P_t = 110$: $r_t = \ln 1.1 = 9.53\%$
- Log returns are the default in statistics of financial markets: they add up over time (next section)

The Difference between r and R

- Taylor expansion of the logarithm around 0
  - $r = \ln(1+R) = R - \frac{R^2}{2} + \frac{R^3}{3} - \dots$
  - ln is concave, so $r < R$ for every $R \neq 0$
- Daily move, $R = 1\%$
  - $r = 0.995\%$, and $R - R^2/2 = 0.995\%$: the gap $R - r$ is 0.5 basis points
- Large move, $R = 10\%$
  - $r = 9.531\%$, and $R - R^2/2 = 9.500\%$: the gap $R - r$ is 46.9 basis points
- A **basis point** (bp) is 0.01 percentage points
- Conclusion: for daily returns $r \approx R$; for large moves and long horizons they differ

Simple vs Log Returns on Real Extremes



- Bitcoin, 12 March 2020: $R = -37.2\%$ but $r = -46.5\%$
- Bitcoin, 7 December 2017: $R = +25.2\%$, $r = +22.5\%$
- S&P 500, 16 March 2020: $R = -12.0\%$, $r = -12.8\%$
- The curve bends down: losses look larger and gains smaller in log returns

SFM_ch1_simple_log_returns

Worked Example: Up 10%, Then Down 10%

- Price path $100 \to 110 \to 99$
  - simple returns: $+10\%$ and -10%; their sum, 0, is misleading
  - compounded: $1.1 \times 0.9 - 1 = -1.0\%$
- Log returns
  - $r_1 = \ln 1.1 = 9.53\%$, $r_2 = \ln 0.9 = -10.54\%$
  - sum: $-1.01\% = \ln(99/100)$, exact

Rule

- Simple returns compound (multiply); log returns add

Question for the Room: Up 50%, Down 50%

- A share gains 50% one year and loses 50% the next
- **What do you think?** Where is the price after two years, relative to the start?
- **Answer**
  - $1.5 \times 0.5 = 0.75$: the wealth changed by -25%
  - log returns: $\ln 1.5 + \ln 0.5 = \ln 0.75 = -28.8\%$
  - a loss of x needs a gain of $x/(1-x)$ to recover: 50% needs 100%

Choosing the Type of Return

Task	Use	Reason
Aggregating over time	log	$r_t(k) = r_t + \cdots + r_{t-k+\mathbf{1}}$, a sum
Portfolio of several assets	simple	$R_p = \sum_i w_i R_i$, exact
Statistical modelling	log	unbounded, closer to a symmetric distribution
Reporting to investors	simple	"the fund gained 10%" is a simple return
Daily data, small moves	either	the difference is about $R^{\mathbf{2}}/2$

- Always say which one you use: a "return of 10%" is ambiguous

Recap: Simple and Log Returns

- $R_t = P_t/P_{t-1} - 1$; $r_t = \ln(P_t/P_{t-1}) = \ln(1 + R_t)$
- $r < R$ for every move; the gap is about $R^2/2$, negligible for daily data
- Log returns add over time; simple returns add across assets
- Gains and losses of the same size do not cancel

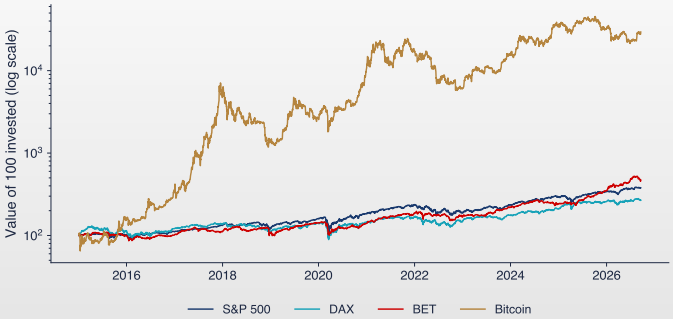
Returns over k Periods

- **Simple**: gross returns multiply
  - $1 + R_t(k) = \dfrac{P_t}{P_{t-k}} = (1 + R_t)(1 + R_{t-1}) \cdots (1 + R_{t-k+1})$
- **Log**: returns add
  - $r_t(k) = \ln \dfrac{P_t}{P_{t-k}} = r_t + r_{t-1} + \cdots + r_{t-k+1}$
- $R_t(k)$, $r_t(k)$: the simple and the log return over the k periods that end at t, from P_{t-k} to P_t
- Consequence: the mean and variance of a k-day log return follow from those of the daily returns
- Weekly, monthly or yearly returns are computed from the prices at the end of each period, not by averaging

