Portfolio Optimization
Financial Data: Stylized Facts

Daniel P. Palomar (2024). *Portfolio Optimization: Theory and Application.*
Cambridge University Press.

portfoliooptimizationbook.com

Outline

Abstract

Different domains in science and engineering are deeply rooted on the specifics of the data. The first step in any endeavor in finance or financial engineering should be to understand financial data. The study and characterization of financial data started flourishing as early as the 1960s and it is now a mature topic in which academics and practitioners have exposed some particularities of the data commonly referred to as stylized facts. These slides take us on a rather visual exploratory analysis of financial data based on empirical market data (Palomar 2024, 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,

    $$y_t \triangleq \log p_t,$$

    preferred for modeling.
    - Enhances mathematical convenience and represents a wider dynamic range.

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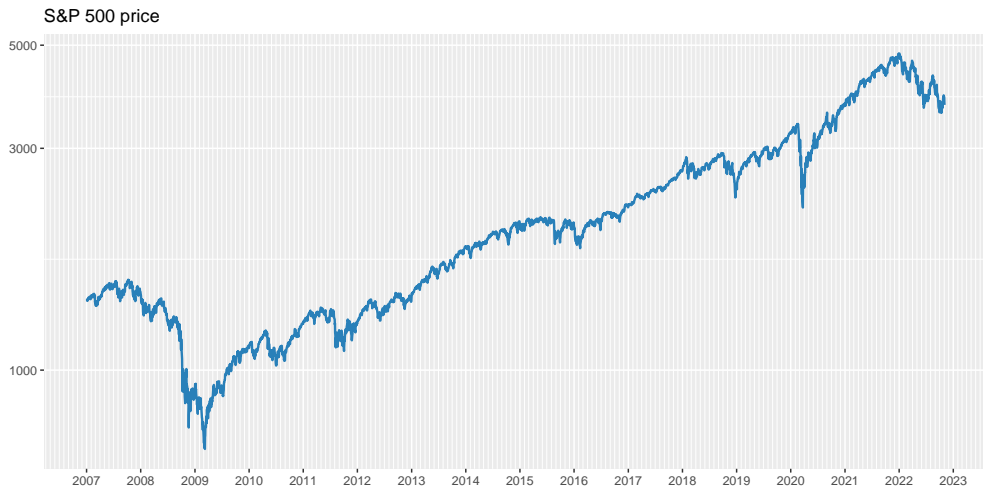   - For multi-asset case: (Lütkepohl 2007; Tsay 2013).

- **Random walk model:**
    - Simplest model for log-prices:

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

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

Prices

Price time series of S&P 500:

S&P 500 price

Price time series of Bitcoin:

BTC price

Returns

- **Price changes or returns:**
  - Returns exhibits stationarity and are suitable for mathematical modeling.
  - Two main types of returns: linear and log returns.
- **Linear return:**
  - Defined as

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

  - Additive among assets, crucial for portfolio return calculations.
  - Facilitates analysis of a portfolio's overall return.
- **Log return:**
  - Defined as

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

  - 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:**
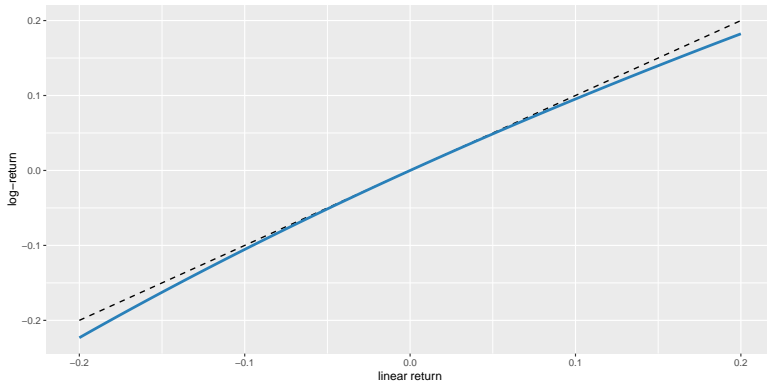 - Simple return and log-return are related as

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

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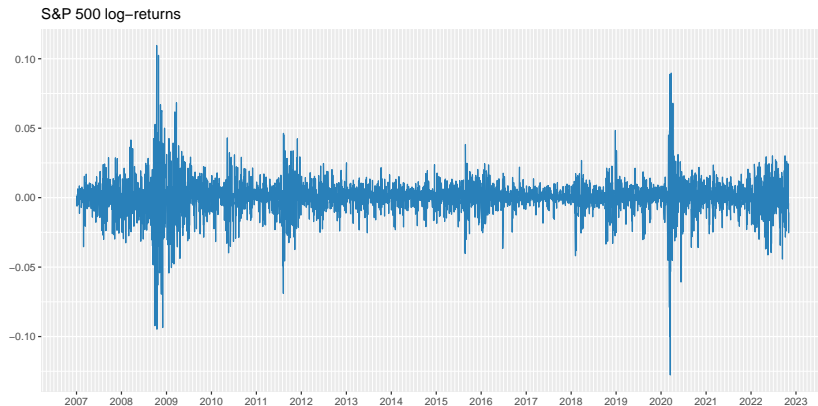 - Accurate for returns less than 5%, then accuracy decreases.

