ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Enhancing Risk Management, Return Forecasting, and Volatility Modeling in Financial Markets Using Non-Gaussian Distributions

An Empirical Study on Intrinsic Valuation and Risk Econometrics of Top-Performing Tech Stocks

Author
Matteo Ferniani

Supervisor:
Giovanni Angelini

Alma Mater Studiorum - Università di Bologna

Degree Class LM-83 - Actuarial and Financial Statistical Sciences

Department of Statistical Sciences

March 2025

Abstract

This research aims to analyze the outperformance of a selected set of high-quality stocks relative to the S&P 500 benchmark, summarized through the VGT ETF. We employ ARMA and ARCH models with non-Gaussian distributions to forecast conditional volatility and returns, ultimately estimating the optimal non-Gaussian Value at Risk (VaR) and Expected Shortfall (ES).

The methodology begins with an in-depth analysis of the financials of three representative stocks from the VGT index—NVIDIA, Microsoft, and Apple. We investigate their performance to understand the drivers behind their outperformance, utilizing various financial indicators, including earnings per share (EPS) growth, return on invested capital (ROIC), and profit margins.

Subsequently, we identify the most suitable distributions for the historical financial time series and develop algorithms to fit the optimal ARMA and ARCH models based on the Akaike Information Criterion (AIC). We compare multiple models, combinations, and distributions to determine the best-fitting approach.

The results indicate that the APARCH(1,1) model with a skewed t-distribution is the most effective for our VGT indicator. In the context of risk modeling, the VaR calculated with a skewed t-distribution demonstrates superior efficiency, while the Expected Shortfall calculated using a generalized error distribution (GED) proves to be the most effective.

In conclusion, advanced ARCH models and non-Gaussian distributions facilitate accurate forecasting of conditional volatility and returns. Our findings suggest that employing non-Gaussian distributions in ARCH modeling yields more efficient estimates of VaR and Expected Shortfall compared to the conventional normal distribution.

Table of Contents

Contents

List of Figures and Tables

List of Figures

List of Tables

Introduction

The focus of this thesis is to analyze the outperformance in terms of returns of selected stocks within the Information Technology sector, using the VGT ETF as a benchmark compared to the S&P 500 index. The analysis aims to determine whether the fundamentals of the companies included in the VGT justify their superior performance and assess the potential for this trend to continue in the future.

The research will begin with a detailed examination of the financial characteristics and performance of the VGT and S&P 500, seeking to identify the underlying reasons for any observed outperformance. Special attention will be given to the distribution of returns for these indices, aiming to determine the most suitable statistical distributions for modeling the financial time series.

Next, the study will evaluate whether the risk/return profile of the VGT ETF surpasses that of the S&P 500, particularly in the context of changing interest rates. The analysis will involve fitting an ARMA-GARCH model with non-Gaussian distributions to the VGT data to identify the best possible model for capturing the conditional volatility of the financial time series.

Subsequently, the research will forecast the conditional volatility and returns of the VGT ETF and develop risk management measures using both Gaussian and non-Gaussian distributions. The effectiveness of each approach will be assessed to construct the most reliable Value at Risk (VaR) and Expected Shortfall (ES) measures.

The thesis is structured as follows: Chapter 1 provides an overview of the relevant literature on financial modelling and the performance analysis of technology sector stocks. Chapter 2 outlines the methodology employed in the analysis, including the selection of models and statistical techniques. Chapter 3 presents the results of the empirical analysis, while Chapter 4 discusses the findings and their implications for future performance predictions and risk management. Finally, Chapter 5 concludes the thesis, summarizing the main findings and suggesting directions for future research.

Background

Key concepts and definitions

Compound Annual Growth Rate (CAGR)

The Compound Annual Growth Rate (CAGR) represents the mean annual growth rate of an investment over a specified period of time, assuming the profits are reinvested at the end of each period. It is a useful measure to compare the growth of different investments or financial metrics over time.

The formula for calculating CAGR is given by:

$$\text{CAGR} = \left(\frac{V_f}{V_i}\right)^{\frac{1}{n}} - 1 \tag{1}$$

where:

- V_f is the final value of the investment.
- V_i is the initial value of the investment.
- n is the number of years.

Return on Invested Capital (ROIC)

Return on Invested Capital (ROIC) is a financial metric used to measure the efficiency of a company in generating profits from its capital. It indicates how well a company is using its capital to generate returns, making it a valuable measure for investors and analysts when assessing the profitability and efficiency of a business.

The formula for calculating ROIC is given by:

$$\text{ROIC} = \frac{\text{Net Operating Profit After Tax (NOPAT)}}{\text{Invested Capital}} \tag{2}$$

where:

- NOPAT is the net operating profit after tax, reflecting the company's operational efficiency.
- Invested Capital is the total capital invested in the business, including equity and debt financing.

Return on Capital Employed (ROCE)

Return on Capital Employed (ROCE) is a financial metric that measures a company's profitability and the efficiency with which its capital is employed. It is calculated by dividing operating profit by capital employed.

$$\text{ROCE} = \frac{\text{Operating Profit}}{\text{Capital Employed}} \times 100 \tag{3}$$

Where:
- **Operating Profit** is the profit earned from regular operations.
- **Capital Employed** is the total assets minus current liabilities.

Return on Assets (ROA)

Return on Assets (ROA) is an indicator of how profitable a company is relative to its total assets. It measures how efficiently management is using its assets to generate earnings.

$$\text{ROA} = \frac{\text{Net Income}}{\text{Total Assets}} \times 100 \tag{4}$$

Where:
- **Net Income** is the profit after taxes and expenses.
- **Total Assets** represent the total resources owned by the company. **Return on Equity**

(ROE)

Return on Equity (ROE) measures the ability of a company to generate profits from its shareholders' equity. It indicates how effectively management is using a company's assets to create profits.

$$\text{ROE} = \frac{\text{Net Income}}{\text{Shareholder's Equity}} \times 100 \tag{5}$$

Where:
- **Net Income** is the profit after taxes and expenses.
- **Shareholder's Equity** is the net assets owned by the shareholders, calculated as total assets minus total liabilities.

Statistical Measures and Financial Metrics

Mean

The mean, or average, is a measure of central tendency that represents the average value of a dataset. It is calculated as:

$$\text{Mean} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{6}$$

where n is the number of observations and x_i are the individual data points.

Median

The median is the middle value of a dataset when arranged in ascending order. It is a measure of central tendency that is less affected by outliers than the mean. For a dataset with n observations:

$$\text{Median} = \begin{cases} x_{\left(\frac{n+1}{2}\right)} & \text{if } n \text{ is odd} \\ \frac{x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}}{2} & \text{if } n \text{ is even} \end{cases} \tag{7}$$

Volatility of Returns

Volatility of returns is typically represented by the standard deviation of the returns. It indicates the degree of variation of returns over a given period:

$$\text{Volatility} = \sigma_r = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(r_i - \text{Mean}_r)^2} \tag{8}$$

where r_i are the returns and Mean_r is the average return.

Skewness

Skewness measures the asymmetry of the probability distribution of a real-valued random variable. It is calculated as:

$$\text{Skewness} = \frac{n}{(n-1)(n-2)}\sum_{i=1}^{n}\left(\frac{x_i - \text{Mean}}{\sigma}\right)^3 \tag{9}$$

Kurtosis

Kurtosis measures the "tailedness" of the probability distribution. It is calculated as:

$$\text{Kurtosis} = \frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum_{i=1}^{n} \left(\frac{x_i - \text{Mean}}{\sigma} \right)^4 - \frac{3(n-1)^2}{(n-2)(n-3)} \tag{10}$$

Sharpe Ratio

The Sharpe Ratio is a measure of risk-adjusted return. It is calculated as:

$$\text{Sharpe Ratio} = \frac{R_p - R_f}{\sigma_p} \tag{11}$$

where R_p is the portfolio return, R_f is the risk-free rate, and σ_p is the standard deviation of the portfolio return.

Drawdown

Drawdown measures the peak-to-trough decline during a specific period. It is calculated as:

$$\text{Drawdown} = \frac{P_{\text{peak}} - P_{\text{trough}}}{P_{\text{peak}}} \tag{12}$$

where P_{peak} is the maximum value and P_{trough} is the minimum value during the drawdown period.

Risk-Free Rate

The risk-free rate is the return on an investment with no risk of financial loss. In the context of finance, it is commonly represented by the yield on long-term government bonds, such as the 10-year U.S. Treasury. This yield is considered a benchmark for determining the risk premium of other investments.

The risk-free rate can be denoted as:

$$R_f = Y_{10Y} \tag{13}$$

where:

- R_f is the risk-free rate.
- Y_{10Y} is the yield on the 10-year U.S. Treasury bond.

This yield is often used in financial models to discount cash flows and evaluate investment opportunities.

Distributions in Financial Returns

In finance, understanding the distribution of return series is crucial for assessing risk, forecasting future returns, and making informed investment decisions. Different distributions can capture the behavior of asset returns, including the presence of skewness, kurtosis, and fat tails. Analyzing these distributions helps in developing risk management strategies and evaluating the performance of financial assets.

Normal Distribution

The normal distribution, also known as the Gaussian distribution, is characterized by its symmetric bell shape. It is widely used in finance due to the Central Limit Theorem, which states that the sum of many independent random variables tends toward a normal distribution.

The probability density function (PDF) of a normal distribution is given by:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{14}$$

where: - μ is the mean, - σ is the standard deviation.

Characteristics: - Symmetrical around the mean. - Defined by two parameters: mean and standard deviation.

t-Distribution

The t-distribution is similar to the normal distribution but has heavier tails, making it useful for small sample sizes and when the population standard deviation is unknown. It is particularly important in hypothesis testing.

The PDF of the t-distribution is given by:

$$f(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}} \tag{15}$$

where: - ν is the degrees of freedom.

Characteristics: - Symmetrical around zero. - Heavier tails than the normal distribution, especially for low degrees of freedom.

Skew-t Distribution

The skew-t distribution extends the t-distribution by allowing for asymmetry (skewness). It is particularly useful for modeling financial returns that exhibit skewness and excess kurtosis.

The PDF of the skew-t distribution is:

$$f(x) = 2t\left(\frac{x-\mu}{\sigma}, \nu\right)\Phi(\lambda\frac{x-\mu}{\sigma}) \tag{16}$$

where: - t is the t-distribution function, - Φ is the cumulative distribution function (CDF) of the standard normal distribution, - λ controls the skewness.

Characteristics: - Asymmetrical; can model left or right skewness. - Useful in finance for capturing the behavior of returns.

Generalized Error Distribution (GED)

The Generalized Error Distribution is a flexible distribution that can model data with varying levels of kurtosis. It is often used in financial applications to fit return data that exhibit leptokurtic (fat-tailed) behavior.

The PDF of the GED is given by:

$$f(x) = \frac{\beta}{2\alpha\Gamma\left(\frac{1}{\beta}\right)} e^{-\left(\frac{|x-\mu|}{\alpha}\right)^{\beta}} \tag{17}$$

where: - α controls scale, - β controls kurtosis.

Characteristics: - Flexible shape; can model a range of tail behaviors. - Can approximate normal distribution when $\beta = 2$.

Normal Inverse Gaussian (NIG) Distribution

The Normal Inverse Gaussian distribution is a versatile distribution that can model financial returns with both skewness and kurtosis. It is particularly useful for capturing the asymmetry and heavy tails commonly observed in financial data.

The PDF of the NIG distribution is:

$$f(x) = \frac{\alpha\delta}{\pi} K_1\left(\alpha\sqrt{\delta^2 + (x-\mu)^2}\right) e^{\delta(x-\mu)} \tag{18}$$

where: - K_1 is the modified Bessel function, - α controls the tail thickness, - δ controls the skewness.

Characteristics: - Can model skewed and heavy-tailed distributions. - Widely used in finance for option pricing and risk management.

QQ Plot

A QQ plot is a graphical tool used to determine if a dataset follows a particular theoretical distribution. It does this by plotting the quantiles of the sample data against the quantiles of the theoretical distribution. If the points in the QQ plot fall approximately along a straight line, this indicates that the sample data follows the specified distribution.

To create a QQ plot, the quantiles of the sample data X and the quantiles of the theoretical distribution Y are plotted against each other. The basic steps for constructing a QQ plot are as follows:

1. **Order the Sample Data**: Sort the sample data $X_1, X_2, \ldots, X_n$.

2. **Calculate Quantiles**: Compute the theoretical quantiles based on the specified distribution.

3. **Plot the Points**: Plot $(Q_{X,i}, Q_{Y,i})$ for $i = 1, 2, \ldots, n$, where $Q_{X,i}$ is the i^{th} quantile of the sample data and $Q_{Y,i}$ is the i^{th} quantile of the theoretical distribution.

The formula for the quantile of the empirical distribution is:

$$Q_{X,i} = X_{(i)} \quad \text{for } i = 1, 2, \ldots, n \tag{19}$$

And for the theoretical distribution, it can be represented as:

$$Q_{Y,i} = F^{-1}\left(\frac{i - 0.5}{n}\right) \tag{20}$$

where: - F^{-1} is the inverse cumulative distribution function (CDF) of the theoretical distribution, - n is the total number of observations.

Interpretation:
- If the points on the QQ plot lie on the reference line (y = x), the data can be considered to follow the specified distribution.
- Deviations from the line indicate departures from the theoretical distribution, such as skewness or kurtosis.

Jarque–Bera Test

The Jarque-Bera test is a statistical test used to assess whether a dataset follows a normal distribution. It evaluates the sample's skewness and kurtosis to determine if they are consistent with those of a normal distribution.

Specifically, the test checks for deviations in the shape of the distribution.

The test statistic for the Jarque-Bera test is given by:

$$JB = \frac{n}{6}\left(S^2 + \frac{(K-3)^2}{4}\right) \tag{21}$$

where: - n is the sample size, - S is the sample skewness, defined as:

$$S = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{X_i - \bar{X}}{s}\right)^3$$

with $\bar{X}$ being the sample mean and s the sample standard deviation, - K is the sample kurtosis, defined as:

$$K = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{X_i - \bar{X}}{s}\right)^4$$

Interpretation:
- Under the null hypothesis, H_0: the data follows a normal distribution.
- The test statistic JB follows a chi-squared distribution with 2 degrees of freedom.
- A higher value of the test statistic indicates a greater deviation from normality. If the p-value is less than a chosen significance level (commonly 0.05), we reject the null hypothesis, concluding that the data is not normally distributed.

Augmented Dickey-Fuller (ADF) Test

The Augmented Dickey-Fuller (ADF) test is a widely used statistical test to assess the presence of a unit root in a univariate time series.

It helps determine whether the time series data is stationary or follows a stochastic trend, which is critical for modeling and forecasting.

The ADF test can be expressed in the following regression form:

$$\Delta y_t = \alpha + \beta t + \rho y_{t-1} + \sum_{i=1}^{p} \phi_i \Delta y_{t-i} + \epsilon_t \tag{22}$$

where: - y_t is the time series at time t, - Δ denotes the difference operator (i.e., $\Delta y_t = y_t - y_{t-1}$), - α is a constant, - βt is a deterministic trend, - ρ is the coefficient of the lagged level of the series, - p is the number of lagged difference terms included, - ϕ_i are the coefficients of the lagged differences, - ϵ_t is a white noise error term.

Null Hypothesis:

- The null hypothesis H_0: the time series has a unit root (i.e., it is non-stationary).
- The alternative hypothesis H_1: the time series is stationary.

Interpretation:

- The test statistic is derived from the estimated ρ parameter.

If the test statistic is less than the critical value from the Dickey-Fuller distribution, we reject the null hypothesis, indicating that the series is stationary.

Ljung-Box Test

The Ljung-Box test is a statistical test used to check for the presence of autocorrelation in a time series.

This test is particularly important in time series analysis and model diagnostics, as it helps assess whether a fitted model captures the underlying data patterns effectively.

The test statistic for the Ljung-Box test is calculated as follows:

$$Q = n(n+2) \sum_{k=1}^{h} \frac{\hat{r}_k^2}{n-k} \tag{23}$$

where: - n is the number of observations, - h is the number of lags being tested, - $\hat{r}_k$ is the sample autocorrelation at lag k.

Null Hypothesis:

- The null hypothesis H_0: There is no autocorrelation in the time series data up to lag h (i.e., the data is white noise).

- The alternative hypothesis H_1: There is significant autocorrelation present in the data.

Interpretation: - The Q statistic follows a chi-squared distribution with h degrees of freedom.

- If the p-value associated with the Q statistic is less than a chosen significance level (commonly 0.05), we reject the null hypothesis, suggesting that the time series exhibits significant autocorrelation.

Returns

To calculate the return from a stock's price, use the following formula:

$$\text{Return} = \frac{P_{\text{end}} - P_{\text{start}}}{P_{\text{start}}} \times 100$$

where:

- P_{end} is the stock price at the end of the period.
- P_{start} is the stock price at the beginning of the period.

The formula calculates the percentage change in the stock price over a given period.

Autocorrelation Function (ACF) Plot

The Autocorrelation Function (ACF) plot is a tool used in time series analysis to measure the correlation between observations at different lags.

It quantifies how the current value of a series is related to its past values.

The autocorrelation at lag k is defined as:

$$\hat{r}_k = \frac{1}{n-k} \sum_{t=k+1}^{n} \frac{(X_t - \bar{X})(X_{t-k} - \bar{X})}{s^2} \tag{24}$$

where: - $\hat{r}_k$ is the sample autocorrelation at lag k, - n is the number of observations, - X_t is the value of the time series at time t, - $\bar{X}$ is the mean of the series, - s^2 is the sample variance.

Interpretation:

- The ACF plot displays $\hat{r}_k$ on the y-axis against lag k on the x-axis. Significant lags are typically highlighted to indicate where autocorrelations are statistically different from zero.

- A rapidly decaying ACF suggests a stationary series, while a slow decay indicates non-stationarity.

Example ACF Plot:

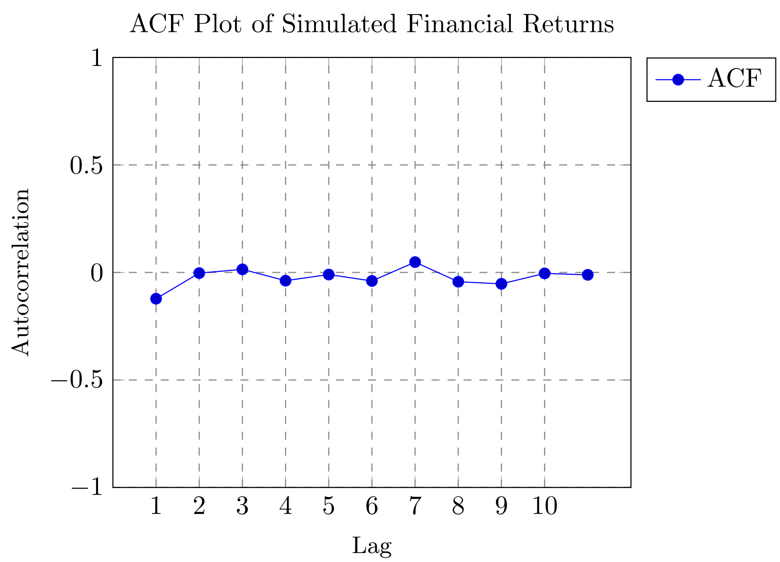


Figure 1: An example ACF plot showing the autocorrelation of simulated financial returns.

ARCH Test

The ARCH test is used to determine whether a time series exhibits autoregressive conditional heteroskedasticity, which means that the variance of the current error term is related to the variances of previous error terms.

This is crucial in financial time series, as it helps in identifying periods of volatility clustering.

The ARCH model is defined as follows:

$$y_t = \mu + \epsilon_t \tag{25}$$

$$\epsilon_t = \sigma_t z_t \quad \text{with} \quad z_t \sim N(0,1) \tag{26}$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \alpha_2 \epsilon_{t-2}^2 + \ldots + \alpha_q \epsilon_{t-q}^2 \tag{27}$$

Where: - y_t is the return at time t, - μ is the mean of the series, - ϵ_t is the error term, - z_t is a white noise error term, - σ_t^2 is the conditional variance at time t, - α_0 is a constant term, - α_i (for $i = 1, 2, \ldots, q$) are the coefficients for the lagged error terms.

To conduct the ARCH test, you typically estimate a linear regression model, then calculate the squared residuals from this model. The null hypothesis of the ARCH test is that the time series does not exhibit ARCH effects (i.e., $\alpha_1 = \alpha_2 = \ldots = \alpha_q = 0$).

A significant result from the test (usually using a LM statistic) indicates that the null hypothesis can be rejected, suggesting that the series exhibits ARCH effects.

ARMA Model

The ARMA model is a widely used statistical model in time series analysis that combines both autoregressive (AR) and moving average (MA) components.

It is effective for modeling and forecasting stationary time series data.

Autoregressive (AR) Model

The autoregressive model of order p (AR(p)) is defined as:

$$y_t = \phi_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + \epsilon_t \tag{28}$$

Where: - y_t is the value of the time series at time t, - ϕ_0 is a constant term, - ϕ_i (for $i = 1, 2, \ldots, p$) are the autoregressive coefficients, - ϵ_t is white noise (a sequence of uncorrelated random variables with mean zero).

Moving Average (MA) Model

The moving average model of order q (MA(q)) is expressed as:

$$y_t = \theta_0 + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \ldots + \theta_q \epsilon_{t-q} + \epsilon_t \tag{29}$$

Where: - y_t is the value of the time series at time t, - θ_0 is a constant term, - θ_i (for $i = 1, 2, \ldots, q$) are the moving average coefficients, - ϵ_t is white noise, as defined previously.

ARMA Model The ARMA model combines both components and is defined as:

$$y_t = \phi_0 + \phi_1 y_{t-1} + \ldots + \phi_p y_{t-p} + \theta_1 \epsilon_{t-1} + \ldots + \theta_q \epsilon_{t-q} + \epsilon_t \tag{30}$$

Where: - p is the order of the AR component, - q is the order of the MA component.

ARMA models are particularly useful for modeling stationary time series data. They can be identified and estimated using various statistical techniques, such as the Box-Jenkins methodology.

ARCH Models

Autoregressive Conditional Heteroskedasticity (ARCH) models are used in time series analysis to model and forecast the volatility of financial returns.

These models allow for time-varying volatility, which is a common phenomenon in financial data, characterized by periods of high and low volatility.

ARCH Model

The basic ARCH model, introduced by Engle (1982), can be defined as:

$$y_t = \mu + \epsilon_t \tag{31}$$

$$\epsilon_t = \sigma_t z_t \quad \text{with} \quad z_t \sim N(0,1) \tag{32}$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \alpha_2 \epsilon_{t-2}^2 + \ldots + \alpha_q \epsilon_{t-q}^2 \tag{33}$$

Where: - y_t is the return at time t, - μ is the mean of the series, - ϵ_t is the error term, - z_t is a white noise error term, - σ_t^2 is the conditional variance at time t, - $\alpha_0 > 0$ and $\alpha_i \geq 0$ for $i = 1, 2, \ldots, q$.

Characteristics: - The ARCH model captures the clustering of volatility, where high volatility events tend to follow other high volatility events.

GARCH Model

The Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model extends the ARCH model by including lagged conditional variances in the equation:

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 \tag{34}$$

Where: - p is the order of the GARCH component, - β_j are the coefficients of the lagged conditional variances.

GJR-GARCH Model

The Glosten-Jagannathan-Runkle GARCH (GJR-GARCH) model introduces an asymmetry in the response of volatility to positive and negative shocks:

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 + \gamma \epsilon_{t-i}^2 I_{\{\epsilon_{t-i} < 0\}} \tag{35}$$

Where: - I is an indicator function that equals 1 when the condition is met.

EGARCH Model

The Exponential GARCH (EGARCH) model allows for a logarithmic specification of the conditional variance, capturing asymmetry and leverage effects:

$$\log(\sigma_t^2) = \alpha_0 + \sum_{i=1}^{q} \alpha_i \left(\frac{\epsilon_{t-i}}{\sigma_{t-i}} \right) + \sum_{j=1}^{p} \beta_j \log(\sigma_{t-j}^2) \tag{36}$$

NGARCH Model

The Nonlinear GARCH (NGARCH) model incorporates nonlinear effects in the volatility dynamics:

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 + \sum_{k=1}^{m} \gamma_k \epsilon_{t-k}^2 \left(\frac{\sigma_{t-k}}{\sigma_{t-k}^2} \right) \tag{37}$$

APARCH Model

The Asymmetric Power ARCH (APARCH) model generalizes the GARCH model to allow for different powers of the shocks:

$$\sigma_t^p = \alpha_0 + \sum_{i=1}^{q} \alpha_i \left(|\epsilon_{t-i}|^\delta + \gamma \epsilon_{t-i} \right) + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^p \tag{38}$$

Where p and δ allow for varying distributions of the errors.

FIGARCH Model

The Fractional Integrated GARCH (FIGARCH) model allows for long memory in the volatility process:

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 + \mathcal{D}(L)(1-L)^d \sigma_t^2 \tag{39}$$

Where $\mathcal{D}(L)$ represents the lag operator and d is the fractional differencing parameter.

Akaike Information Criterion (AIC)

The Akaike Information Criterion (AIC) is a widely used metric for model selection in statistical analysis. It helps to identify the best-fitting model among a set of candidates by balancing model fit and complexity.

The AIC penalizes models with more parameters to prevent overfitting, thus encouraging simpler models that still capture the underlying data structure.

The AIC is calculated using the following formula:

$$\text{AIC} = 2k - 2\ln(L) \tag{40}$$

Where: - k is the number of parameters in the model, - L is the maximum value of the likelihood function for the model.

When comparing different models, the one with the lowest AIC value is considered the best. This approach allows researchers to evaluate trade-offs between model accuracy and complexity effectively. AIC does not provide an absolute measure of fit but rather serves as a relative tool for model comparison.

In practice, the AIC can be applied to various types of models, including linear regression, time series models, and more complex statistical models. By assessing multiple models, researchers can select one that achieves a good balance of explanatory power and parsimony.

Conditional Volatility

Conditional volatility refers to the volatility of a financial time series that can change over time, depending on past values of the series.

In financial markets, volatility is not constant and tends to cluster, meaning that high-volatility periods are often followed by more high-volatility periods and vice versa. This characteristic makes modeling conditional volatility crucial for understanding and forecasting financial returns.

In the context of ARCH models, conditional volatility at time t can be represented as:

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 \tag{41}$$

Where: - σ_t^2 is the conditional variance (volatility squared) at time t, - α_0 is a constant term, - α_i are coefficients that determine the contribution of past error terms ϵ_{t-i} to the current volatility, - q is the order of the ARCH model.

Forecasting conditional volatility using ARCH models provides several benefits:

1. **Risk Management**: Understanding and predicting volatility helps investors manage risk by adjusting their portfolios based on anticipated market fluctuations.

2. **Option Pricing**: Conditional volatility is crucial for pricing options and other derivatives, as it affects the implied volatility and, consequently, option premiums.

3. **Financial Decision-Making**: Accurate forecasts of volatility allow traders and analysts to make informed decisions regarding asset allocation, hedging strategies, and investment timing.

4. **Regulatory Compliance**: Financial institutions often need to assess their exposure to volatility for compliance with regulations regarding capital adequacy and risk assessment.

By leveraging ARCH models to forecast conditional volatility, practitioners can enhance their analytical capabilities and improve the robustness of their financial models.

Standardized Residuals from ARCH Models

Standardized residuals are a critical component of time series analysis, particularly in the context of ARCH models. They provide insights into the model fit and the underlying assumptions about the distribution of residuals.

The standardized residuals, denoted as z_t, can be calculated from the residuals ϵ_t and the conditional standard deviation σ_t obtained from the ARCH model:

$$z_t = \frac{\epsilon_t}{\sigma_t} \tag{42}$$

Where: - $\epsilon_t = y_t - \mu$ is the residual at time t (the difference between the observed value and the mean), - σ_t is the conditional standard deviation (the square root of the conditional variance) estimated by the ARCH model.

Standardized residuals are useful for several reasons:

1. **Residual Analysis**: By analyzing the standardized residuals, researchers can assess the goodness-of-fit of the ARCH model. If the model is appropriately specified, the standardized residuals should resemble a white noise process, exhibiting no patterns or serial correlation.

2. **Model Validation**: Standardized residuals can be plotted to evaluate the model's assumptions, such as homoscedasticity (constant variance) and normality. Any deviations from these assumptions may indicate that the model needs refinement or that a different model might be more appropriate.

3. **Comparison of Models**: When comparing different ARCH models, standardized residuals can provide a basis for evaluation. Models that yield standardized residuals with better characteristics (e.g., closer to a normal distribution) may be preferred.

4. **Detection of Outliers**: Standardized residuals can help identify outliers in the data. Large standardized residuals (typically those greater than 3 or less than -3) may suggest the presence of anomalies in the underlying data.

By utilizing standardized residuals from ARCH models, analysts can enhance their understanding of model performance, validate model assumptions, and improve the robustness of financial forecasting.

Root Mean Square Error (RMSE) and Mean Absolute Error (MAE)

Root Mean Square Error (RMSE)

Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) are commonly used metrics for evaluating the accuracy of predictive models, particularly in the context of regression analysis and time series forecasting.

RMSE measures the average magnitude of the errors between predicted and observed values. It is sensitive to large errors due to the squaring of the residuals, making it useful for highlighting significant discrepancies.

The RMSE is calculated using the following formula:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{43}$$

Where: - n is the number of observations, - y_i is the observed value, - $\hat{y}_i$ is the predicted value.

Mean Absolute Error (MAE)

MAE provides the average of the absolute differences between predicted and observed values. Unlike RMSE, it treats all errors equally, which can make it more interpretable.

The MAE is calculated using the following formula:

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{44}$$

Where: - n is the number of observations, - y_i is the observed value, - $\hat{y}_i$ is the predicted value.

Both RMSE and MAE provide valuable insights into the performance of predictive models. RMSE is often preferred when larger errors are particularly undesirable, while MAE is useful when all errors should be treated equally. Depending on the context and specific requirements of the analysis, either metric can be selected to assess model accuracy.

Value at Risk (VaR)

Value at Risk (VaR) is a statistical measure used to assess the risk of loss on a portfolio of financial assets.

It provides an estimate of the potential loss in value of an asset or portfolio over a specified time period for a given confidence interval.

VaR is widely used in risk management, regulatory frameworks, and financial reporting.

VaR can be expressed mathematically as follows:

$$\text{VaR}_\alpha = -Q_\alpha(X) \tag{45}$$

Where: - VaR_α is the Value at Risk at a confidence level α (e.g., 95- $Q_\alpha(X)$ is the α-quantile of the loss distribution X.

For example, if the VaR at the 95% confidence level is $1 million, it means there is a 5% probability that the portfolio could lose more than $1 million over the specified time period (e.g., one day).

