Portfolio Optimization

Financial Data: Stylized Facts

Daniel P. Palomar (2025). *Portfolio Optimization: Theory and Application.* Cambridge University Press.

portfoliooptimizationbook.com

Latest update: 2025-09-15

Outline

Executive Summary

- Understanding the specifics of the data is fundamental in any scientific and engineering domain.
- The first step in finance and financial engineering is to understand financial data.
- The study and characterization of financial data flourished in the 1960s and is now a mature topic.
- Academics and practitioners have identified particularities of the data known as **stylized facts**.
- These slides provide a visual exploratory analysis of financial data based on empirical market data, following (Palomar 2025, chap. 2).

Outline

Stylized Facts in Finance

Stylized Facts in Financial Markets: Characteristics observed across various instruments, markets, and time periods.

Lack of Stationarity

- Financial time series statistics change over time.
- Past returns are not reliable indicators of future performance.

Volatility Clustering

- Large price changes tend to follow large price changes, and small changes follow small changes.
- Documented by Mandelbrot (1963) and Fama (1965).

Absence of Autocorrelations

- Returns often show insignificant autocorrelations.
- Supported by the efficient-market hypothesis (Fama, 1970).

Stylized Facts in Finance

Heavy Tails

- Financial data distributions do not conform to Gaussian distributions.
- Exhibit heavy tails, indicating more extreme outcomes than predicted by Gaussian models.

Gain/Loss Asymmetry

- The distribution of returns is not symmetric.
- Indicates a difference in behavior between gains and losses.

Positive Correlation of Assets

- Returns often positively correlated due to market movements.
- Assets tend to move together with the market.

Outline

Prices

Asset Pricing in Financial Markets

- Asset price denoted by p_t, with t representing discrete time periods.
- Time periods can range from minutes to years.

Logarithmic Transformation

- Logarithm of prices is preferred for modeling:

$$y_t \triangleq \log p_t.$$

- Enhances mathematical convenience and represents a wider dynamic range.

Recommended Textbooks

- For financial data modeling: (Meucci 2005; Tsay 2010; Ruppert and Matteson 2015).
- For multi-asset case: (Lütkepohl 2007; Tsay 2013).

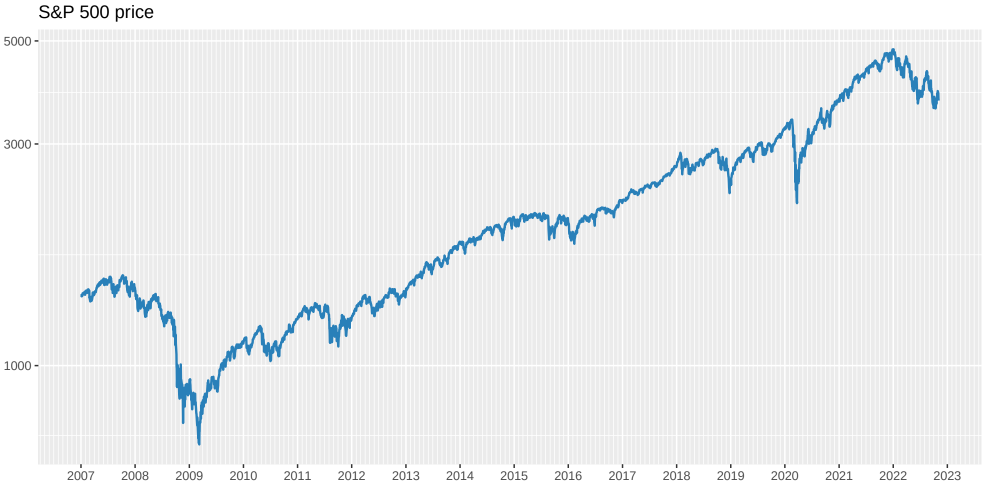
Random Walk Model: Simplest model for log-prices:

$$y_t = \mu + y_{t-1} + \epsilon_t$$

where μ represents the drift and ϵ_t is the i.i.d. random noise.

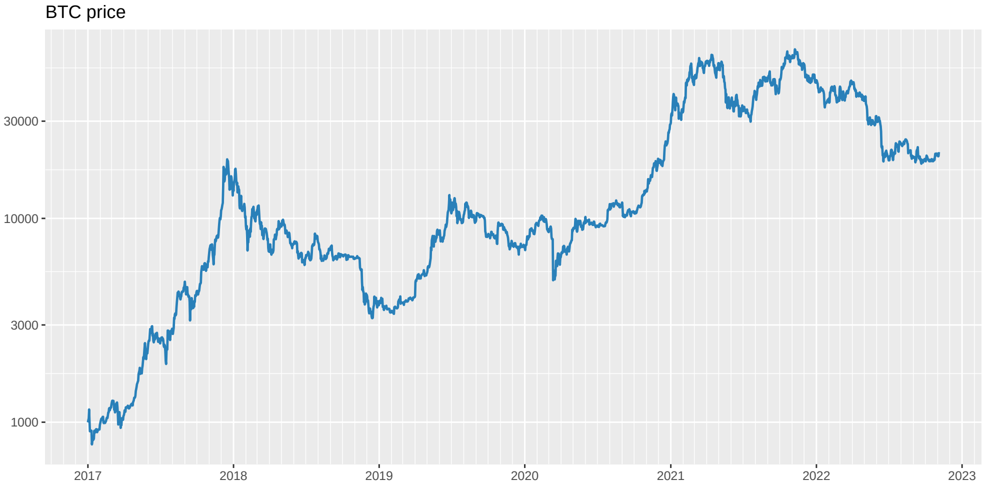
Prices

Price time series of S&P 500:



Prices

Price time series of Bitcoin:



Returns

Price Changes or Returns

- Returns exhibit stationarity and are suitable for mathematical modeling.
- Two main types of returns: linear and log returns.

Linear Return:

$$r_t^{\text{lin}} \triangleq \frac{p_t - p_{t-1}}{p_{t-1}} = \frac{p_t}{p_{t-1}} - 1$$

- Additive among assets, crucial for portfolio return calculations.
- Facilitates analysis of a portfolio's overall return.

Log Return:

$$r_t^{\text{log}} \triangleq y_t - y_{t-1} = \log\left(\frac{p_t}{p_{t-1}}\right)$$

- Additive along the time domain, simplifying time series modeling.
- Stationary according to the random walk model:

$$r_t^{\text{log}} = y_t - y_{t-1} = \mu + \epsilon_t$$

Returns

Relationship Between Returns

- Simple return and log-return are related as

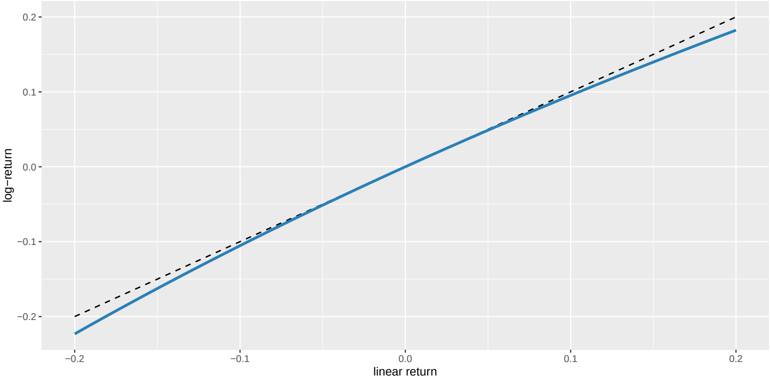
$$r_t^{\text{log}} = \log\left(1 + r_t^{\text{lin}}\right)$$

- Approximation: $r_t^{\text{log}} \approx r_t^{\text{lin}}$ for small r_t^{lin}.
- Accurate for returns less than 5%.