Cumulative Return and Wealth

- Wealth after T periods, starting from W_0: $W_T = W_0 \prod_{t=1}^{T}(1 + R_t) = W_0 \exp\left(\sum_{t=1}^{T} r_t\right)$
- $\prod_{t=1}^{T}$: the product of the T gross returns; $\exp(x) = e^x$ undoes the logarithm
- **Cumulative simple return**: $\text{CSR}_T = W_T/W_0 - 1$
  - what the investor gained, in percent
- **Cumulative log return**: $\text{CLR}_T = \ln(W_T/W_0) = \sum_t r_t$
  - link: $\text{CSR}_T = e^{\text{CLR}_T} - 1$
- Log scale on charts: equal vertical distances are equal percentage changes

Growth of 100 Invested, 2015–2026



- 100 invested on 5 January 2015, the first day with prices for all four series (the CAGR tables start each series on its own first day of 2015); final values: S&P 500 379, DAX 267, BET 470, Bitcoin 29 475
- Bitcoin: cumulative simple return 29 375%, cumulative log return 5.69; BET: 370% and 1.55
- Without the log scale, every series except Bitcoin would look flat

SFM_ch1_portfolio_returns

Portfolio Returns

- Portfolio with weights $w_1, \ldots, w_n$, $\sum_i w_i = 1$, set at the start of the period
  - R_i, r_i: the simple and the log return of asset $i = 1, \ldots, n$ over the period; w_i: the share of wealth in asset i
  - **simple**: $R_p = \sum_i w_i R_i$, exact
  - **log**: $r_p = \ln\left(\sum_i w_i e^{r_i}\right) \neq \sum_i w_i r_i$
- Example: 60% in asset 1 ($R_1 = 10\%$), 40% in asset 2 ($R_2 = -5\%$)
  - $R_p = 0.6 \times 10\% + 0.4 \times (-5\%) = 4.0\%$; exact $r_p = \ln(1 + R_p) = 3.92\%$
  - weighted log returns: $0.6 \times 9.53\% + 0.4 \times (-5.13\%) = 3.67\%$, too low
- Weights drift when prices move: a fixed-weight portfolio must be **rebalanced**

SFM_ch1_portfolio_returns

Real Data: a TLV–SNP Portfolio

- 50% Banca Transilvania, 50% OMV Petrom, rebalanced daily, 2015–2026
- One day: exact log return vs weighted sum of log returns
  - mean absolute gap 0.38 basis points; largest gap 13.9 basis points, on 31 May 2016
- Over the whole period the small gaps accumulate
  - exact cumulative log return 2.56, so 1 leu grows to 13.0 lei
  - weighted log returns give 2.47, so only 11.8 lei
- Growth rate: portfolio 24.5% a year; TLV 27.3%; SNP 19.7%
- Rule: build portfolios with simple returns, then take logs if needed

SFM_ch1_portfolio_returns

Recap: Multi-Period and Portfolio Returns

- Over time: multiply gross returns or add log returns
- Wealth $W_T = W_0 e^{\mathrm{CLR}_T}$; $\mathrm{CSR}_T = e^{\mathrm{CLR}_T} - 1$
- Across assets: $R_p = \sum_i w_i R_i$ exactly; the log version is only an approximation
- Small daily approximation errors become visible over a decade

Annualising the Mean and the Volatility

- A = number of observations per year: about 252 for exchanges, 365 for crypto, 52 for weekly, 12 for monthly data
- μ, σ: the mean and the standard deviation of the returns at the frequency of the data (for example, daily)
- **Mean**: log returns add, so $\mu_{\text{year}} = A\,\mu$
  - the same rule is used for the mean of simple returns (arithmetic annualisation)
- **Volatility** = standard deviation of returns: $\sigma_{\text{year}} = \sqrt{A}\,\sigma$, the **square-root-of-time rule**
  - valid when returns are uncorrelated with constant variance: then $\text{Var}(r_1 + \cdots + r_A) = A\,\sigma^2$, and the square root gives $\sqrt{A}\,\sigma$
- S&P 500: daily $\sigma = 1.12\%$, so $\sigma_{\text{year}} = 1.12\% \times 15.87 = 17.7\%$
  - Bitcoin: $\sqrt{365} = 19.10$ gives 66.7%; the stock-market factor $\sqrt{252}$ would give only 55.4%

The Compound Annual Growth Rate

- **CAGR** (compound annual growth rate): the constant yearly rate that turns P_0 into P_T in Y years
  - $\text{CAGR} = \left(P_T/P_0\right)^{1/Y} - 1$, with Y in calendar years
- Link with log returns: CAGR $=$ exp(annual mean log return) $- 1$
- S&P 500, 2015–2026: mean log return 11.2% a year, $e^{11.2\%} - 1 = 11.9\%$, the CAGR is 11.9%
- The CAGR uses only the first and the last price: it says nothing about the path in between
- Also called the geometric mean return