This measure helps investors and risk managers to understand the potential for extreme losses and to allocate capital accordingly.

There are several methods to calculate VaR, including:
- **Parametric Approach**: Assumes a normal distribution of returns, using the mean and standard deviation to derive VaR.
- **Historical Simulation**: Uses historical returns to simulate potential future losses, allowing for the estimation of VaR without distributional assumptions.
- **Monte Carlo Simulation**: Involves generating a large number of random scenarios for returns based on estimated parameters, allowing for the calculation of VaR from simulated outcomes.

Selecting an appropriate method for calculating VaR depends on the specific context, data availability, and the characteristics of the portfolio being analyzed.

Expected Shortfall (ES)

Expected Shortfall (ES), also known as Conditional Value at Risk (CVaR), is a risk measure that quantifies the expected loss on an investment in the worst-case scenarios beyond the Value at Risk (VaR) threshold.

It provides a more comprehensive view of risk compared to VaR by taking into account the magnitude of losses that exceed the VaR level.

The Expected Shortfall at a confidence level α can be expressed mathematically as:

$$\mathrm{ES}_\alpha = E[L \mid L > \mathrm{VaR}_\alpha] \tag{46}$$

Where: - ES_α is the Expected Shortfall at confidence level α, - L is the loss distribution, - VaR_α is the Value at Risk at confidence level α.

For example, if the Expected Shortfall at the 95% confidence level is $1.5 million, it means that in the worst 5% of scenarios, the average loss is $1.5 million or more. This measure helps risk managers and investors to better understand the tail risk of their portfolios and to make more informed decisions regarding capital allocation and risk mitigation strategies.

1. **Tail Risk Assessment**: Unlike VaR, which only indicates the threshold loss, ES provides information about the expected loss in scenarios where losses exceed that threshold, making it a more robust measure for assessing tail risk.
2. **Regulatory Compliance**: Financial institutions often utilize ES for regulatory purposes as it aligns with the requirement to evaluate extreme risk events comprehensively.
3. **Risk Management**: ES assists in risk management by allowing firms to quantify potential extreme losses, enabling more effective capital reserves and risk mitigation strategies.

Expected Shortfall is a crucial tool for risk assessment, offering deeper insights into potential extreme losses that could impact financial stability.

Importance of Calculating VaR and Expected Shortfall Failure Rates

The failure rates of Value at Risk (VaR) and Expected Shortfall (ES) are crucial for effective risk management. Below are some key reasons for calculating these failure rates:

- **Performance Evaluation of Risk Models**
  - **Model Validation:** Assessing VaR and ES failure rates helps validate the effectiveness of risk models. A model that consistently underestimates risk may require recalibration or replacement.
  - **Accuracy of Risk Estimates:** By comparing actual losses to the predicted values, firms can gauge how accurately their models are capturing risk. High failure rates indicate that the models may not be capturing the underlying distribution of returns adequately.
- **Risk Management and Decision-Making**
  - **Resource Allocation:** Understanding the failure rates allows organizations to allocate capital more effectively. If failure rates are high, firms may need to hold more capital reserves to cover potential losses.
  - **Strategic Decisions:** Knowledge of risk estimates and their failure rates informs strategic decisions, such as entering new markets, launching new products, or adjusting trading strategies.
- **Regulatory Compliance**
  - **Meeting Regulatory Standards:** Financial institutions often face regulatory requirements that mandate robust risk management practices. Calculating and reporting VaR and ES failure rates helps ensure compliance with these regulations.
  - **Stress Testing:** Regulatory bodies often require firms to conduct stress tests and scenario analyses. Monitoring failure rates aids in these assessments, ensuring that firms are prepared for adverse conditions.
- **Identifying Tail Risk**
  - **Understanding Extreme Events:** VaR only provides information about potential losses at a specific confidence level, whereas ES measures the average loss in the worst-case scenarios beyond the VaR threshold. Calculating failure rates highlights how often extreme events occur and their impact.
  - **Improving Risk Profiles:** By analyzing the frequency and impact of losses beyond the VaR level, firms can better understand their exposure to tail risks and take appropriate measures to mitigate them.

$$\text{VaR Failure Rate} = \frac{N_{\text{fail}}}{N_{\text{total}}} \tag{47}$$

Where:

- N_{fail} is the number of instances where actual losses exceed the VaR estimate.
- N_{total} is the total number of observations or periods assessed.

$$\text{ES Failure Rate} = \frac{N_{\text{ES fail}}}{N_{\text{total}}} \tag{48}$$

Where:

- $N_{\text{ES fail}}$ is the number of instances where actual losses exceed the ES estimate.
- N_{total} is the total number of observations or periods assessed.

Description of the research context

Vanguard Information Technology ETF (VGT)

The Vanguard Information Technology ETF (VGT) is an exchange-traded fund that aims to track the performance of a benchmark index representing the investment returns of information technology stocks.

The fund follows the MSCI US Investable Market Information Technology 25/50 Index, providing broad exposure to the U.S. technology sector across large-, mid-, and small-cap companies.

Investment Approach

VGT employs a passively managed, full-replication strategy whenever possible, ensuring the fund closely mirrors the index.

When regulatory constraints prevent full replication, a sampling approach is used to approximate the key characteristics of the index. The fund remains fully invested, minimizing net tracking error through low expenses.

Benchmark: MSCI US IMI Information Technology 25/50 Index

The MSCI US IMI Information Technology 25/50 Index serves as the benchmark for VGT, comprising U.S. companies in the technology sector, including technology software and services, hardware and equipment, and semiconductor manufacturers.

The index uses the Global Industry Classification Standard (GICS®) to classify companies, ensuring comprehensive representation of the target sector.

The 25/50 constraints ensure diversification by imposing investment limits on regulated investment companies, in accordance with the US Internal Revenue Code.

Importance of VGT Stocks

The top companies in VGT, such as Apple, Microsoft, NVIDIA, Broadcom, Oracle, Salesforce, Advanced Micro Devices, Adobe, Accenture, and Cisco Systems, play a critical role in the global technology landscape. These firms are leaders in their respective fields, including software development, cloud computing, semiconductor manufacturing, digital services, and IT consulting. Their products and services drive significant portions of the digital economy, from consumer electronics to enterprise solutions.

The importance of these companies extends beyond financial markets. They influence technological progress, set industry standards, and shape consumer behavior worldwide. For instance, Apple and Microsoft lead in software and hardware innovations, while NVIDIA drives advancements in artificial intelligence (AI) and graphics processing. Companies like Oracle and Salesforce provide essential cloud-based business solutions, while firms such as AMD and Broadcom are central to semiconductor development. These companies' financial performance often serves as a barometer for the broader sector and can significantly impact market sentiment.

Sector Overview and Dynamics

The technology sector is characterized by rapid change, driven by innovation, evolving consumer preferences, and economic factors. Currently, several macroeconomic and geopolitical dynamics are influencing the sector:

Tariff War Between the U.S. and China The ongoing trade tensions between the U.S. and China, including tariffs on technology products, have major implications for companies in the sector. Many tech firms rely on complex global supply chains with significant components and assembly processes based in China. The tariffs can increase production costs, disrupt supply chains, and affect profitability, particularly for semiconductor and hardware manufacturers. However, the tensions have also prompted firms to diversify their supply chains, explore alternative manufacturing locations, and invest in domestic production to reduce dependency on Chinese markets.

Interest Rate Cuts and Monetary Policy Changes in interest rates and monetary policy have a considerable effect on the technology sector. Low interest rates make it cheaper for companies to borrow for investments in research and development, expansion, and acquisitions. This environment is favorable for growth-driven tech firms, which often rely on substantial funding for innovation. Interest rate cuts can also boost consumer spending, which benefits companies that sell consumer technology products. However, if interest rates rise, it can lead to higher borrowing costs and affect stock valuations, particularly for high-growth firms with future-oriented earnings projections.

Sector Trends

The technology sector continues to evolve with several prominent trends:

- **Artificial Intelligence and Machine Learning**: Increasing adoption of AI technologies across various industries, including healthcare, finance, and manufacturing, is driving demand for software, services, and specialized hardware.
- **Cloud Computing and Digital Transformation**: Companies continue to migrate to cloud-based infrastructure and services, which enhances productivity and flexibility. This trend benefits cloud service providers and firms offering digital business solutions.
- **Semiconductor Innovation**: With the rising demand for chips in devices, data centers, and electric vehicles, semiconductor companies are investing in advanced manufacturing processes and next-generation chip designs.
- **Cybersecurity**: As businesses increasingly operate online, the demand for cybersecurity solutions is growing, creating opportunities for companies offering advanced security services and software.

These dynamics underscore the critical role of the technology sector in the global economy, influencing everything from consumer behavior to industrial practices. The top companies in VGT represent a broad spectrum of innovation and are essential for driving future economic growth.

Models used

ARMA-GARCH Model with Non-Gaussian Distribution

The ARMA-GARCH model combines the Autoregressive Moving Average (ARMA) model for modeling the mean of a time series and the Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model for modeling the variance. When using a non-Gaussian distribution, such as the Generalized Error Distribution (GED), the model can better capture the heavy tails and skewness often present in financial returns.

The ARMA(p, q)-GARCH(r, s) model is defined as:

$$y_t = \mu + \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{j=1}^{q} \theta_j \epsilon_{t-j} + \epsilon_t, \tag{49}$$

$$\epsilon_t = \sigma_t z_t, \tag{50}$$

$$\sigma_t^2 = \omega + \sum_{i=1}^{r} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{s} \beta_j \sigma_{t-j}^2, \tag{51}$$

where y_t is the return series at time t, μ is the mean, ϕ_i and θ_j are the AR and MA coefficients, σ_t^2 is the conditional variance, ω, α_i, and β_j are the GARCH model parameters, and z_t follows a non-Gaussian distribution, such as GED.

Characteristics:

- Captures volatility clustering in financial returns.
- Allows for modeling non-Gaussian error distributions, providing flexibility for heavy tails.
- Suitable for financial time series exhibiting conditional heteroskedasticity.

Example: ARMA(1,1)-APARCH(1,1) Model with GED Distribution

The ARMA(1,1)-APARCH(1,1) model extends the GARCH framework by incorporating an Asymmetric Power ARCH (APARCH) structure, allowing for asymmetry and power transformations in the conditional variance equation.

The model is defined as:

$$y_t = \mu + \phi_1 y_{t-1} + \theta_1 \epsilon_{t-1} + \epsilon_t, \tag{52}$$

$$\epsilon_t = \sigma_t z_t, \tag{53}$$

$$\sigma_t^\delta = \omega + \alpha_1 \left(|\epsilon_{t-1}| - \gamma\epsilon_{t-1}\right)^\delta + \beta_1 \sigma_{t-1}^\delta, \tag{54}$$

where: - δ is the power parameter. - γ captures the leverage effect, allowing for asymmetric responses to positive and negative shocks. - z_t follows the GED distribution.

Characteristics: - The APARCH model's flexibility enables better modeling of asymmetries in financial time series.

- The use of GED allows capturing fat tails in the distribution of returns.
- Suitable for financial data with non-linear patterns in volatility.

These descriptions will help set the foundation for discussing the models used in your thesis and their application to financial time series data.

Research motivations

The motivation behind this research is driven by the need to understand and enhance financial modeling in the context of technology sector stocks. This study aims to determine whether the observed outperformance of tech stocks is supported by strong financial fundamentals and sustained growth potential. By evaluating key financial indicators, we can assess if these stocks' returns are justified or if the market is overly optimistic about their future prospects.

Another key aspect of this research is identifying the most suitable non-Gaussian distribution for modeling financial returns. The heavy tails and skewness often present in return series make traditional Gaussian assumptions inadequate for accurately predicting future behavior. This study will focus on finding the best models for both conditional mean and conditional variance, which are critical for forecasting volatility and returns in a more reliable manner.

To achieve these goals, we will develop an algorithm that systematically finds the optimal combination of ARMA orders for the mean equation, the best type of GARCH model, its corresponding parameters, and the most appropriate distribution. This comprehensive approach ensures that the chosen models capture the complexities of financial time series.

Ultimately, this research aims to improve risk management practices by building risk measures that account for the true distributional characteristics of financial returns. Using more efficient distributions than the commonly assumed Gaussian can lead to more accurate Value-at-Risk and Expected Shortfall estimates, thus enhancing the reliability of risk forecasts and decision-making.

Methodology

1.1 Price performance

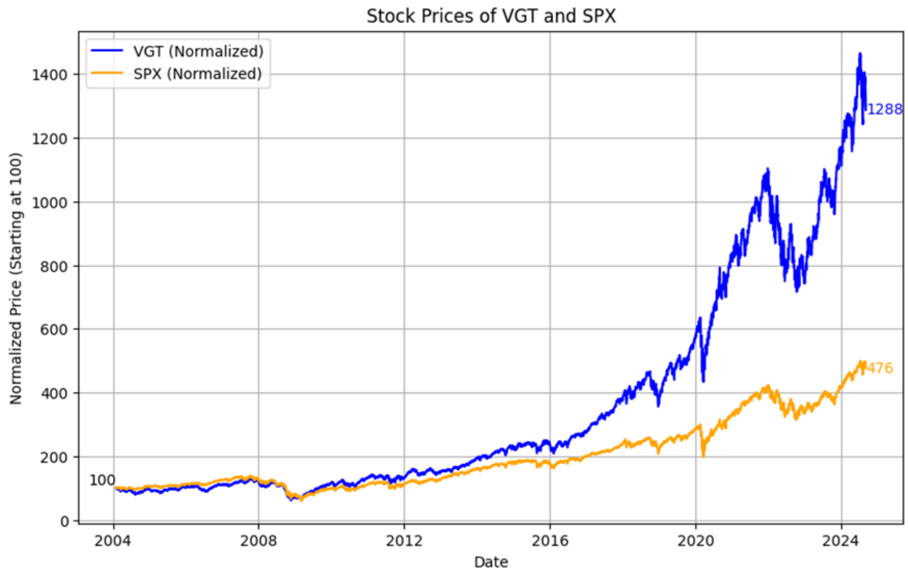


Figure 2: Normalized VGT and SPX prices

The chart shows the normalized price performance of the VGT (Vanguard Information Technology ETF) and the SPX (S&P 500 Index) from 2004 to 2024. Starting at an equal value of 100, VGT significantly outperforms SPX, reaching a value of 1288 compared to SPX's 476 by 2024. This indicates that technology stocks have experienced much higher growth compared to the broader market, reflecting the sector's strong price appreciation and potential for higher returns over the long term. This trend aligns with the broader narrative of tech sector dominance in recent years

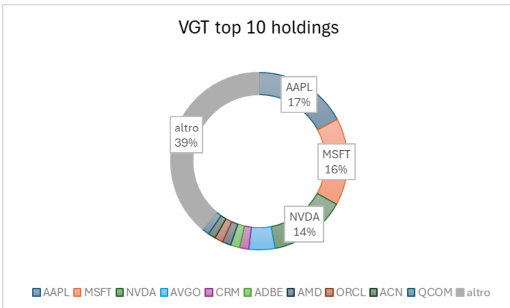
1.1.1 CAGR and Total Return

Index	Start Date	End Date	CAGR (%)	Total Return (%)
VGT	2004-02-01	2024-09-08	13.21	1187.69
SPX	2004-02-01	2024-09-08	7.87	376.40

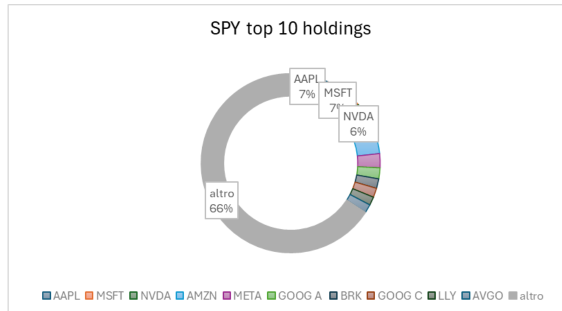
Table 1: CAGR and Total Return for VGT and SPX

The Vanguard Information Technology ETF (VGT) demonstrated a remarkable Total Return of 1187.69% from February 1, 2004, to September 8, 2024, significantly outperforming the S&P 500 Index (SPX), which had a Total Return of 376.40% over the same period. This substantial difference in returns, coupled with the higher CAGR of VGT (13.21% compared to SPX's 7.87%), underscores the superior performance and growth potential of the technology sector relative to the broader market.

1.2 Holdings



(a) VGT top 10 holdings



(b) SPX top 10 holdings

Figure 3: VGT and SPX top 10 holdings

The Vanguard Information Technology ETF (VGT) is heavily concentrated in a few major technology companies, with Apple (17.22%), Microsoft (15.84%), and NVIDIA (14.07%) making up nearly half of its total holdings. This concentration reflects the significant influence of these tech giants on the overall performance of the ETF. Other notable holdings include Broadcom (4.75%) and Salesforce (1.68%), with the remaining 38.93% spread across various other tech companies. [1]

In contrast, the SPDR S&P 500 ETF (SPY) has a more diversified portfolio, though it also includes significant positions in Apple (6.86%), Microsoft (6.78%), and NVIDIA (6.03%). The top ten holdings in SPY are more varied, including companies from different sectors such as Amazon (3.61%), Meta Platforms (2.55%), and Alphabet's Class A and C shares (1.98% and 1.66%, respectively). The “other” category in SPY is much larger at 65.76%, indicating a broader diversification across the S&P 500 index. [2]

[1]Vanguard Information Technology ETF (VGT) — investor.vanguard.com/investment-products
[2]SPDR S&P 500 ETF Trust (SPY) — investor.vanguard.com/investment-products

1.2.1 Returns of NVDA, MSFT, AAPL and SPX

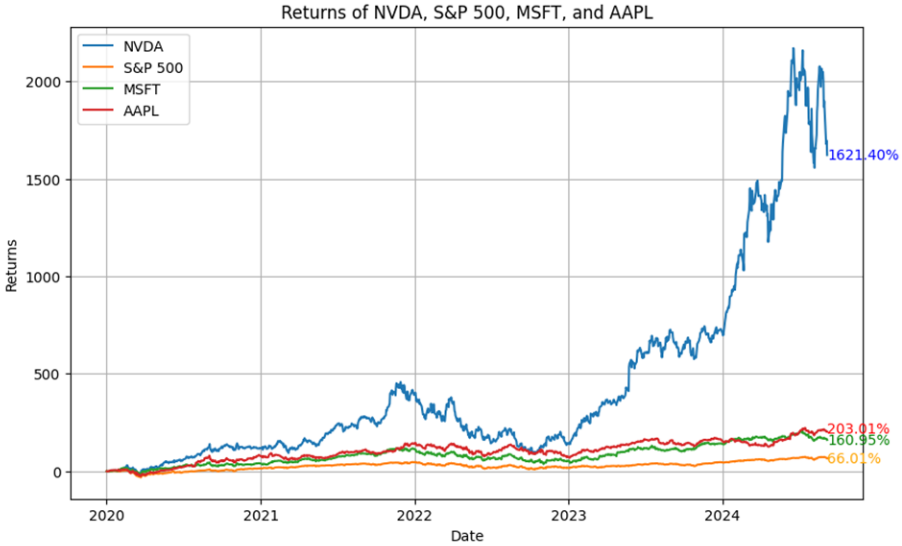


Figure 4: Tot Returns of NVDA, MSFT, AAPL and SPX

The chart illustrates the returns of NVIDIA (NVDA), the S&P 500 Index, Microsoft (MSFT), and Apple (AAPL) from 2020 to 2024. NVIDIA shows a dramatic outperformance compared to the other assets, achieving a return of 1621.40%, reflecting its rapid growth and strong market momentum, particularly in recent years. In contrast, the S&P 500's return is much more modest at 66.01%, indicating relatively steady growth for the broader market. Microsoft and Apple also outperform the S&P 500, with returns of 160.95% and 203.01% respectively, but still fall significantly short of NVIDIA's exceptional performance. This highlights the substantial impact of individual tech stocks, especially NVIDIA, on overall market dynamics and the importance of stock selection in achieving superior returns [1]

[1] YahooFinance

1.3 Sector Analysis

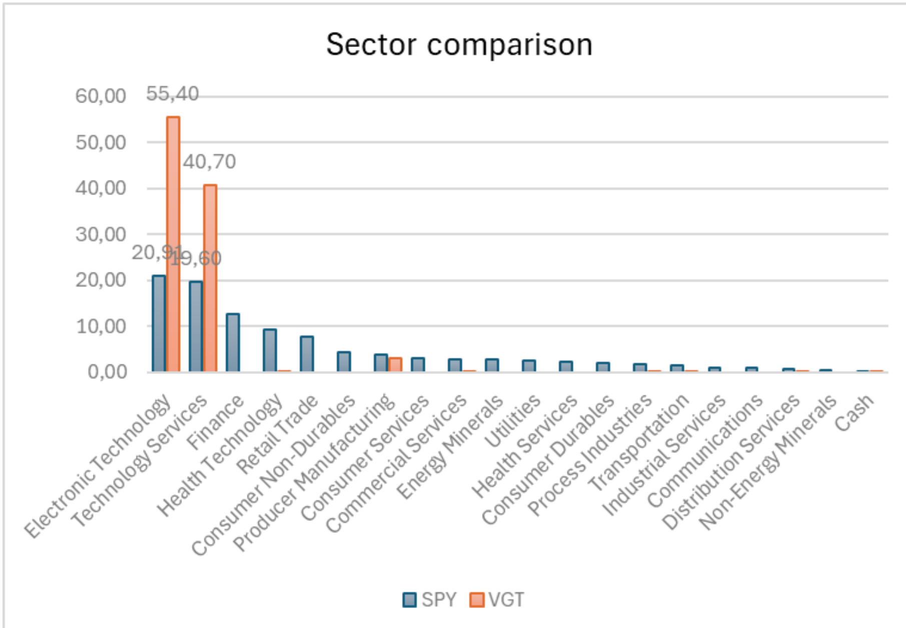


Figure 5: Sector Comparison

The sector allocation of the Vanguard Information Technology ETF (VGT) and the SPDR S&P 500 ETF (SPY) reveals significant differences in their investment strategies. VGT is heavily concentrated in the technology sector, with 55.40% in Electronic Technology and 40.70% in Technology Services. This focus underscores VGT's commitment to the technology industry, with minimal exposure to other sectors. [1]

In contrast, SPY offers a more diversified portfolio. While it also has substantial investments in Electronic Technology (20.91%) and Technology Services (19.60%), it includes significant allocations in Finance (12.71%), Health Technology (9.17%), and Retail Trade (7.81%). This diversification across multiple sectors reflects SPY's broader market approach, aiming to mirror the performance of the S&P 500 index. [2]

[1] Vanguard Information Technology ETF (VGT) — investor.vanguard.com/investment-products
[2] SPDR S&P 500 ETF Trust (SPY) — investor.vanguard.com/investment-products

1.3.1 Return of Selected Sectors

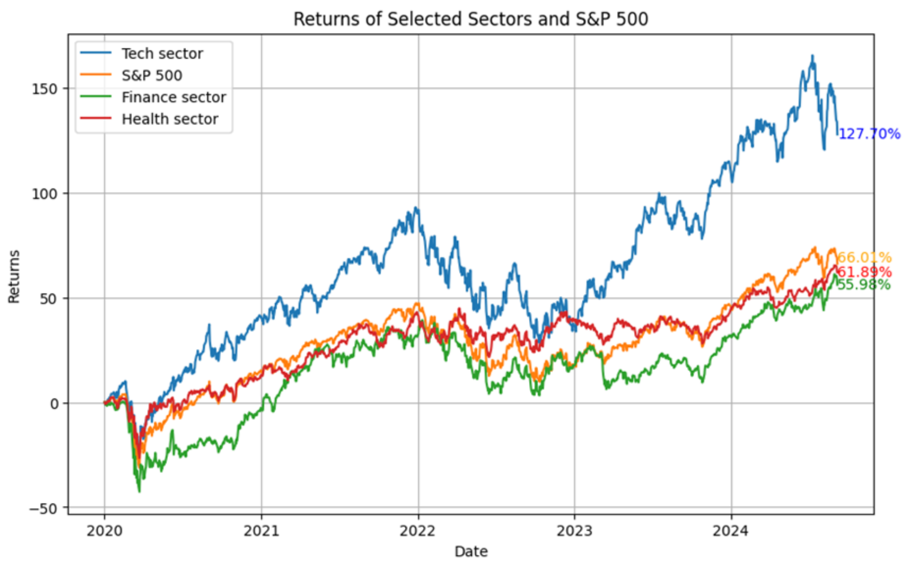


Figure 6: Return of Selected Sectors

The chart compares the returns of the tech sector, S&P 500, finance sector, and health sector from 2020 to 2024. The tech sector shows the highest growth, with a return of 127.70%, substantially outperforming the other sectors. The S&P 500, which represents the broader market, has a return of 66.01%, indicating that the overall market growth is driven significantly by the tech sector. The health sector and finance sector exhibit similar performances, with returns of 61.88% and 55.98%, respectively. This suggests that while these sectors have rebounded since 2020, their growth has been more moderate compared to the robust gains observed in the tech sector. The chart illustrates the tech sector's strong influence on overall market returns, emphasizing its role in driving market outperformance during this period. [1]

[1] YahooFinance

1.4 Market Cap Comparison

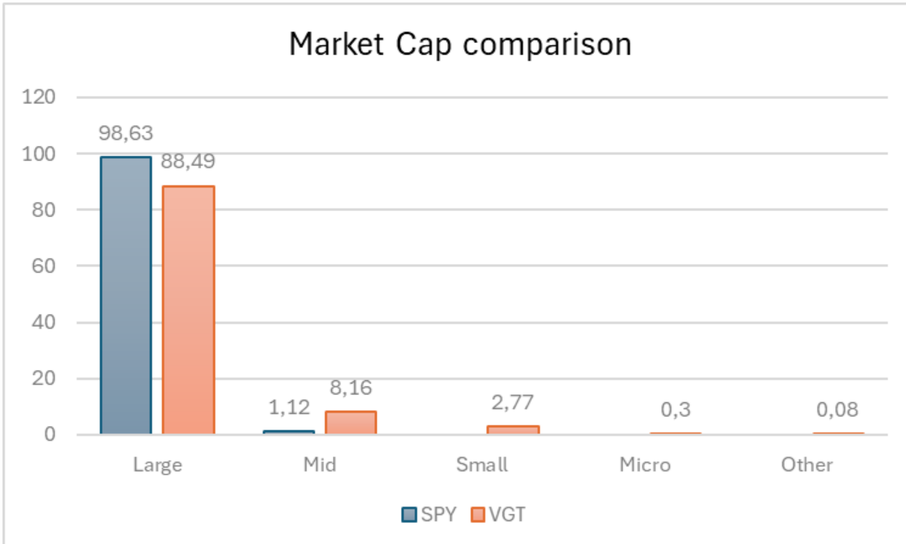


Figure 7: Market Cap Comparison

The graph presents the market capitalization distribution for two ETFs: SPY (S&P 500 ETF) and VGT (Vanguard Information Technology ETF). SPY is heavily weighted towards large-cap stocks, with 98.63% of its holdings in this category. This reflects its focus on the largest companies in the S&P 500. In comparison, VGT also has a significant portion in large-cap stocks at 88.49%, but this is slightly less than SPY, indicating a broader inclusion of smaller companies within the tech sector. [1]

When it comes to mid-cap stocks, SPY has a minimal allocation of 1.12%, emphasizing its large-cap orientation. On the other hand, VGT allocates a higher percentage to mid-cap stocks at 8.16%, suggesting a more diversified approach within the tech industry.

SPY does not include small-cap stocks, maintaining its focus on larger, more established companies. In contrast, VGT includes 2.77% in small-cap stocks, providing exposure to smaller, potentially high-growth tech companies. For micro-cap stocks, SPY has no allocation, consistent with its large-cap strategy, while VGT has a small allocation of 0.3%, adding a slight element of high-risk, high-reward investments. [2]

Both SPY and VGT have negligible allocations to other categories, with SPY at 0% and VGT at 0.08%.

[1] Vanguard Information Technology ETF (VGT) — investor.vanguard.com/investment-products
[2] SPDR S&P 500 ETF Trust (SPY) — investor.vanguard.com/investment-products

1.4.1 Returns by Market Cap

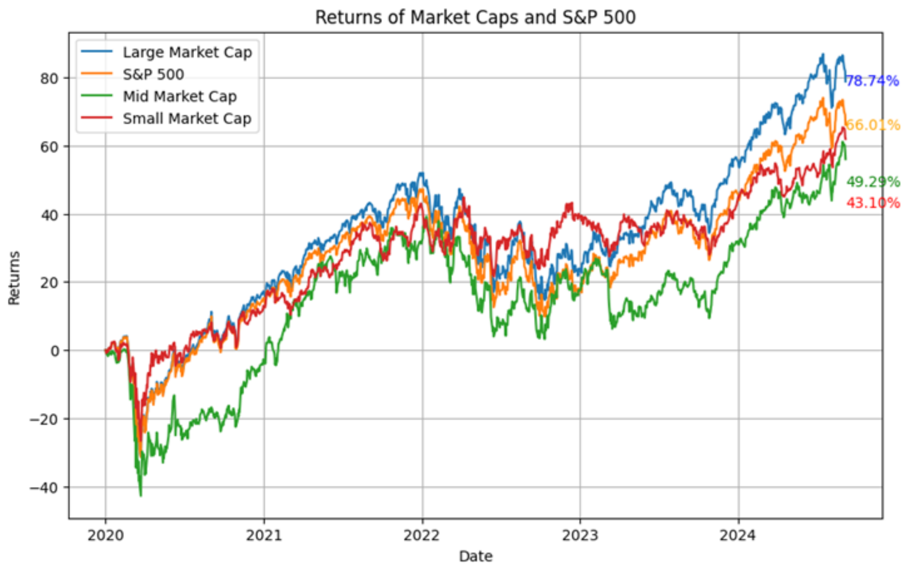


Figure 8: Returns by Market Cap

The image depicts the performance of various market cap categories (Large, Mid, Small) and the S&P 500 over a period spanning from 2020 to 2024. A clear upward trend is evident across all categories, suggesting a generally favorable market environment during this timeframe. [1]

Key Observations:

Large Market Cap: The largest market cap category consistently outperforms the others, indicating that larger companies have generally exhibited stronger growth and stability.

S&P 500: As a broad market index, the S&P 500's performance provides a benchmark. It closely tracks the Large Market Cap category, suggesting that larger companies dominate the overall market.

Mid and Small Market Caps: While these categories also show upward trends, they exhibit greater volatility compared to Large Market Cap and S&P 500. This suggests that smaller companies may be more susceptible to market fluctuations and economic conditions.