- **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:

S&P 500 log–returns

- 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:

BTC log-returns

- 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:**
  - Gaussian distribution may not be suitable in all domains.
  - 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:
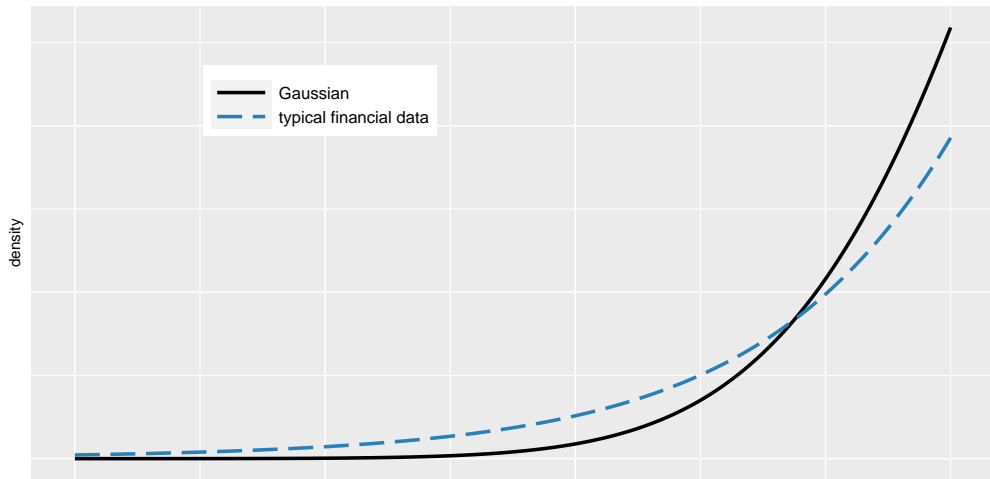
- **Impact of skewness and kurtosis:**
  - Skewness and kurtosis contribute to the likelihood of extreme negative returns.
  - Significant for investors holding the asset, as it affects risk assessment.

- **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.
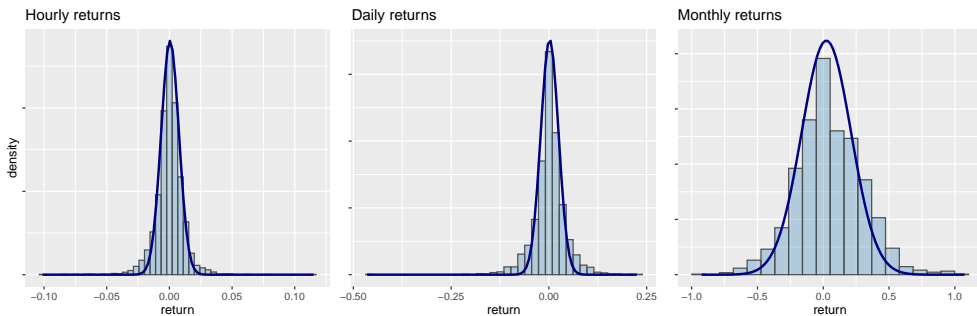 - Indicates higher probability of extreme outcomes than predicted by Gaussian models.

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

Histogram of Bitcoin log-returns at different frequencies (with Gaussian fit):
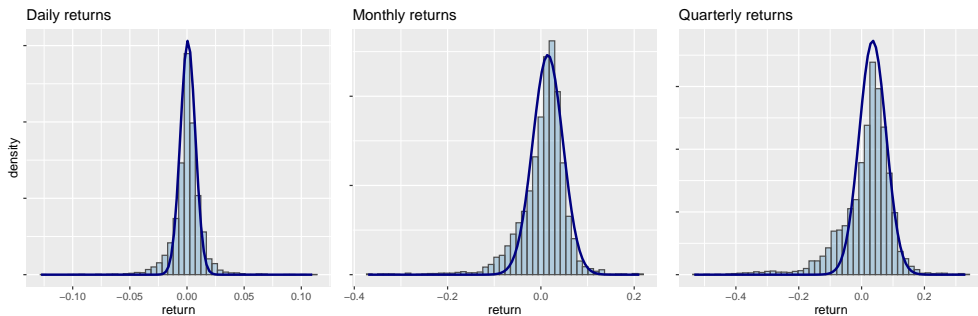
Non-Gaussianity: Asymmetry and heavy tails

- **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 or skewness[*]

- **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.
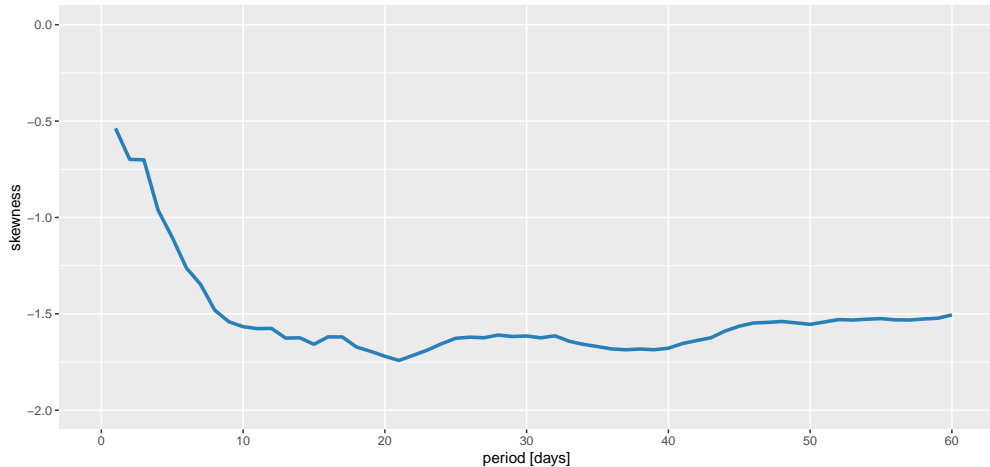 - **Bitcoin skewness analysis (2017-2022):**
    - Shows a similar trend to S&P 500.
    - Skewness of Bitcoin is closer to zero, indicating more symmetry.