Practical Implications

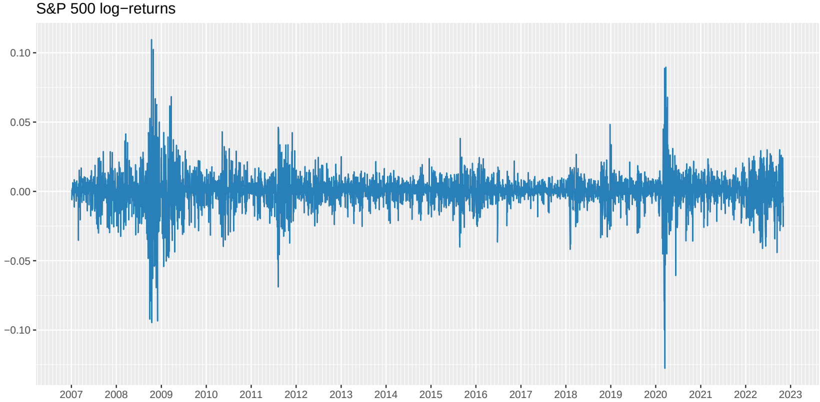
- Linear returns are preferred for portfolio analysis.
- Log returns are favored for time series modeling and mathematical convenience.

Approximation of log-return versus linear return:



Returns

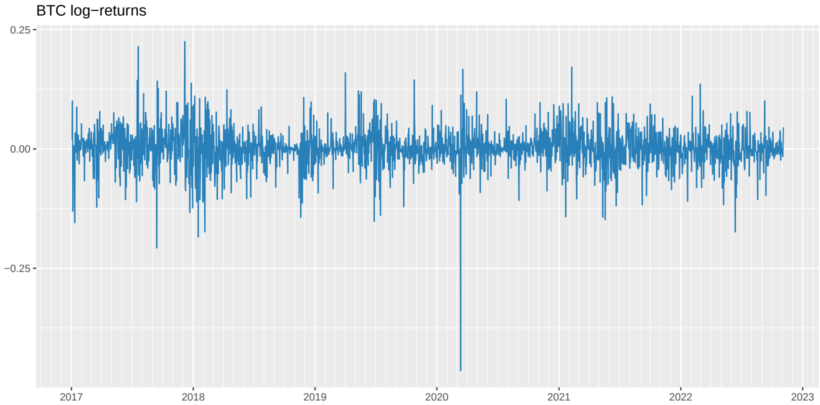
Daily log-return time series of S&P 500:



- High-volatility period during global financial crisis in 2008, as well as the high peak in volatility in early 2020 due to the COVID-19 pandemic.

Returns

Daily log-return time series of Bitcoin:



- Bitcoin flash crash on March 12, 2020, with a drop close to 50% in a single day.

Returns

Volatility Comparison

Volatility is a measure of the dispersion of returns for a given security or market index.

Annualized Volatility Calculation

- For S&P 500: ~21%
- For Bitcoin: ~78%

Volatility Interpretation

- 12% to 20%: considered low volatility.
- Above 30%: considered extremely volatile.

Asset Class Volatility

- S&P 500: classified as low volatility.
- Bitcoin: classified as extremely volatile.

Outline

Non-Gaussianity: Asymmetry and Heavy Tails

Gaussian Distribution Overview

- Commonly used for continuous random variables.
- Characterized by mean (μ) and variance (σ^2).

Probability Distribution Function (pdf)

- Gaussian pdf is given by:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2},$$

- μ is the mean, σ^2 is the variance.

Limitations and Higher-Order Moments

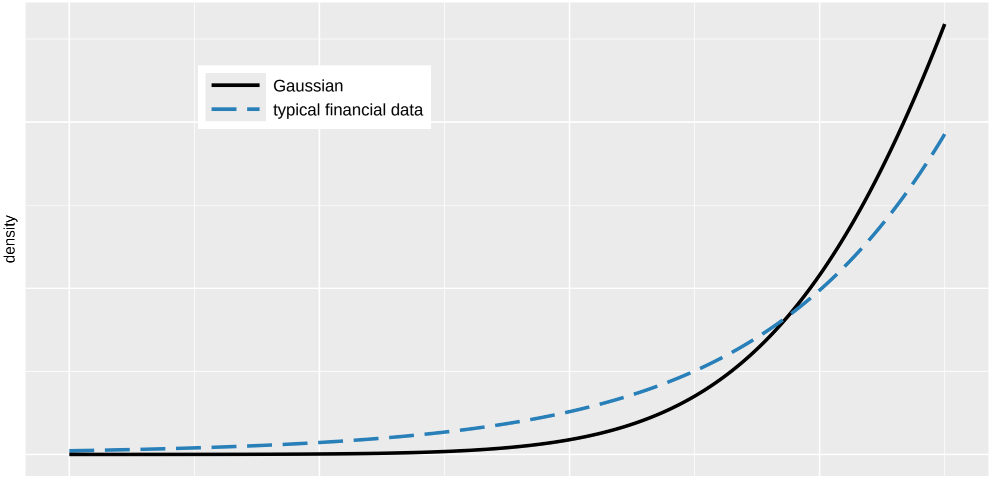
- Financial systems and radar signals often exhibit non-Gaussian characteristics.
- Higher-order moments are necessary for accurate characterization.

Skewness and Kurtosis

- *Skewness*: Measures asymmetry of the distribution.
- *Kurtosis*: Measures tail thickness, indicating tail decay relative to Gaussian distribution.

Non-Gaussianity: Asymmetry and Heavy Tails

Left tail of Gaussian and typical financial data distributions:



Non-Gaussianity: Asymmetry and Heavy Tails

Impact of Skewness and Kurtosis

- Skewness and kurtosis contribute to the likelihood of extreme negative returns.
- Significant for investors holding the asset, as it affects risk assessment.

Financial Data Distribution vs. Gaussian

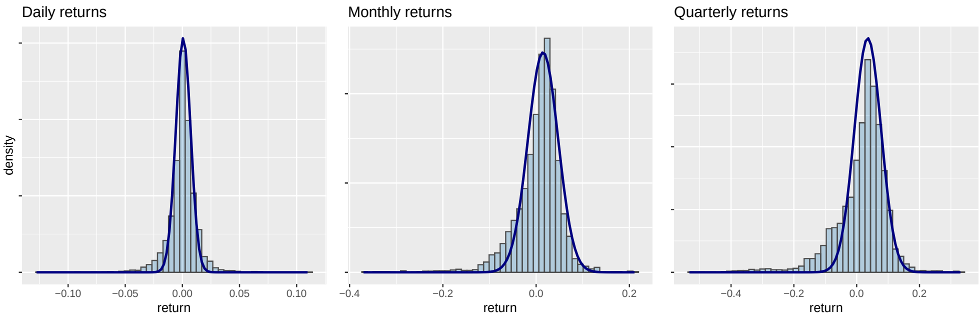
- Financial data return distributions often have fatter left tails compared to Gaussian.
- Illustrated in empirical data comparisons.

Heavy Tails in Distributions

- Distributions with slower tail decay than Gaussian are termed heavy, fat, or thick tails.
- Indicates higher probability of extreme outcomes than predicted by Gaussian models.