[1] YahooFinance

1.5 ROIC

ROIC (Return on Invested Capital) plays a crucial yet often underappreciated role in equity valuation, going beyond mere growth expectations. Companies with high ROIC typically possess stronger competitive advantages and better management, leading to sustainable value creation through the compounding effect. These firms can reinvest their earnings into high-return opportunities, driving superior long-term shareholder value compared to companies with lower ROIC.

Unlike traditional metrics like P/E ratios or PEG ratios, which may not adequately account for high-growth companies' valuation complexities, ROIC provides a more reliable measure of a business's quality and profitability. High ROIC indicates the presence of superior products or business models, while low ROIC reflects limited growth prospects due to fewer profitable reinvestment opportunities.

Sustainable ROIC combined with growth offers a powerful formula for valuation, especially when the market underestimates a company's true ROIC potential. Evaluating both long-term ROIC and growth prospects provides a more comprehensive approach to identifying undervalued firms and understanding a company's real long-term value.

[1]

[1] ROIC – The Underappreciated Variable in Valuation — Kennedy capital management

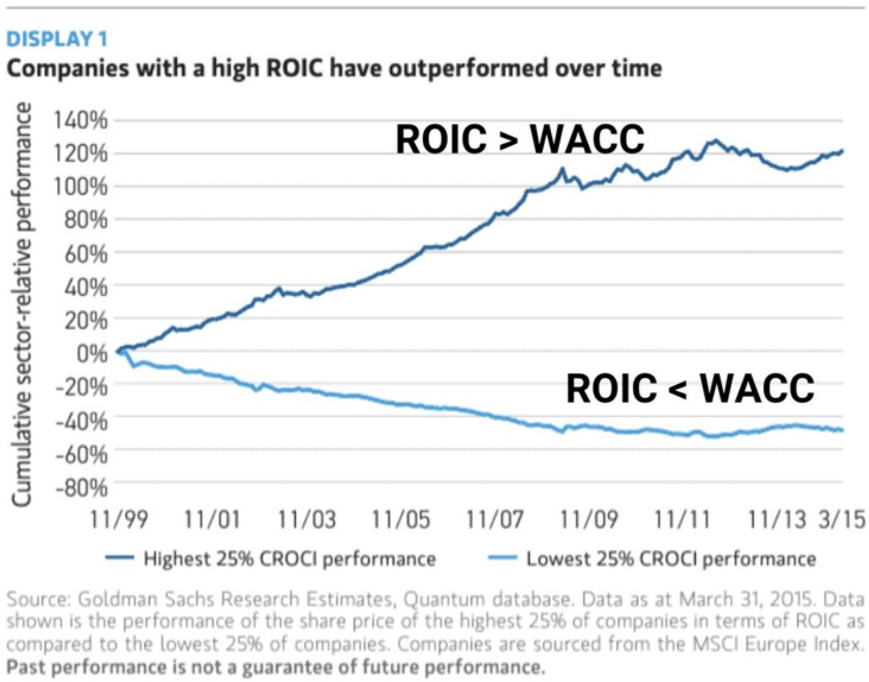


Figure 9: ROIC vs WACC

[1] Outperformance: Companies with a high ROIC (ROIC ¿ WACC) have consistently outperformed those with a low ROIC (ROIC ¡ WACC) over the time period depicted (November 1999 to March 2015).

The chart shows the cumulative performance relative to the sector average. This means that the high ROIC companies have consistently generated returns above the sector benchmark.

[1]Goldamn Sachs Global Investment Research

1.6 Top 3 holdings efficiency

1.6.1 NVDA efficiency

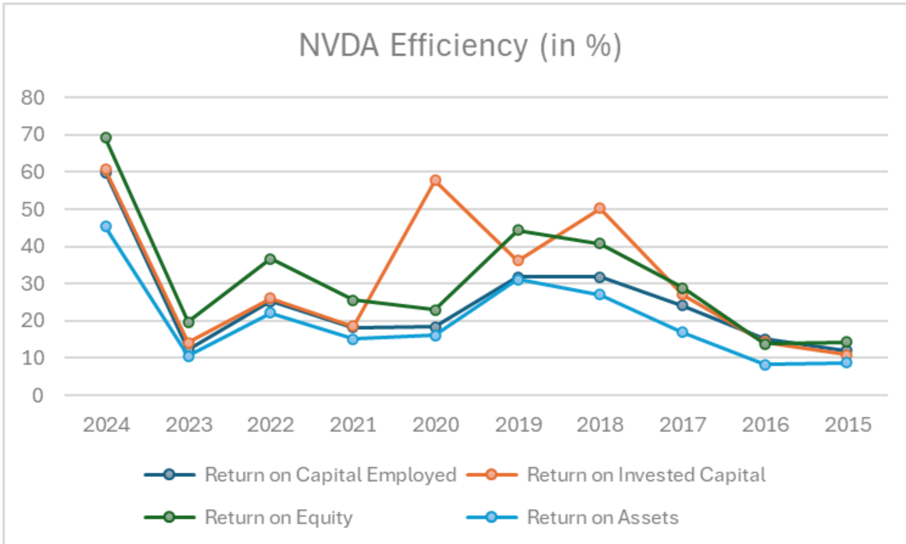


Figure 10: NVDA ROIC, ROCE, ROA and ROE

The chart shows NVIDIA's efficiency measures from 2015 to 2024, covering metrics like Return on Capital Employed (ROCE), Return on Invested Capital (ROIC), Return on Equity (ROE), and Return on Assets (ROA). Notably, ROIC and ROE surged in the last year (2024), reaching peaks around 60–70%, likely reflecting a period of strong performance and efficient capital utilization.

1.6.2 AAPL efficiency

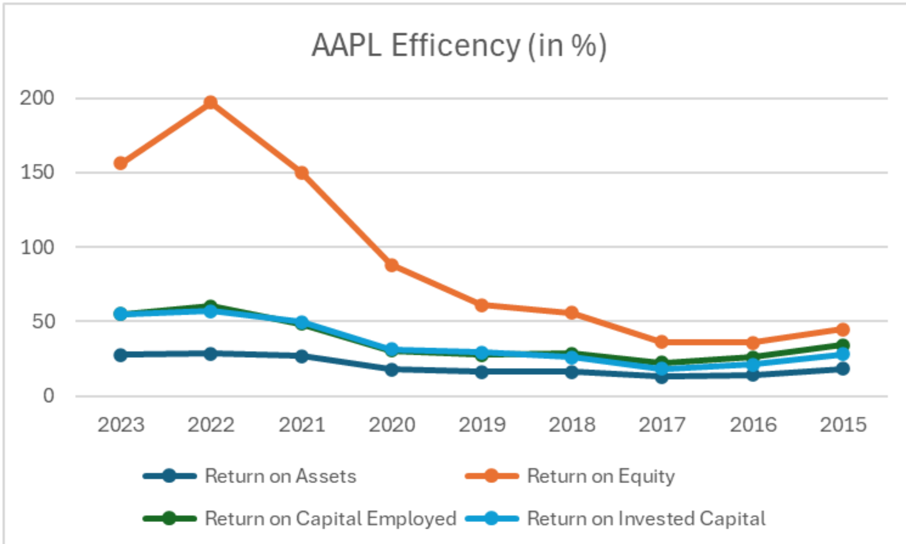


Figure 11: AAPL ROIC, ROCE, ROA and ROE

The chart illustrates Apple's efficiency measures from 2024 to 2015, including Return on Assets (ROA), Return on Equity (ROE), Return on Capital Employed (ROCE), and Return on Invested Capital (ROIC).ROE stands out with higher values, peaking above 100% around 2022, reflecting strong shareholder returns during this period. ROCE and ROIC stand above at 50%.
[1]

[1] DiscountingCashFlow.com

1.6.3 MSFT efficiency

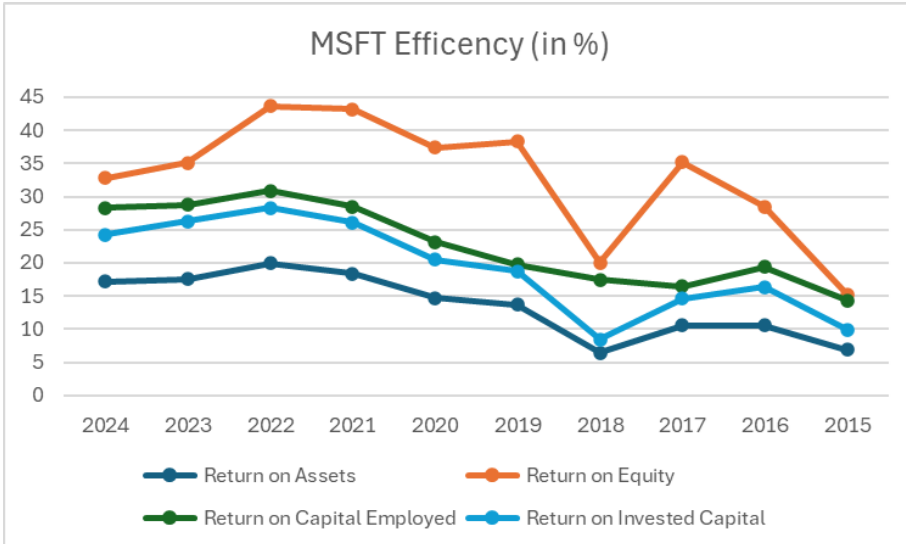


Figure 12: MSFT ROIC, ROCE, ROA and ROE

[1]

The chart illustrates Microsoft's efficiency measures from 2024 to 2015, including Return on Assets (ROA), Return on Equity (ROE), Return on Capital Employed (ROCE), and Return on Invested Capital (ROIC). The trend is positive over the time and at 2024 ROIC and ROCE stand at 25-30%.

[1] DiscountingCashFlow.com

1.6.4 ROIC: S&P 500 Performance

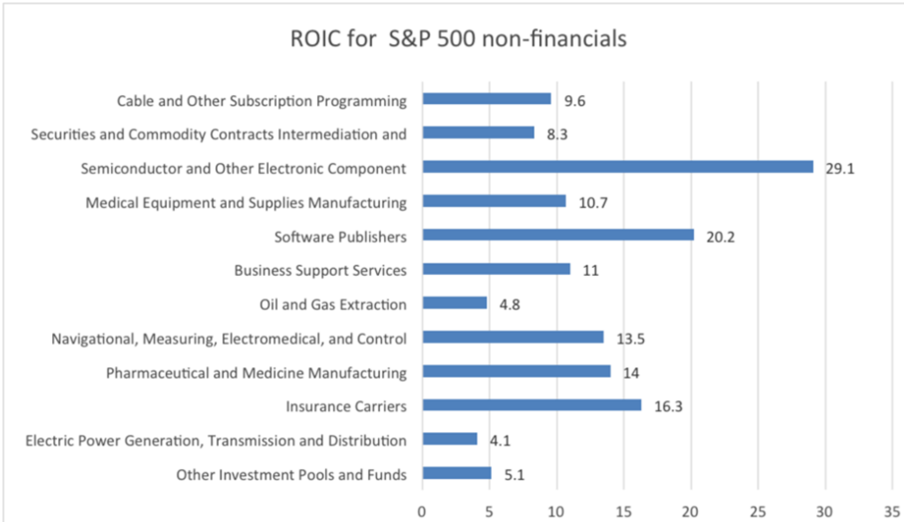


Figure 13: ROIC: S&P 500 Performance

[1]

Return on Invested Capital (ROIC) is a vital metric for assessing corporate performance. It quantifies the profit earned for every dollar invested in the company by both bondholders and shareholders. This profitability ratio is essential and is closely monitored by investors.

Among various sectors, the highest ROIC is found in the Semiconductor and Other Electronic Component Manufacturing industry, closely followed by Software Publishers. Conversely, the sectors with the lowest ROIC are Electric Power Generation and Oil and Gas Extraction.

[1] Economatica - Value reports — valuereports.economatica.com/roic-sp-500-performance

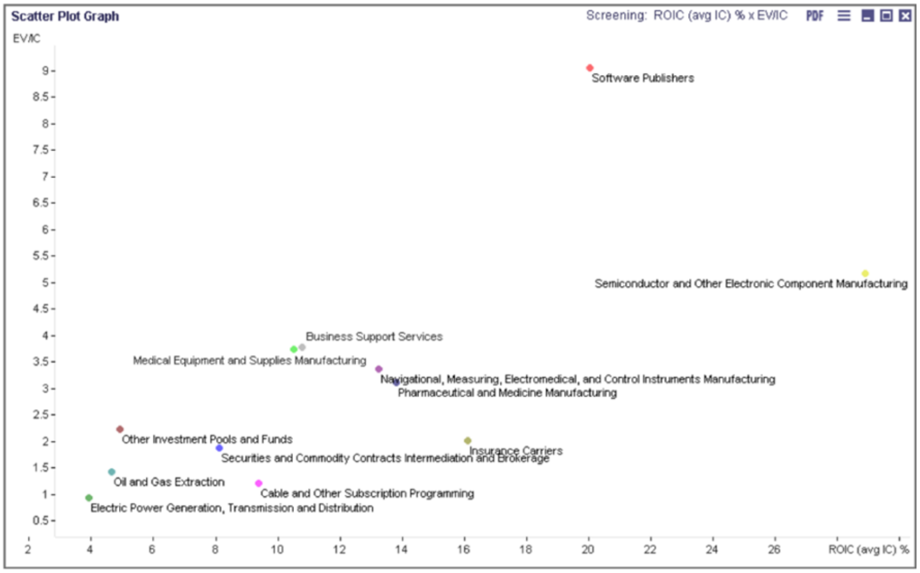


Figure 14: ROIC: S&P 500 Performance

The relationship between Enterprise Value to Invested Capital (EV/IC) and Return on Invested Capital (ROIC) provides valuable insights into a company's valuation and profitability. EV/IC serves as an alternative measure to the Price to Book Value (P/B) ratio, offering a different perspective on how a company is valued relative to its invested capital.

A higher EV/IC ratio may indicate that investors are willing to pay a premium for the company's future cash flows, especially if the ROIC is also high. Conversely, if ROIC is low while the EV/IC is high, it could suggest that the company is not generating sufficient returns on its investments relative to its valuation. [1]

It is evident that the Return on Invested Capital (ROIC) of the selected companies typically exceeds the average ROIC of the S&P 500, even when comparing specific sectors.

[1] Economatica - Value reports — valuereports.economatica.com/roic-sp-500-performance

1.7 Free Cash Flow and Earnings Per Share

1.7.1 NVDA per share data

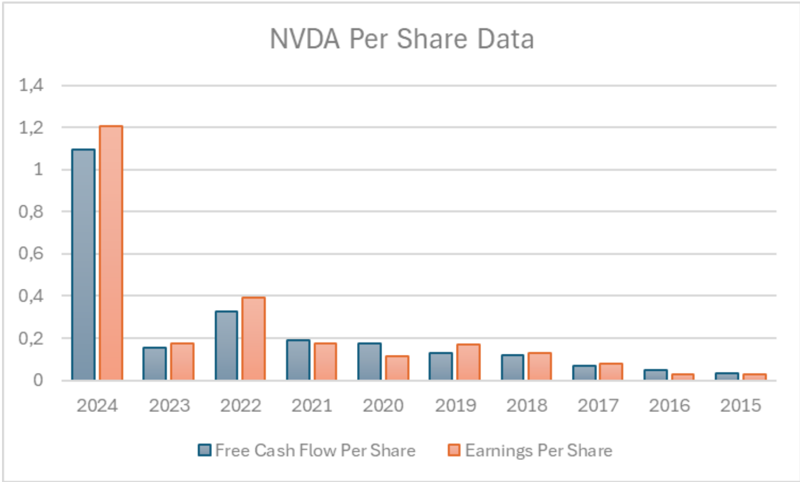


Figure 15: NVDA per share data

1.7.2 AAPL per share data

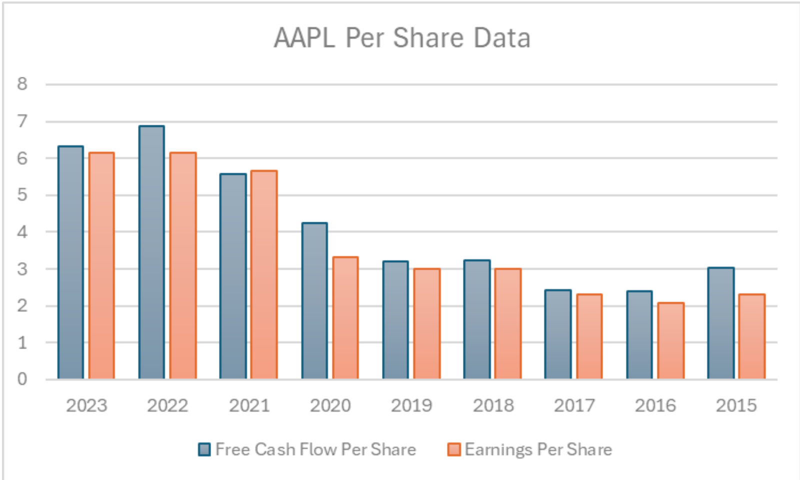


Figure 16: AAPL per share data

[1]

[1] DiscountingCashFlow.com

1.7.3 MSFT per share data

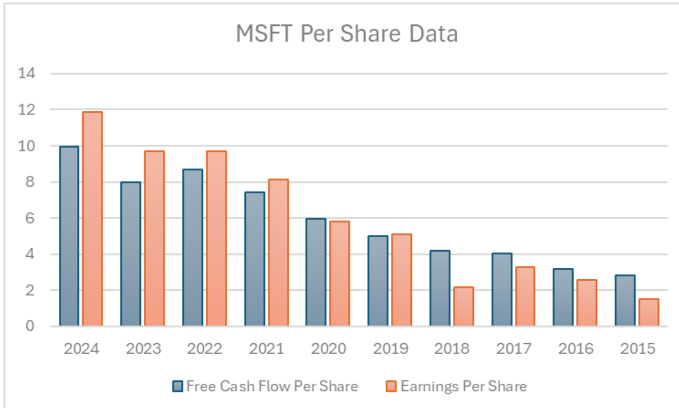


Figure 17: MSFT per share data

Earnings Per Share (EPS) and Free Cash Flow (FCF) per share are two fundamental metrics often used to assess the intrinsic value of a stock. These metrics measure a company's profitability and cash flow generation, respectively, which are key factors that influence investors' perception of a company's worth.

- **EPS as a Driver of Intrinsic Value**
  - **Profitability**: EPS reflects a company's profitability by dividing its net income by the number of outstanding shares. Higher EPS generally indicates stronger financial performance and greater earnings potential.
  - **Investor Expectations**: Investors often use EPS as a benchmark to compare companies within the same industry. A company with consistently increasing EPS may attract more investor interest and drive up its stock price.
  - **Dividend Payouts**: Many companies use EPS as a basis for determining dividend payouts. Higher EPS can lead to higher dividend payments, which can be a significant factor in attracting investors.
- **FCF per Share as a Driver of Intrinsic Value**
  - **Cash Flow Generation**: FCF per share measures a company's ability to generate cash flow after accounting for capital expenditures. It provides a more accurate picture of a company's financial health than EPS, which can be influenced by non-cash accounting items.
  - **Investment Opportunities**: Strong FCF per share indicates that a company has excess cash to reinvest in growth opportunities, pay down debt, or return capital to shareholders through dividends or buybacks.
  - **Sustainable Growth**: A company with consistent FCF per share is more likely to sustain its growth and profitability over the long term.

1.7.4 EPS Annual Growth SP500 vs Big Tech

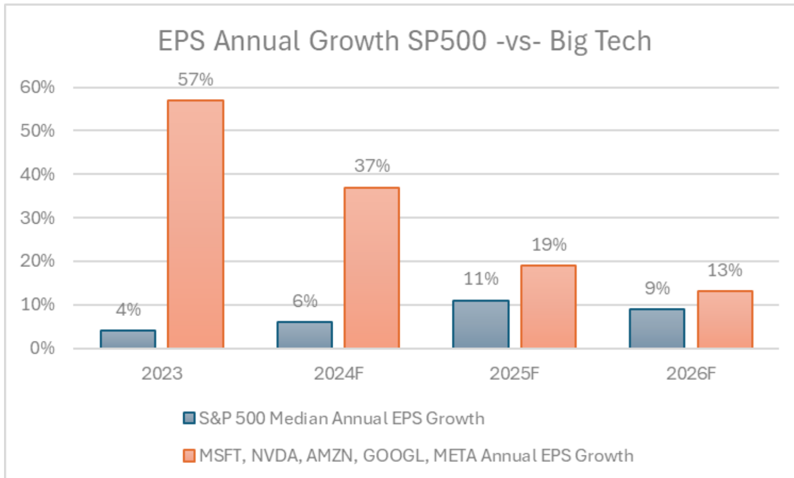


Figure 18: EPS Annual Growth SP500 vs Big Tech

The graph highlights that the selected tech companies are expected to grow their earnings at a much faster rate than the median of the S&P 500 over the given period, reflecting their strong performance and growth potential in the tech sector [1]

[1] Visualcapitalist.com

1.8 S&P500 Margins

1.8.1 S&P500 Operating Margin

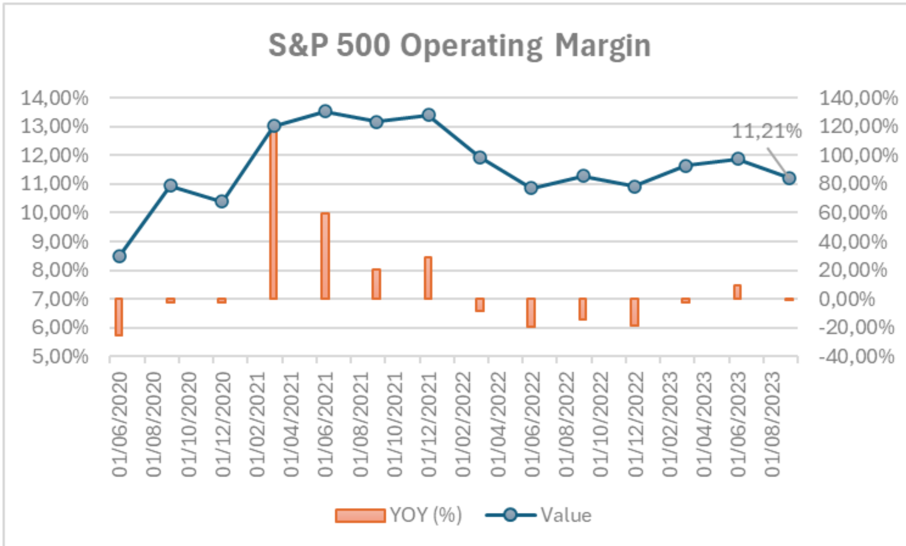


Figure 19: S&P500 Operating Margin

As of September 30, 2023, the S&P 500 operating margin stands at 11.21% [1]

1.8.2 S&P500 Revenue Growth and Net Profit Margin

The S&P 500 has demonstrated steady growth and profitability over recent years. Key financial metrics include:

- **Revenue Growth:** The 5-year average growth rate is 6.8%, while the 10-year average growth rate is 5.1%.
- **Net Profit Margin:** The 5-year average net profit margin stands at 11.5%.

[2]

[1] GuruFocus.com
[2] FactSet.com — earnings insight

1.9 Income Statement Analysis

1.9.1 NVDA Income Statement

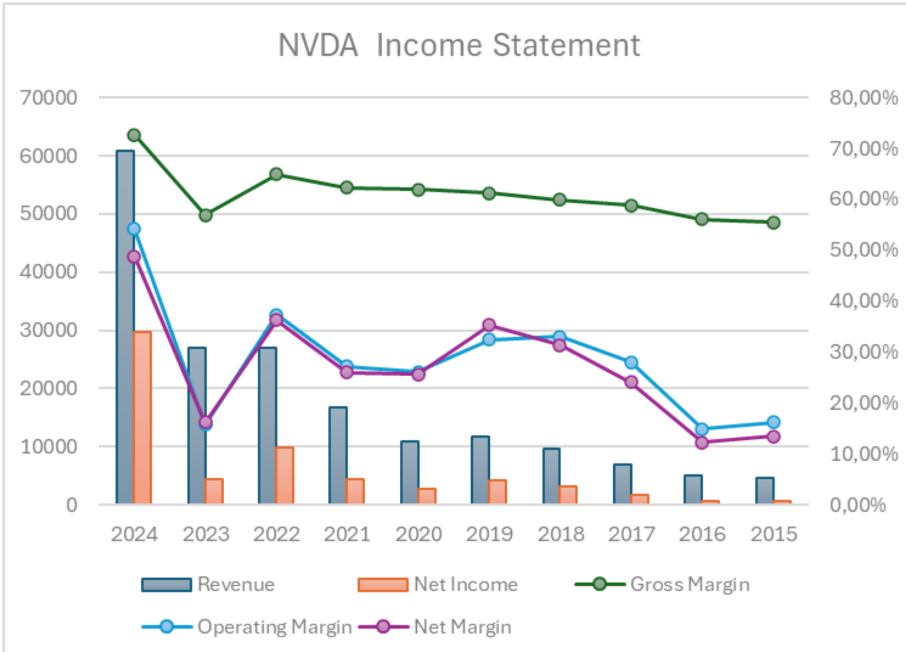


Figure 20: NVDA Income Statement

Revenue Growth: NVIDIA has experienced significant revenue growth over the past decade. Revenue has increased steadily, with a notable acceleration in recent years.

Net Income: Net income has followed a similar trend to revenue, with a significant increase in recent years.

Gross Margin: Gross margin has remained relatively stable over the period, hovering around 60%.

Operating Margin: Operating margin has fluctuated over the years but has generally shown an upward trend.

Strong Financial Performance: NVIDIA's financial performance has been impressive, with consistent revenue growth and increasing profitability.

Market Leadership: NVIDIA's strong financial performance may be attributed to its position as a market leader in the semiconductor industry, particularly in graphics processing units (GPUs). [1]

[1] DiscountingCashFlow.com

1.9.2 AAPL Income Statement

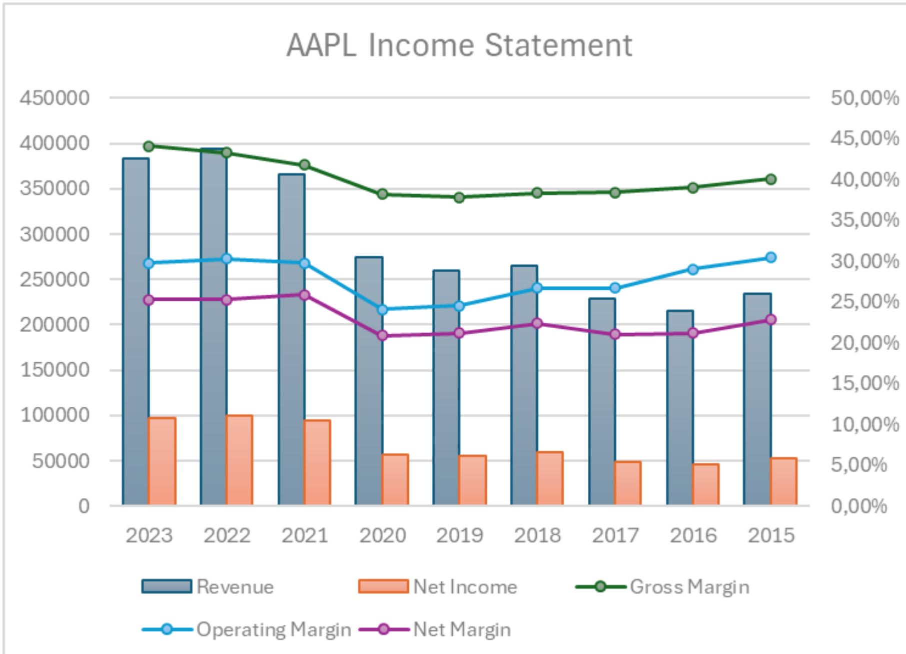


Figure 21: AAPL Income Statement

Apple has experienced consistent revenue growth over the past decade. Revenue has increased steadily, with a particularly strong acceleration in recent years.

Net income has followed a similar trend to revenue, with a significant increase in recent years.

Gross margin has remained relatively stable over the period, hovering around 40%.

Operating Margin and Net Margin: Both operating margin and net margin have fluctuated over the years but have generally shown an upward trend.

Strong Financial Performance: Apple's financial performance has been impressive, with consistent revenue growth and increasing profitability.

Market Leadership: Apple's strong financial performance may be attributed to its position as a market leader in the smartphone, tablet, and computer industries. [1]

[1] DiscountingCashFlow.com

1.9.3 MSFT Income Statement

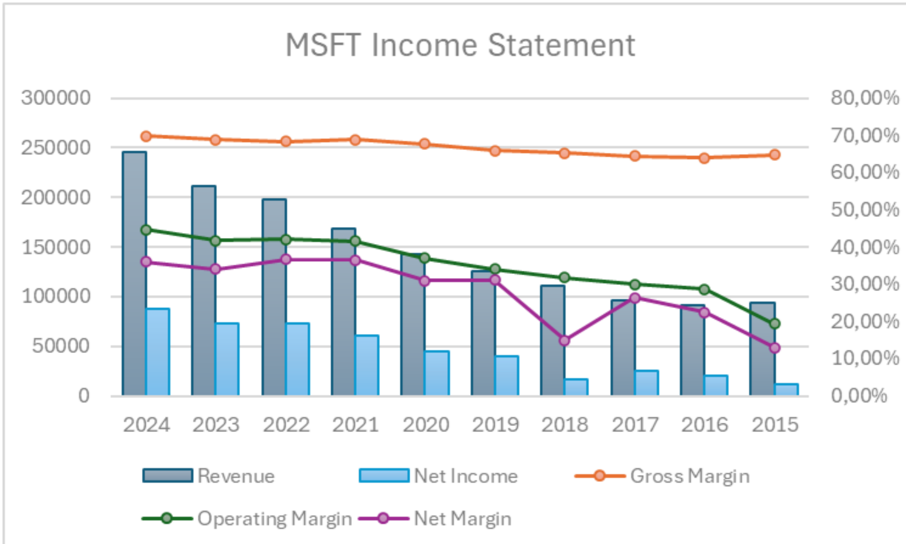


Figure 22: MSFT Income Statement

Microsoft has experienced consistent revenue growth over the past decade. Revenue has increased steadily, with a particularly strong acceleration in recent years.

Net income has followed a similar trend to revenue, with a significant increase in recent years.

Gross margin has remained relatively stable over the period, hovering around 70%.

Operating Margin and Net Margin: Both operating margin and net margin have fluctuated over the years but have generally shown an upward trend.

Strong Financial Performance: Microsoft's financial performance has been impressive, with consistent revenue growth and increasing profitability.