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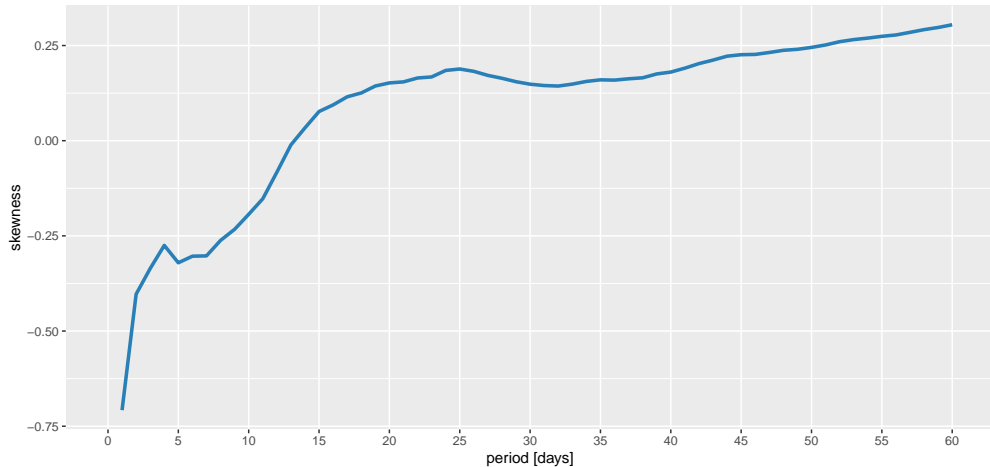 - Cryptocurrencies, represented by Bitcoin, tend to be more symmetric than stocks, such as those in the S&P 500.

Skewness of S&P 500 log-returns:
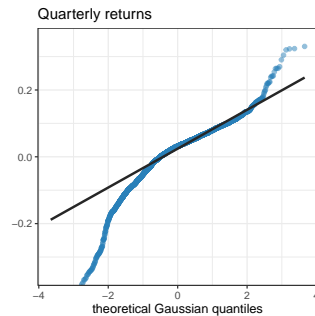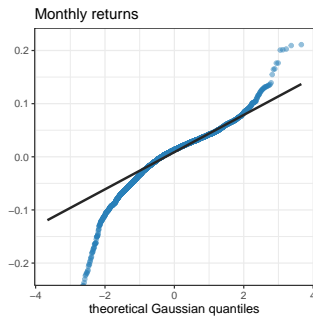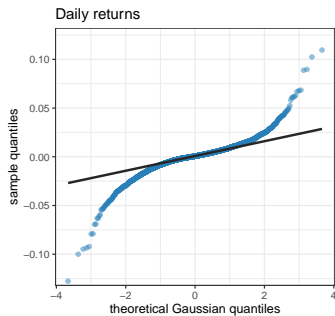
Skewness of Bitcoin log-returns:

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:**
    - Deviation in both left and right tails from the line of equality.
    - Indicates presence of heavy tails in S&P 500 log-returns.
  - **Bitcoin Q-Q plots:**
    - Similar deviations in tails observed.
    - Confirms 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.
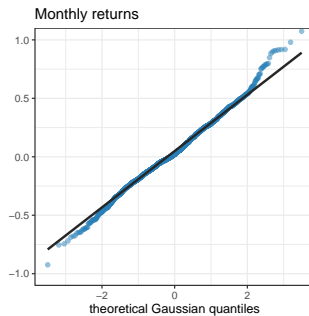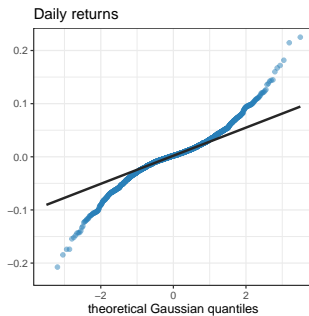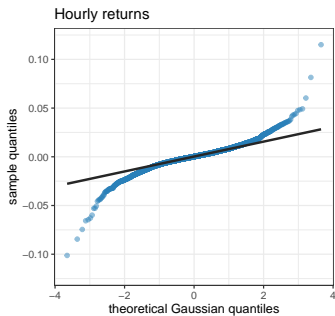
Q-Q plots of S&P 500 log-returns at different frequencies:

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

Non-Gaussianity: Heavy-tailness or kurtosis*

- **Understanding kurtosis:**
  - Measures "tailedness" of a distribution.
  - Gaussian distribution kurtosis: 3.
  - Higher kurtosis indicates heavier tails.
  - Excess kurtosis: 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.
    - Indicates smaller kurtosis compared to S&P 500.
- **Comparative insights:**
  - Initial observation suggests cryptocurrencies might be less heavy-tailed than stocks.
  - Excess kurtosis across periods:
    - **2017-2019:** 5.41 for the S&P 500 and 3.46 for Bitcoin.
    - **2020:** Notable increase in Bitcoin's heavy-tailed behavior.
    - **2021-2022:** 0.95 for the S&P 500 and 2.34 Bitcoin.
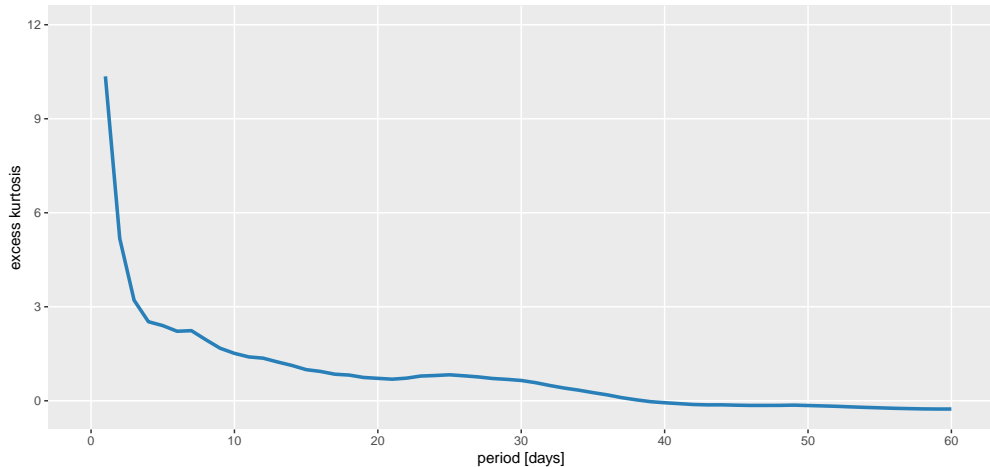 - 2020 marked a significant divergence, with Bitcoin showing more extreme heavy tails.

Excess kurtosis of S&P 500 log-returns:
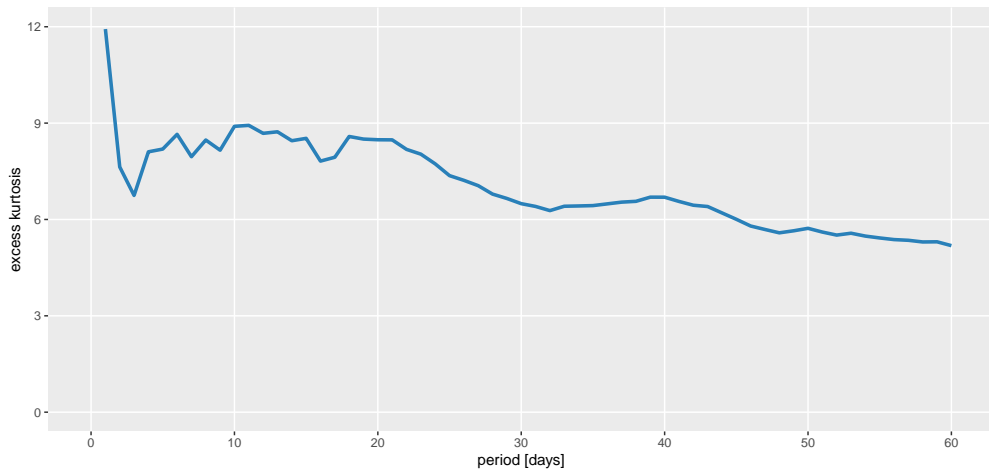
Excess kurtosis of Bitcoin log-returns:

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.

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.

Q-Q plots of S&P 500 log-returns versus different candidate distributions:
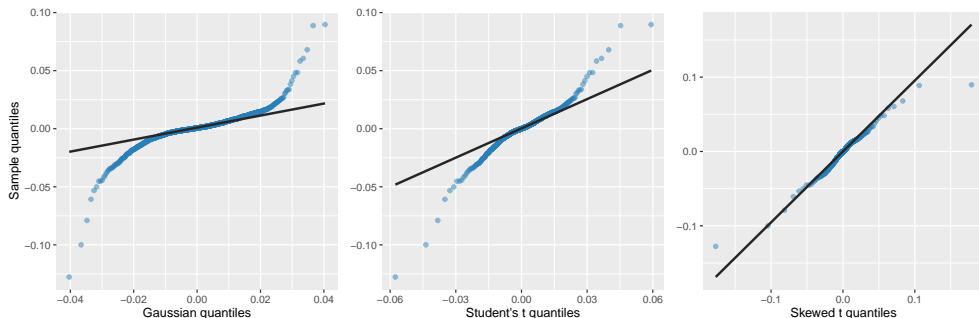
- **Visual inspection via Q-Q plots:**
    - Comparison of empirical data against Gaussian, Student t, and Skewed t distributions.
    - Confirms skewed t distribution as a good fit for the data.

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:**
  - **S&P 500 autocorrelation analysis:**
    - No significant autocorrelation detected that could be used for forecasting.
    - ACF plot shows lags within the statistically insignificant level, except at lag 0.
  - **Bitcoin autocorrelation analysis:**
    - Similar to S&P 500, no significant autocorrelation found.
    - Hourly returns also show no significant autocorrelations.
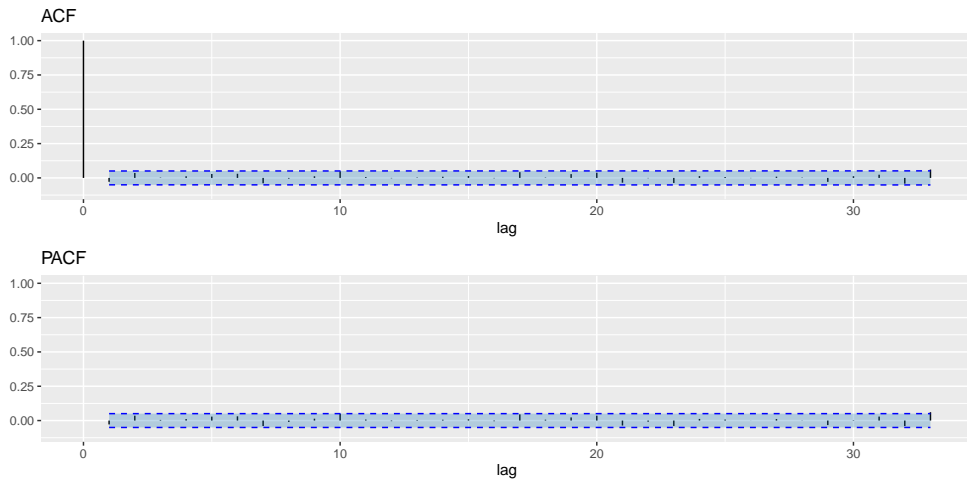
Linear structure in returns