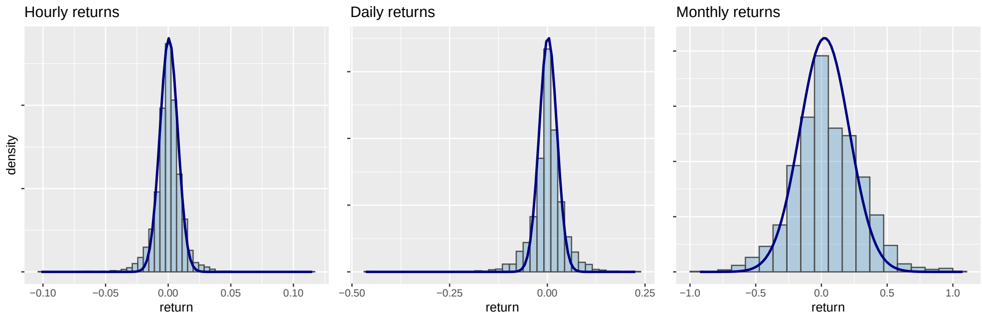
Non-Gaussianity: Asymmetry and Heavy Tails

Histogram of S&P 500 log-returns at different frequencies (with Gaussian fit):



Non-Gaussianity: Asymmetry and heavy tails

Histogram of Bitcoin log-returns at different frequencies (with Gaussian fit):



Non-Gaussianity: Asymmetry and Heavy Tails

Histogram Analysis of S&P 500 Log-Returns

- Displays log-returns at daily, monthly, and quarterly frequencies.
- Tails of the histogram are heavier/thicker than Gaussian distribution.
- Histogram exhibits asymmetry.

Histogram Analysis of Bitcoin Log-Returns

- Also shows clear heavy tails, indicating a deviation from Gaussian.
- Asymmetry is present but less pronounced compared to S&P 500.

Beyond Histograms

- Histograms offer a basic visual inspection of distribution characteristics.
- Other plot of skewness and kurtosis provide clearer insights into asymmetry and heavy-tail properties.

Non-Gaussianity: Asymmetry and Heavy Tails

Understanding Skewness

- Skewness measures the asymmetry of a distribution around its mean.
- Zero skewness indicates symmetry.
- Negative skew: thick tail on the left.
- Positive skew: thick tail on the right.
- Defined as the third standardized moment: $\mathbb{E}\left[\left(\frac{X-\mu}{\sigma}\right)^3\right]$.

Skewness in Financial Data

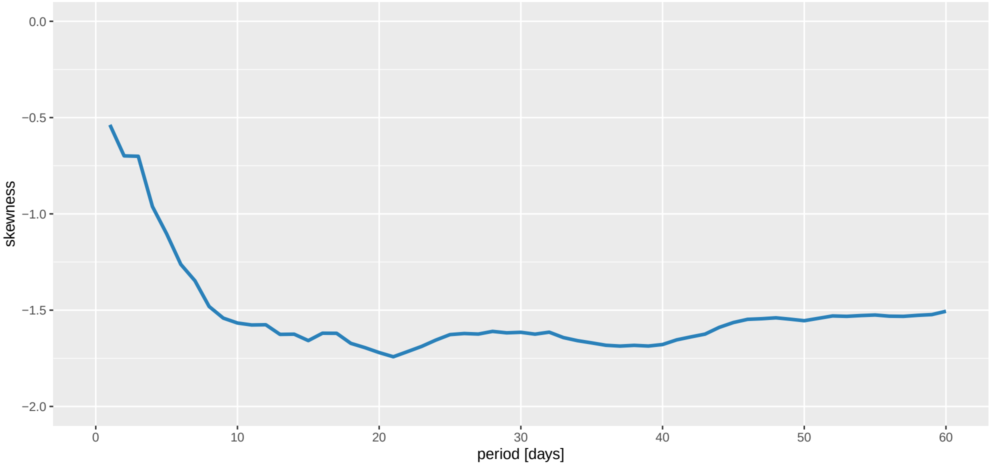
- S&P 500 skewness analysis (2007-2022): skewness decreases as the return period increases from one day to ten days, then saturates.
- Bitcoin skewness analysis (2017-2022): shows a similar trend to S&P 500, and its skewness is closer to zero, indicating more symmetry.

Comparative Insights

Cryptocurrencies, represented by Bitcoin, tend to be more symmetric than stocks, such as those in the S&P 500.

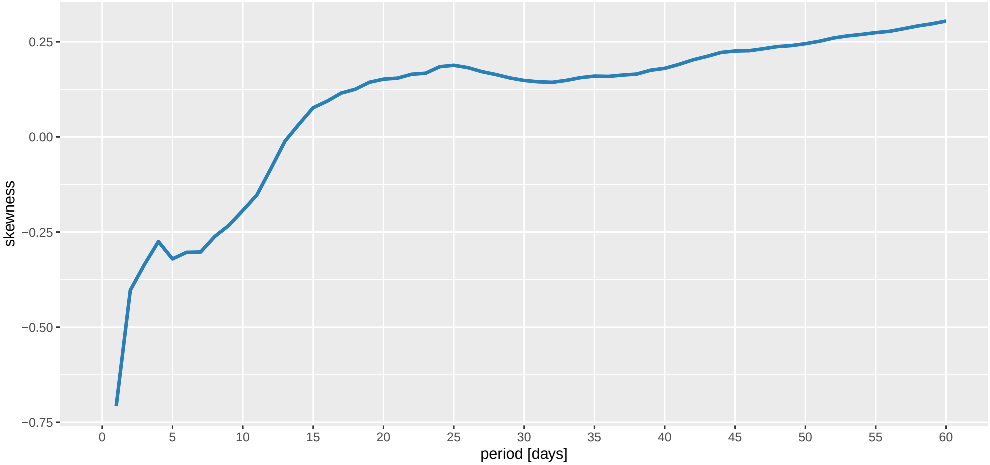
Non-Gaussianity: Asymmetry or Skewness*

Skewness of S&P 500 log-returns:



Non-Gaussianity: Asymmetry or Skewness*

Skewness of Bitcoin log-returns:



Non-Gaussianity: Heavy-Tailness or Kurtosis

Q-Q Plots for Tail Assessment

- Q-Q plots compare quantiles of two distributions.
- Useful for assessing tail behavior relative to Gaussian distribution.

Analysis of Financial Data Q-Q Plots

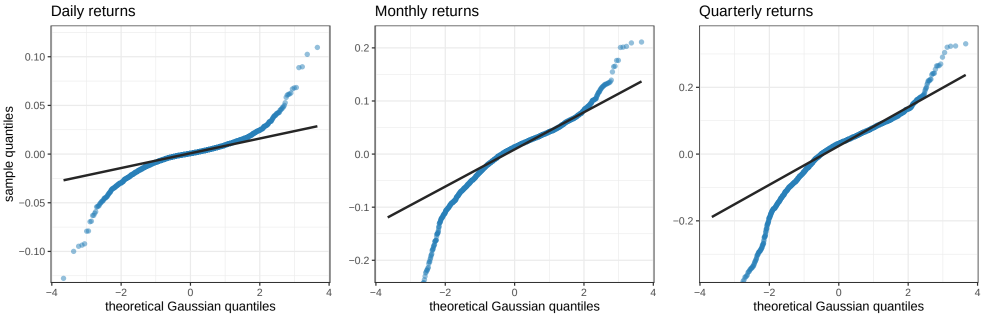
- S&P 500 Q-Q plots show deviation in both left and right tails from the line of equality, indicating the presence of heavy tails in S&P 500 log-returns.
- Bitcoin Q-Q plots show similar deviations in tails, confirming heavy tails in Bitcoin log-returns.

Interpretation of Deviations

- Deviations from the straight line in a Q-Q plot signal departure from Gaussian tail behavior.
- Both S&P 500 and Bitcoin exhibit more extreme returns than a Gaussian distribution would predict.