Precision of the Estimated Mean

- Standard error of the annual mean: $\mathrm{SE}(\hat{\mu}_{\text{year}}) = \sigma_{\text{year}}/\sqrt{Y}$, with Y years of data
  - $\hat{\mu}_{\text{year}}$: the estimated annual mean (a hat marks an estimate from the sample)
  - it depends on the number of **years**, not on the number of observations
  - 95% confidence interval: $\hat{\mu}_{\text{year}} \pm 1.96\,\mathrm{SE}$, with 1.96 the 97.5% quantile of the standard Normal distribution
- S&P 500, 11.7 years: annual mean log return 11.2%, SE 5.2%
  - 95% CI (confidence interval): [1.1%, 21.4%]
- Bitcoin: $47.4\% \pm 1.96 \times 19.5\%$
  - the interval is [9.2%, 85.6%]: very wide
- Volatility, by contrast, is estimated precisely from daily data

SFM_ch1_descriptive_stats

Question for the Room: More Data, Better Mean?

- You replace 10 years of daily S&P 500 data by 10 years of 5-minute data: 78 bars in a 6.5-hour trading day, so 78 times more observations
- **What do you think?** Does the standard error of the annual mean return become smaller?
- **Answer**
  - No: the mean over 10 years is fixed by the first and last price, $\hat{\mu} = \ln(P_T/P_0)/Y$
  - only more **years** reduce its standard error, as $1/\sqrt{Y}$
  - the volatility estimate, in contrast, does improve with higher frequency

Recap: Annualisation

- Mean $\times A$, volatility $\times\sqrt{A}$, with A the actual observations per year
- $\sqrt{A}$ assumes uncorrelated returns with constant variance
- CAGR $= (P_T/P_0)^{1/Y} - 1$, the exponential of the annual mean log return minus 1
- Mean returns are imprecise: SE $= \sigma_{\text{year}}/\sqrt{Y}$

Location and Spread

- Sample $r_1, \ldots, r_n$ of daily log returns (in %)
- **Mean**: $\bar{r} = \frac{1}{n} \sum_{t=1}^{n} r_t$, the average daily return
- **Variance**: $\hat{\sigma}^2 = \frac{1}{n-1} \sum_{t=1}^{n} (r_t - \bar{r})^2$; **standard deviation** $\hat{\sigma} = \sqrt{\hat{\sigma}^2}$
  - in finance the standard deviation of returns is called **volatility**
- **Quantile** q_α: the value below which a fraction α of returns lie
  - median $q_{0.5}$; the 1% quantile $q_{0.01}$ is the basis of VaR 1% (Chapter 10)
- **Minimum** and **maximum**: the worst and best day, always with their dates

Shape: Skewness and Kurtosis

- **Skewness**: $\hat{S} = \frac{1}{n}\sum_t (r_t - \bar{r})^3 / \hat{\sigma}^3$
  - the cube keeps the sign of each deviation, so large losses push $\hat{S}$ below 0 and large gains above 0
  - $S = 0$: symmetric; $S < 0$: a longer left tail, large losses more frequent than large gains
- **Kurtosis**: $\hat{K} = \frac{1}{n}\sum_t (r_t - \bar{r})^4 / \hat{\sigma}^4$; **excess kurtosis** $\hat{K} - 3$
  - the fourth power makes the extreme days dominate $\hat{K}$
  - the Normal distribution has $K = 3$, so excess kurtosis 0
  - excess kurtosis > 0: **heavy tails**, extreme days more frequent than under the Normal distribution
- Both are sensitive to single extreme days: report them with the extremes
- Heavy tails and asymmetry are **stylised facts** of returns Cont (2001); formal tests in Chapter 2

Daily Log Returns: Indices and Bitcoin, 2015–2026

Series	Mean	Std. dev.	Skew.	Exc. kurt.	Minimum (date)	Maximum
S&P 500	0.045	1.12	−0.64	15.8	−12.8 (16 Mar 2020)	9.1
DAX	0.032	1.20	−0.56	10.0	−13.1 (12 Mar 2020)	10.4
BET	0.053	0.98	−1.41	17.8	−11.9 (19 Dec 2018)	6.8
BET-TR	0.080	0.98	−1.43	18.1	−11.9 (19 Dec 2018)	6.7
Bitcoin	0.130	3.49	−0.73	12.2	−46.5 (12 Mar 2020)	22.5

- All values in % except skewness and excess kurtosis; each series on its own calendar
- Bitcoin: daily standard deviation about three times that of the stock indices
- All series: negative skewness and large excess kurtosis
- The worst days cluster in March 2020 (COVID-19) and on 19 December 2018 for the BET