Autocorrelation of S&P 500 daily log-returns:
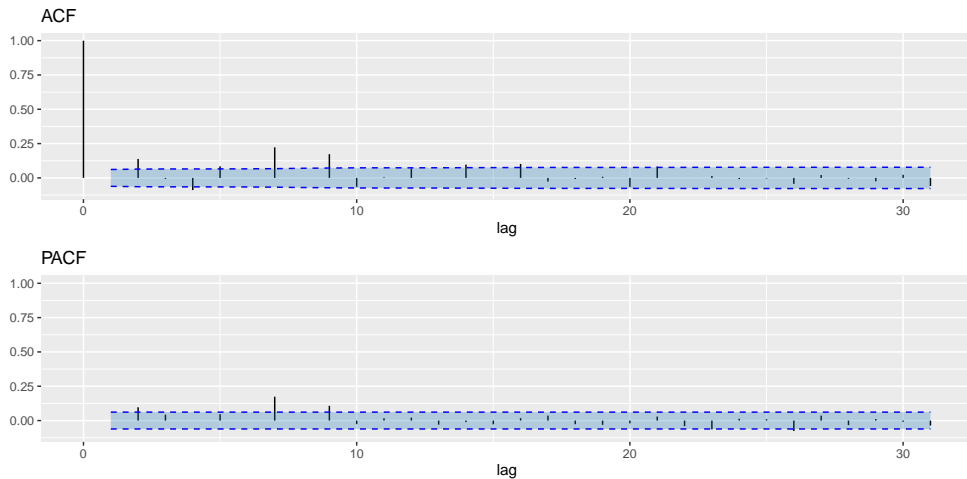
Linear structure in returns

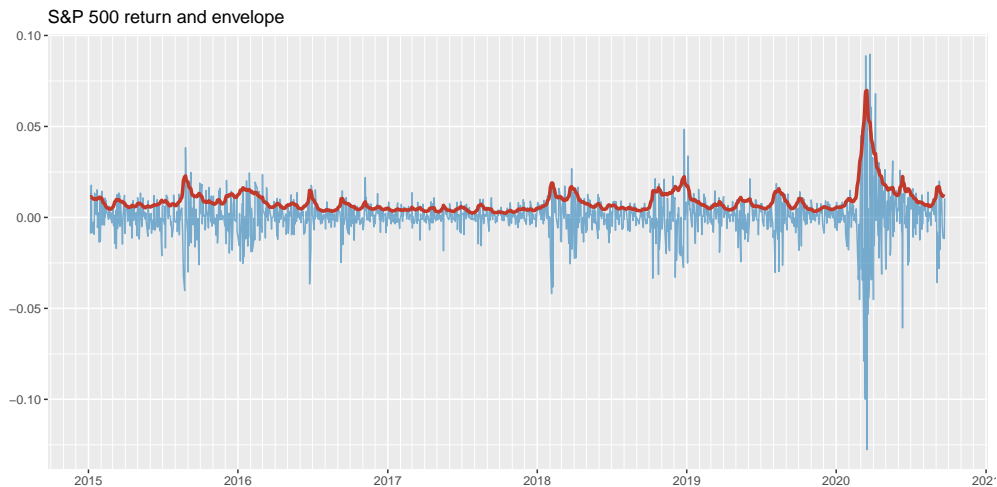Autocorrelation of Bitcoin daily log-returns:

Nonlinear structure in returns

- **Temporal structure beyond autocorrelations:**
  - Absence of significant autocorrelations does not imply lack of temporal structure.
  - Volatility envelope reveals time-varying standard deviation, indicating structure.
- **Volatility clustering:**
  - Periods of high volatility are followed by high volatility, and low by low.
  - Indicates that volatility, rather than returns themselves, may have predictable patterns.
- **Empirical evidence of volatility clustering:**
  - **S&P 500 volatility clustering:**
    - Log-returns and volatility envelope demonstrate volatility clustering.
    - Slow changes in volatility envelope suggest predictability.
  - **Bitcoin volatility clustering:**
    - Similar observations of volatility clustering in Bitcoin log-returns.
    - Indicates presence of temporal structure in volatility across different assets.
- **Implications for forecasting:**
  - Direct forecasting of returns may be challenging, volatility patterns offer some potential.
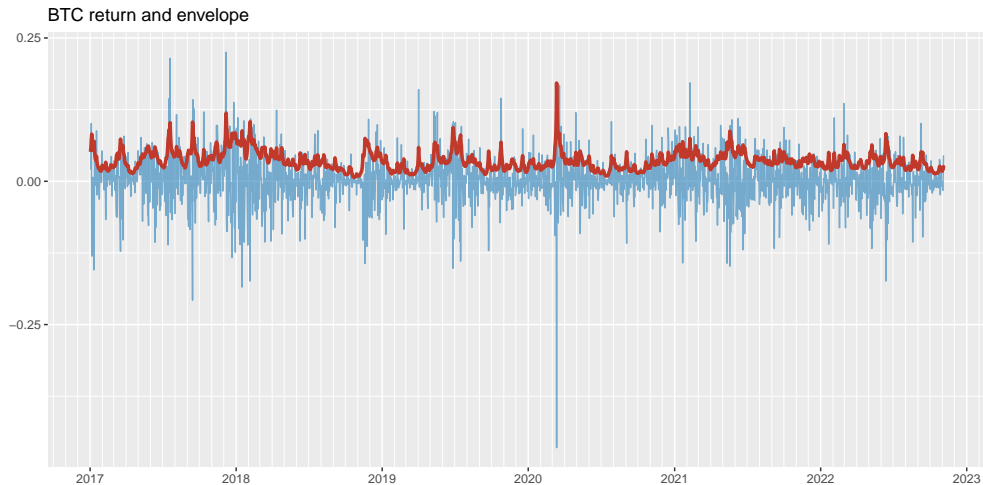 - Recognizing and forecasting volatility clustering can enhance trading strategies and risk management.