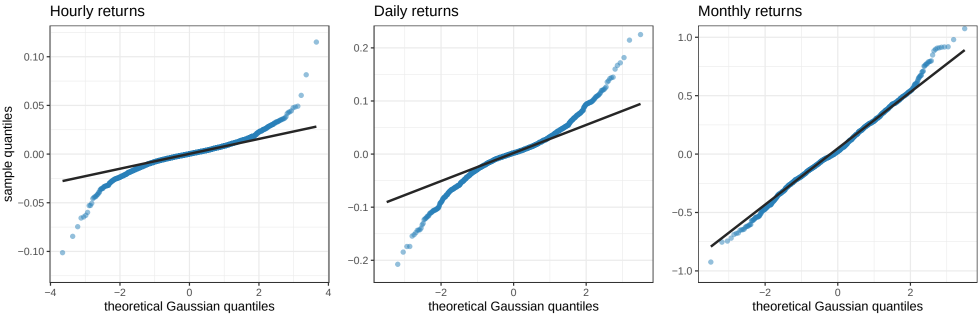
Non-Gaussianity: Heavy-Tailness or Kurtosis

Q-Q plots of S&P 500 log-returns at different frequencies:



Non-Gaussianity: Heavy-Tailness or Kurtosis

Q-Q plots of Bitcoin log-returns at different frequencies:



Non-Gaussianity: Heavy-Tailness or Kurtosis*

Understanding Kurtosis

- Kurtosis measures the “tailedness” of a distribution.
- The kurtosis of a Gaussian distribution is 3.
- Higher kurtosis indicates heavier tails.
- Excess kurtosis is the kurtosis value minus 3.

Kurtosis in Financial Data

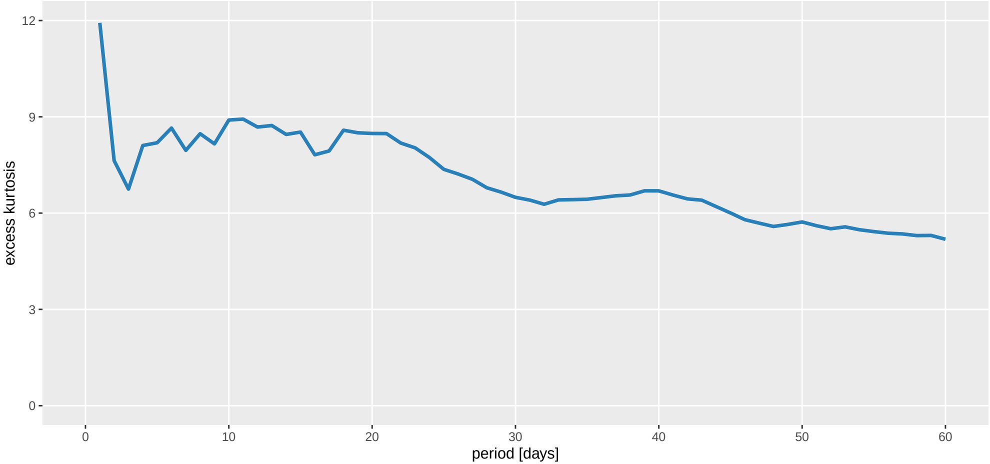
- S&P 500 kurtosis analysis (2007-2022): excess kurtosis decreases rapidly with period increase, then saturates around 6 to 8.
- Bitcoin kurtosis analysis (2017-2022): excess kurtosis decreases rapidly to less than 3 with period increase, which indicates a smaller kurtosis compared to the S&P 500.

Comparative Insights

- Initial observation suggests that cryptocurrencies might be less heavy-tailed than stocks.
- During 2017-2019, the excess kurtosis was 5.41 for the S&P 500 and 3.46 for Bitcoin.
- The year 2020 marked a significant divergence, with a notable increase in Bitcoin's heavy-tailed behavior.
- During 2021-2022, the excess kurtosis was 0.95 for the S&P 500 and 2.34 for Bitcoin.

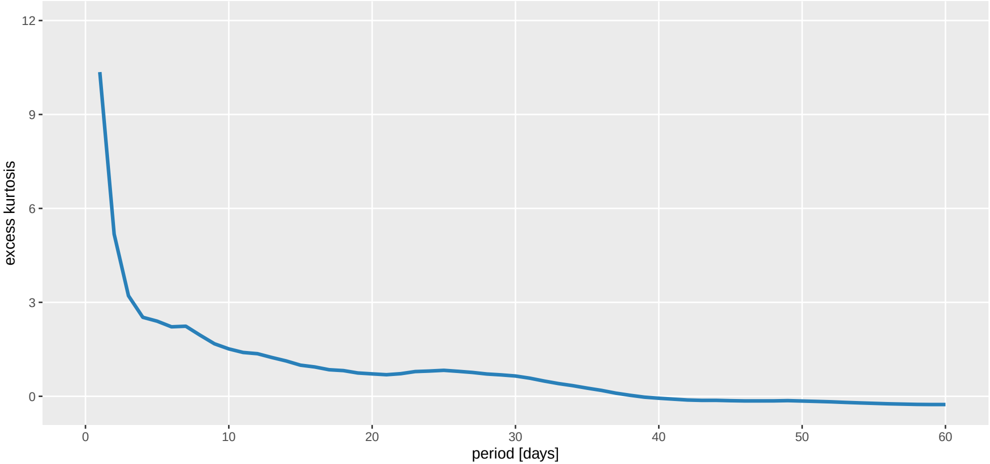
Non-Gaussianity: Heavy-Tailness or Kurtosis*

Excess kurtosis of S&P 500 log-returns:



Non-Gaussianity: Heavy-Tailness or Kurtosis*

Excess kurtosis of Bitcoin log-returns:



Non-Gaussianity: Statistical Tests*

Statistical Tests for Financial Data Characterization

- Financial data exhibit skewness and kurtosis.
- Assessing fit with mean, variance, skewness, and kurtosis requires statistical tests.

Anderson–Darling Statistic

- Measures fit of data to a specific distribution.
- Lower values indicate a better fit.
- Hypotheses:
  - H0: Data follow the specified distribution.
  - H1: Data do not follow the specified distribution.

***p*-Value Interpretation**

- Used to decide if data come from the chosen distribution.
- Thresholds typically range from 0.01 to 0.05.
- Small p-value (< 0.05): strong evidence to reject H0.

Non-Gaussianity: Alternative Distributions*

Results of Anderson–Darling test on financial data, suporting the skewed t distribution:

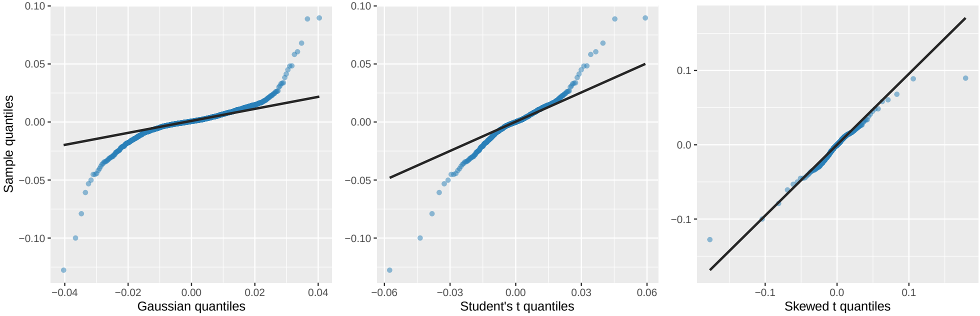
Distribution	Anderson-Darling test	\$p\$-value
Gaussian	55.315	4.17e-07
Student \$t\$	5.4503	0.001751
Skewed \$t\$	2.3208	0.06161

Anderson–Darling Test Results

- Tested distributions: Gaussian, Student t (heavy tails), Skewed t (skewness and heavy tails).
- Skewed t distribution fits S&P 500 data well for 2015-2020.