Market Leadership: Microsoft's strong financial performance may be attributed to its position as a market leader in software, cloud computing, and gaming. [1]

[1] DiscountingCashFlow.com

1.10 Balance Sheet Analysis

1.10.1 NVDA Balance Sheet

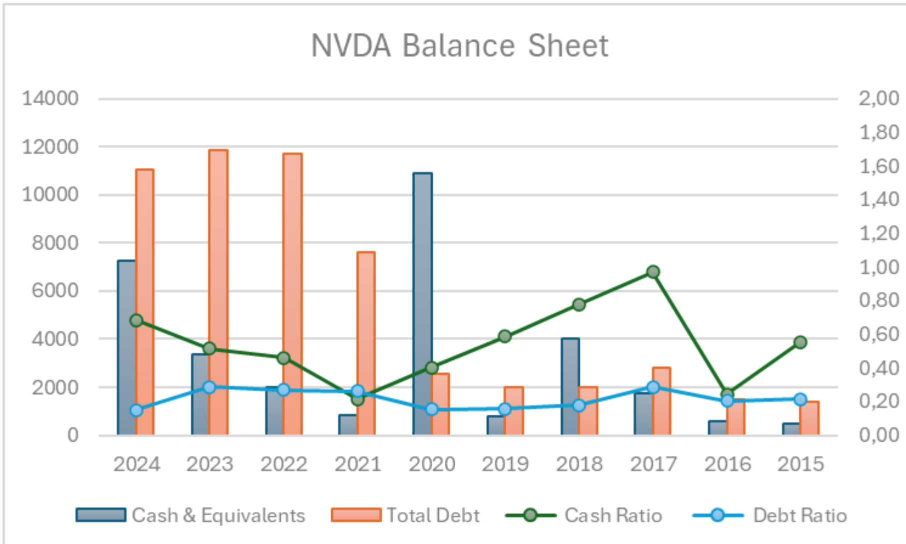


Figure 23: NVDA Balance Sheet

Cash & Equivalents: NVIDIA has consistently maintained a significant amount of cash and equivalents throughout the period.

Total Debt: Total debt has remained relatively low, suggesting a conservative approach to financing.

Cash Ratio: The cash ratio, calculated as cash and equivalents divided by current liabilities, has fluctuated over the years. However, it has generally remained above 1.0, indicating that NVIDIA has sufficient liquidity to cover its short-term obligations.

Debt Ratio: The debt ratio, calculated as total debt divided by total assets, has also fluctuated but has remained relatively low. This suggests a healthy capital structure and a limited reliance on debt financing.

Financial Strength: NVIDIA's strong cash position and low debt levels indicate a solid financial foundation.

Investment Flexibility: The significant amount of cash and equivalents provides NVIDIA with flexibility to invest in research and development, acquisitions, or other strategic initiatives.

Reduced Financial Risk: A low debt ratio can reduce the company's financial risk and improve its creditworthiness, making it easier to access financing if needed.

Liquidity: The cash ratio above 1.0 indicates that NVIDIA has sufficient liquidity to meet its short-term obligations, reducing the risk of financial distress.
[1]

[1] DiscountingCashFlow.com — Google Gemini

1.10.2 AAPL Balance Sheet

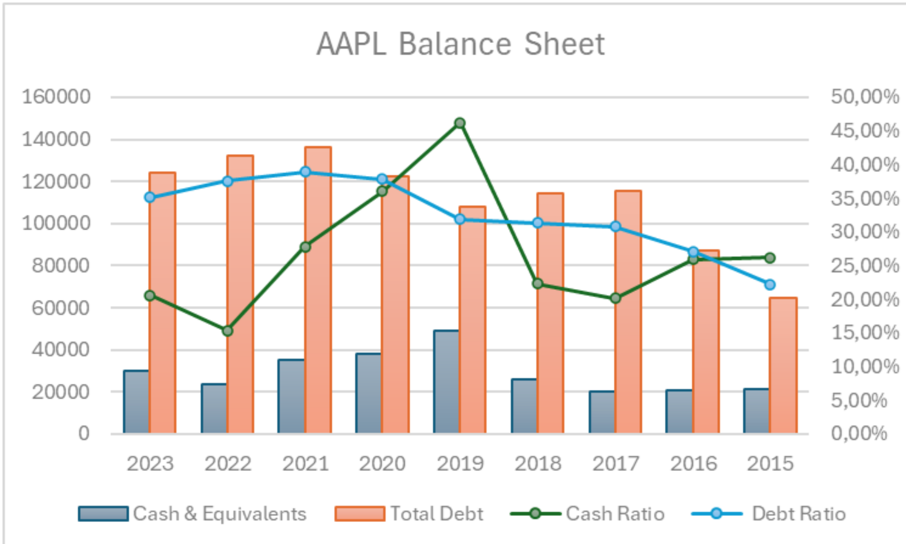


Figure 24: AAPL Balance Sheet

Cash & Equivalents: Apple has consistently maintained a significant amount of cash and equivalents throughout the period.

Total Debt: Total debt has remained relatively low, suggesting a conservative approach to financing.

Cash Ratio: The cash ratio, calculated as cash and equivalents divided by current liabilities, has fluctuated over the years. However, it has generally remained above 1.0, indicating that Apple has sufficient liquidity to cover its short-term obligations.

Debt Ratio: The debt ratio, calculated as total debt divided by total assets, has also fluctuated but has remained relatively low. This suggests a healthy capital structure and a limited reliance on debt financing.

Financial Strength: Apple's strong cash position and low debt levels indicate a solid financial foundation.

Investment Flexibility: The significant amount of cash and equivalents provides Apple with flexibility to invest in research and development, acquisitions, or other strategic initiatives.

Reduced Financial Risk: A low debt ratio can reduce the company's financial risk and improve its creditworthiness, making it easier to access financing if needed.

Liquidity: The cash ratio above 1.0 indicates that Apple has sufficient liquidity to meet its short-term obligations, reducing the risk of financial distress.
[1]

[1] DiscountingCashFlow.com — Google Gemini

1.10.3 MSFT Balance Sheet

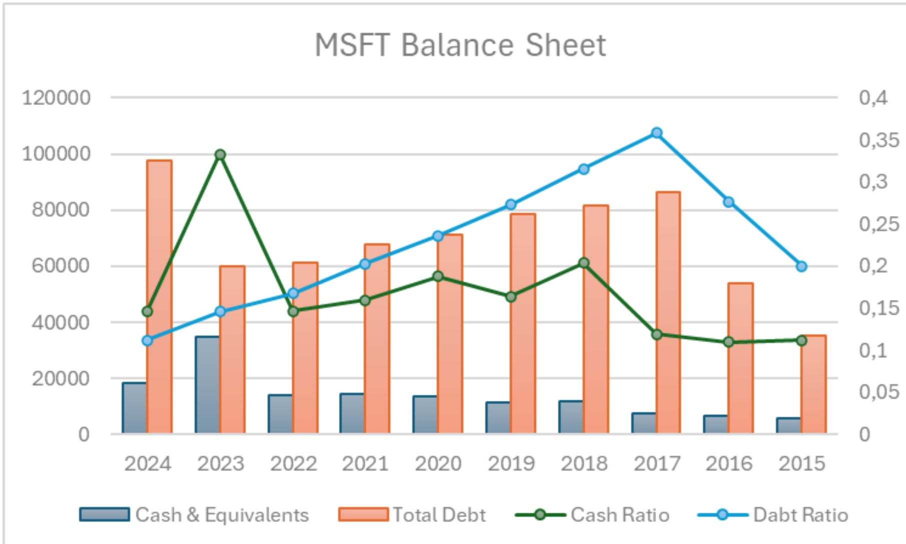


Figure 25: MSFT Balance Sheet

Cash & Equivalents: Microsoft has consistently maintained a significant amount of cash and equivalents throughout the period.

Total Debt: Total debt has remained relatively low, suggesting a conservative approach to financing. This low debt level can reduce financial risk and improve the company's creditworthiness.

Cash Ratio: The cash ratio, calculated as cash and equivalents divided by current liabilities, has fluctuated over the years. However, it has generally remained above 1.0, indicating that Microsoft has sufficient liquidity to cover its short-term obligations.

Debt Ratio: The debt ratio, calculated as total debt divided by total assets, has also fluctuated but has remained relatively low.

Financial Strength: Microsoft's strong cash position and low debt levels indicate a solid financial foundation.

Investment Flexibility: The significant amount of cash and equivalents provides Microsoft with flexibility to invest in research and development, acquisitions, or other strategic initiatives.
Reduced Financial Risk: A low debt ratio can reduce the company's financial risk and improve its creditworthiness, making it easier to access financing if needed.
Liquidity: The cash ratio above 1.0 indicates that Microsoft has sufficient liquidity to meet its short-term obligations, reducing the risk of financial distress.
[1]

[1] DiscountingCashFlow.com — Google Gemini

1.11 Cash Flow Statement Analysis

1.11.1 NVDA Cash Flow Statement

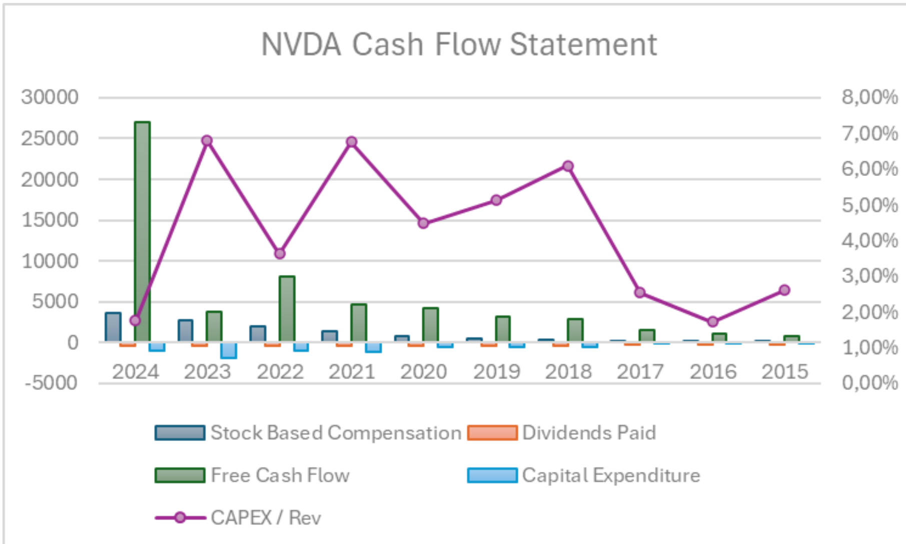


Figure 26: NVDA Cash Flow Statement

Stock-Based Compensation: Stock-Based Compensation has increased significantly over the past few years, suggesting that NVIDIA has relied heavily on equity-based compensation to attract and retain talent.

Dividends Paid: Dividends Paid have remained relatively low, indicating that NVIDIA has prioritized reinvesting its cash flow into the business.

Free Cash Flow: Free Cash Flow has fluctuated over the years but has shown a positive trend, especially in the last year.

Capital Expenditure: Capital Expenditure has also fluctuated, with periods of higher spending likely associated with investments in research and development, manufacturing facilities, or acquisitions.

CAPEX/Revenue: The ratio of Capital Expenditure to Revenue has varied over time. A higher ratio suggests that NVIDIA is investing a larger portion of its revenue in growth initiatives.

Growth Strategy: NVIDIA's focus on stock-based compensation and reinvesting cash flow suggests a growth-oriented strategy, prioritizing talent acquisition and business expansion.

Financial Health: The positive trend in Free Cash Flow indicates a healthy financial position and the ability to generate cash for future investments.

Investment Priorities: The fluctuations in Capital Expenditure highlight NVIDIA's strategic approach to investments, allocating resources to areas deemed essential for growth.
[1]

[1] DiscountingCashFlow.com — Google Gemini

1.11.2 AAPL Cash Flow Statement

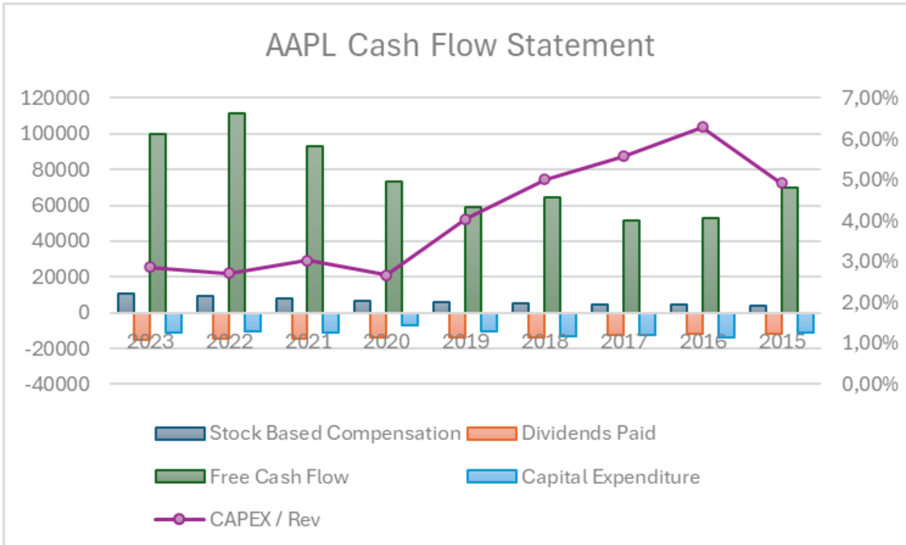


Figure 27: AAPL Cash Flow Statement

Stock-Based Compensation: Stock-Based Compensation has increased significantly over the past few years, suggesting that Apple has relied heavily on equity-based compensation to attract and retain talent.

Dividends Paid: Dividends Paid have increased steadily, indicating a growing commitment to shareholder returns.

Free Cash Flow: Free Cash Flow has fluctuated over the years but has generally shown a positive trend.

Capital Expenditure: Capital Expenditure has also fluctuated, with periods of higher spending likely associated with investments in research and development, manufacturing facilities, or acquisitions.

CAPEX/Revenue: The ratio of Capital Expenditure to Revenue has varied over time, decreasing from a 6% to less than 3% indicating Apple efficiency.

Growth Strategy: Apple's focus on stock-based compensation and increasing dividend payouts suggests a balanced approach to growth and shareholder returns.

Financial Health: The positive trend in Free Cash Flow indicates a healthy financial position and the ability to generate cash for future investments. [1]

[1] DiscountingCashFlow.com — Google Gemini

1.11.3 MSFT Cash Flow Statement

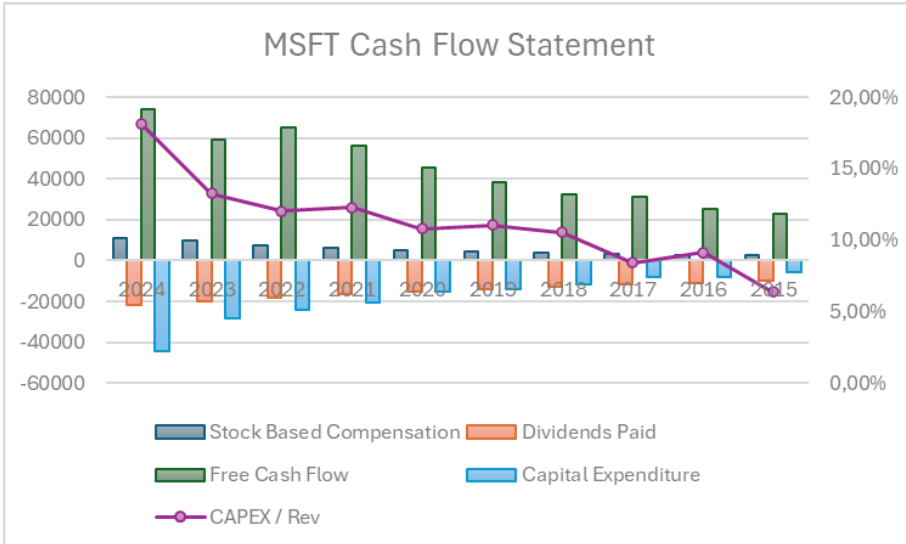


Figure 28: MSFT Cash Flow Statement

Stock-Based Compensation: Stock-Based Compensation has increased significantly over the past few years, suggesting that Microsoft has relied heavily on equity-based compensation to attract and retain talent.

Dividends Paid: Dividends Paid have increased steadily, indicating a growing commitment to shareholder returns.

Free Cash Flow: Free Cash Flow has fluctuated over the years but has shown a positive trend, indicating that Microsoft has been generating more cash from operations than it has been spending.

Capital Expenditure: Capital Expenditure has also fluctuated, but has shown a positive trend likely associated with investments in research and development, manufacturing facilities, especially cloud and AI development.

CAPEX/Revenue: The ratio of Capital Expenditure to Revenue has grown over time, suggesting that Microsoft is investing a larger portion of its revenue in growth initiatives.

Growth Strategy: Microsoft's focus on stock-based compensation and increasing dividend payouts suggests a balanced approach to growth and shareholder returns.

Financial Health: The positive trend in Free Cash Flow indicates a healthy financial position and the ability to generate cash for future investments.
[1]

[1]DiscountingCashFlow.com — Google Gemini

1.12 Capital-light businesses

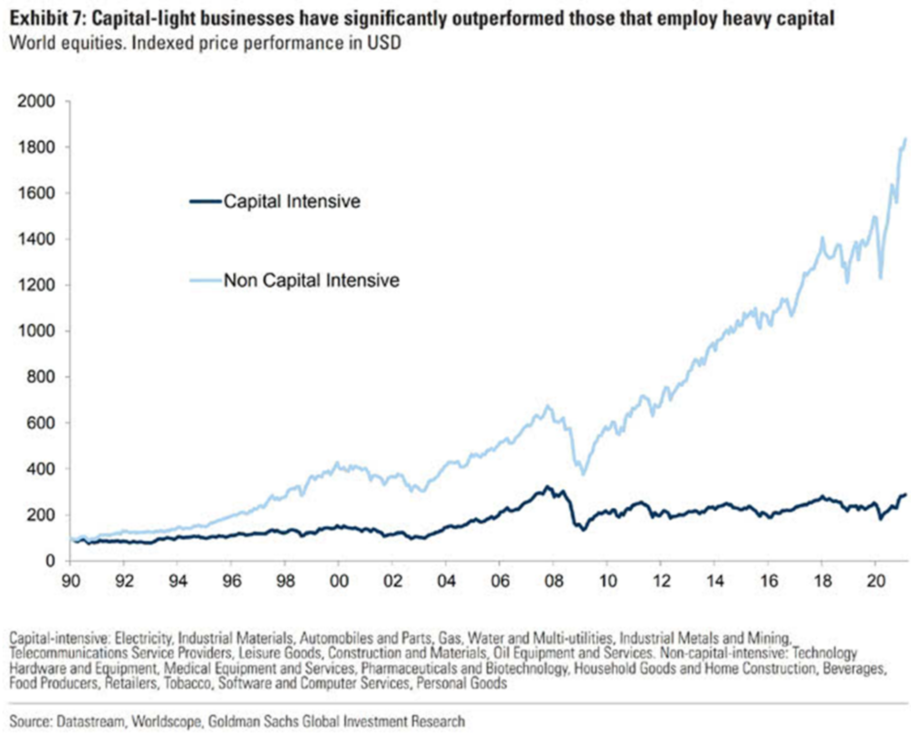


Figure 29: Capital-light vs capital-intensive businesses

There are several reasons why capital-light businesses often outperform capital-intensive businesses over time:

1. **Scalability:** Capital-light businesses are typically easier to scale. They require fewer physical assets and can expand their operations without significant upfront investments.

2. **Flexibility:** Capital-light businesses are more flexible and adaptable to changing market conditions. They can pivot their business models, introduce new products or services, or enter new markets with relative ease.

3. **Lower Fixed Costs:** Capital-light businesses have lower fixed costs compared to capital-intensive businesses. This means they are less susceptible to economic downturns and have a higher margin of safety. When revenue declines, they are better able to maintain profitability.

4. **Reduced Risk:** Capital-light businesses are often less risky than capital-intensive businesses. They are less exposed to the risks associated with asset depreciation, obsolescence, and fluctuations in commodity prices.

5. **Higher Return on Investment (ROI):** Capital-light businesses can often achieve higher returns on investment compared to capital-intensive businesses. This is because they require less capital to generate the same level of revenue.

6. **Innovation and Disruption:** Capital-light businesses are often at the forefront of innovation and disruption. They can leverage technology to develop new products and services, enter new markets, and disrupt traditional business models.

7. **Reduced Environmental Impact:** Capital-light businesses tend to have a lower environmental impact compared to capital-intensive businesses. They often require less energy, resources, and infrastructure, which can reduce their carbon footprint and improve their sustainability.

[1]

1.13 Taking into consideration the Bubble

1.14 Tech share of the US stock market

Figure 30: Tech's growing share of the US stock market

The information technology sector now accounts for 32% of the S&P 500's total market value, its highest share since 2000 when it reached nearly 35%. Microsoft, Apple, and Nvidia alone make up over 20% of the index.
[2]

[1] Goldman Sachs Global Investment Research — Google Gemini
[2] Echoes of dotcom bubble haunt AI-driven US stock market — Reuters By Lewis Krauskopf

Figure 31: Tech sector's Forward 12 month Price-to-Earnings ratio

Despite this significant presence, tech stocks are currently valued more conservatively than during the peak of the dot-com bubble. They trade at 31 times forward earnings, compared to up to 48 times in 2000, based on Datastream data.
[1]

1.14.1 S&P500 and S&P500 Equal Weight

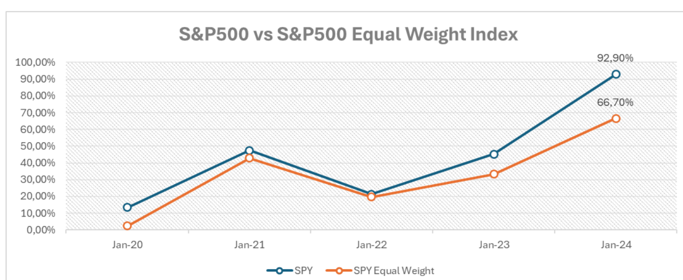


Figure 32: SP500 vs SP500 Equal Weight Index

[1]Echoes of dotcom bubble haunt AI-driven US stock market — Reuters By Lewis Krauskopf

The chart illustrates the performance comparison between the S&P 500 (represented by SPY) and the S&P 500 Equal Weight Index from January 2020 to January 2024. Over this period, the traditional S&P 500 index significantly outperformed the equal-weighted version, with a total return of 92.9% compared to 66.7% for the equal-weight index.

This performance disparity can be attributed to the traditional index's heavier weighting toward large-cap, high-quality companies, particularly in the technology sector, which has seen strong growth in recent years. In contrast, the equal-weighted index gives the same importance to smaller and potentially less stable companies, which may not have performed as well, leading to lower returns.

Investing in an equal-weighted index involves more exposure to smaller or less established firms, which can increase risk. These companies might be more vulnerable to economic downturns, market volatility, or other challenges, potentially resulting in lower returns or higher losses during periods of market stress. Thus, while diversification through equal weighting can reduce reliance on the performance of a few large firms, it may also lead to higher risk by including companies of varying quality levels.
[1]

1.14.2 Top 10 S&P500 Companies by Decade

Top 10 S&P500 companies 1990	
Company	**% of Index**
IBM	2,9
Exxon Mobil	2,9
General Electric	2,3
Philip Morris	2,2
Royal Dutch Shell	1,9
Bristol-Myers Squibb	1,6
Merck & Co	1,6
Walmart	1,6
AT&T	1,5
Coca-Cola	1,4

Top 10 S&P500 companies 2000	
Company	**% of Index**
General Electric	4,1
Exxon Mobil	2,6
Pfizer	2,5
Cisco Systems	2,4
Citigroup	2,2
Walmart	2
Microsoft	2
American International	2
Merck & Co	1,8
Intel	1,7

Top 10 S&P500 companies 2010	
Company	**% of Index**
Exxon Mobil	3,2
Apple	2,6
Microsoft	1,8
General Electric	1,7
Chevron	1,6
IBM	1,6
Procter & Gamble	1,6
AT&T	1,5
Johnson & Johnson	1,5
JPMorgan Chase	1,5

Top 10 S&P500 companies 2024	
Company	**% of Index**
Apple	7
Nvidia	6,4
Microsoft	6,4
Alphabet	6,2
Amazon	3,8
Meta	2,4
Eli Lilly	1,8
Broadcom	1,6
Tesla	1,4
JPMorgan Chase	1,2

Figure 33: Top 10 S&P500 Companies by Decade

[1]S&P 500 vs. S&P 500 Equal Weight Index — Visual Capitalist By Marcus Lu

Comparing the top companies in the S&P 500 between 1990 and 2024 reveals important differences in the quality, growth potential, and competitive advantages of these firms, reflecting broader shifts in the economy and market dynamics.

- **1990: Stability and Tangible Assets**
  The leading companies in 1990 were largely characterized by stability and the ownership of tangible assets. Firms like Exxon Mobil, Royal Dutch Shell, and General Electric had significant physical infrastructure and global operations in energy and industrial sectors. Their competitive advantage often lay in economies of scale, established brand recognition, and control over supply chains. Companies like Philip Morris and Coca-Cola had strong pricing power and brand loyalty, especially in consumer staples.

  The potential for growth among these firms was relatively modest, as many operated in mature industries with slower growth rates. Although they provided consistent cash flows and dividend payouts, innovation was slower, and the ability to scale up was often limited by physical constraints and regulatory factors. Their health was tied closely to macroeconomic cycles, commodity prices, and shifts in consumer habits.

- **2024: High Growth, Intangible Assets, and Innovation**
  The top companies in 2024, predominantly from the technology sector, are defined by their substantial growth potential and reliance on intangible assets like intellectual property, software, and data. Firms such as Apple, Microsoft, Nvidia, and Alphabet possess strong competitive advantages in the form of network effects, technological innovation, and proprietary ecosystems. Their dominance in areas like cloud computing, AI, and digital platforms has created high barriers to entry for competitors, enhancing their market positions.

  The growth potential of today's top companies is significantly higher, driven by continuous innovation and expanding digital markets. For instance, Nvidia's leadership in AI and graphics processing provides considerable opportunities in emerging fields such as autonomous driving and machine learning. Similarly, Microsoft's cloud services and Apple's integrated hardware-software ecosystem fuel recurring revenue streams and customer loyalty. The tech giants' balance sheets are typically healthier, with substantial cash reserves, low debt levels, and high-profit margins, which further support ongoing innovation and adaptability.

- **Competitive Advantage: Then vs. Now**
  In 1990, competitive advantage was often derived from physical assets, supply chain efficiency, and brand loyalty in mature industries. Today, the advantage lies in technological leadership, platform dominance, and data ownership, with companies leveraging scale and digital ecosystems to drive growth and defend their market positions. The firms in 2024 not only benefit from first-mover advantages but also possess the agility to pivot toward new markets and technologies quickly.

- **Quality and Health Comparison**
  The top companies in 2024 are arguably more dynamic and adaptable, benefiting from the ability to scale quickly with lower capital expenditure due to the nature of software and digital services. They also enjoy higher growth rates and are less constrained by physical limitations, making their long-term prospects more favorable. However, the reliance on a concentrated group of high-growth tech companies introduces more volatility, and any downturn in the tech sector can have a larger impact on the overall market.

In contrast, the 1990 companies, while less growth-oriented, provided more consistent returns and stability through economic cycles, as they were less sensitive to rapid changes in technology or market trends. The shift from tangible to intangible assets has fundamentally changed what constitutes "quality" in a leading company, favoring firms that can continuously innovate and maintain leadership in rapidly evolving industries.

The analysis highlights how shifts in the nature of assets and industry focus over the past few decades have influenced what is considered a leading company's essential characteristics. [1]

1.15 Valuation Today vs Yesterday

1.15.1 S&P500 Technology Forward PE Relative to that of the S&P500 and PEG Ratio

The PEG (Price/Earnings to Growth) ratio and the Forward P/E (Price/Earnings) ratio are both valuation metrics used to assess a company's stock, but they focus on different aspects:

- **Forward P/E Ratio:**
  - The Forward P/E ratio measures a company's current stock price relative to its estimated future earnings per share (EPS). It uses forecasted earnings for the next 12 months, providing an outlook on how the market values the company's earnings growth.
  - The formula for the Forward P/E ratio is:

$$\text{Forward P/E} = \frac{\text{Current Stock Price}}{\text{Estimated Future EPS}}$$

  - This metric gives investors an idea of whether a stock is overvalued or undervalued based on future earnings expectations. However, it does not take into account the company's growth rate, which can lead to misleading conclusions if growth expectations are high or low.
- **PEG Ratio:**
  - The PEG ratio enhances the P/E ratio by incorporating the company's expected earnings growth rate. It provides a more comprehensive view by considering how much investors are paying for each unit of earnings growth.
  - The formula for the PEG ratio is:

$$\text{PEG} = \frac{\text{Forward P/E}}{\text{Earnings Growth Rate (\%)}}$$

  - By including the growth rate, the PEG ratio helps identify whether a stock is fairly valued given its growth prospects. A PEG ratio below 1 generally suggests the stock may be undervalued, considering its growth, while a PEG ratio above 1 might indicate overvaluation.

[1] How the Top S&P 500 Companies Have Changed Over Time — Visual Capitalist By Marcus Lu

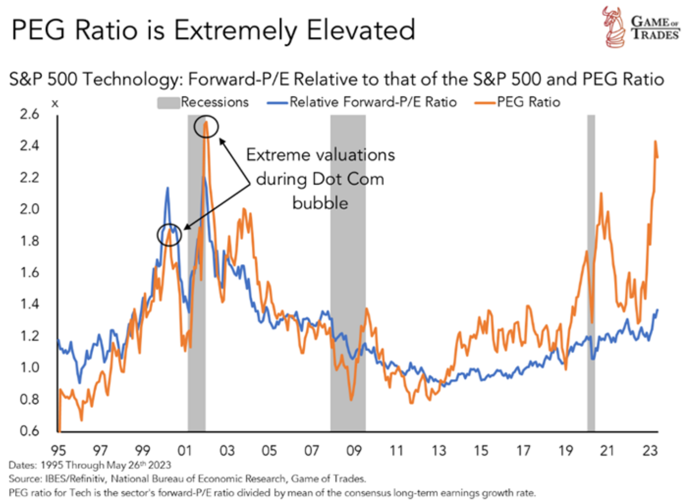


Figure 34: S&P500 Technology Forward PE Relative to that of the S&P500 and PEG Ratio

The S&P 500 Technology sector's Forward P/E (Price/Earnings) ratio relative to the overall S&P 500 provides insight into how the market values the future earnings of technology companies compared to the broader market. Here's what this relative valuation can indicate:

- **Growth Expectations**
  A higher Forward P/E ratio for the Technology sector compared to the overall S&P 500 suggests that investors anticipate higher earnings growth in technology companies relative to the rest of the market. It reflects optimism about future innovations, increased demand for tech products, or new revenue streams.