Volatility clustering in S&P 500:

S&P 500 return and envelope

Nonlinear structure in returns

Volatility clustering in Bitcoin:

BTC return and envelope

- **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.
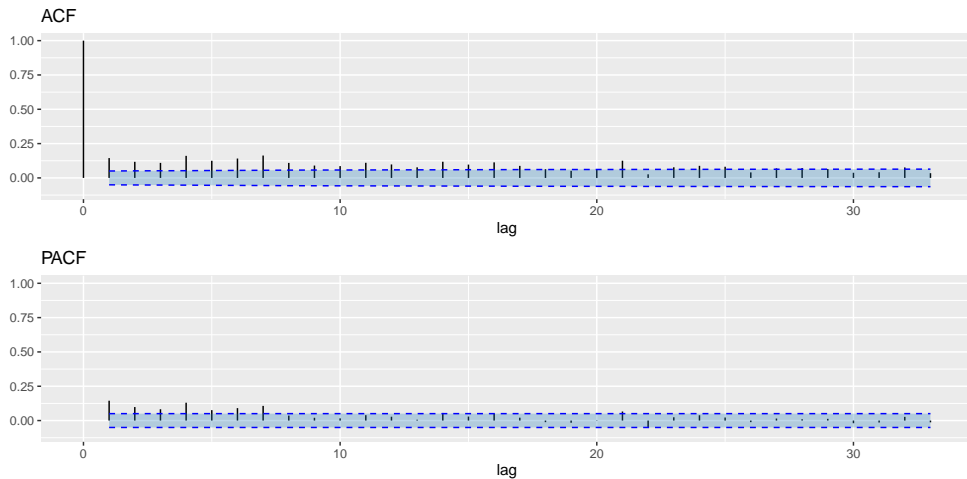 - Provides insight into the magnitude of returns, irrespective of direction.

Autocorrelation of absolute value of S&P 500 daily log-returns:
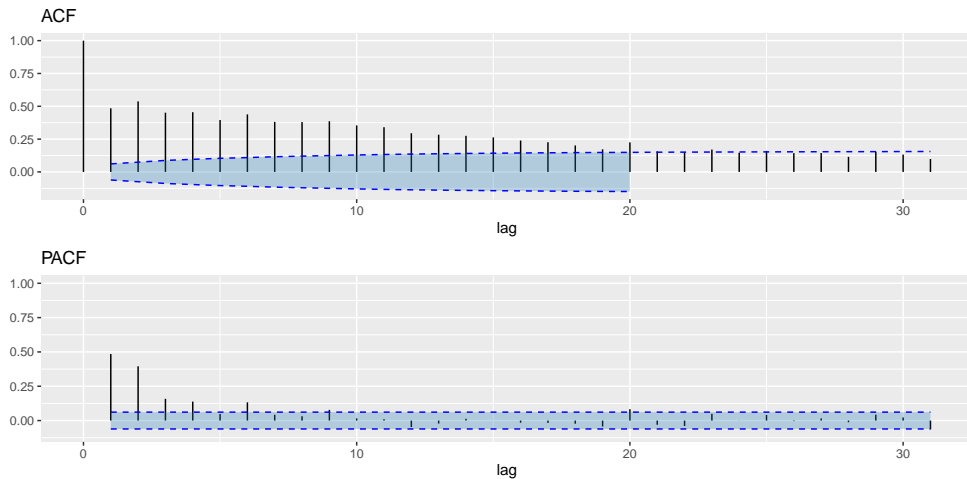
Nonlinear structure in returns[*]

Autocorrelation of absolute value of Bitcoin daily log-returns:
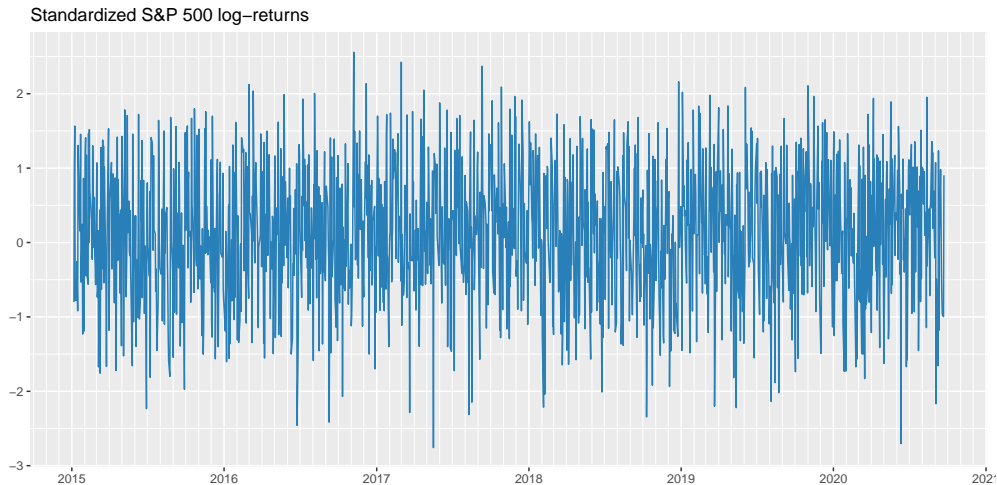
Nonlinear structure in returns[*]