Non-Gaussianity: Alternative Distributions*

Q-Q plots of S&P 500 log-returns versus different candidate distributions:



Visual inspection via Q-Q plots:

- Comparison of empirical data against Gaussian, Student t, and Skewed t distributions.
- Confirms skewed t distribution as a good fit for the data.

Outline

Efficient-Market Hypothesis

Efficient-Market Hypothesis (EMH) Overview

- States that share prices reflect all available information.
- Argues against the possibility of consistently generating "alpha".
- Suggests stocks always trade at fair value, making undervalued or overvalued purchases impossible.
- Implies that higher returns can only be achieved through riskier investments.

Controversy Surrounding EMH

- EMH is a foundational yet highly debated concept in finance.
- Critics argue that it's possible to find undervalued stocks and predict market trends.
- Evidence against EMH includes successful investors and funds that have outperformed the market consistently.

Efficient-Market Hypothesis

Implications of EMH

- If true, neither technical nor fundamental analysis can consistently produce risk-adjusted excess returns.
- Only insider information could lead to significant risk-adjusted returns.
- Promotes the idea of investing in low-cost, passive portfolios as a more effective strategy.

Opposition to EMH

- Some argue for the feasibility of beating the market through strategic portfolio design.
- The existence of successful market-beating investors and funds challenges the EMH.

Temporal Analysis in Finance

- Investigates whether returns are independent and identically distributed (i.i.d.) or exhibit temporal structure.
- Essential for understanding the feasibility of forecasting returns or prices.
- Relevant textbooks: (Tsay 2010; Cowpertwait and Metcalfe 2009; Ruppert and Matteson 2015).

Linear Structure in Returns

Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF)

- Measure linear dependency in time series data.
- ACF: correlation of a signal with its past values.
- PACF: correlation of a signal with its past values, excluding effects of intermediate lags.

EMH and Temporal Dependency

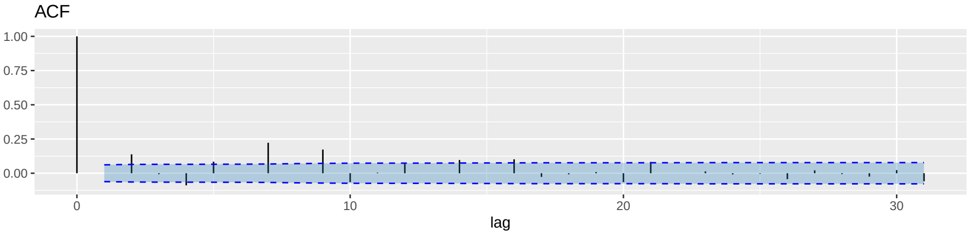
- EMH suggests that there should be no significant autocorrelation in financial time series.
- If EMH holds true, exploiting autocorrelation for forecasting is not feasible.

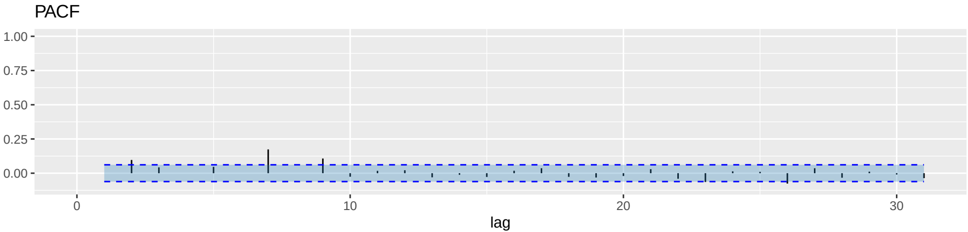
Empirical Findings for S&P 500 and Bitcoin

- For the S&P 500, no significant autocorrelation is detected that could be used for forecasting.
- Its ACF plot shows lags within the statistically insignificant level, except at lag 0.
- For Bitcoin, similar to the S&P 500, no significant autocorrelation is found.
- Hourly returns for Bitcoin also show no significant autocorrelations.

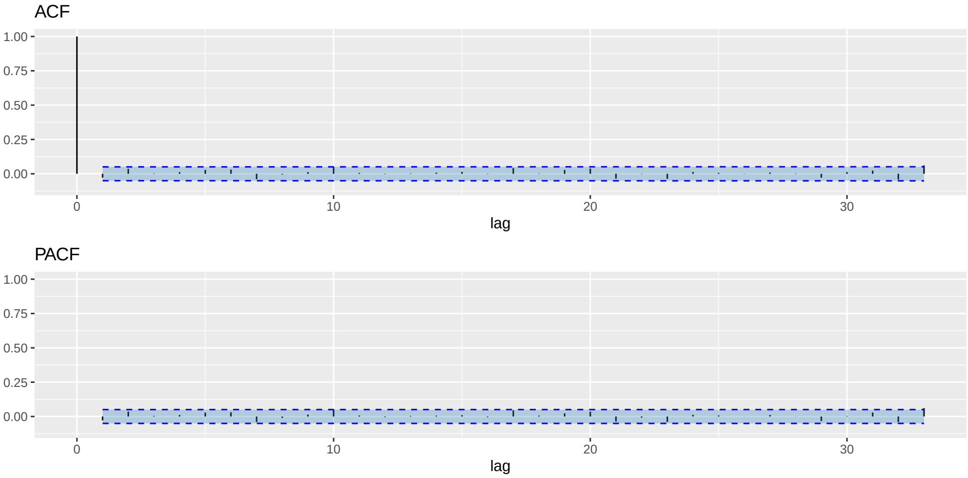
Linear Structure in Returns

Autocorrelation of S&P 500 daily log-returns:





Autocorrelation of Bitcoin daily log-returns:



Nonlinear Structure in Returns

Temporal Structure Beyond Autocorrelations

- Absence of significant autocorrelations does not imply a lack of temporal structure.
- The volatility envelope reveals time-varying standard deviation, indicating structure.

Volatility Clustering

- Periods of high volatility are followed by high volatility, and low by low.
- This indicates that volatility, rather than returns themselves, may have predictable patterns.

Empirical Evidence of Volatility Clustering

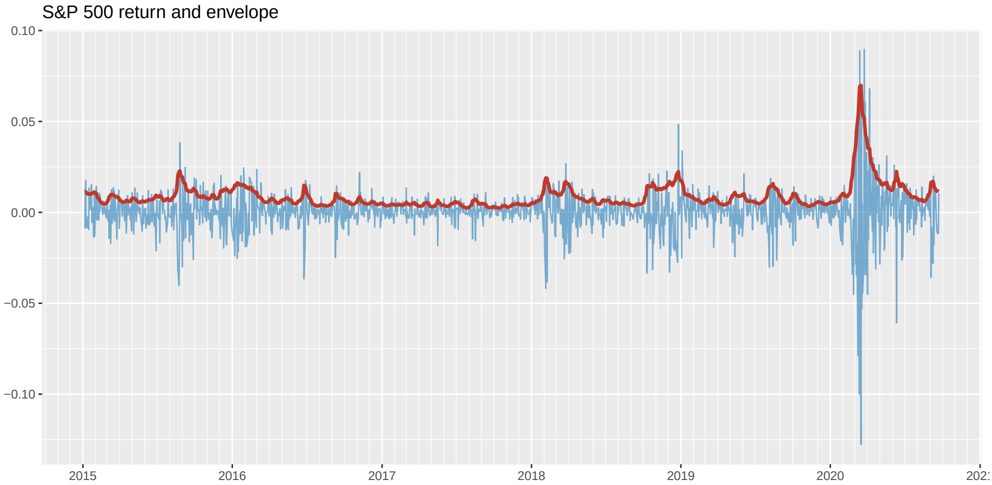
- S&P 500 volatility clustering: the volatility envelope of the returns exhibits clustering.
- Bitcoin volatility clustering: also volatility clustering is observed..