SFM_ch1_descriptive_stats

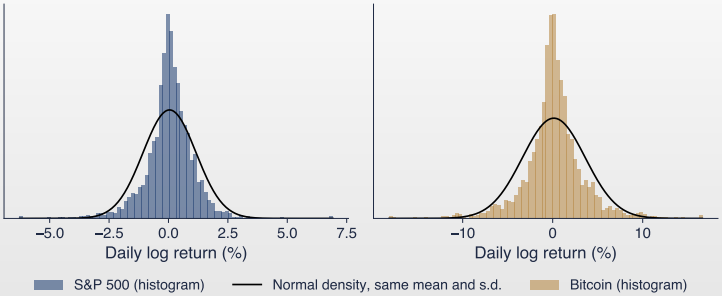
Daily Log Returns: Five BVB Stocks, 2015–2026

Series	Mean	Std. dev.	Skew.	Exc. kurt.	Minimum (date)	Maximum
Banca Transilvania (TLV)	0.106	1.69	−0.55	25.2	−22.2 (19 Dec 2018)	20.5
OMV Petrom (SNP)	0.079	1.54	−0.66	9.2	−15.0 (9 Mar 2020)	10.0
BRD	0.075	1.61	−0.51	9.2	−18.2 (19 Dec 2018)	10.3
Nuclearelectrica (SNN)	0.118	1.53	0.29	5.2	−9.4 (9 Mar 2020)	11.3
Transgaz (TGN)	0.091	1.55	0.01	5.3	−10.1 (19 Dec 2018)	11.1

- Single stocks are more volatile than the BET index: diversification at work
- Banca Transilvania and BRD lost most on 19 December 2018, after the announcement of a tax on bank assets, adopted as OUG 114/2018
- Excess kurtosis differs widely between stocks: compare TLV with SNN
- Data issue: the EODHD adjusted close of TLV on 30 May 2016 is wrong (−15.0%, then +20.5% on 31 May); kept here, without these two days the excess kurtosis is 18.1
- Crashes on the BVB have their own statistics Pele & Mazurencu-Marinescu (2012)

SFM_ch1_descriptive_stats

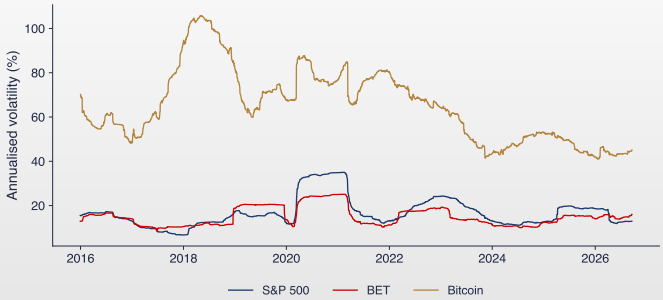
Histograms vs the Normal Distribution



- Black line: the Normal density with the same mean and standard deviation; horizontal axis cut at the 0.1% and 99.9% quantiles, the days beyond them are counted in the outermost bars
- S&P 500: 79.6% of days lie within one standard deviation of the mean (Normal: 68.3%)
- Beyond three standard deviations: 1.39% of days vs 0.27% under the Normal distribution, 5.2 times more

SFM_ch1_descriptive_stats

Time-Varying Volatility



- Rolling one-year volatility, annualised: S&P 500 between 7% and 35%; Bitcoin between 41% and 106%
- One number for the whole sample hides calm and turbulent periods
- Volatility estimators and GARCH models: Chapters 8 and 9

SFM_ch1_descriptive_stats

Recap: Descriptive Statistics

- Report mean, standard deviation, skewness, excess kurtosis, quantiles and extremes with dates
- Daily returns: mean near 0, volatility 1.0–1.7% a day (Bitcoin 3.5%), heavy tails, often negative skewness
- Extreme days are several times more frequent than under the Normal distribution
- Volatility is not constant over time

Limits of a Single Performance Indicator

- Return alone rewards risk taking: a leveraged bet can show a high return for years
- Three questions an investor asks
  - how much did it grow? (CAGR)
  - how bumpy was the ride? (volatility, downside deviation)
  - how deep was the worst fall? (maximum drawdown)
- **Risk-adjusted** indicators divide a return by a risk measure: Sharpe, Sortino, Calmar
- All are estimates from one sample: each has an estimation error