- **Perceived Risk or Volatility**
  A premium in the Technology sector's Forward P/E could also imply that investors are willing to accept higher risk for the potential of higher returns in the tech industry. The technology sector is often considered more volatile, with companies exposed to rapid changes in consumer preferences, regulation, or technological advancement. If the relative Forward P/E is closer to that of the broader S&P 500, it might indicate that the risk associated with technology companies is perceived to be in line with the overall market.

- **Market Sentiment and Sector Rotation**
  When the Technology sector's Forward P/E relative to the S&P 500 increases, it can signal strong market sentiment towards tech stocks, possibly driven by a favorable economic environment for growth stocks or recent positive news specific to the sector.

[1]

[1]The Tech Sector: A Comparative Analysis of Valuations from the Dot Com Era to Today — Bravo Research

1.16 AI Parenthesis

1.16.1 Tech Manufacturers by R&D Budget Changes 2022–2023

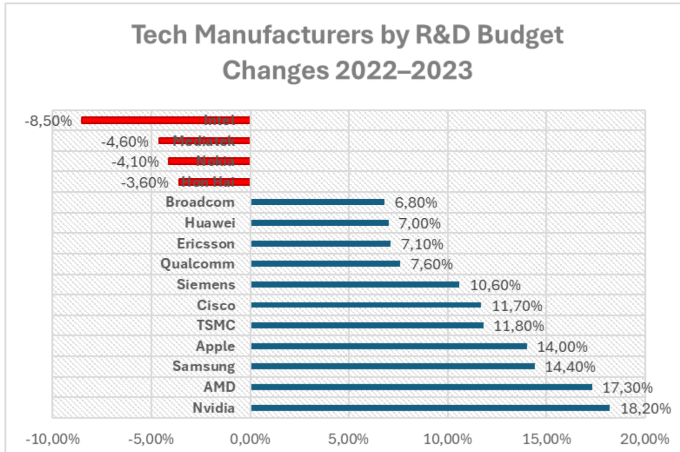


Figure 35: Tech Manufacturers by R&D Budget Changes 2022–2023

The graph showcases the growth in research and development (R&D) spending among leading hardware producers between 2022 and 2023. It highlights which companies are prioritizing innovation and expanding their technological capabilities, especially in the context of AI, semiconductors, and advanced hardware.

- **Top Performers in R&D Growth**
  Nvidia leads with an 18.2% increase in R&D spending, reflecting its commitment to maintaining leadership in AI and GPU technology. The rapid growth in AI-driven applications and the increasing demand for powerful computing hardware likely drive this investment.

  AMD follows closely at 17.3%, as it aims to keep pace with Nvidia, particularly in the AI and high-performance computing markets. The company's strategy to release new AI chips annually underscores its push to enhance its competitive position.

  Other significant R&D increases are seen in Samsung and Apple (both at approximately 14%). For Samsung, this reflects a focus on semiconductors and display technologies, while Apple's R&D growth aligns with its expansion into custom silicon and augmented reality development.

- **Implications**
  The substantial RD spending growth among top hardware producers reflects a strong emphasis on advancing AI, semiconductor technology, and next-generation computing solutions. Leaders like Nvidia, AMD, and Apple are investing aggressively to maintain their competitive edge in high-growth markets. In contrast, companies that are cutting back on RD may face challenges staying at the forefront of innovation, particularly as technological advancements continue to accelerate.

  The graph suggests a clear divide between firms prioritizing rapid technological progress and those facing restructuring or other challenges that limit their RD investments. This divide could have long-term implications for market leadership and innovation in the hardware industry.

[1]

1.16.2 Generative AI Tools

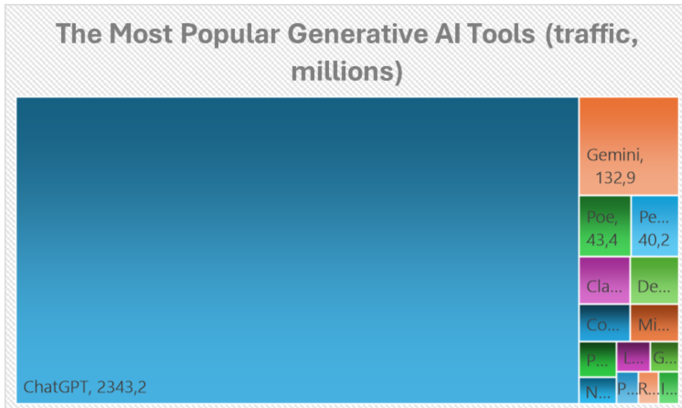


Figure 36: The Most Popular Generative AI Tools

The graph highlights the popularity of generative AI tools based on web traffic, with ChatGPT leading the field by a significant margin. In March 2024, ChatGPT recorded 2.3 billion visits, far outpacing its competitors. This dominance aligns with reports that the tool had over 200 million weekly active users by August 2024. Microsoft, which owns a 49% stake in OpenAI, benefits greatly from ChatGPT's widespread adoption, potentially reinforcing its position in the AI market through integrated products like Copilot, which also appears on the list with 26 million visits.

Gemini, owned by Google, is the second most popular tool with 133 million visits. This indicates substantial user engagement, although it remains well behind ChatGPT. The competition between ChatGPT and Gemini reflects a broader rivalry between Microsoft and Google in the AI space. Each company is leveraging its AI assets to integrate generative tools into their ecosystems, aiming to dominate both consumer and enterprise markets.

Other notable AI tools, like Claude (32M visits) and Midjourney (25M visits), also show significant traffic, indicating that while ChatGPT commands the largest share of the market, there is still room for various specialized AI tools, whether focused on language, image generation, or video.

[2]

[1]AIRanked: Tech Manufacturers by R&D Investment Change in 2023 — Visual Capitalist By Pallavi Rao

[2]Ranked: The Most Popular Generative AI Tools in 2024 — Visual Capitalist By Kayla Zhu

1.17 Timing the Bubble

Throughout history, major technological breakthroughs have spurred significant increases in investment, influencing the broader business cycle. A prime example is the era of computer and internet adoption, which necessitated widespread investment across various areas, including semiconductors, fiber-optic networks, and software development. This wave of investment was primarily concentrated in the eight years following the early '90s recession, leading up to the end of the dot-com bubble. During this time, mounting optimism about the technology fueled a period of substantial capital expenditure, resulting in the highest average annual growth in business investment over any eight-year span in the past 60 years. The surge drove a business cycle expansion that relied more heavily on investment than previous cycles and was accompanied by a significant asset-price bubble.

Looking ahead, AI has the potential to deliver economic gains comparable to—or even greater than—those of computers, suggesting we may be on the cusp of a similarly transformative investment boom. As we discuss further below, the upcoming investment ramp-up could be even more rapid and less tied to realized productivity improvements than in the '90s, due to the concentrated nature of AI-related capital expenditures among a few cash-rich companies. [1]

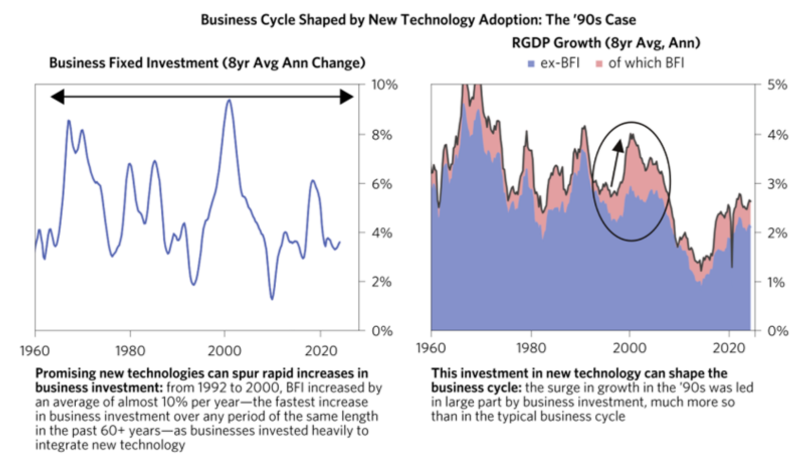


Figure 37: Business Fixed Investment Cycle and RGDP

[1] Are We on the Brink of an AI Investment Arms Race? — BridgeWater By Greg Jensen, Josh Moriarty

Corporate investment often lags behind broader economic trends, reacting to changes in demand for products. However, it can become more forward-looking when businesses or investors have strong incentives and the capacity to take speculative risks. These conditions appear to be present for AI today.

The incentives for rapid AI adoption are clear, despite uncertainty about its full economic impact. Many expect AI to create significant economic value, presenting substantial opportunities for early adopters and risks for those who fall behind. If AI boosts productivity growth by one to three percentage points over the next decade, it could add trillions of dollars to the economy.

Currently, AI's role in productivity growth is limited, with only a few applications, such as software engineering, showing significant impact. The Census Bureau's data suggests that just 5% of companies use AI regularly, and less than 0.4% report moderate automation of labor tasks. Macroeconomic data also shows little productivity growth in sectors where AI adoption is advancing most quickly. However, broad consensus on AI's potential is not necessary to drive investment, as long as some key players—mainly large corporations in tech and major investors—are willing to make significant bets on its future. [1]

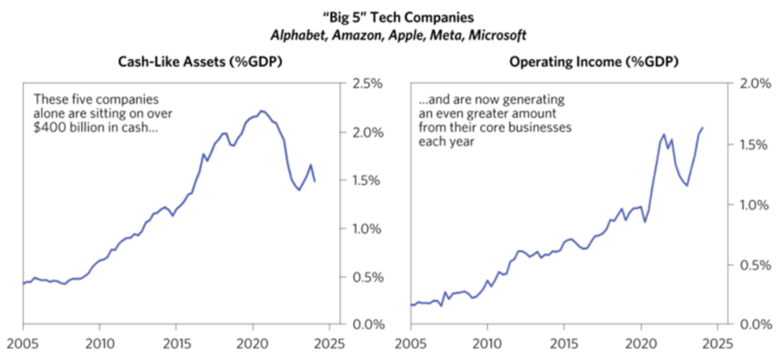


Figure 38: Cash Assets and operating Income of Big 5

These five companies alone are sitting on over $400 billion in cash and are now generating an even greater amount from their core businesses each year

The US corporate sector is broadly well-positioned to make bets on AI: American companies are sitting on secularly high levels of cash and profits are also around historic highs as a share of the economy

[1] Are We on the Brink of an AI Investment Arms Race? — BridgeWater By Greg Jensen, Josh Moriarty

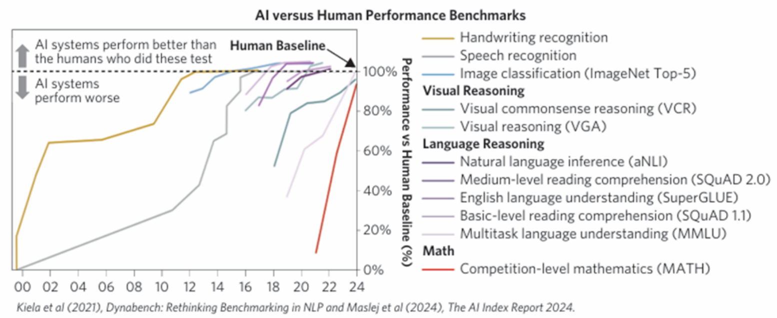


Figure 39: AI vs Human Performance Benchmarks

A significant surge in investment is likely on the horizon, with the potential for a much broader boom in AI-related spending. Up until now, AI investment has been substantial in dollar terms but concentrated among a limited group of players, mainly large tech companies and a few smaller startups. This trend is expected to continue, with spending by these key players accelerating.

The progress of AI technology has been remarkable. It took 12 years for AI to achieve human-level performance in handwriting recognition, but since then, it has demonstrated similar capabilities in increasingly complex areas. Starting with speech and image recognition, AI then advanced to visual reasoning and, more recently, language reasoning.

As the challenges have grown, the speed of AI's progress has outpaced expectations, accelerating more dramatically than a simple graph can illustrate. Today, AI systems are even reaching human-level proficiency in competitive math tasks, showcasing their rapid and ongoing evolution. [1]

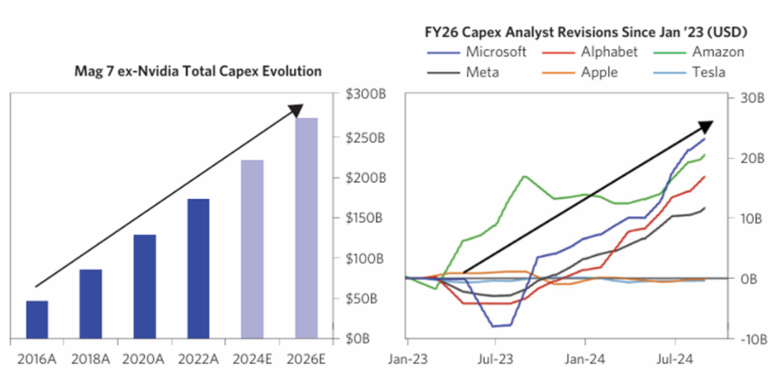


Figure 40: Capex Surge

[1]Is an AI Bubble Ahead of Us or Behind Us? — BridgeWater by GREG JENSEN ATUL NARAYAN ALEX GREENE LAUREN SIMON

We are on the verge of a major surge in capital expenditure.

So far, the stock market gains have largely favored companies central to the initial AI infrastructure buildout, with limited expectations priced in for those likely to benefit from broader AI adoption.

From a cyclical standpoint, current economic conditions appear supportive of continued growth in equity markets.

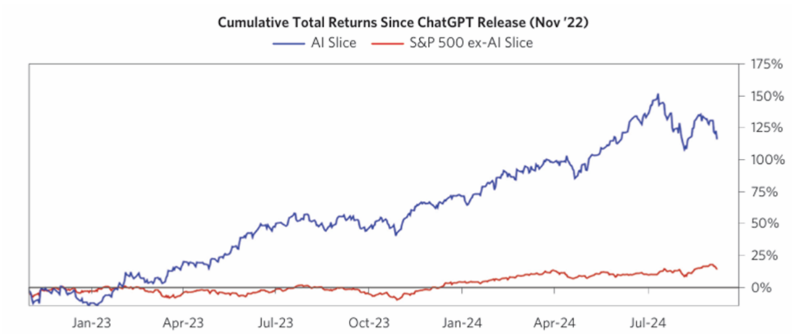


Figure 41: AI vs ex-AI Cumulative Total Returns

Markets Aren't Yet Pricing In the Widespread Potential of AI

Currently, the stock market reflects AI's impact mainly in a narrow group of companies involved in AI infrastructure, with limited expectations for a broader range of beneficiaries. This trend became evident following the release of ChatGPT in November 2022, which introduced generative AI to a wide audience and sparked increased investor interest. The concentration of the rally isn't surprising, considering the uncertainties surrounding the technology's future and the potential winners and losers. However, as the market begins to recognize and anticipate the wider impact of AI, stock pricing is likely to expand to include a broader array of companies. [1]

[1] Is an AI Bubble Ahead of Us or Behind Us? — BridgeWater by GREG JENSEN ATUL NARAYAN ALEX GREENE LAUREN SIMON

Macro Conditions Are Setting Up an Accommodative Policy Backdrop for Further AI Investment

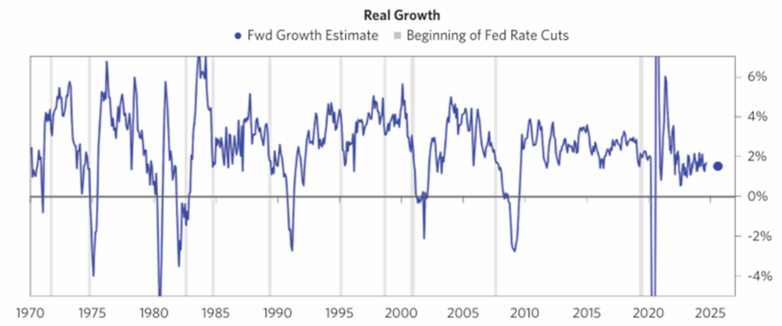


Figure 42: US Real Growth

Macro conditions are creating a supportive environment for increased AI investment. Unlike typical Fed rate-cut cycles, which often coincide with economic slowdowns that pressure equities, the current situation is different. The public sector has taken a leading role in driving an income-led economic cycle, with limited private sector credit excesses, helping maintain economic resilience even during the Fed's aggressive tightening phase. Now, as inflation eases, the Fed is positioned to start reducing rates proactively, while economic conditions remain stable. This presents an unusual opportunity for rate cuts to support equities at a time when cash flows and investor sentiment are still relatively strong.

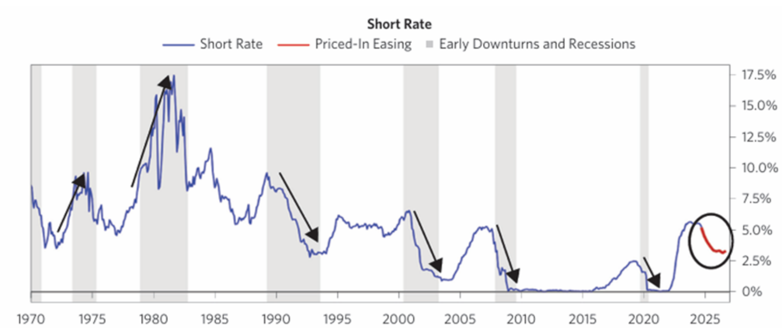


Figure 43: US Short Rate

This shift towards resilient growth is happening at a time when business balance sheets are robust, with ample cash reserves and significant borrowing capacity. Companies have plenty of resources available to invest in AI, and the prospect of lower capital costs provides an additional incentive to accelerate their investment plans. [1]

[1]Is an AI Bubble Ahead of Us or Behind Us? — BridgeWater by GREG JENSEN ATUL NARAYAN ALEX GREENE LAUREN SIMON

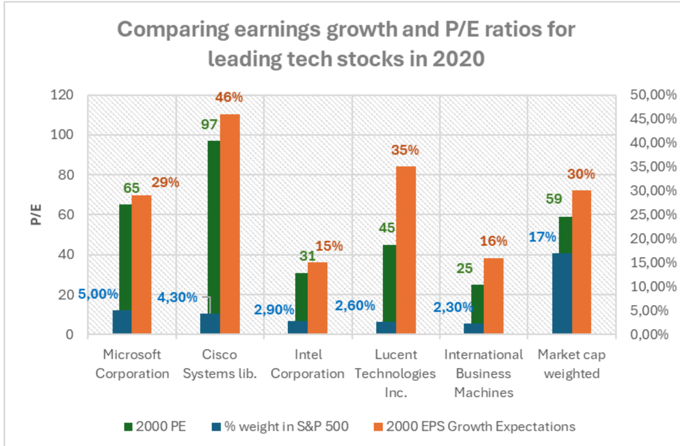


Figure 44: Comparing earnings growth and P/E ratios for leading tech stocks in 2020

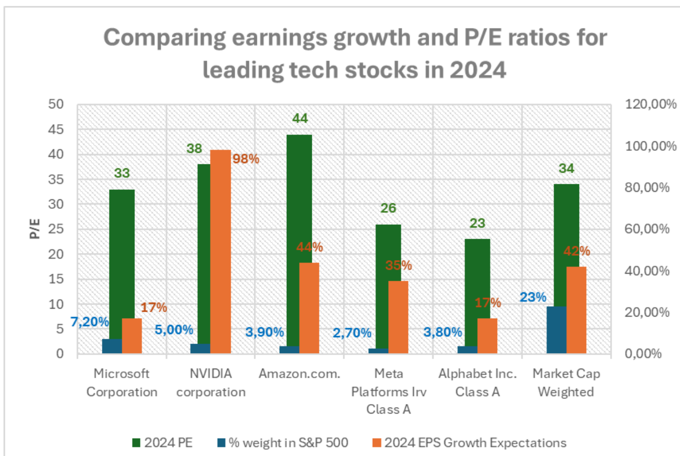


Figure 45: Comparing earnings growth and P/E ratios for leading tech stocks in 2024

Analyzing Tech Giants: A Shift in Market Dynamics

The provided table offers a fascinating comparison of leading tech companies' earnings growth expectations and P/E ratios in 2000 and 2024. Several key trends emerge:

- **Increased Market Cap Weight:**

  In 2024, the combined market cap weight of the listed tech companies is significantly higher than in 2000. This indicates a substantial growth in their market influence and dominance. This increased weight suggests that these companies play a more pivotal role in the overall market performance, influencing broader market trends.

- **Lower P/E Ratios:**

  Despite higher growth expectations in 2024, the overall P/E ratio for the market cap-weighted average is lower compared to 2000. This could be attributed to various factors, including:

  - Increased investor sophistication and demand for higher returns.
  - A more mature tech industry with established business models.
  - A shift in investor preferences towards value stocks or other asset classes.

- **Higher Growth Expectations:**

  In 2024, the growth expectations for these companies are generally higher than in 2000. This suggests that investors anticipate continued innovation, expansion, and strong earnings growth from these tech giants. This increased growth expectation, coupled with lower P/E ratios, presents an intriguing investment opportunity.

[1]

Figure 46: NASDAQ Comparison

The graph compares the performance of three groups of stocks:

- **NASDAQ 100 from January 1995 to January 2002 (blue line):**

  This line shows the dramatic rise and subsequent sharp decline, characteristic of the dot-com bubble. The index surged to nearly 1,100% growth before collapsing around 2000-2002, reflecting the boom and bust of technology stocks during that period.

[1] AI: Are we in another dot-com bubble? — substack by Kelvin Mu

- **AI Leaders (MSFT, NVDA, AMZN, META, GOOGL) from January 2019 to March 2024 (dotted line):**

  The line represents the performance of leading AI-related companies. The growth pattern here suggests a steady increase, indicating optimism around AI adoption and infrastructure investment. However, it lacks the extreme volatility seen in the dot-com era.

- **NASDAQ 100 from January 2019 to March 2024 (orange line):**

  This line shows the broader index's performance, which has generally followed the AI leaders but with less dramatic growth, indicating that the broader tech market has not benefited as narrowly or intensely from recent trends.

[1]

1.18 VGT Stocks Fundamentals and Price

1.18.1 NVDA Fundamentals and Price

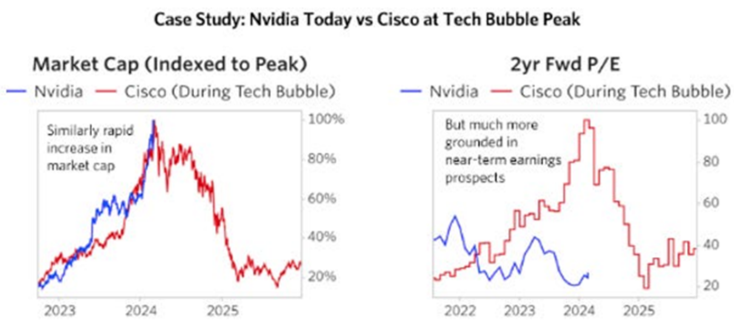


Figure 47: NASDAQ vs Cisco 1

Similary rapid increase in market cap, but much more grounded in near-term earnigs prospects [2]

[1]AI: Are we in another dot-com bubble? — substack by Kelvin Mu
[2]Research Insights — BrideWater

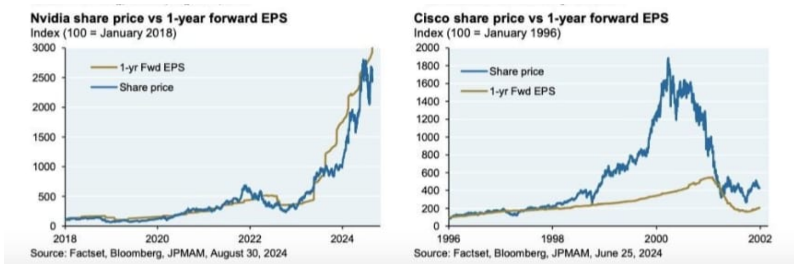


Figure 48: NASDAQ vs Cisco 2

A comparison between Nvidia share price vs 1-year forward EPS, share price and Cisco's.

- **Share Price Outperformance:**

  Both Nvidia and Cisco have experienced significant share price appreciation over the respective periods. However, Nvidia's share price has consistently outperformed Cisco's, particularly in recent years.

- **1-Year Forward EPS:**

  While both companies have seen increases in their 1-year forward EPS, Nvidia's EPS growth has been more substantial and sustained. This suggests that investors have placed a higher premium on Nvidia's future earnings potential.

- **Valuation Gap:**

  Despite Nvidia's superior earnings growth, its share price has significantly outpaced its EPS growth. This indicates that investors are willing to pay a higher valuation premium for Nvidia, possibly due to its innovative products, strong market position, and future growth prospects.

[1]

[1] Factset, Bloomberg, JPMAM, Google Gemini

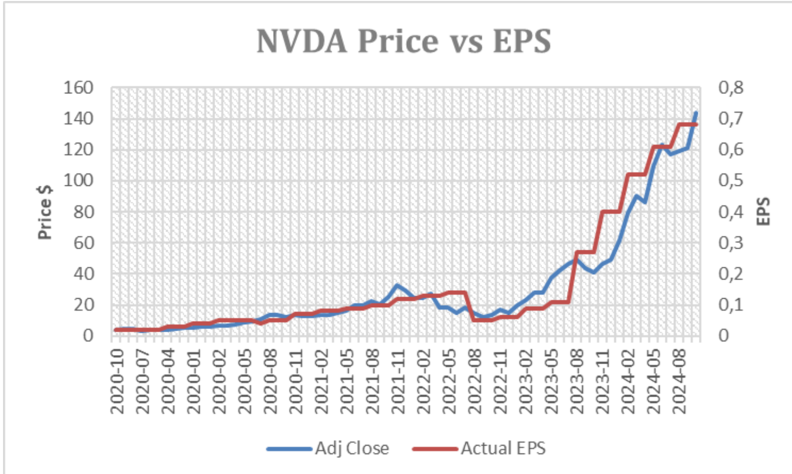


Figure 49: NVDA Price vs EPS

Earnings Growth Justifying the Stock Price: If the EPS is growing in line with the stock price, it implies that the company's profitability is increasing, justifying the higher valuation. Investors are rewarding Nvidia with a higher stock price because the company is delivering better earnings, which reflects stronger financial performance.

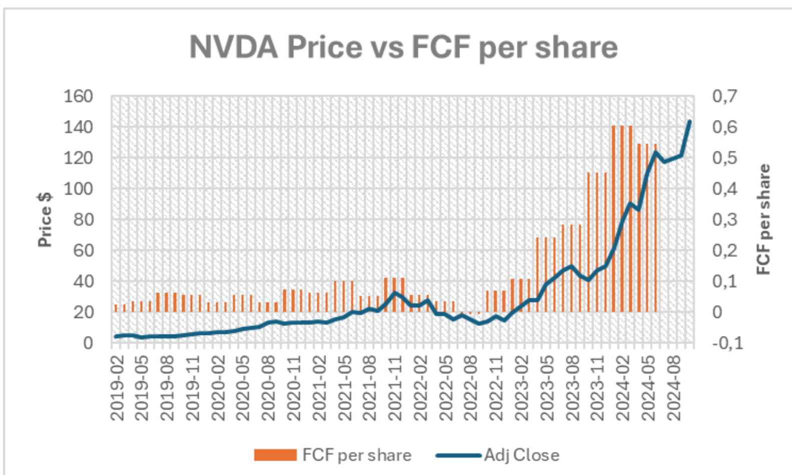


Figure 50: NVDA Price vs FCF per share

Cash Flow Growth Supporting Valuation: If FCF per share is growing along with the stock price, it indicates that Nvidia is generating more cash from its operations, which supports the higher valuation. It shows that the company's ability to generate cash is improving, justifying the rise in the stock price. [1]

[1] DiscountingCashFlow, Tickernomics, YahooFinance

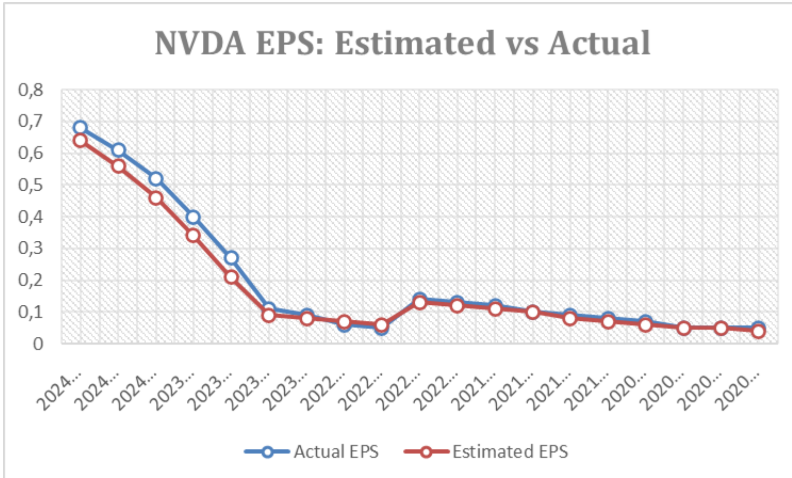


Figure 51: NVDA Actual EPS vs Estimated EPS

Positive Market Reaction: When actual EPS exceeds estimated EPS, it often leads to a positive reaction in the stock market. Investors view this as a sign of strong business performance, which can boost confidence and drive the stock price higher. The market tends to reward companies that outperform expectations because it indicates that the company is executing well, potentially generating better returns than initially anticipated.

Increased Valuation Multiples: Surpassing EPS estimates can lead to higher valuation multiples, such as the price-to-earnings (P/E) ratio. When Nvidia repeatedly beats expectations, it signals to the market that the company is experiencing strong growth, which may justify a higher valuation. Investors may be willing to pay a premium for the stock because of the demonstrated earnings momentum and the potential for future earnings surprises.