Implications for Forecasting

- Direct forecasting of returns may be challenging, but volatility patterns offer some potential.
- Recognizing and forecasting volatility clustering can enhance trading strategies and risk management.

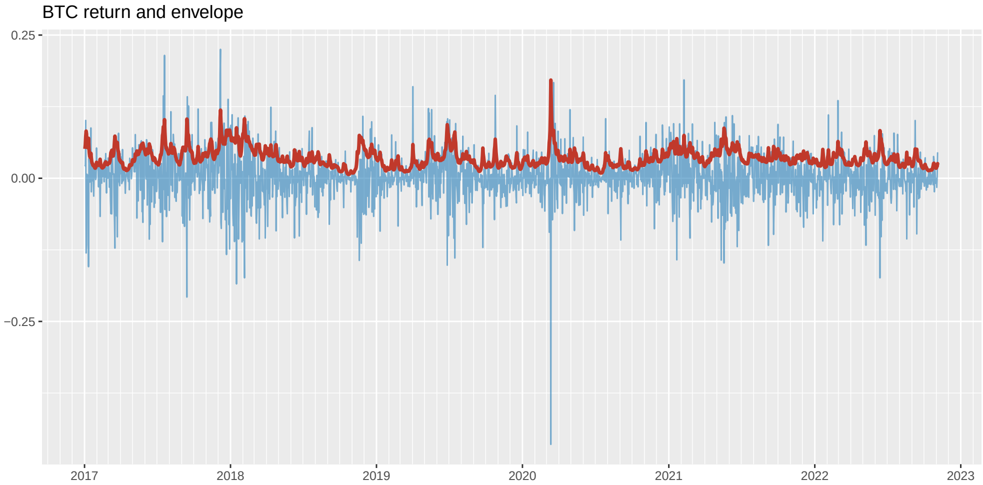
Nonlinear Structure in Returns

Volatility clustering in S&P 500:



Nonlinear Structure in Returns

Volatility clustering in Bitcoin:



Nonlinear Structure in Returns*

Understanding Autocorrelation Limitations

- Autocorrelation assesses linear dependencies, missing nonlinear relationships.
- Nonlinear dependencies in financial time series can be crucial.

Nonlinear Dependencies and Machine Learning

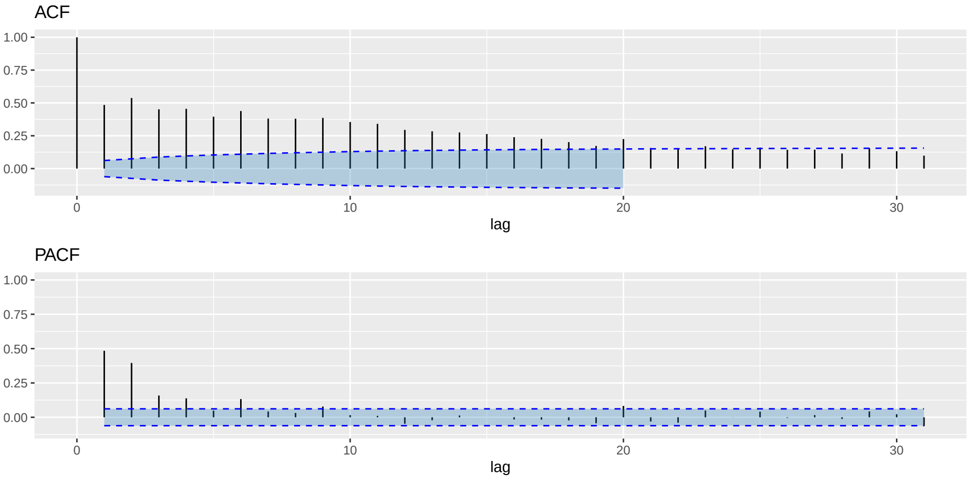
- Machine learning offers tools to uncover and leverage nonlinear dependencies (López de Prado 2018).
- Nonlinear analysis can reveal hidden patterns not detected by traditional methods.

Autocorrelation of Absolute Returns

- Analyzing autocorrelation of absolute returns can expose volatility clustering.
- Provides insight into the magnitude of returns, irrespective of direction.

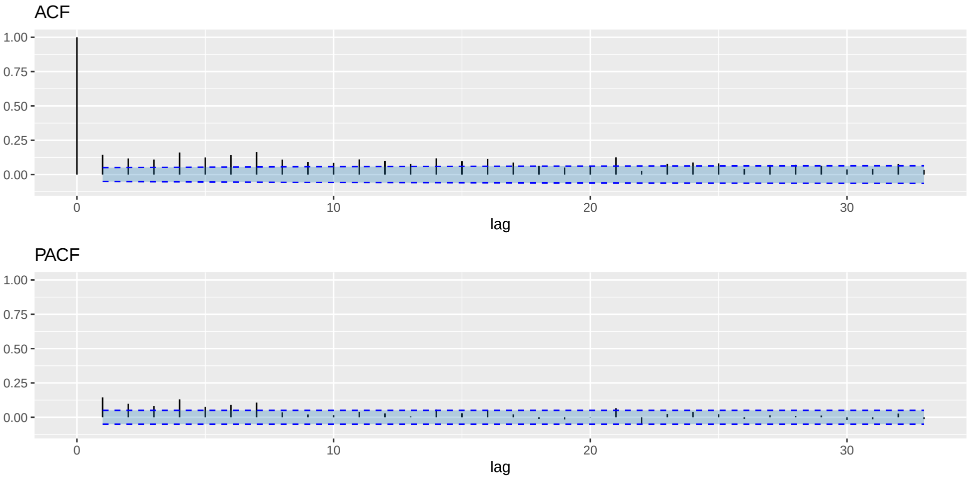
Nonlinear Structure in Returns*

Autocorrelation of absolute value of S&P 500 daily log-returns:



Nonlinear Structure in Returns*

Autocorrelation of absolute value of Bitcoin daily log-returns:



Nonlinear Structure in Returns*

Empirical Findings on Absolute Returns

- Significant autocorrelation is observed in the absolute values of S&P 500 log-returns, which indicates the presence of temporal structure in volatility.
- Bitcoin absolute returns autocorrelation shows significant autocorrelation, though less pronounced than the S&P 500, which suggests that volatility clustering is a common feature across different assets.

Implications for Financial Analysis

- The presence of significant autocorrelation in absolute returns highlights the importance of considering both linear and nonlinear dependencies.
- This insight can improve forecasting models and risk management strategies by accounting for volatility patterns.

Standardized Returns: Removing Volatility Clustering

- Standardized returns are obtained by dividing original returns by their volatility.
- This process aims to remove volatility clustering from the time series.

Nonlinear Structure in Returns*

Benefits of Standardized Returns

- Creates a time series with more uniform volatility.
- Facilitates the analysis of returns independent of their volatility patterns.