- **Empirical findings on absolute returns:**
  - **S&P 500 absolute returns autocorrelation:**
    - Significant autocorrelation observed in the absolute values of log-returns.
    - Indicates presence of temporal structure in volatility.
  - **Bitcoin absolute returns autocorrelation:**
    - Shows significant autocorrelation, though less pronounced than S&P 500.
    - Suggests 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.

Nonlinear structure in returns[*]

- **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.
- **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:**
  - **S&P 500 standardized returns:**
    - Illustration shows removal of volatility clustering, resulting in a more uniform series.
  - **Bitcoin standardized returns:**
    - Similar process applied to Bitcoin, 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).
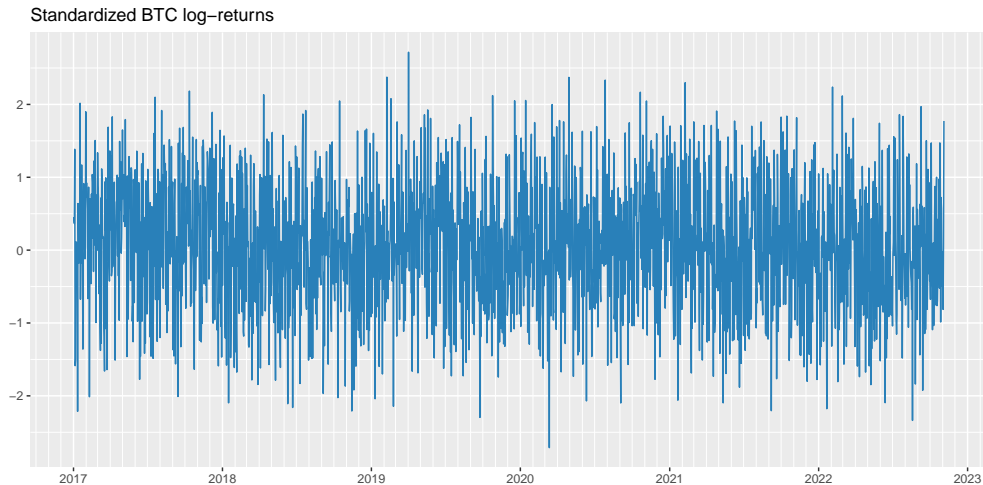
Nonlinear structure in returns[*]

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

Standardized S&P 500 log−returns

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

Standardized BTC log-returns

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:**
  - **Fully correlated assets ($\rho = 1$):**
    - No diversification benefit.
    - Portfolio variance remains the same as individual asset variance.
  - **Uncorrelated assets ($\rho = 0$):**
    - Diversification reduces portfolio variance to half.
    - Portfolio volatility is $\sqrt{0.5}$.
  - **Negatively correlated assets ($\rho < 0$):**
    - Diversification benefit increases with negative correlation.
    - 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 risk exposure of another asset.

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

Volatility of 1/N portfolio

Asset structure: Correlation matrix
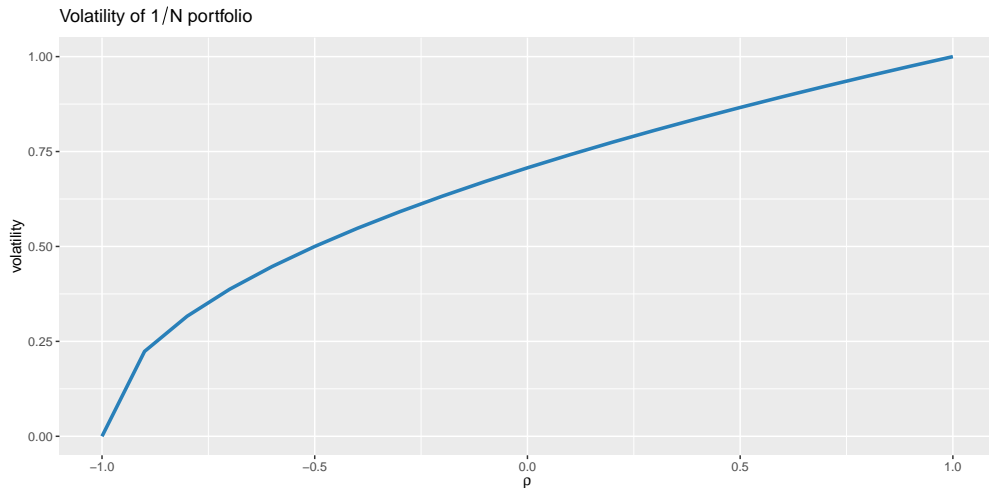
- **Correlation matrix heatmaps for financial assets:**
  - Heatmaps visualize the correlation among assets.
  - Diagonal elements represent self-correlation (always 1).
  - Off-diagonal elements show correlations between different assets.
- **Observations from S&P 500 stocks:**
  - Correlations among 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.