Upward Revisions in Future Estimates: When actual EPS consistently beats estimates, analysts may revise their future earnings forecasts upwards, raising expectations for the company's future performance. These upward revisions can further fuel stock price appreciation as higher growth projections become factored into the stock's valuation. [1]

[1] DiscountingCashFlow, YahooFinance

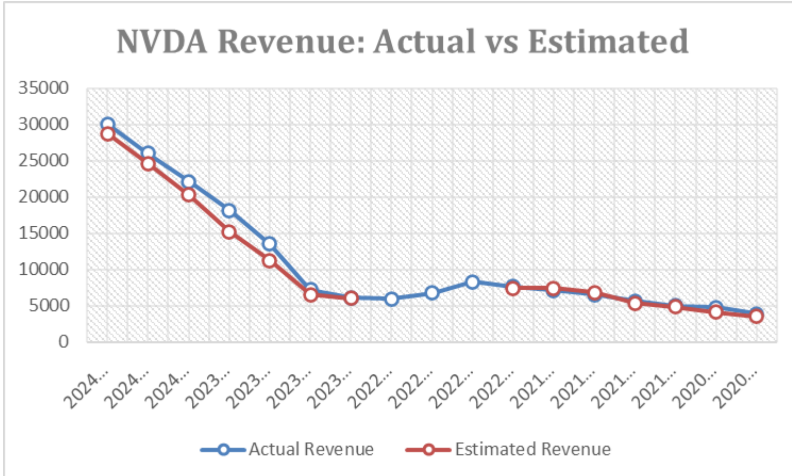


Figure 52: NVDA Actual vs Estimated Revenue

Positive Market Reaction: Beating revenue estimates often leads to a positive response from the stock market, as it indicates stronger-than-expected demand for the company's products or services. Investors typically view higher-than-expected revenue as a sign of business strength, which can lead to increased buying interest and drive the stock price higher.

Improved Growth Outlook: Higher actual revenue than estimated suggests that Nvidia's growth trajectory is stronger than anticipated. This could be due to increased sales of its products, successful expansion into new markets, or higher adoption of its technology. As a result, analysts and investors may revise their growth forecasts upward, leading to a higher valuation for the stock.

Increased Valuation Multiples: Surpassing revenue estimates can lead to higher valuation multiples, such as the price-to-sales (P/S) ratio. When a company delivers better-than-expected top-line growth, investors may be willing to pay a premium for the stock, believing that the company has a strong business model and promising future prospects. [1]

[1] DiscountingCashFlow, YahooFinance

1.18.2 AAPL Fundamentals and Price

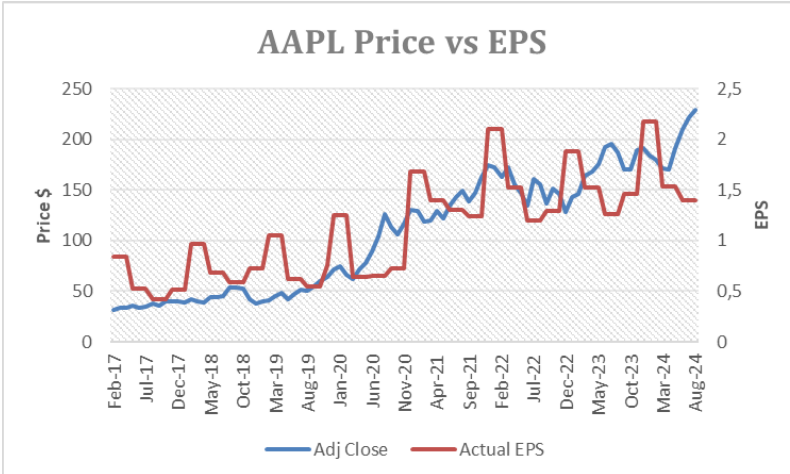


Figure 53: AAPL Price vs EPS

Positive Correlation: There's a general positive correlation between the stock price and EPS. As EPS increases, the stock price tends to rise, and vice versa.

Price Outpacing EPS: The stock price often moves ahead of EPS, suggesting that investor expectations and future growth potential are driving the price increase.

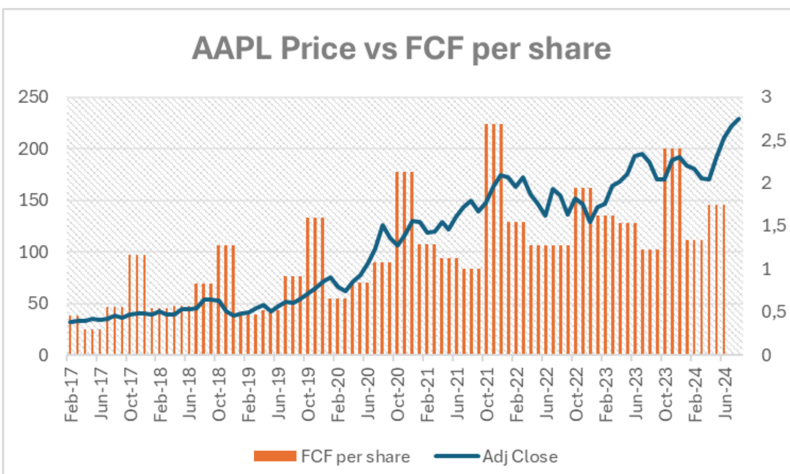


Figure 54: AAPL Price vs FCF per share

Positive Correlation: There's a general positive correlation between the stock price and FCF per share. As FCF per share increases, the stock price tends to rise, and vice versa.

Price Outpacing FCF: The stock price often moves ahead of FCF per share, suggesting that investor expectations and future growth potential are driving the price increase. [1]

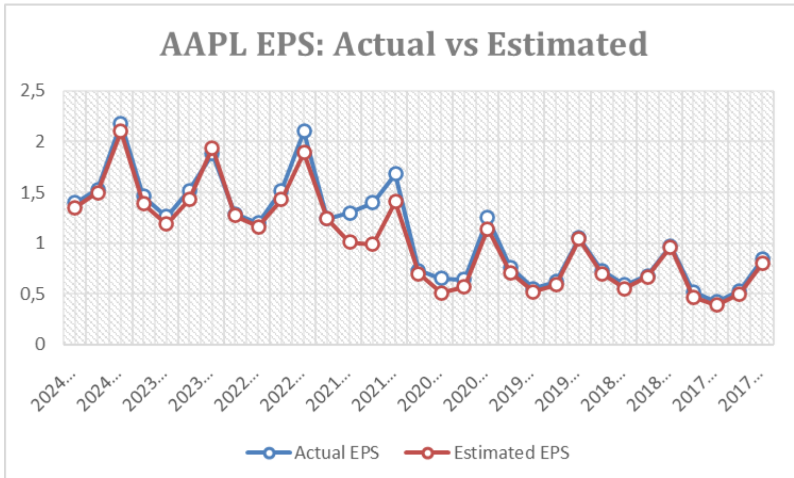


Figure 55: AAPL Actual EPS vs Estimated EPS

Apple consistently beats analyst estimates, but often by a very narrow margin. This suggests that analysts are generally accurate in their forecasts, and Apple's consistent outperformance may be priced into the stock.

Implications for Stock Price:

- **Limited Upside Surprise:**

  Since Apple consistently meets or slightly exceeds expectations, there may be limited upside potential for the stock price based solely on earnings surprises.

- **High Expectations:**

  The market's high expectations for Apple may put pressure on the company to continue delivering strong performance to justify its valuation. Any significant miss in earnings could lead to a sharp decline in the stock price.

- **Valuation Premium:**

  Apple's premium valuation may be justified by its strong brand, innovative products, and dominant market position. However, the company needs to continue delivering consistent growth to maintain this premium.

[2]

[1] DiscountingCashFlow, YahooFinance
[2] DiscountingCashFlow, YahooFinance, Google Gemini

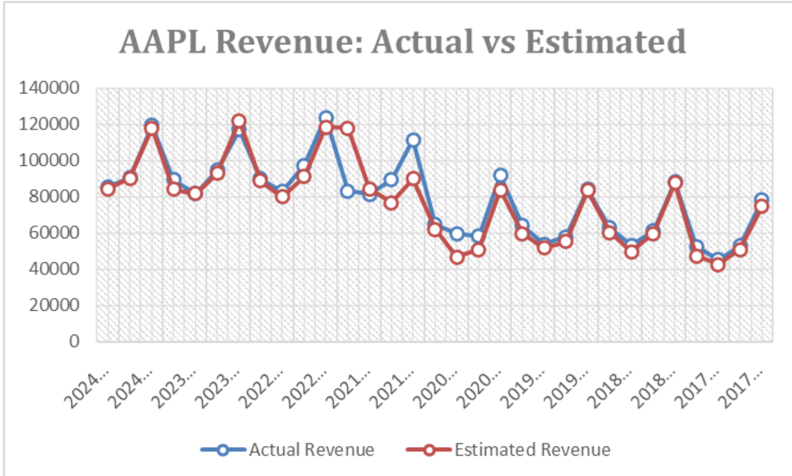


Figure 56: AAPL Actual vs Estimated Revenue

Revenue are more in line with analyst estimates. [1]

1.18.3 MSFT Fundamentals and Price

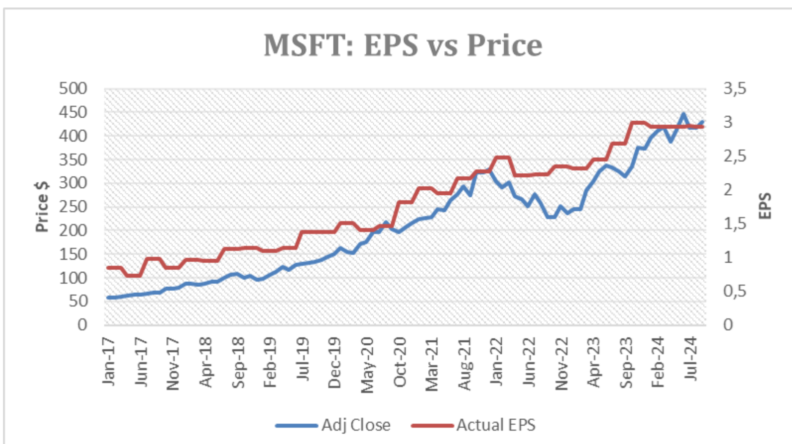


Figure 57: MSFT Price vs EPS

Similar comments can be made for MSFT as for NVDA, with a more gradual surge in both price and EPS for MSFT.[2]

[1]DiscountingCashFlow, YahooFinance
[2]DiscountingCashFlow, YahooFinance

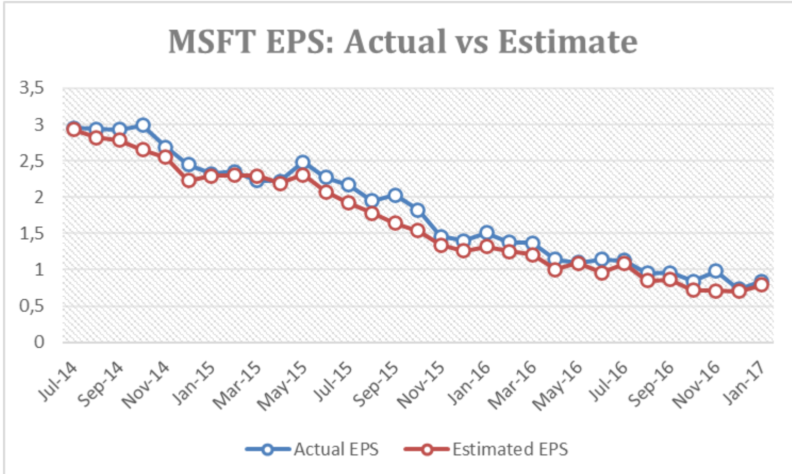


Figure 58: MSFT Actual EPS vs Estimated EPS

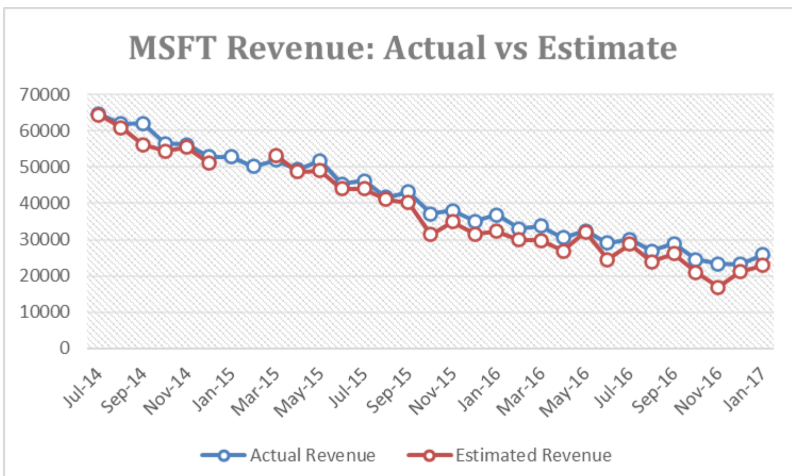


Figure 59: MSFT Actual vs Estimated Revenue

Similar comments can be made for MSFT as for NVDA about beating analysts' expectations. [1]

[1] DiscountingCashFlow, YahooFinance

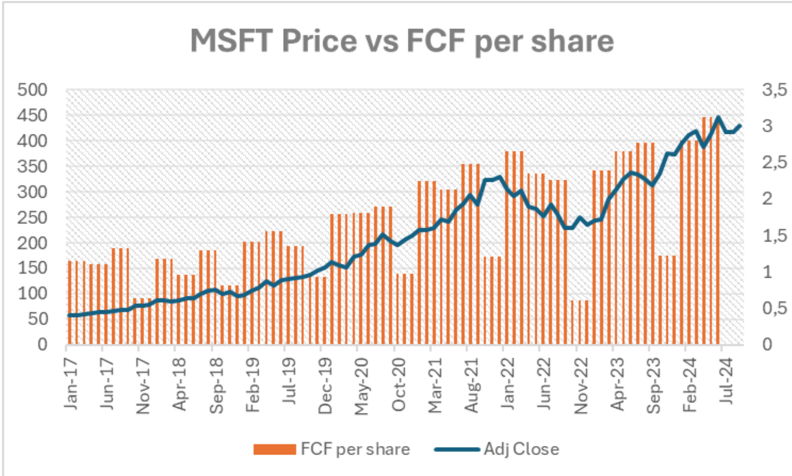


Figure 60: MSFT Price vs FCF per share

Similar comments can be made for MSFT as for NVDA, with a more gradual surge in both price and FCF per share for MSFT. [1]

1.19 US GDP by Industry

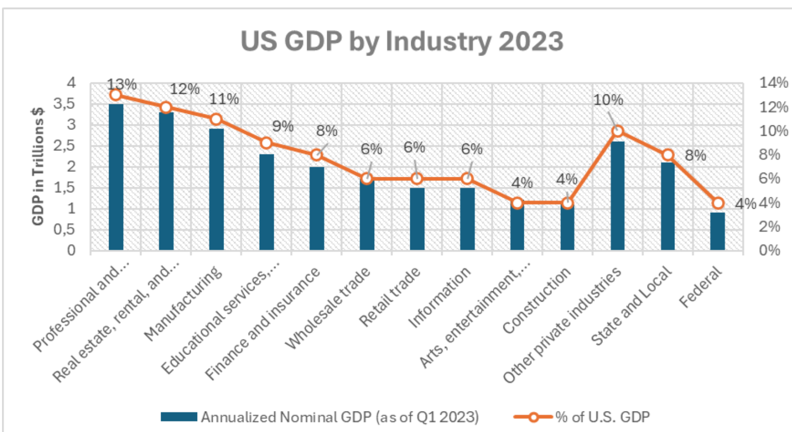


Figure 61: US GDP by Industry 2023

In the Wikipedia article "Sectors of the US Economy," data is available up to 2009. It shows that in the year 2000, the largest share of the US GDP was made up of Finance, Real Estate, and Insurance, accounting for over 20%. Business services ranked fifth, with just over 10%. Manufacturing was in second place, with slightly less than 15%, followed by Wholesale & Retail Trade in third

[1] DiscountingCashFlow, YahooFinance

and Government in fourth, each contributing around 12-13%. Today, Professional and Business Services account for 13% of the US GDP. Finance, Insurance, and Real Estate still make up 20%, Wholesale & Retail Trade 12%, Manufacturing 11%, and Government 12%. [1] [2]

Table 2: Compound Annual Rate of Output Growth by Industry Sector (2022-2032P)

Industry	**Sector**	**Growth Rate (2022-32P)**
Software publishers	Information	5.2%
Computing infrastructure, data processing	Information	3.9%
Wireless telecom carriers	Information	3.6%
Home health care services	Health care	3.6%
Oil and gas extraction	Mining	3.5%

It is projected that the segments with the highest compound annual growth rate (2022–2032) will be Software, Computing and Data Services, and Wireless Telecommunications, all within the Information Technology sector. Health Care and Mining are also expected to demonstrate strong growth. [3]

[1] Archivo:Sectors of US Economy as Percent of GDP 1947-2009 — Wikipedia
[2] US GDP by Industry 2023 — Visual Capitalist
[3] US GDP by Industry 2023 — Visual Capitalist

1.20 First Part Conclusion

In the first part of this research, we compared the VGT ETF, representing high-quality big tech stocks, with the S&P 500 benchmark. The aim was to understand whether the outperformance of high-quality tech stocks over recent decades is justified and, crucially, whether it may continue in the future.

To provide context, VGT achieved a compound annual growth rate (CAGR) of 13.21% from 2004 to 2024, compared to the S&P 500's CAGR of 7.87%, with total returns of 1187.69% versus 376.40%, respectively. Volatility and return risks will be analyzed in later sections.

One factor in VGT's superior performance can be attributed to its concentrated holdings in high-performing stocks. We focused on the top three holdings of each index as of Q2 2024. For VGT, these were AAPL (17%), MSFT (16%), and NVDA (14%), compared to the S&P 500's respective weights in AAPL (7%), MSFT (7%), and NVDA (6%). The total returns of these top holdings from 2020 to Q2 2024 were as follows: AAPL (203%), MSFT (160%), and NVDA (1621%), while the S&P 500 as a whole returned just 66%. This concentration in high-performing stocks has clearly contributed to VGT's outperformance.

In addition, VGT's sector concentration is nearly 100% in Electronic Technology and Technology Services, sectors that have significantly outpaced the S&P 500's performance. Market capitalization distribution, however, was found to have little relevance in explaining the observed outperformance.

Another important aspect of the analysis involved efficiency metrics, specifically Return on Capital Employed (ROCE), Return on Equity (ROE), Return on Assets (ROA), and especially Return on Invested Capital (ROIC). According to a study by Kennedy Capital Management, companies with high ROIC tend to have stronger competitive advantages and better management, leading to sustainable value creation through the compounding effect. Unlike traditional metrics such as the P/E or PEG ratios, which may fall short in assessing high-growth companies, ROIC offers a more reliable measure of a business's quality and profitability. High ROIC typically indicates strong products or business models, while low ROIC reflects limited growth potential due to fewer profitable reinvestment opportunities. Sustainable ROIC, combined with growth, provides a powerful basis for valuation, especially when the market underestimates a company's true ROIC potential. Comparing the ROIC of the top holdings against the S&P 500 by sector further highlights one of the reasons for VGT's outperformance.

The third factor we examined is earnings per share (EPS) and free cash flow (FCF) per share. Assuming that, in the long run, stock prices follow EPS and FCF per share, we can see the significant growth in EPS and FCF per share in VGT's top three holdings compared to the S&P 500 benchmark.

Analyzing key profitability metrics further emphasizes the difference. The S&P 500's operating margin for Q2 2023 was 11.21%, with a five-year average revenue growth of 6.8% and a net profit margin of 11.5%. In contrast, AAPL, NVDA, and MSFT demonstrated higher profitability, growth, liquidity, and free cash flow, as well as stronger debt positions, reflecting their stronger fundamentals.

In summary, these big tech stocks excel in both fundamentals and price performance.

The goal of the second segment of this analysis is to evaluate whether recent enthusiasm around big tech and artificial intelligence has created a bubble, and whether the last two decades, especially recent years, present a biased view.

The tech sector's share of the U.S. stock market is approaching levels seen in 2000, yet the economic landscape has shifted. The service sector, where big tech companies predominantly operate, now represents a substantial portion of U.S. GDP. The top 10 companies in the S&P 500 have also changed dramatically since 2000. Back then, General Electric led with a 4.1% weighting, followed by Exxon Mobil (2.6%) and Pfizer (2.5%). By 2024, the leaders are Apple (7%), Nvidia (6.4%), and Microsoft (6.4%).

The current forward P/E ratio is lower than it was in the 2000s, suggesting that today's stock price growth is more supported by earnings. In addition, big tech companies are now making substantial investments in R&D, particularly in AI, signaling a focus on future growth rather than short-term performance.

Bridgewater's research has contributed insights into understanding the current state of the AI-driven bubble in big tech stocks. High capital expenditures, underpinned by remarkable advances in AI capabilities, are occurring in companies with record amounts of cash, minimal fixed business investments, and unprecedented profitability. These factors suggest that the AI-driven growth cycle may still be in its early stages.

Furthermore, current macroeconomic conditions—with expansive monetary and fiscal policies—suggest that we could be at the start of a new growth phase.

The final section of this segment provides a visualization that compares fundamentals supporting stock prices now and in the 2000s. Our conclusion is that, unlike in the 2000 tech bubble, today's big tech companies are highly profitable, cash-generating giants. Back then, these companies were emerging players; now, they have a material impact on GDP. [1]

[1] data from 2004-02-01 to 2024-09-08 ROIC – The Underappreciated Variable in Valuation — Kennedy capital management Goldamn Sachs Global Investment Research Economatica - Value reports — valuereports.economatica.com/roic-sp-500-performance EPS and FCF per share "theory": Koyfin - Qualtrim Vanguard — Investors BridgeWater — Research & Insights This content was written with the assistance of Google Gemini and OpenAI

1.21 VGT and S&P500 Returns Analysis

1.21.1 Return Plot

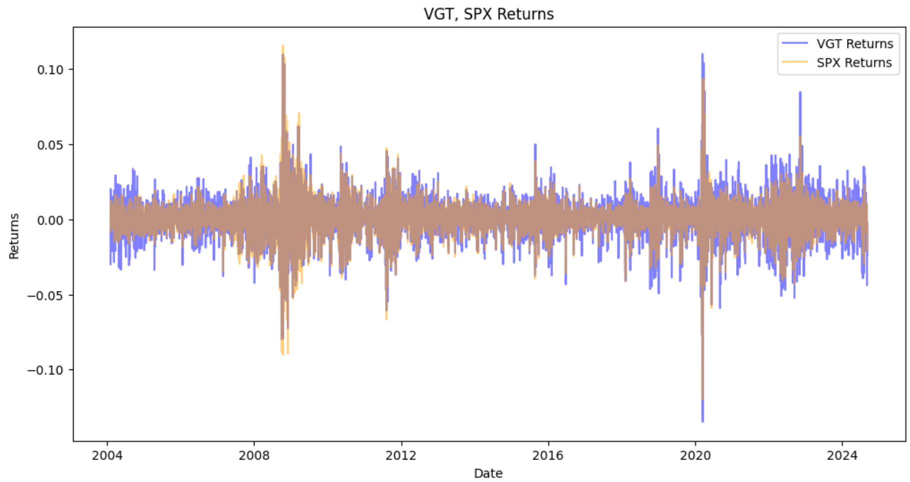


Figure 62: VGT and S&P500 Returns plot

[1]

1.21.2 Return Statistics

Table 3: VGT and SPX Statistics

Statistic	VGT (%)	SPX (%)
Mean	0.06	0.04
Standard Deviation	1.41	1.20
Skewness	-0.14	-0.26
Excess Kurtosis	7.10	12.94
Max	11.01	11.58
Min	-13.49	-11.98

- **Mean Returns:** VGT (0.06%) has a slightly higher mean daily return compared to the S&P 500 (0.04%). This suggests that over the time period you're analyzing, the VGT has on average provided slightly better daily returns compared to the broader market.
- **Volatility (Standard Deviation):** VGT's standard deviation (1.41%) is higher than SPX's (1.20%). VGT is more volatile than the S&P 500, meaning its returns fluctuate more from day to day. This is typical for a technology-focused fund, as the tech sector is known for higher volatility compared to the more diversified S&P 500.

[1] YahooFinance

- **Skewness:** Both distributions show negative skewness:
  - **VGT (-0.14):** Slightly negative skew indicates a small probability of extreme negative returns, but it is closer to zero, indicating a more balanced distribution of returns compared to SPX.
  - **SPX (-0.26):** The S&P 500 has a more pronounced negative skew, suggesting that it experienced more frequent extreme negative returns than VGT. In general, negative skewness implies that there are more extreme losses than extreme gains.
- **Excess Kurtosis:** VGT (7.10) and SPX (12.94) both exhibit high kurtosis, with SPX showing much higher kurtosis. High kurtosis means both distributions have fat tails, indicating a higher likelihood of extreme events (either gains or losses) than a normal distribution would predict. SPX's much higher kurtosis (12.94) suggests that the S&P 500 experienced more extreme daily return events compared to VGT. This is a bit surprising, as VGT is often seen as a more volatile index, but it indicates that the broad market had more frequent "shock" events in the dataset you're analyzing.

1.21.3 Sharpe Ratios

Table 4: Sharpe Ratios for VGT and SPX in Different Rate Scenarios

Scenario	**VGT Sharpe Ratio**	**SPX Sharpe Ratio**
Today Rates Scenario	0.0317	0.0190
Fed (End 2025) Scenario	0.0344	0.0222

- **Today's Interest Rate Scenario:** VGT (Sharpe Ratio: 0.0317) outperforms SPX (Sharpe Ratio: 0.0190) in today's higher interest rate environment. VGT's stronger risk-adjusted performance suggests that despite its sector-specific exposure (technology), it has been able to provide better returns relative to the volatility it carries compared to the broad market. SPX's lower Sharpe ratio reflects weaker risk-adjusted returns, which could be due to broader market volatility, earnings pressure, or sensitivity to macroeconomic factors in the higher rate environment. The relatively low Sharpe ratios indicate that both funds have faced challenges in delivering strong excess returns over the risk-free rate in a higher-rate scenario, though VGT has performed better.
- **Fed's Expected 2025 Scenario (lower rates):** Both VGT (Sharpe Ratio: 0.0344) and SPX (Sharpe Ratio: 0.0222) show improvements in their Sharpe ratios under the expected lower interest rate scenario by 2025. VGT continues to outperform SPX in this scenario, with its Sharpe ratio increasing from 0.0317 to 0.0344, indicating that VGT's tech exposure remains attractive in a lower-rate environment. The improvement suggests that as borrowing costs decrease and growth sectors like technology benefit, VGT's risk-adjusted returns become even more favorable. SPX's Sharpe ratio also improves from 0.0190 to 0.0222, but the improvement is not as significant as VGT's. This indicates that while a broad-market index like SPX benefits from lower rates, its risk-adjusted performance still lags behind VGT.

The current risk-free rate is 3.741%, based on the U.S. 10-Year Treasury yield as reported by CNBC on September 21, 2024. [1]

The Federal Reserve's target rate is projected to be 2.75% by the end of 2025, according to the FedWatch tool from the CME Group. [2]

[1] U.S. 10 Year Treasury, CNBC 21/09/2024

[2] FedWatch - CME Group

1.21.4 Positive and Negative Returns

Table 5: Return Statistics for VGT and SPX

Statistic	**VGT (%)**	**SPX (%)**
Mean Return	0.06	0.04
Mean Positive Return	0.94	0.74
Max Return	11.01	11.58
Mean Negative Return	-1.05	-0.80
Min Return	-13.49	-11.98

- **Mean Return:** Both VGT (0.06%) and SPX (0.04%) have relatively low mean returns, with VGT slightly outperforming SPX.
- **Mean Positive Return:** VGT shows a higher mean positive return (0.94%) compared to SPX (0.74%), indicating that VGT tends to perform better on days with positive returns.
- **Max Return:** SPX has a slightly higher maximum return (11.58%) than VGT (11.01%), suggesting that SPX experienced a stronger positive outlier at its peak.
- **Mean Negative Return:** VGT has a more negative mean return on its down days (-1.05%) compared to SPX (-0.80%), implying that VGT tends to fall more sharply when its returns are negative.
- **Min Return:** VGT also shows a larger negative outlier, with a minimum return of -13.49% versus SPX's -11.98%, highlighting greater downside risk for VGT.

1.22 Volatility Analysis

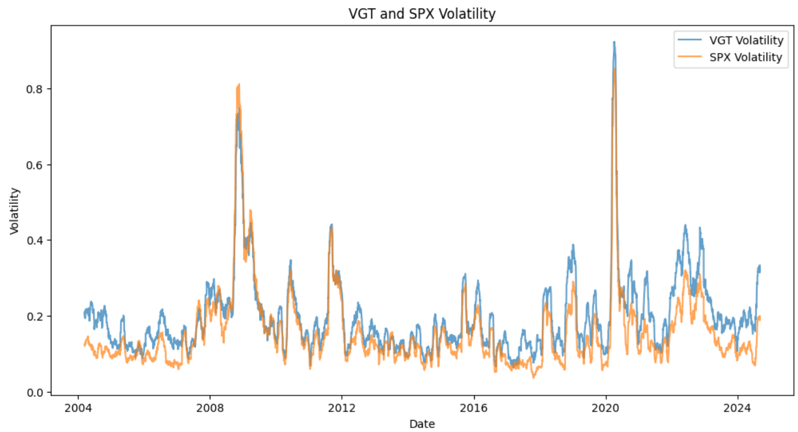


Figure 63: VGT and S&P500 Volatility plot

Computing the standard deviation of daily returns for a 30-day rolling window. This means that for each day, the standard deviation is calculated based on the returns of the previous 30 days. The 30-day volatility is annualized, assuming there are approximately 252 trading days in a year.

1.22.1 Upside and Downside Volatility

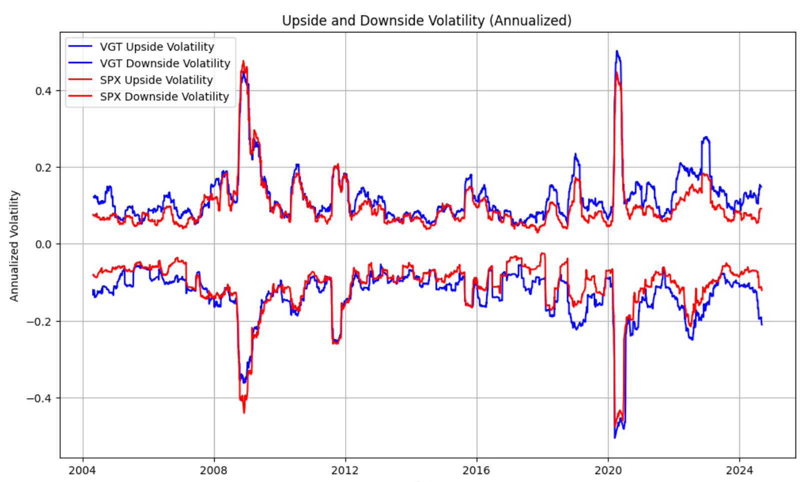


Figure 64: Upside and Downside Volatility plot

Table 6: Numerical Summary for VGT and SPX Rolling Volatilities

Volatility Type	Mean	Max	Min
VGT Upside	0.1208	0.5013	0.0462
VGT Downside	0.1368	0.5052	0.0535
SPX Upside	0.0978	0.4759	0.0288
SPX Downside	0.1134	0.4752	0.0240

- **VGT Upside vs. Downside Volatility:**
  - VGT shows slightly higher downside volatility (mean of 0.1368) compared to its upside volatility (mean of 0.1208).
  - This suggests that VGT tends to experience larger price fluctuations when moving downwards, indicating higher risk during negative market movements.
- **SPX Upside vs. Downside Volatility:**
  - Similarly, SPX also has higher downside volatility (mean of 0.1134) compared to upside volatility (mean of 0.0978).
  - However, both figures are lower than VGT's, implying that SPX is less volatile overall in both directions.
- **Maximum Volatility:**
  - Both VGT and SPX show comparable maximum volatilities.
  - However, VGT consistently has higher minimum volatilities, meaning it generally experiences more frequent and larger price swings compared to SPX.