Empirical Application to Financial Data

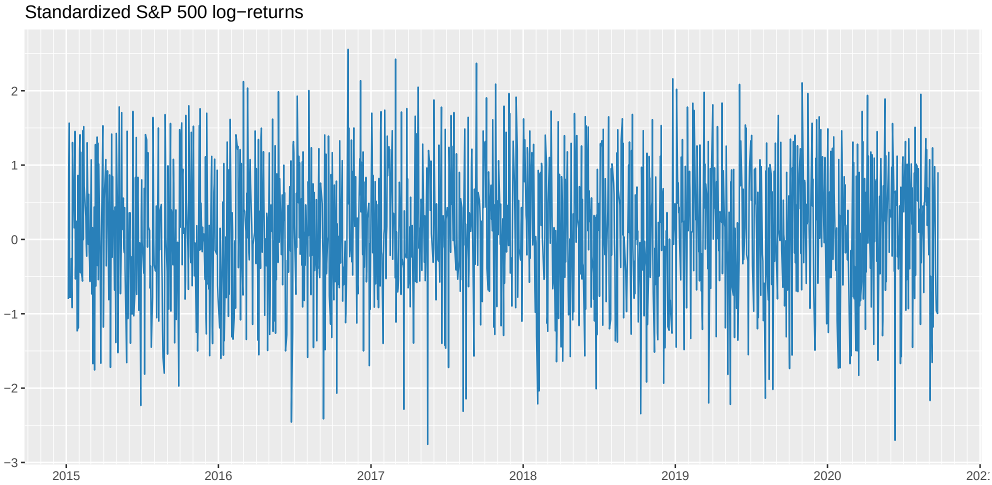
- For S&P 500 standardized returns, illustration shows the removal of volatility clustering, resulting in a more uniform series.
- For Bitcoin standardized returns, a similar process is applied, demonstrating effectiveness across different assets.

Implications for Financial Analysis

- Standardized returns provide a clearer view of the underlying return dynamics, free from the influence of volatility clustering.
- This approach can enhance the accuracy of models that assume homoscedasticity (constant volatility).

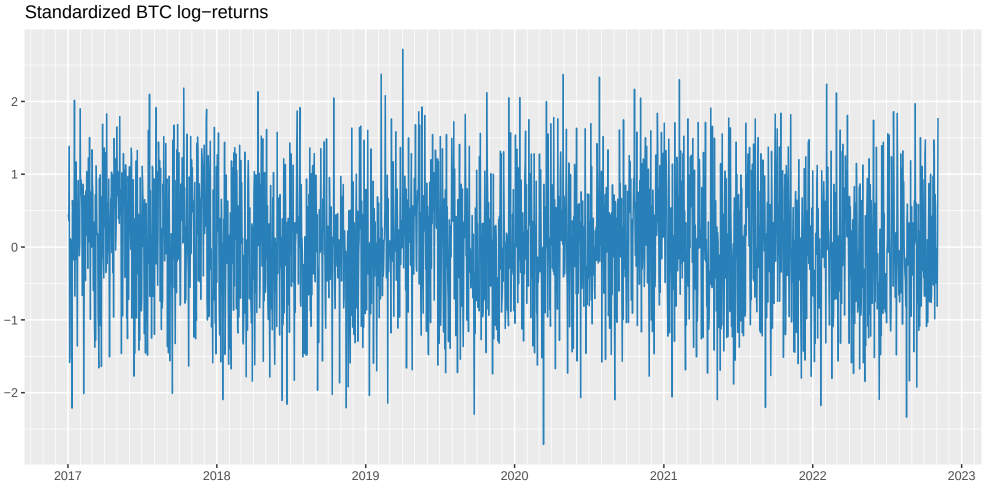
Nonlinear Structure in Returns*

Standardized S&P 500 log-returns after factoring out the volatility envelope:



Nonlinear Structure in Returns*

Standardized Bitcoin log-returns after factoring out the volatility envelope:



Outline

Asset Structure: Effect of Correlation

Cross-Sectional Structure in Asset Modeling

- Assets should not be modeled independently due to interdependencies.
- Joint modeling is crucial for accurate risk assessment in portfolios.

Importance of Asset Correlation

- Diversification relies on the correlation between assets.
- High correlation between assets can limit the benefits of diversification.

Effect of Correlation on Portfolio Volatility

- For fully correlated assets ($\rho = 1$), there is no diversification benefit and the portfolio variance remains the same as the individual asset variance.
- For uncorrelated assets ($\rho = 0$), diversification reduces the portfolio variance to half and the portfolio volatility is $\sqrt{0.5}$.
- For negatively correlated assets ($\rho < 0$), the diversification benefit increases with negative correlation, so the more negative the correlation, the greater the risk reduction.

Asset Structure: Effect of Correlation

Search for Low Correlation

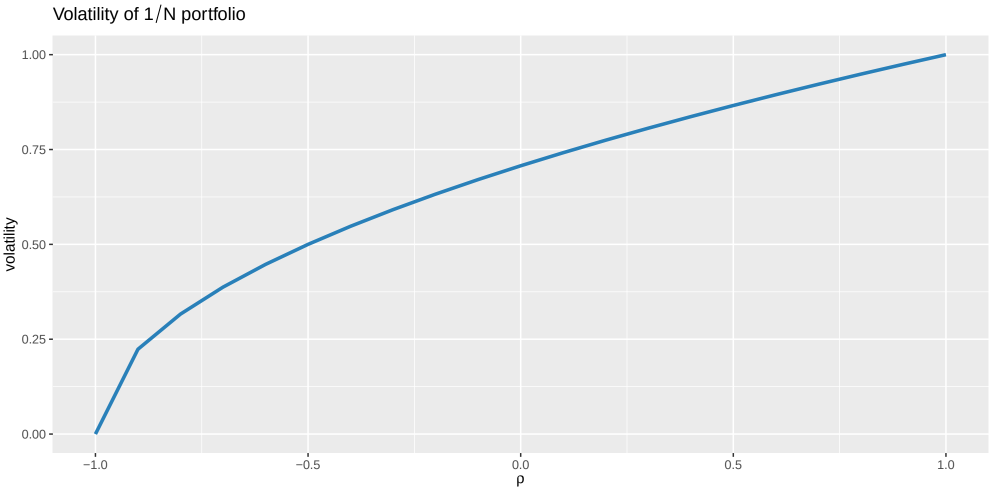
- High correlation among assets is common, making uncorrelated assets valuable for diversification.
- Identifying uncorrelated or negatively correlated assets is a key goal in portfolio management.

Synthetic Assets for Hedging

- Fully negatively correlated assets ($\rho = -1$) can be constructed synthetically for hedging purposes.
- Synthetic hedging assets are designed to offset the risk exposure of another asset.

Asset structure: Effect of correlation

Effect of asset correlation on volatility for a 2-asset portfolio:



Asset Structure: Correlation Matrix

Correlation Matrix Heatmaps for Financial Assets

- Heatmaps visualize the correlation among assets.
- Diagonal elements represent self-correlation (always 1).
- Off-diagonal elements show correlations between different assets.

Observations from S&P 500 Stocks

- Correlations among stocks are generally weaker than self-correlation.
- No significant negative correlations observed in the heatmap.