The Sharpe Ratio

- **Sharpe ratio** Sharpe (1966), Sharpe (1994): excess return per unit of total risk
  - $\mathrm{SR} = \dfrac{\mu - r_f}{\sigma}$; μ mean return, r_f risk-free rate, σ volatility
  - from daily to annual: $\mathrm{SR}_{\mathrm{year}} = \sqrt{A}\,\mathrm{SR}_{\mathrm{day}}$
- Here: mean of daily simple returns, $r_f = 0$
  - the assets are in RON, EUR and USD, each with its own risk-free rate
  - S&P 500: daily SR 0.046, annual SR 0.72
- Same Sharpe ratio, same return per unit of risk, whatever the leverage



William F. Sharpe, Nobel Prize in Economics 1990

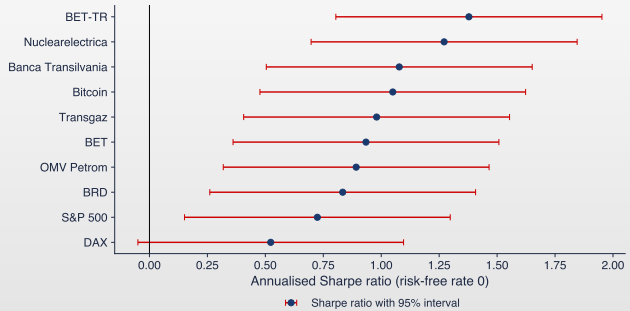
Photo: Larry D. Moore (2007); CC BY 4.0; Wikimedia Commons

Estimation Error of the Sharpe Ratio

- Standard error for i.i.d. (independent, identically distributed) returns Lo (2002)
  - $\mathrm{SE}(\widehat{\mathrm{SR}}) \approx \sqrt{(1 + \mathrm{SR}^2/2)/n}$ per period; annual: multiply by $\sqrt{A}$
  - $\widehat{\mathrm{SR}}$: the Sharpe ratio estimated from the sample; n: the number of returns; A: returns per year
  - roughly $1/\sqrt{Y}$ for the annual Sharpe ratio: about 0.3 with 12 years of data
- S&P 500, 2015–2026: SR $= 0.72$, SE 0.29
  - 95% CI [0.15, 1.30]
- Heavy tails and volatility clustering make the true error larger than this formula
- Choosing the best of many strategies inflates the winner's Sharpe ratio Bailey & López de Prado (2014)

SFM_ch1_performance

Sharpe Ratios with 95% Intervals, 2015–2026



- Highest point estimate: BET-TR
- All intervals overlap: the ranking is uncertain; overlap is not a formal test, which compares the difference of two Sharpe ratios with its own standard error
- The DAX interval contains 0
- Interval widths are almost equal: they depend mainly on the length of the sample

SFM_ch1_performance

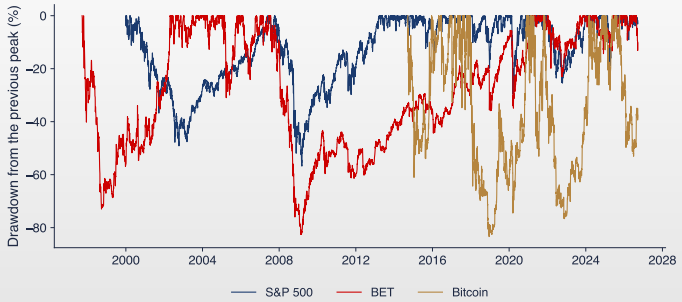
The Sortino Ratio

- Investors dislike losses, not gains: the volatility treats both the same
- **Downside deviation** below a target τ (here $\tau = 0$)
  - $\sigma_D = \sqrt{\frac{1}{n}\sum_t \min(R_t - \tau, 0)^2}$, only days below the target count
  - $\min(R_t - \tau, 0)$: the shortfall below the target on day t, zero on days above it; n: all days of the sample
- **Sortino ratio** Sortino & van der Meer (1991): $\text{Sortino} = \frac{\mu - \tau}{\sigma_D}$, annualised like the Sharpe ratio
- For a symmetric distribution $\sigma_D \approx \sigma/\sqrt{2}$, so Sortino $\approx 1.4\times$ Sharpe
- S&P 500, 2015–2026: Sharpe 0.72, Sortino 1.02

Drawdown and Maximum Drawdown

- **Drawdown**: the fall from the highest price reached so far
  - $\mathrm{DD}_t = \dfrac{P_t}{\max_{s \le t} P_s} - 1 \le 0$; zero at a new maximum
  - $\max_{s \le t} P_s$: the highest price over all days s up to day t
- **MDD** (maximum drawdown): $\mathrm{MDD} = \min_t \mathrm{DD}_t$, the worst peak-to-trough fall
  - also report the dates: peak, trough and **recovery** (back to the old peak)
- A loss x needs a gain $x/(1-x)$ to recover
  - 10% needs 11%; 20% needs 25%; 50% needs 100%; 80% needs 400%
- The MDD depends on the path and grows with the length of the sample Magdon-Ismail et al. (2004)