1.23 VGT and S&P500 Drawdown

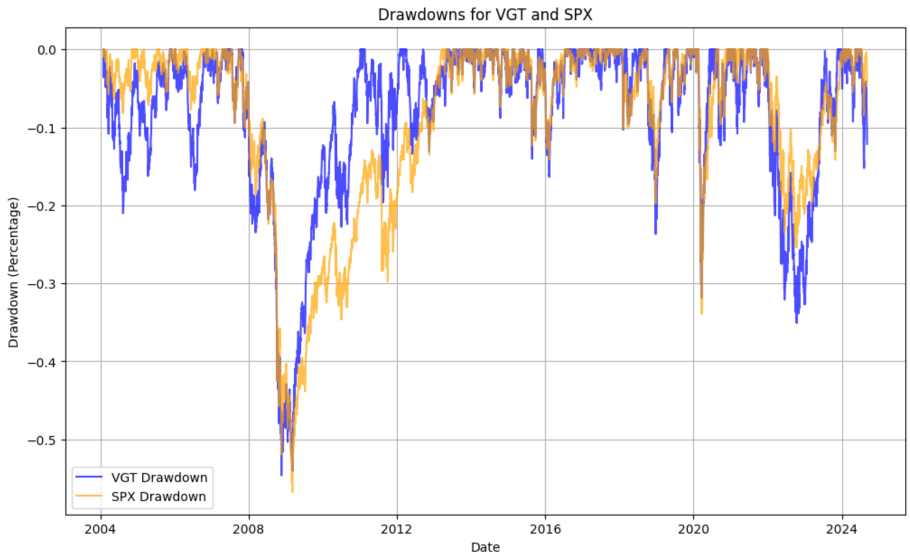


Figure 65: Drawdown plot

Table 7: Drawdown Summary for VGT and SPX

Drawdown Type	VGT	SPX
Worst Drawdown	-54.63%	-56.78%
Mean Drawdown	-8.21%	-8.76%

- **Worst Drawdown:**
  - Both VGT (-54.63%) and SPX (-56.78%) have experienced significant historical losses.
  - SPX has a slightly worse maximum drawdown, indicating a deeper historical decline during market downturns.
- **Mean Drawdown:**
  - The mean drawdown for SPX (-8.76%) is marginally higher than VGT's (-8.21%).
  - This suggests that, on average, SPX experiences slightly deeper declines compared to VGT.

1.24 VGT and S&P500 Daily Return Distribution

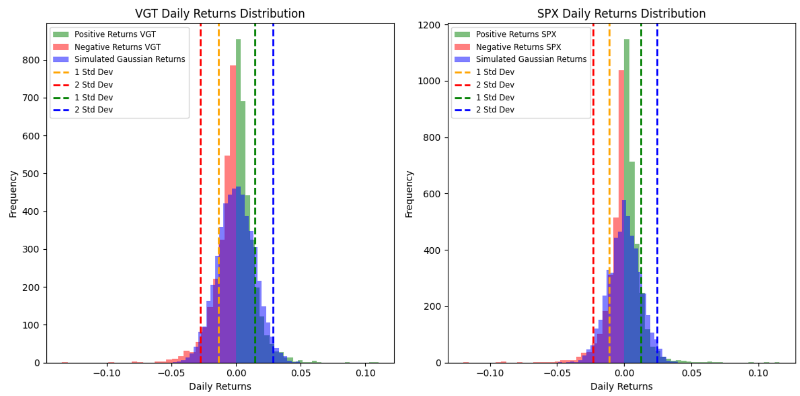


Figure 66: VGT and S&P500 Daily Return Distribution

1.25 Distribution of Financial Returns

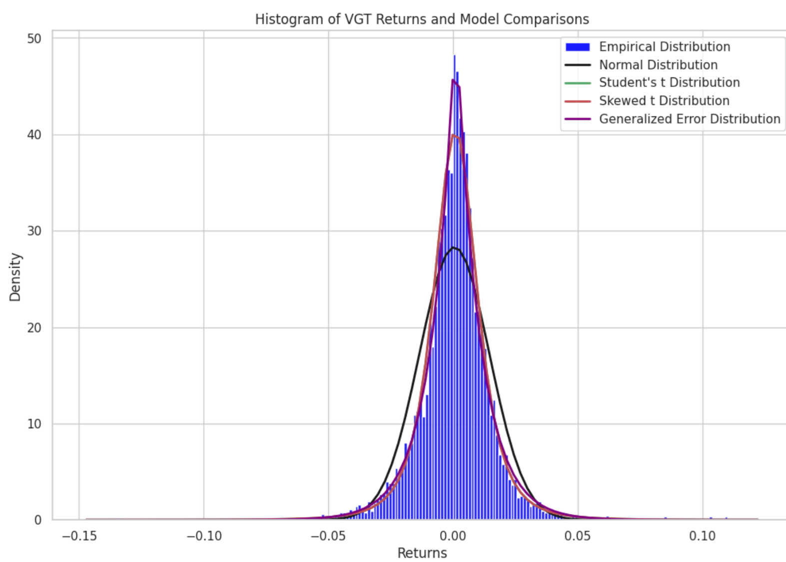


Figure 67: Histogram of VGT Returns and Model Comparisons

The aim of this analysis is to identify the most suitable statistical distribution for modeling financial returns, specifically for VGT and the S&P 500. We will compare the Normal distribution, Generalized Error Distribution (GED), Student's t-distribution, and Skewed t-distribution against the original histogram of returns.

The histogram will serve as a baseline to evaluate how well each distribution captures the characteristics of financial returns, particularly regarding fat tails and skewness. While the Normal distribution is commonly used, it may not adequately account for extreme market movements. In contrast, the GED and Student's t-distribution can better accommodate these characteristics, while the Skewed t-distribution allows for modeling asymmetry in the data.

Through this comparative analysis, we aim to enhance our understanding of return behavior and improve risk management strategies in financial markets.

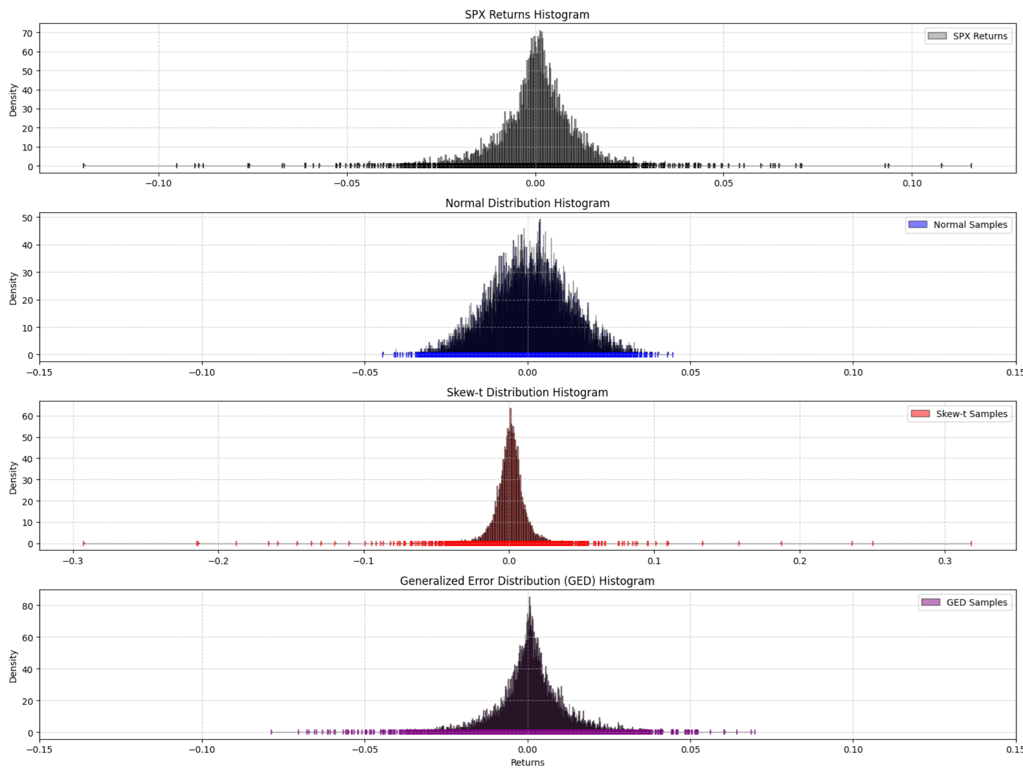


Figure 68: Fit Distributions for S&P500

Table 8: Statistical Summary for SPX Returns and Sample Distributions

Statistic	SPX Returns	Normal Samples	Skew-t Samples	GED Samples
Mean	0.0004	0.0005	0.0007	0.0010
Median	0.0007	0.0006	0.0008	0.0008
Standard Deviation	0.0120	0.0120	0.0143	0.0113
Variance	0.0001	0.0001	0.0002	0.0001
Skewness	-0.2607	0.0073	0.1077	-0.1557
Kurtosis	12.9419	-0.0451	86.4032	4.6706

1. **SPX Returns:** - The SPX returns exhibit a mean of **0.0004**, indicating a slight average positive return. The median is slightly higher at **0.0007**, suggesting that the return distribution may be right-skewed, as the mean is less than the median. - The **standard deviation** of **0.0120** signifies moderate volatility, while the **variance** of **0.0001** further confirms this level of dispersion in the returns. - The **skewness** of **-0.2607** indicates a slight leftward skew, meaning that there are more frequent small positive returns than negative ones, but larger negative returns can occur. - The **kurtosis** of **12.9419** is notably high, suggesting that the distribution has fat tails, indicating a higher likelihood of extreme returns (both positive and negative) compared to a normal distribution.

2. **Normal Samples:** - The Normal distribution has a mean of **0.0005**, which is marginally higher than the SPX returns. The median of **0.0006** also aligns closely, indicating a symmetrical distribution. - With a standard deviation of **0.0120** and variance of **0.0001**, the Normal distribution shows similar volatility levels to the SPX. - The **skewness** of **0.0073** suggests that the Normal distribution is nearly symmetric, lacking the tail behavior observed in the SPX returns. - The **kurtosis** of **-0.0451** indicates a platykurtic distribution, meaning it has lighter tails compared to the SPX. This implies that the Normal distribution does not adequately capture the risk of extreme returns that the SPX data shows.

3. **Skew-t Samples:** - The mean of **0.0007** and median of **0.0008** are both higher than those of SPX, indicating a tendency for more significant positive returns in this distribution. - The standard deviation of **0.0143** reveals increased volatility compared to SPX, suggesting more considerable fluctuations in returns. - The **skewness** of **0.1077** indicates a slight rightward skew, implying a tendency for larger positive returns relative to SPX. - The extraordinarily high kurtosis of **86.4032** reflects a distribution with extremely fat tails, suggesting that the likelihood of extreme returns is much higher than in the SPX returns, thus emphasizing greater risk.

4. **GED Samples:** - The mean of **0.0010** and median of **0.0008** indicates a higher average return compared to SPX, suggesting a potential bias toward positive outcomes. - The standard deviation of **0.0113** indicates similar volatility levels to SPX, while the variance of **0.0001** confirms this. - The **skewness** of **-0.1557** indicates a slight leftward skew, though less pronounced than that of the SPX returns, suggesting fewer extreme negative outcomes. - The **kurtosis** of **4.6706** indicates a leptokurtic distribution, suggesting heavier tails than the Normal distribution but less extreme than that of the Skew-t distribution. This highlights that GED can capture some risk of extreme events better than Normal but not as comprehensively as

the Skew-t distribution.

Conclusion: In summary, the SPX returns distribution is characterized by moderate volatility and fat tails, indicative of the risks inherent in equity markets. The Normal distribution, while simpler, fails to capture the extreme behaviors of SPX returns. In contrast, the Skew-t distribution presents the most significant potential for extreme outcomes, highlighting its applicability in risk management contexts. The GED offers a middle ground, capturing some level of extreme behavior while maintaining manageable complexity. Each distribution provides unique insights that can enhance understanding and modeling of financial returns. [1]

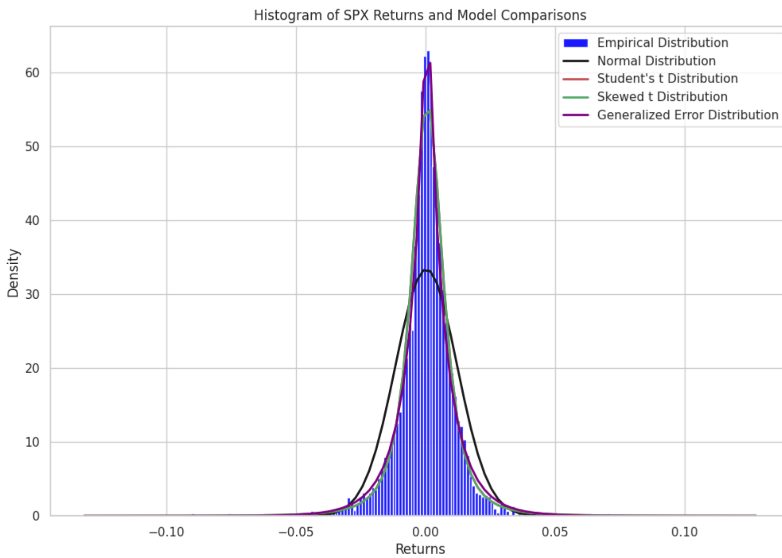


Figure 69: Histogram of S&P500 Returns and Model Comparisons

[1] written with the assistance of OpenAI

1.26 Kurtosis and Skewness

Table 9: Kurtosis and Skewness for VGT and SPX

Statistic	VGT	SPX
Kurtosis	10.10	15.94
Interpretation	Likely has heavy tails	Likely has heavy tails
Skewness	-0.14	-0.26
Interpretation	Negative skewness (left tail is longer)	Negative skewness (left tail is longer)

1.27 Normality of Returns

- **Q-Q Plot for VGT Returns**
  - The plot shows some deviation from the red reference line, particularly at the tails (both left and right). This indicates that the distribution of VGT returns has heavier tails than a normal distribution.
  - The left tail (lower quantiles) has points that fall below the reference line, suggesting more extreme negative returns than expected under normality (negative skewness).
  - Similarly, the right tail (upper quantiles) shows points above the reference line, indicating more extreme positive returns, contributing to the fat-tail behavior.
- **Q-Q Plot for SPX Returns**
  - The SPX returns plot also shows deviation from the normal line, but with more pronounced discrepancies at the tails compared to VGT.
  - The left tail exhibits larger deviations below the line, indicating more extreme negative returns, consistent with negative skewness.
  - The right tail also shows points above the line, suggesting heavy tails and a higher likelihood of extreme positive returns.

1.27.1 Jarque-Bera Test for Returns

Table 10: Jarque-Bera Test Results for VGT and SPX

Test	VGT	SPX
Jarque-Bera Statistic	10896.43	36237.44
p-value	0.0000	0.0000
Interpretation	Does not follow a normal distribution, indicating potential heavy tails	Does not follow a normal distribution, indicating potential heavy tails

1.28 ADF Test

Table 11: ADF Test Results for VGT and SPX

Test	**VGT**	**SPX**
ADF Statistic	-17.9542	-17.9214
p-value	2.83e-30	2.91e-30

The Augmented Dickey-Fuller (ADF) test is used to test for stationarity in a time series.

A p-value close to zero (much less than a common significance level of 0.05) indicates strong evidence against the null hypothesis. Both series have extremely low p-values, suggesting that we can reject the null hypothesis and conclude that both VGT and SPX are stationary.

Table 12: Ljung-Box Test Results for VGT and SPX

Test	**VGT**	**SPX**
Ljung-Box Statistic (lag=10)	101.137	131.240
p-value	3.23×10^{-17}	2.61×10^{-23}

The Ljung-Box test is used to determine whether there are significant autocorrelations in a time series at lags up to a specified number (in this case, 10 lags)

Both VGT and SPX exhibit significant autocorrelation in their returns over the tested lags

1.29 Autocorrelation Function (ACF)

Table 13: ACF of VGT and SPX Returns (First 10 Lags)

Lag	VGT ACF	SPX ACF
1	-0.0783	-0.1222
2	-0.0045	-0.0030
3	0.0072	0.0144
4	-0.0349	-0.0383
5	-0.0004	-0.0097
6	-0.0487	-0.0396
7	0.0506	0.0481
8	-0.0617	-0.0432
9	0.0569	0.0530
10	-0.0094	-0.0044
11	0.0105	-0.0111

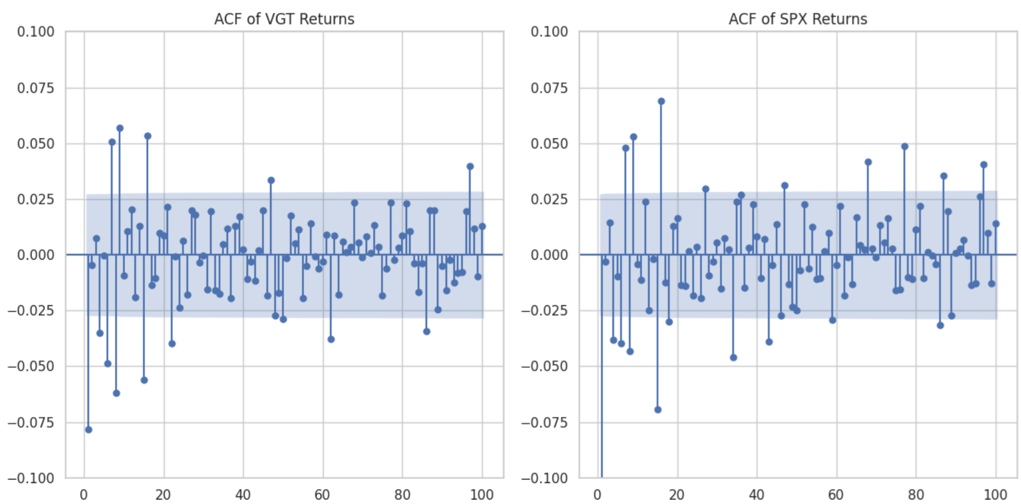


Figure 70: ACF of SPX and VGT Returns

- **VGT Returns**
  - The ACF plot for VGT returns shows a significant positive autocorrelation at lag 1, indicating that returns from the previous period have a strong positive influence on current returns.
  - There is also evidence of some autocorrelation at higher lags, suggesting a more complex pattern of dependence.

- **SPX Returns**
  - The ACF plot for SPX returns shows a similar pattern to VGT, with a significant positive autocorrelation at lag 1 and some evidence of autocorrelation at higher lags.
  - However, the autocorrelation for SPX returns appears to decay more quickly, suggesting a weaker and shorter-lived dependence.
- **Stationarity**
  - The presence of significant autocorrelation at higher lags suggests that both VGT and SPX returns may not be strictly stationary.
  - Stationarity is a key assumption in many statistical models, and non-stationarity can affect the validity of statistical tests and forecasts.
- **Mean Reversion**
  - The positive autocorrelation at lag 1 indicates a tendency for returns to revert back to their mean.
  - This suggests that if returns deviate significantly from the mean, they are likely to eventually return to the average level.

1.30 Algorithm for Identifying the Optimal ARIMA Model

The model selection process involves searching over a range of ARIMA parameters, with p and q ranging from 0 to 3. The optimal model is chosen based on the lowest Akaike Information Criterion (AIC), ensuring a stationary model that best fits the data.

Table 14: Optimal ARMA Model

Parameter	Value
Best ARMA model order	(0, 1)
Best AIC	-29502.44

For modelling purposes Returns have been rescaled, moltiplying them by 100

1.30.1 Optimal ARMA model summary

Table 15: SARIMAX Results for Optimal ARMA Model (VGT)

Parameter	Value
Dependent Variable	Adj Close
Number of Observations	5184
Model	ARIMA(0, 0, 1)
Log Likelihood	-9118.983
AIC	18243.966
BIC	18263.626
HQIC	18250.844
Covariance Type	opg
Coefficients	
Constant	0.0593
MA.L1	-0.0794
σ^2	1.9744
Standard Errors	
Constant (std err)	0.018
MA.L1 (std err)	0.008
σ^2 (std err)	0.020
Statistics	
z (Constant)	3.235
z (MA.L1)	-10.026
z (σ^2)	98.835
P-values	
P-value (Constant)	0.001
P-value (MA.L1)	0.000
P-value (σ^2)	0.000
Diagnostics	
Ljung-Box (Q)	0.00
Prob(Q)	0.98
Jarque-Bera (JB)	9233.49
Prob(JB)	0.00
Heteroskedasticity (H)	1.25
Prob(H) (two-sided)	0.00
Skew	-0.20
Kurtosis	9.53

1.30.2 Residual Analysis

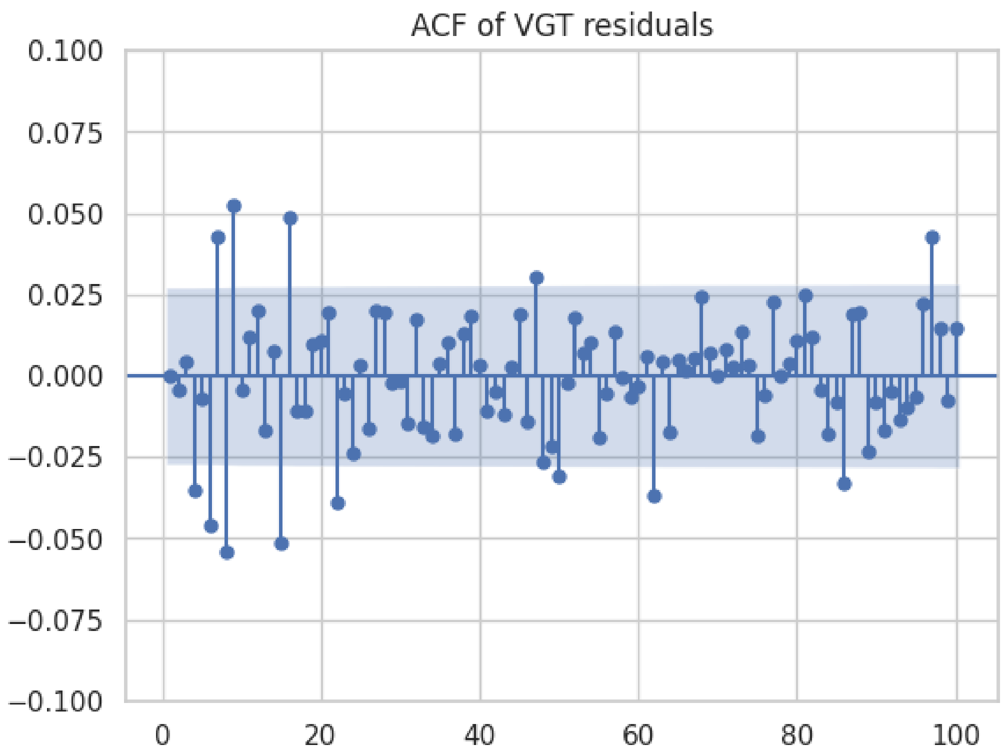


Figure 71: ACF for VGT Residuals

- **No Significant Autocorrelation:** The ACF plot shows no strong significant autocorrelation at any lag. This indicates that the residuals are mostly uncorrelated, suggesting that the model has decently captured the dependence structure in the data.
- **Randomness:** The random scatter of the ACF coefficients within the confidence bands suggests that the residuals are white noise, which is a desirable property for a well-specified model.
- **Model Adequacy:** The absence of strong significant autocorrelation in the residuals is an indication that the chosen model is adequate in capturing the underlying dynamics of the VGT returns.

1.30.3 Ljung Box Test

Table 16: Ljung-Box Statistics (lag 10)

lb_stat	lb_pvalue
56.8073	1.448738e-08

- **Null Hypothesis:** The null hypothesis of the Ljung-Box test is that there is no autocorrelation in the residuals up to a specified lag.
- **P-value:** If the p-value is less than a chosen significance level (e.g., 0.05), we reject the null hypothesis and conclude that there is evidence of autocorrelation in the residuals.

1.30.4 ARCH Test

The result of the ARCH test is given by:

p-value $= 2.456217285493036 \times 10^{-253}$

The rejection of the null hypothesis indicates that there is significant evidence of conditional heteroskedasticity in the data.

1.31 Returns Squared

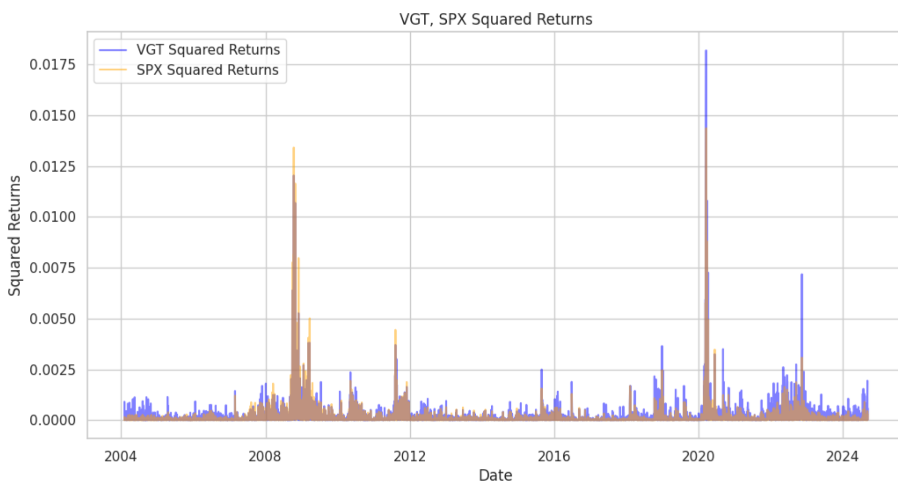


Figure 72: VGT and S&P500 Squared Residuals

Squared returns of the SPX and VGT financial series, utilized to detect ARCH behavior

1.32 Algorithm for Identifying the Optimal ARCH Model

This algorithm is designed to identify the optimal ARCH (Autoregressive Conditional Heteroskedasticity) model based on the following criteria:

1. **Input Models:** The algorithm accepts various GARCH (Generalized Autoregressive Conditional Heteroskedasticity) types, including:

   - GARCH
   - GJR-GARCH (Glosten-Jagannathan-Runkle GARCH)
   - EGARCH (Exponential GARCH)
   - HARCH (Heterogeneous ARCH)
   - APARCH (Asymmetric Power ARCH)
   - FIGARCH (Fractionally Integrated GARCH)

   Each model can have orders ranging from 1 to 4, allowing for flexibility in capturing the dynamics of the financial series.

2. **Distributions:** The algorithm allows the selection of different distributional assumptions for the residuals, including:

   - Normal distribution
   - Student's t-distribution
   - Skewed t-distribution
   - Generalized Error Distribution (GED)
   - Normal Inverse Gaussian distribution (NIG)

   This provides a comprehensive framework for modeling the underlying data characteristics.

3. **Selection Method:** The optimal model is determined using the Akaike Information Criterion (AIC), a widely-used metric for model selection that balances model fit and complexity. The model with the lowest AIC value is selected as the optimal ARCH model, ensuring an effective representation of the data while avoiding overfitting.

The algorithm identified the following optimal models based on the provided criteria:

1. **Initial Model Selection:** The first run of the algorithm selected the **APARCH(1,3)** model with a **skewed t-distribution**. This model was deemed optimal for capturing the conditional heteroskedasticity present in the data.

2. **Subsequent Analysis:** Upon rerunning the algorithm, the **FIGARCH(1,1)** model with a **skewed t-distribution** was selected. This indicates a shift in the model that best fits the data after further analysis.

3. **Pre-Rescaling Model Findings:** Before rescaling the returns of the VGT series, the algorithm initially suggested the **EGARCH(1,1)** and **EGARCH(2,2)** models, both employing GED distributions. These results highlight the potential variations in model selection based on the data scaling process.

4. **Comprehensive Evaluation:** To ensure a thorough understanding and validation of the results, we will analyze all four models: **APARCH(1,3)**, **FIGARCH(1,1)**, **EGARCH(1,1)**, and **EGARCH(2,2)**. This comprehensive evaluation will help in confirming the robustness of the model selections and their respective performance in capturing the dynamics of the financial series.

1.32.1 APARCH(1,3)-skewt dist

Dep. Variable:	Adj Close
R-squared:	0.000
Mean Model:	Constant Mean
Adj. R-squared:	0.000
Vol Model:	Power ARCH
Log-Likelihood:	-8124.01
Distribution:	Standardized Skew Student's t
AIC:	16266.0
Method:	Maximum Likelihood
BIC:	16325.0
No. Observations:	5184
Df Residuals:	5183
Df Model:	1

	coef	**std err**	**t**	**P¿—t—**	**95.0% Conf. Int.**
mu	0.0946	1.472e-02	6.427	1.302e-10	[6.577e-02, 0.123]

	coef	**std err**	**t**	**P¿—t—**	**95.0% Conf. Int.**
omega	0.0237	5.603e-03	4.237	2.267e-05	[1.276e-02, 3.472e-02]
alpha[1]	0.1083	1.476e-02	7.336	2.197e-13	[7.933e-02, 0.137]
beta[1]	0.8906	0.230	3.879	1.050e-04	[0.441, 1.341]
beta[2]	1.6231e-09	0.361	4.500e-09	1.000	[-0.707, 0.707]
beta[3]	1.6225e-09	0.169	9.592e-09	1.000	[-0.332, 0.332]
delta	1.7043	0.242	7.050	1.795e-12	[1.230, 2.178]

	coef	**std err**	**t**	**P¿—t—**	**95.0% Conf. Int.**
eta	7.6664	0.770	9.955	2.392e-23	[6.157, 9.176]
lambda	-0.1174	1.736e-02	-6.761	1.374e-11	[-0.151, -8.335e-02]

The coefficients β_2 and β_3 are not statistically significant.