Cryptocurrency Correlations

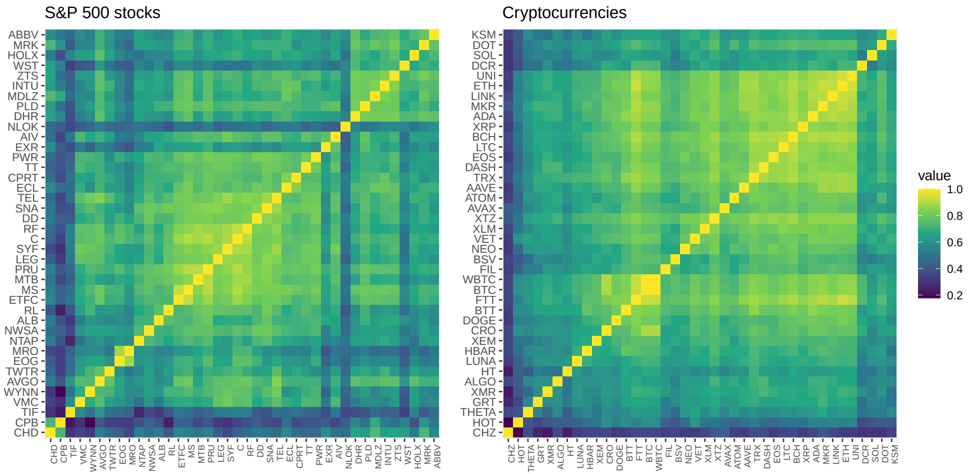
- Similar to S&P 500 stocks, off-diagonal correlations are weaker.
- Exception: BTC and WBTC show full correlation by definition.

Implications for Portfolio Management

- Understanding asset correlations is crucial for diversification strategies.
- Even within a diversified portfolio, correlations can limit risk reduction.
- Identifying assets with low or negative correlations can enhance portfolio resilience.
- The case of BTC and WBTC highlights the importance of understanding the nature of assets in portfolio construction.

Asset Structure: Correlation Matrix

Correlation matrix of returns for stocks and cryptocurrencies:



Asset Structure: Distribution of Correlations

Cross-Correlation Observations

- Histograms confirm that cross-correlations among assets are predominantly nonnegative.
- Positive correlations are a common characteristic in both stock and cryptocurrency markets.

Market Movement Influence

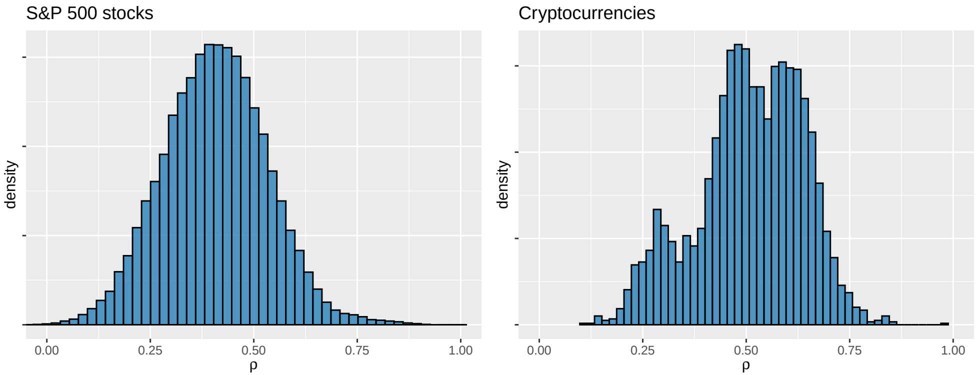
- Assets often move in tandem with the market, leading to positive correlations.
- Market trends can have a significant impact on the correlation structure of assets.

Implications for Investment Strategies

- Positive correlations must be considered when constructing diversified portfolios.
- The presence of positive correlations can affect the effectiveness of diversification in risk management.
- Investors may seek assets with lower correlations or alternative investments to achieve better diversification.

Asset Structure: Distribution of Correlations

Histogram of correlations among returns of stocks and cryptocurrencies:



Asset Structure: Eigenvalues of Covariance Matrix

Factor Model Structure in Asset Correlations

- Eigenvalues of the correlation matrix often show a distinct pattern.
- Few large eigenvalues and many smaller ones suggest a factor model structure.

Eigenvalue Distribution Insights

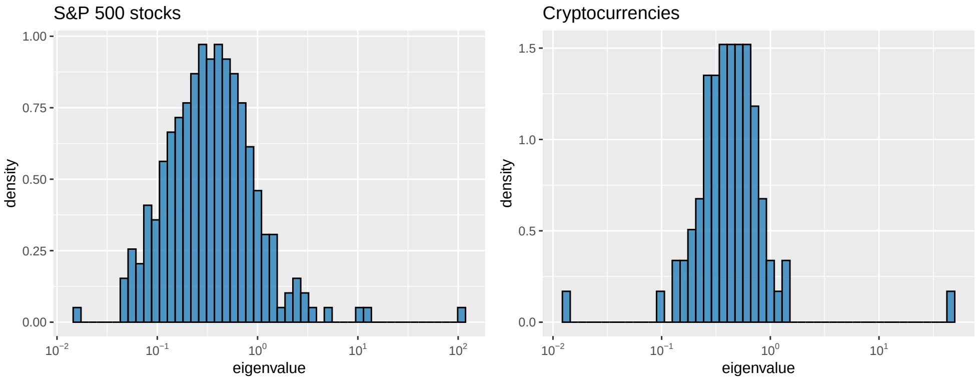
- Large eigenvalues represent common factors affecting multiple assets.
- Smaller eigenvalues indicate idiosyncratic or asset-specific factors.

Empirical Eigenvalue Analysis

- For S&P 500 stocks, one eigenvalue dominates, likely representing the overall market factor.
- A few other significant eigenvalues may represent industry or style factors.
- The remaining eigenvalues are much smaller, indicating asset-specific influences.
- For the top 82 cryptocurrencies, a similar pattern with one predominant eigenvalue indicates a strong market factor is also present.

Asset Structure: Eigenvalues of Covariance Matrix

Histogram of correlation matrix eigenvalues of stocks and cryptocurrencies:



Asset Structure: Eigenvalues of Covariance Matrix

Factor Model Implications

- Dominant eigenvalue corresponds to the market index, explaining a large portion of variance.
- The presence of a few larger eigenvalues supports the concept of multi-factor models in asset pricing.
- Factor models can simplify portfolio risk assessment and management.

Visualization of Eigenvalues

- A histogram on a logarithmic scale highlights the disparity between the largest and smaller eigenvalues.
- Reinforces the factor model structure in both traditional and cryptocurrency markets.

Outline

Summary

Financial data display unique characteristics known as stylized facts, with the most prominent ones including:

- **Lack of Stationarity**: The statistics of financial data change over time significantly and any attempt of modeling will have to continuously adapt.
- **Volatility Clustering**: This is perhaps the most visually apparent aspect of financial time series. There is a myriad of models in the literature that can be utilized for forecasting.
- **Heavy Tails**: The distribution of financial data is definitely not Gaussian and this constitutes a significant departure from many traditional modeling approaches.
- **Strong Asset Correlation**: The goal in investing is to discover assets that are not strongly correlated, which is a daunting task due to the naturally occurring strong asset correlation.

References I

Cowpertwait, Paul S. P., and Andrew V. Metcalfe. 2009. *Introductory Time Series with R*. Springer.

López de Prado, M. 2018. *Advances in Financial Machine Learning*. Wiley.

Lütkepohl, H. 2007. *New Introduction to Multiple Time Series Analysis*. Springer.

Meucci, A. 2005. *Risk and Asset Allocation*. Springer.

Palomar, D. P. 2025. *Portfolio Optimization: Theory and Application*. Cambridge University Press.

Ruppert, D., and S. D. Matteson. 2015. *Statistics and Data Analysis for Financial Engineering: With R Examples*. 2nd ed. Springer.

Tsay, R. S. 2010. *Analysis of Financial Time Series*. 3rd ed. John Wiley & Sons.

References II

———. 2013. *Multivariate Time Series Analysis: With R and Financial Applications*. John Wiley & Sons.