Drawdowns over the Full History



- S&P 500: MDD −56.8%, 9 Oct 2007 to 9 Mar 2009, recovered on 28 Mar 2013
- BET (since 1997): MDD −82.5%, 24 Jul 2007 to 25 Feb 2009, recovered only on 16 Mar 2021
- Bitcoin: MDD −83.4%, 16 Dec 2017 to 15 Dec 2018

SFM_ch1_performance

Reading the Drawdown Chart

- Depth
  - after a drawdown of -82.5%, the BET needed $+473\%$ to recover; Bitcoin, $+502\%$
- Duration
  - the BET needed 13.6 years to regain its 2007 peak
  - share of days more than 20% below the peak: S&P 500 27%, BET 56%, Bitcoin 66%
- Volatility and Sharpe ratio ignore the order of returns; the drawdown does not
- In the comparison period 2015–2026, the MDD of each stock index happened in February–March 2020; Bitcoin and several BVB stocks had theirs at other dates

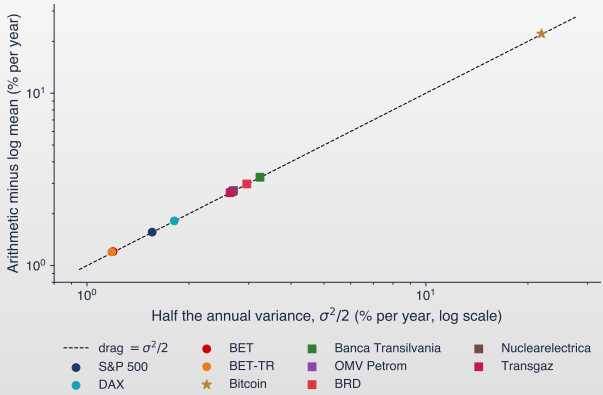
The Calmar Ratio

- **Calmar ratio**: growth per unit of the worst fall, $\text{Calmar} = \dfrac{\text{CAGR}}{|\text{MDD}|}$
- Popular with fund managers: the MDD is the loss an investor actually lives through
- S&P 500, 2015–2026: CAGR 11.9%, MDD -33.9%
  - Calmar $= 11.9/33.9 = 0.35$
- A single event decides the denominator: the Calmar ratio is even noisier than the Sharpe ratio
- Compare Calmar ratios only over the same period

Volatility Drag

- For a log return with mean m and variance σ^2, the simple return has mean $\approx m + \sigma^2/2$
  - so the growth rate is $\mu_{\text{log}} \approx \mu_{\text{arith}} - \sigma^2/2$
  - μ_{arith}: mean of the simple returns; μ_{log}: mean of the log returns, the growth rate of wealth
  - the gap $\sigma^2/2$ is the **volatility drag** Booth & Fama (1992); it comes from the concavity of ln (Jensen's inequality)
- Two assets with the same arithmetic mean, 10% a year
  - $\sigma = 5\%$: drag 0.12%, growth 9.88% a year; after 10 years 1 leu becomes 2.68
  - $\sigma = 30\%$: drag 4.50%, growth 5.50% a year; after 10 years 1 leu becomes 1.73
- Doubling the volatility quadruples the drag

Volatility Drag on Real Data



- Each point: one asset, 2015–2026
- Vertical: arithmetic minus log annual mean; horizontal: $\sigma^2/2$
- All points lie on the line: the formula works on real data
- Bitcoin: arithmetic mean 69.5%, log mean 47.4%, drag 22.1 percentage points; $\sigma^2/2 = 21.9\%$

SFM_ch1_volatility_drag

Question for the Room: Which Bitcoin Return?

- Bitcoin, 2015–2026: arithmetic mean of simple returns 69.5% a year; CAGR 60.6%
- **What do you think?** Which number describes the growth of an investor's wealth?
- **Answer**
  - the CAGR: it is the rate that actually turned the first price into the last
  - the arithmetic mean is the expected return of a single year, not the growth over many years
  - the volatility drag of 22.1 points separates the arithmetic mean from the mean log return, 47.4%; the CAGR is $e^{47.4\%} - 1 = 60.6\%$

Performance Indicators: Indices and Bitcoin, 2015–2026

	CAGR %	Vol. %	Sharpe	Sortino	MDD %	Calmar	Drag %
S&P 500	11.9	17.6	0.72	1.02	−33.9	0.35	1.56
DAX	8.5	19.0	0.52	0.73	−38.8	0.22	1.82
BET	14.1	15.4	0.93	1.28	−31.1	0.45	1.21
BET-TR	22.1	15.4	1.38	1.92	−31.1	0.71	1.20
Bitcoin	60.6	66.2	1.05	1.55	−83.4	0.73	22.13