Correlation matrix of returns for stocks and cryptocurrencies:
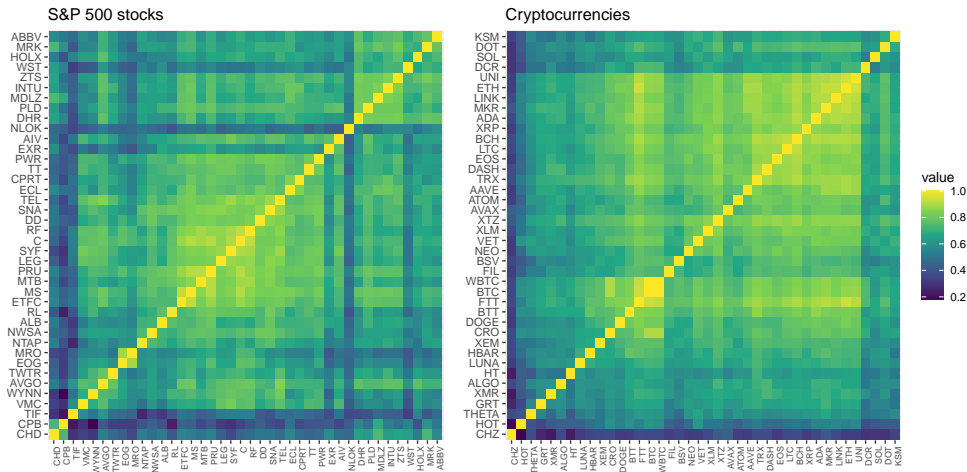
- **Cross-correlation observations:**
  - Histograms confirm that cross-correlations among assets are predominantly nonnegative.
  - Positive correlations are a common characteristic in both stock and cryptocurrency markets.

- **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.

Histogram of correlations among returns of stocks and cryptocurrencies:

Asset structure: Eigenvalues covariance matrix
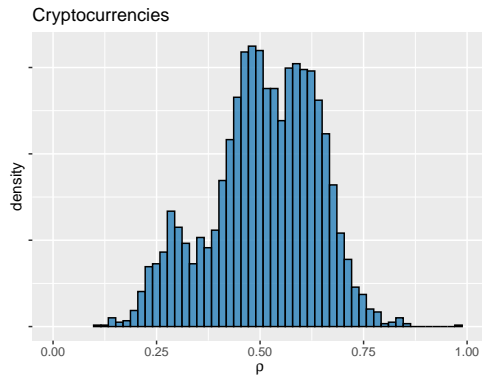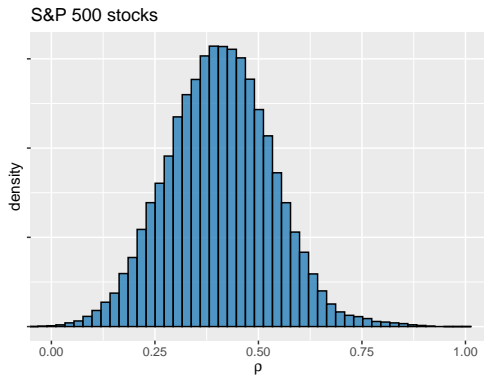
- **Factor model structure in asset correlations:**
  - Eigenvalues of the correlation matrix often show a distinct pattern.
  - Few large eigenvalues and many smaller ones suggest a factor model structure.
- **Eigenvalue distribution insights:**
  - Large eigenvalues represent common factors affecting multiple assets.
  - Smaller eigenvalues indicate idiosyncratic or asset-specific factors.
- **Empirical eigenvalue analysis:**
  - **S&P 500 stocks:**
    - One eigenvalue dominates, likely representing the overall market factor.
    - A few other significant eigenvalues may represent industry or style factors.
    - Remaining eigenvalues are much smaller, indicating asset-specific influences.
  - **Top 82 cryptocurrencies:**
    - Similar pattern with one predominant eigenvalue.
    - Indicates a strong market factor also present in the cryptocurrency market.

Histogram of correlation matrix eigenvalues of stocks and cryptocurrencies:
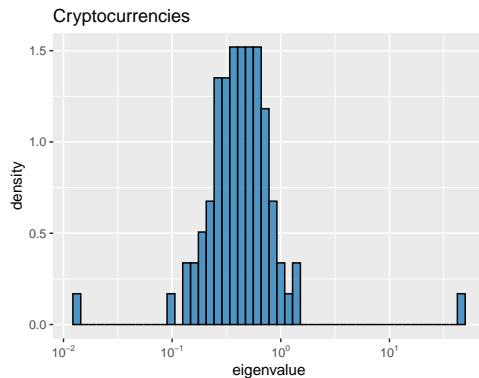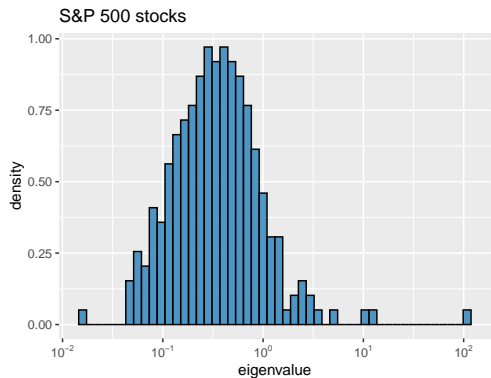
- **Factor model implications:**
  - Dominant eigenvalue corresponds to market index, explaining a large portion of variance.
  - Presence of a few larger eigenvalues supports the concept of multi-factor models in asset pricing.
  - Factor models can simplify portfolio risk assessment and management.

- **Visualization of eigenvalues:**
  - 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. 2024. *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.

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