- Annualised with the actual frequency of each series; Vol.: volatility of daily simple returns; Sharpe and Sortino with $r_f = 0$
- Bitcoin: by far the highest CAGR, the highest volatility and the deepest drawdown
- BET-TR vs BET: same risk, dividends lift every return-based indicator

SFM_ch1_performance

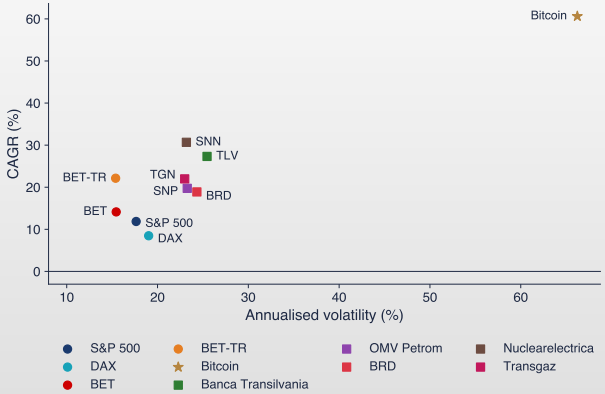
Performance Indicators: BVB Stocks, 2015–2026

	CAGR %	Vol. %	Sharpe	Sortino	MDD %	Calmar	Drag %
Banca Transilvania (TLV)	27.3	25.5	1.08	1.59	−38.9	0.70	3.25
OMV Petrom (SNP)	19.7	23.3	0.89	1.29	−45.4	0.43	2.73
BRD	18.9	24.3	0.83	1.21	−35.6	0.53	2.98
Nuclearelectrica (SNN)	30.7	23.2	1.27	2.00	−36.8	0.83	2.69
Transgaz (TGN)	22.0	23.0	0.98	1.49	−44.8	0.49	2.65

- Adjusted prices: dividends included
- Single stocks: higher volatility and deeper drawdowns than the BET index
- Remember the standard error, about 0.29 for each Sharpe ratio: small differences are not significant

SFM_ch1_performance

Risk and Return, 2015–2026



- Horizontal: annualised volatility; vertical: CAGR
- Best by CAGR: Bitcoin; by Sharpe: BET-TR; by Sortino: SNN; by Calmar: SNN
- Lowest volatility: BET-TR; shallowest MDD: BET-TR (−31.115%) and BET (−31.124%), practically tied
- The ranking depends on the indicator: always say which one you use

SFM_ch1_performance

Recap: Performance Indicators

- CAGR (growth), volatility (risk), Sharpe and Sortino (return per unit of risk)
- MDD and Calmar: the worst fall, which the investor actually lives through
- Volatility drag: $\mu_{\text{log}} \approx \mu_{\text{arith}} - \sigma^2/2$, growth is below the arithmetic mean
- Every indicator is an estimate: compare differences with their standard errors

An Open Research Question

- **Do Sharpe ratios persist?**
  - if a BVB stock had a high Sharpe ratio in 2015–2020, did it also have one in 2021–2026?
  - if not, ranking funds or stocks by past Sharpe ratios is of little use to an investor
- Why it is open: short samples, few stocks, and a standard error of about 0.4 for a Sharpe ratio estimated on half of our sample
- Related evidence: many published return patterns do not survive new data Harvey, Liu & Zhu (2016)
- AI tools can speed up such a study; they do not replace checking it Wang et al. (2023)

Possible Uses of AI Tools

- **Literature**: list papers on the persistence of performance and summarise their methods
- **Code**: draft a Python function for rolling Sharpe ratios on the course data
- **Robustness**: propose other windows, other indicators (Sortino, Calmar), other markets
- **Explanation**: turn a table of results into a first draft of the interpretation
- Example prompt
  - `Write a Python function that computes the annualised Sharpe ratio of daily simple returns on non-overlapping 3-year windows, with the Lo (2002) standard error.`

What to Check

- Data: adjusted close for stocks, the same calendar for all assets, the right period
- Formulas: the annualisation factor ($\sqrt{252}$ vs $\sqrt{365}$), simple vs log returns, n vs $n-1$
- Numbers: recompute two or three values on small examples or with a second method
- References: every cited paper must exist; check the DOI
- Selection: if you tried ten windows, say so; a backtest tried many times overfits Bailey et al. (2014)

Project Idea

- **Question**: do Sharpe ratios of BVB stocks persist from one period to the next?
  - data: adjusted closes of 10–15 BVB stocks, 2010–2026, from the course data
- Steps
  - split the sample into two halves; compute Sharpe ratios and their standard errors in each
  - measure persistence with the rank correlation between the two halves
  - repeat with Sortino and Calmar ratios
- 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
Simple return	$R_t = P_t/P_{t-\mathbf{1}} - 1$
Log return	$r_t = \ln(P_t/P_{t-\mathbf{1}}) = \ln(1+R_t)$
k-period returns	$1 + R_t(k) = \prod_{j=\mathbf{0}}^{k-\mathbf{1}}(1+R_{t-j}), \quad r_t(k) = \sum_{j=\mathbf{0}}^{k-\mathbf{1}} r_{t-j}$
Portfolio	$R_p = \sum_i w_i R_i$
Annualisation	$\mu_{\mathbf{year}} = A\,\mu, \quad \sigma_{\mathbf{year}} = \sqrt{A}\,\sigma$
CAGR	$(P_T/P_{\mathbf{0}})^{\mathbf{1}/Y} - 1$
Sharpe, Sortino	$(\mu - r_f)/\sigma, \quad (\mu-\tau)/\sigma_D$
MDD, Calmar	$\min_t\left(P_t/\max_{s\le t} P_s - 1\right), \quad \mathrm{CAGR}/\lvert\mathrm{MDD}\rvert$
Volatility drag	$\mu_{\mathbf{log}} \approx \mu_{\mathbf{arith}} - \sigma^{\mathbf{2}}/2$
SE of mean, of Sharpe	$\sigma_{\mathbf{year}}/\sqrt{Y}, \quad \sqrt{(1+\mathrm{SR}^{\mathbf{2}}/2)/n}$

Self-Assessment

- **Question**: a fund reports daily log returns with mean 0.05% and standard deviation 1%; what are its annual mean and volatility?
  - **Answer**: $252 \times 0.05\% = 12.6\%$ and $\sqrt{252} \times 1\% = 15.9\%$
- **Question**: why is the arithmetic mean return of Bitcoin far above its CAGR?
  - **Answer**: the volatility drag $\sigma^2/2$ is large when σ is large
- **Question**: an index falls by 60%; what gain does it need to recover?
  - **Answer**: $0.6/0.4 = 150\%$
- Next: Chapter 2, classical distributions and stylised facts of returns

References (1/2)

- Bailey, D.H., Borwein, J.M., López de Prado, M., Zhu, Q.J. (2014). Pseudo-mathematics and financial charlatanism: the effects of backtest overfitting on out-of-sample performance. *Notices of the American Mathematical Society*, 61(5), 458–471.
- Bailey, D.H., López de Prado, M. (2014). The deflated Sharpe ratio: correcting for selection bias, backtest overfitting, and non-normality. *Journal of Portfolio Management*, 40(5), 94–107.
- Booth, D.G., Fama, E.F. (1992). Diversification returns and asset contributions. *Financial Analysts Journal*, 48(3), 26–32.
- Borak, S., Härdle, W.K., López-Cabrera, B. (2013). *Statistics of Financial Markets: Exercises and Solutions*, 2nd ed. Springer.
- Brown, S.J., Goetzmann, W., Ibbotson, R.G., Ross, S.A. (1992). Survivorship bias in performance studies. *Review of Financial Studies*, 5(4), 553–580.
- Campbell, J.Y., Lo, A.W., MacKinlay, A.C. (1997). *The Econometrics of Financial Markets*. Princeton University Press.
- Cont, R. (2001). Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*, 1(2), 223–236.
- Dickey, D.A., Fuller, W.A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 74(366), 427–431.
- Franke, J., Härdle, W.K., Hafner, C.M. (2019). *Statistics of Financial Markets: An Introduction*, 5th ed. Springer.

References (2/2)

- Harvey, C.R., Liu, Y., Zhu, H. (2016). . . . and the cross-section of expected returns. *Review of Financial Studies*, 29(1), 5–68.
- Lo, A.W. (2002). The statistics of Sharpe ratios. *Financial Analysts Journal*, 58(4), 36–52.
- Magdon-Ismail, M., Atiya, A.F., Pratap, A., Abu-Mostafa, Y.S. (2004). On the maximum drawdown of a Brownian motion. *Journal of Applied Probability*, 41(1), 147–161.
- Pele, D.T., Mazurencu-Marinescu, M. (2012). Modelling stock market crashes: the case of Bucharest Stock Exchange. *Procedia – Social and Behavioral Sciences*, 58, 533–542.
- Sharpe, W.F. (1966). Mutual fund performance. *Journal of Business*, 39(S1), 119–138.
- Sharpe, W.F. (1994). The Sharpe ratio. *Journal of Portfolio Management*, 21(1), 49–58.
- Sortino, F.A., van der Meer, R. (1991). Downside risk. *Journal of Portfolio Management*, 17(4), 27–31.
- 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.