- ω: This is the unconditional variance parameter. It represents the long-run average variance of the series.
- α_1: This is the ARCH parameter. It captures the impact of the lagged squared residual on the current conditional variance. A higher α_1 indicates a stronger effect of past shocks on current volatility.
- β_1, β_2, β_3: These are the GARCH parameters. They capture the impact of past conditional variances on the current conditional variance. Higher β values indicate a stronger persistence of volatility.
- δ: This is the asymmetry parameter in the APARCH model. It captures the asymmetric effect of positive and negative shocks on volatility. A positive δ indicates that negative shocks have a greater impact on volatility than positive shocks.
- η: This is the shape parameter of the standardized skew Student's t distribution. It controls the tail thickness of the distribution. A higher η indicates heavier tails, which can be important for modeling extreme events.
- λ: This is the skewness parameter of the standardized skew Student's t distribution. It controls the asymmetry of the distribution. A negative λ indicates a left-skewed distribution (more negative outliers).

1.32.2 FIGARCH(1,1)-skewt dist

Dep. Variable:	Adj Close
R-squared:	0.000
Mean Model:	Constant Mean
Adj. R-squared:	0.000
Vol Model:	FIGARCH
Log-Likelihood:	-8117.73
Distribution:	Standardized Skew Student's t
AIC:	16249.5
Method:	Maximum Likelihood
BIC:	16295.3
No. Observations:	5184
Df Residuals:	5183
Df Model:	1

	coef	**std err**	**t**	**P¿—t—**	**95.0% Conf. Int.**
mu	0.0920	1.472e-02	6.251	4.085e-10	[6.315e-02, 0.121]

	coef	**std err**	**t**	**P¿—t—**	**95.0% Conf. Int.**
omega	0.0403	1.191e-02	3.385	7.124e-04	[1.697e-02, 6.367e-02]
phi	0.1248	4.029e-02	3.097	1.956e-03	[4.581e-02, 0.204]
d	0.5591	7.941e-02	7.041	1.914e-12	[0.403, 0.715]
beta	0.6106	7.878e-02	7.750	9.163e-15	[0.456, 0.765]

	coef	**std err**	**t**	**P¿—t—**	**95.0% Conf. Int.**
eta	7.4741	0.759	9.844	7.253e-23	[5.986, 8.962]
lambda	-0.1202	1.751e-02	-6.864	6.718e-12	[-0.154, -8.584e-02]

- ω: This is the unconditional variance parameter. It represents the long-run average variance of the series.
- ϕ: This is the fractional integration parameter. It captures the degree of long memory in the conditional variance. A value of d between 0 and 0.5 indicates long memory, while a value of 0 indicates short memory.
- d: This is the fractional integration parameter, as explained above.
- β: This is the GARCH parameter. It captures the impact of past conditional variances on the current conditional variance. A higher β indicates a stronger persistence of volatility.
- η: This is the shape parameter of the standardized skew Student's t distribution. It controls the tail thickness of the distribution. A higher η indicates heavier tails, which can be important for modeling extreme events.

- λ: This is the skewness parameter of the standardized skew Student's t distribution. It controls the asymmetry of the distribution. A negative λ indicates a left-skewed distribution (more negative outliers).

1.32.3 EGARCH(1,1)-ged dist

Table 17: Constant Mean - EGARCH Model Results

Dep. Variable:	Adj Close
R-squared:	0.000
Mean Model:	Constant Mean
Adj. R-squared:	0.000
Vol Model:	EGARCH
Log-Likelihood:	-8145.51
Distribution:	Generalized Error Distribution
AIC:	16301.0
Method:	Maximum Likelihood
BIC:	16333.8
No. Observations:	5184
Df Residuals:	5183
Df Model:	1

	coef	**std err**	**t**	**P¿—t—**	**95.0% Conf. Int.**
Mean Model					
mu	0.1174	1.379e-02	8.514	1.673e-17	[9.037e-02, 0.144]
Volatility Model					
omega	0.0164	3.532e-03	4.646	3.391e-06	[9.486e-03, 2.333e-02]
alpha[1]	0.2147	1.914e-02	11.214	3.468e-29	[0.177, 0.252]
beta[1]	0.9783	4.402e-03	222.209	0.000	[0.970, 0.987]
Distribution					
nu	1.3731	4.116e-02	33.358	5.535e-244	[1.292, 1.454]

- μ**:** This coefficient represents the constant mean in the model. It is the expected value of the dependent variable (Adj Close) in the absence of any other factors.
- ω**:** This is the unconditional variance parameter. It represents the long-run average variance of the series.
- α_1**:** This is the ARCH parameter. It captures the impact of the lagged squared residual on the current conditional variance. A higher α_1 indicates a stronger effect of past shocks on current volatility.
- β_1**:** This is the GARCH parameter. It captures the impact of past conditional variances on the current conditional variance. A higher β_1 indicates a stronger persistence of volatility.

- ν**:** This is the shape parameter of the Generalized Error Distribution (GED) used to model the innovations. A higher ν indicates heavier tails in the distribution.

1.32.4 EGARCH(2,2)-ged dist

Table 18: Constant Mean - EGARCH Model Results

Dep. Variable:	Adj Close
R-squared:	0.000
Mean Model:	Constant Mean
Adj. R-squared:	0.000
Vol Model:	EGARCH
Log-Likelihood:	-8143.25
Distribution:	Generalized Error Distribution
AIC:	16300.5
Method:	Maximum Likelihood
BIC:	16346.4
No. Observations:	5184
Date:	Fri, Oct 11 2024
Df Residuals:	5183
Time:	17:47:07
Df Model:	1

	coef	**std err**	**t**	**P¿—t—**	**95.0% Conf. Int.**
Mean Model					
μ	0.1169	1.378e-02	8.483	2.193e-17	[8.992e-02, 0.144]
Volatility Model					
ω	0.0316	6.901e-03	4.579	4.667e-06	[1.808e-02, 4.513e-02]
α_1	0.1877	2.412e-02	7.782	7.133e-15	[0.140, 0.235]
α_2	0.2260	2.488e-02	9.081	1.073e-19	[0.177, 0.275]
β_1	0.0734	6.483e-02	1.132	0.258	[-5.367e-02, 0.200]
β_2	0.8842	6.398e-02	13.820	1.930e-43	[0.759, 1.010]
Distribution					
ν	1.3772	4.109e-02	33.512	3.183e-246	[1.297, 1.458]

- ω**:** This is the unconditional variance parameter. It represents the long-run average variance of the series.
- α_1, α_2**:** These are the ARCH parameters. They capture the impact of the lagged squared residuals on the current conditional variance. Higher α values indicate a stronger effect of past shocks on current volatility.
- β_1, β_2**:** These are the GARCH parameters. They capture the impact of past conditional variances on the current conditional variance. Higher β values indicate a stronger persistence

of volatility.

- **ν:** This is the shape parameter of the Generalized Error Distribution (GED) used to model the innovations. A higher ν indicates heavier tails in the distribution.

1.32.5 Best Model (AIC)

Table 19: AIC Comparison of VGT Models

Model	**AIC**
APARCH(2,3)-skewt	16262.0170
FIGARCH(1,1)-skewt	16249.4542
EGARCH(1,1)-ged	16301.0170
EGARCH(2,2)-ged	16300.5098

Best Model: FIGARCH(1,1)-skewt with AIC: 16249.4542

1.32.6 ARCH-LM Test

Table 20: ARCH-LM Test Results for VGT Models

Model	**p-value**	**Statistic**
APARCH(1,1)-skewt	0.0000	1359.8133
FIGARCH(1,1)-skewt	0.0000	1359.5451
EGARCH(1,1)-ged	0.0000	1361.8750
EGARCH(2,2)-ged	0.0000	1361.8381

Models' residuals are conditionally heteroskedastic

1.32.7 APARCH(1,1)-skewt dist

Table 21: Constant Mean - Power ARCH Model Results

Dep. Variable:	Adj Close
R-squared:	0.000
Mean Model:	Constant Mean
Adj. R-squared:	0.000
Vol Model:	Power ARCH
Log-Likelihood:	-8124.01
Distribution:	Standardized Skew Student's t
AIC:	16262.0
Method:	Maximum Likelihood
BIC:	16307.9
No. Observations:	5184
Df Residuals:	5183
Df Model:	1

	coef	**std err**	**t**	**P¿—t—**	**95.0% Conf. Int.**
Mean Model					
μ	0.0946	1.466e-02	6.455	1.086e-10	[6.589e-02, 0.123]
Volatility Model					
ω	0.0237	5.454e-03	4.352	1.346e-05	[1.305e-02, 3.443e-02]
α_1	0.1083	1.190e-02	9.101	8.958e-20	[8.494e-02, 0.132]
β_1	0.8906	1.082e-02	82.280	0.000	[0.869, 0.912]
δ	1.7041	0.235	7.248	4.234e-13	[1.243, 2.165]
Distribution					
η	7.6657	0.768	9.982	1.836e-23	[6.161, 9.171]
λ	-0.1174	1.735e-02	-6.763	1.350e-11	[-0.151, -8.336e-02]

The APARCH model has been re-specified with orders $(1,1)$ due to the low significance of β_2 and β_3.

1.32.8 GARCH Models Conditionally Volatility

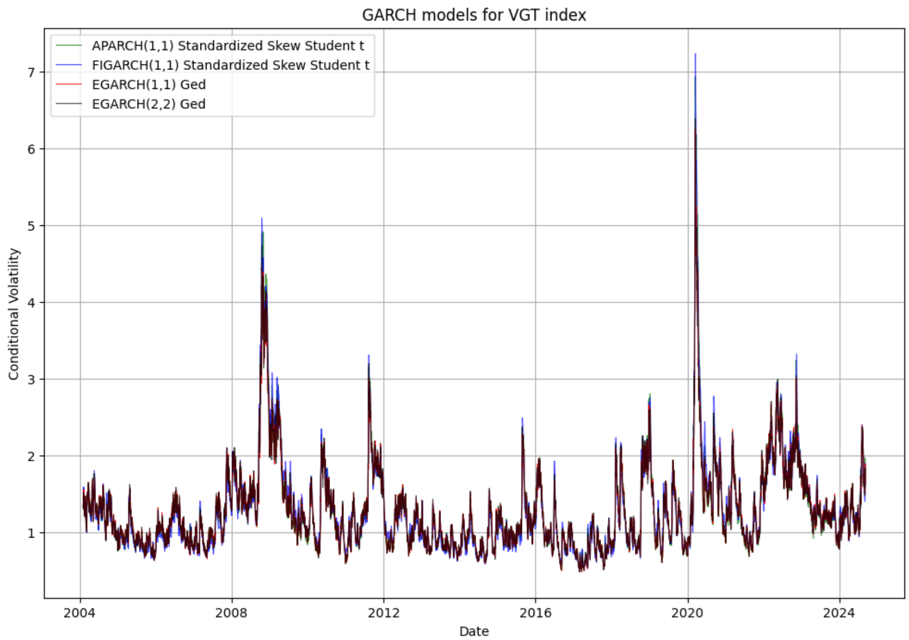


Figure 73: GARCH Models Conditionally Volatility

[1]

[1]Code developed in Google Colab and reviewed with OpenAI

1.33 GARCH Residuals Analysis

1.33.1 GARCH Models Standardized Residuals plot

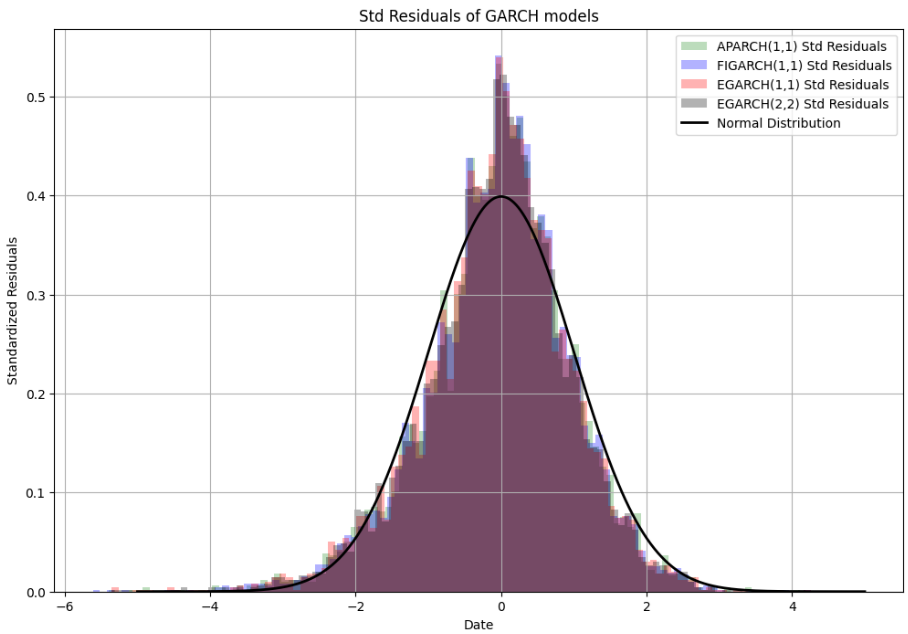


Figure 74: GARCH Models Standardized Residuals

In GARCH models, the residuals are standardized by dividing them by the conditional standard deviation [1]

[1] Code developed in Google Colab and reviewed with OpenAI

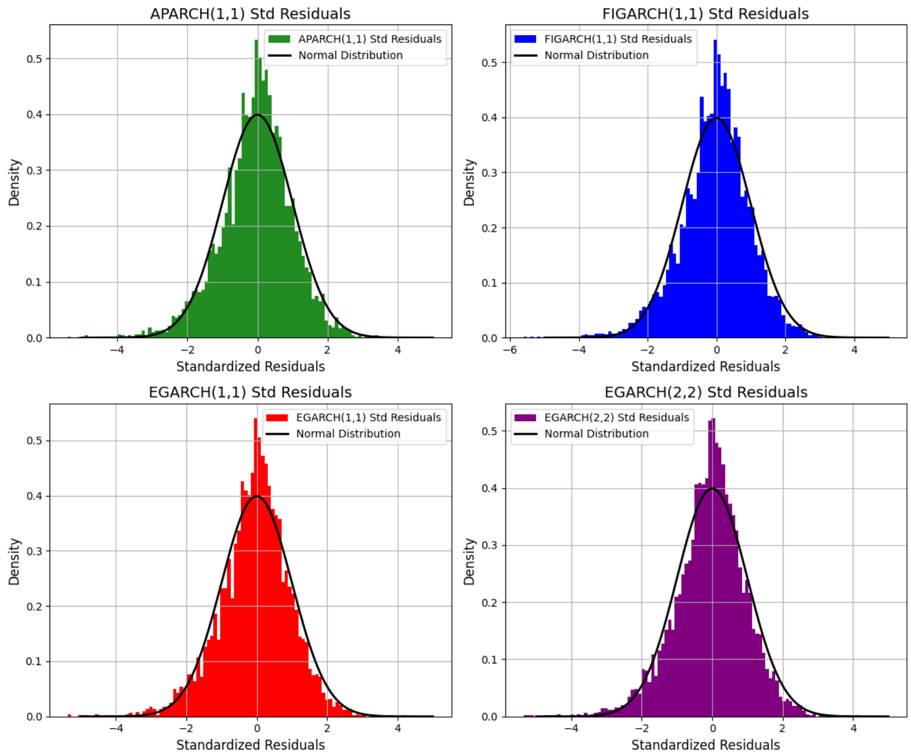


Figure 75: GARCH Models Standardized Residuals compared

1.33.2 Jarque-Bera Test for Standardized Residuals

Table 22: Jarque-Bera Test Results for VGT Models

Model	**Jarque-Bera Statistic**	**p-value**
APARCH(1,1)	502.5212	0.0000
FIGARCH(1,1)	526.4604	0.0000
EGARCH(1,1)	554.5026	0.0000
EGARCH(2,2)	525.3348	0.0000

The null hypothesis is rejected and concluded that the data is significantly non-normal

1.33.3 Ljung-Box Test for Standardized Residuals

Table 23: Ljung-Box Test Results for VGT Models

Model	**Ljung-Box Statistic**	**p-value**
APARCH(1,1)	11.4591	0.3229
FIGARCH(1,1)	12.1627	0.2743
EGARCH(1,1)	12.3337	0.2633
EGARCH(2,2)	12.1398	0.2758

Best Model based on Ljung-Box Test: APARCH(1,1) with p-value: 0.3229
The p-value is greater than 0.05, indicating that the residuals have no significant autocorrelation.

1.33.4 QQ Plot for GARCH Models Standardized Residuals

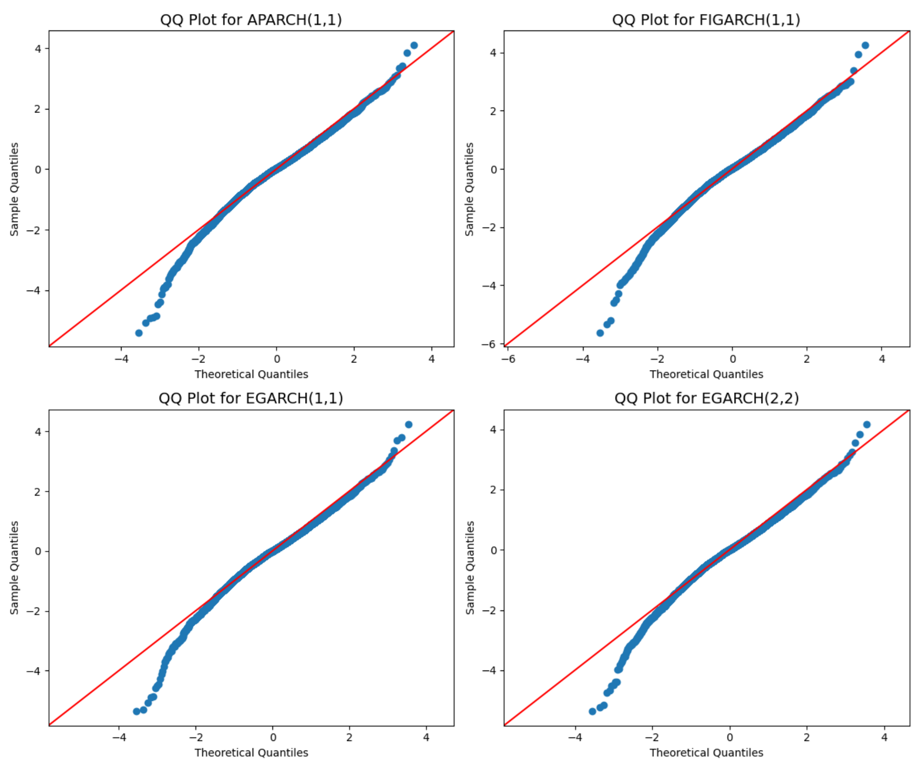


Figure 76: QQ Plot for GARCH Models Standardized Residuals

- **APARCH(1,1)**: The Q-Q plot shows a relatively good fit to the normal distribution. The points are close to the diagonal line, indicating that the residuals are approximately normally distributed. However, there are some deviations, particularly in the tails, suggesting that the distribution may have slightly heavier tails than a normal distribution.
- **FIGARCH(1,1)**: The Q-Q plot for FIGARCH(1,1) also shows a reasonable fit to the normal distribution. The points are generally close to the diagonal line, but there are some deviations, especially in the tails.
- **EGARCH(1,1)**: The Q-Q plot for EGARCH(1,1) shows a similar pattern to the other models. The points are close to the diagonal line, but there are some deviations, particularly in the tails.
- **EGARCH(2,2)**: The Q-Q plot for EGARCH(2,2) shows a slightly better fit to the normal distribution compared to the other models. The points are closer to the diagonal line, and there are fewer deviations in the tails.

1.34 Fitting FIGARCH and APARCH Models to ARMA Residuals

1.34.1 Why Fit Volatility Models to ARMA Residuals

- By first fitting an ARMA model, the primary linear structure is removed, leaving residuals that may still exhibit non-linear patterns in volatility.
- FIGARCH and APARCH are then applied to these residuals to explicitly model the conditional heteroskedasticity (changing variance over time).
- This two-step approach helps in better understanding the time-varying nature of volatility in the data, capturing both short-term patterns (through ARMA) and long-term volatility dynamics or asymmetric effects (through FIGARCH/APARCH).

1.34.2 APARCH Model to ARMA Residuals

Table 24: APARCH Model to ARMA Residuals

Dep. Variable:	None
Mean Model:	Constant Mean
Vol Model:	Power ARCH
Distribution:	Standardized Skew Student's t
Method:	Maximum Likelihood
No. Observations:	5184
R-squared:	0.000
Adj. R-squared:	0.000
Log-Likelihood:	-8125.51
AIC:	16265.0
BIC:	16310.9
Df Residuals:	5183
Df Model:	1

	Coef	**Std Err**	**t**	**P¿—t—**	**95.0% Conf. Int.**
Mean Model					
μ	0.0373	0.01494	2.497	0.01254	[0.008015, 0.06656]
Volatility Model					
ω	0.0227	0.005352	4.243	2.201e-05	[0.01222, 0.03320]
$\alpha[1]$	0.1069	0.01164	9.182	4.249e-20	[0.08409, 0.130]
$\beta[1]$	0.8921	0.01086	82.131	0.000	[0.871, 0.913]
δ	1.7115	0.232	7.378	1.605e-13	[1.257, 2.166]
Distribution					
η	7.8052	0.791	9.864	5.939e-23	[6.254, 9.356]
λ	-0.1374	0.01723	-7.976	1.506e-15	[-0.171, -0.104]

Mean Model Coefficients

- μ (Mu): The estimated constant mean is 0.0373, significant at the 5% level (p-value = 0.01254), indicating that there is a statistically significant mean level in the data.

Volatility Model Coefficients

- ω (Omega): Represents the baseline variance. The estimated value is 0.0227, significant at a very low p-value, suggesting a meaningful long-run variance.
- $\alpha[1]$ (Alpha[1]): Measures the impact of past squared residuals (lagged shocks) on current volatility. A high and significant value (0.1069) indicates that past volatility significantly affects current volatility.
- $\beta[1]$ (Beta[1]): Indicates the persistence of volatility. The high value (0.8921) suggests strong persistence, meaning past volatility influences future volatility for an extended period.
- δ (Delta): Power parameter in the Power ARCH model. Its significant value (1.7115) implies that a non-linear transformation of the absolute residuals improves the volatility modeling.

Distribution Parameters

- η (Eta): Shape parameter of the skew Student's t-distribution. A significant and large value (7.8052) indicates heavier tails than the normal distribution, capturing extreme events.
- λ (Lambda): Skewness parameter, with a significant negative value (-0.1374), implying asymmetry in the residual distribution.

The model effectively captures the time-varying nature of volatility in the data, with significant coefficients for past shocks, volatility persistence, and the power term. The use of a skew Student's t-distribution helps to account for the heavy tails and skewness observed in financial time series data.

1.34.3 FIGARCH Model to ARMA Residuals

Table 25: Constant Mean - FIGARCH Model Results

Dep. Variable:	None
R-squared:	0.000
Mean Model:	Constant Mean
Adj. R-squared:	0.000
Vol Model:	FIGARCH
Log-Likelihood:	-8119.04
Distribution:	Standardized Skew Student's t
AIC:	16252.1
Method:	Maximum Likelihood
BIC:	16298.0
No. Observations:	5184
Df Residuals:	5183
Df Model:	1

Mean Model	**coef**	**std err**	**t**	**P¿—t—**	**95.0% Conf. Int.**
mu	0.0344	1.487e-02	2.312	2.078e-02	[5.235e-03, 6.354e-02]

Volatility Model	**coef**	**std err**	**t**	**P¿—t—**	**95.0% Conf. Int.**
omega	0.0389	1.204e-02	3.235	1.217e-03	[1.534e-02, 6.252e-02]
phi	0.1282	3.998e-02	3.207	1.343e-03	[4.984e-02, 0.207]
d	0.5487	7.698e-02	7.128	1.020e-12	[0.398, 0.700]
beta	0.6033	7.967e-02	7.573	3.642e-14	[0.447, 0.759]

Distribution	**coef**	**std err**	**t**	**P¿—t—**	**95.0% Conf. Int.**
eta	7.6270	0.782	9.757	1.725e-22	[6.095, 9.159]
lambda	-0.1410	1.743e-02	-8.089	6.008e-16	[-0.175, -0.107]

Looking at the model parameters, the mean model coefficient (mu) is significant, with a p-value of 0.02078, suggesting that the estimated mean return is statistically different from zero. In the volatility model, all parameters, including omega, phi, d, and beta, are statistically significant, indicating that past volatility and shocks significantly influence future volatility. Specifically, the values of d and beta suggest a strong persistence in volatility, characteristic of financial time series data.

The distributional parameters, eta and lambda, are also significant and provide insight into the skewness and kurtosis of the returns, enhancing the model's ability to fit the data adequately.

While the FIGARCH Model on the ARMA Residuals captures essential aspects of volatility, the low R-squared values and high information criteria highlight the necessity for further exploration of model specifications.

1.34.4 Conditional Volatility APARCH-ARMA and simple APARCH

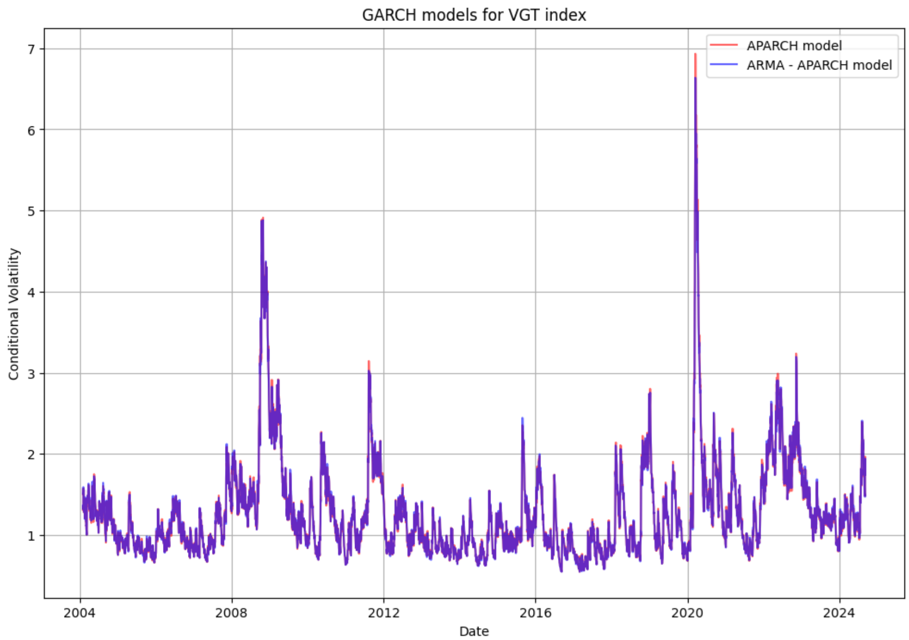


Figure 77: Conditional Volatility APARCH-ARMA and simple APARCH

Both models generate forecasts of conditional volatility, which represents the expected level of volatility at each point in time. The key findings are as follows:

- **APARCH Model:** The APARCH model produces a generally higher level of conditional volatility compared to the ARMA-APARCH model. This suggests that the APARCH model captures more extreme volatility events or exhibits a higher degree of persistence in volatility.
- **ARMA-APARCH Model:** The ARMA-APARCH model exhibits a lower overall level of conditional volatility. This may indicate a less volatile forecast or a different interpretation of the underlying volatility dynamics.
- **Volatility Clustering:** Both models capture the phenomenon of volatility clustering, where periods of high volatility tend to be followed by other periods of high volatility. This is a common characteristic of financial time series.

1.34.5 Conditional Volatility FIGARCH-ARMA and simple FIGARCH

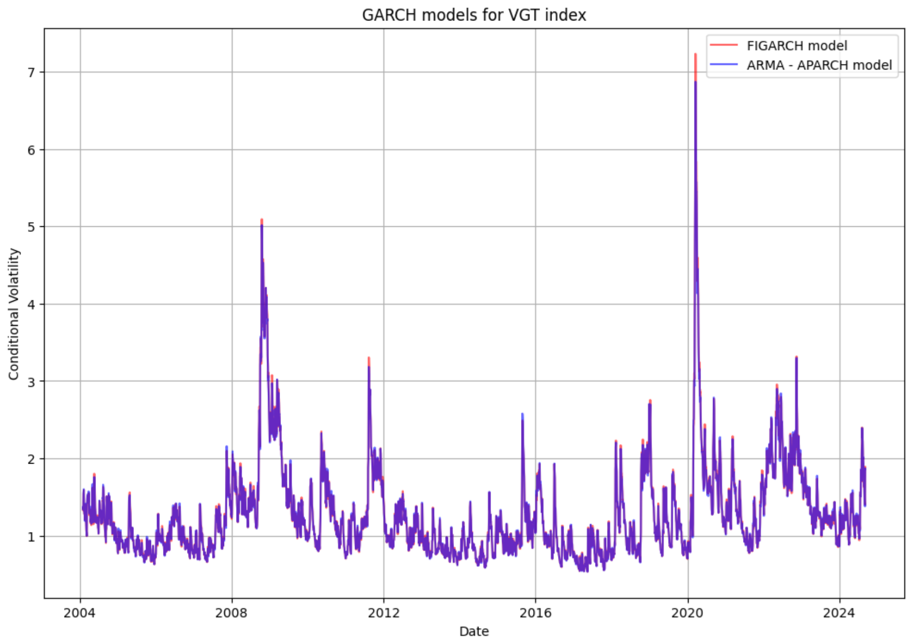


Figure 78: Conditional Volatility FIGARCH-ARMA and simple FIGARCH

Both models generate forecasts of conditional volatility, which represents the expected level of volatility at each point in time. The key findings are as follows:

- **FIGARCH Model:** The FIGARCH model exhibits a generally higher level of conditional volatility compared to the ARMA-APARCH model. This suggests that the FIGARCH model captures more extreme volatility events or exhibits a higher degree of persistence in volatility.
- **ARMA-APARCH Model:** The ARMA-APARCH model exhibits a lower overall level of conditional volatility. This may indicate a less volatile forecast or a different interpretation of the underlying volatility dynamics.
- **Volatility Clustering:** Both models capture the phenomenon of volatility clustering, where periods of high volatility tend to be followed by other periods of high volatility. This is a common characteristic of financial time series.

1.34.6 AIC for ARMA-APARCH and simple APARCH

Model	AIC
ARMA-APARCH	16265.024
APARCH	16262.017

Table 26: AIC values for ARMA-APARCH and APARCH models

1.34.7 AIC for ARMA-FIGARCH and simple FIGARCH

Model	AIC
ARMA-FIGARCH	16252.088
FIGARCH	16249.454

Table 27: AIC values for ARMA-FIGARCH and FIGARCH models

The provided tables present the Akaike Information Criterion (AIC) values for different GARCH models: ARMA-APARCH, APARCH, ARMA-FIGARCH, and FIGARCH. AIC is a model selection criterion that penalizes models with more parameters to avoid overfitting. A lower AIC value generally indicates a better-fitting model.

- **ARMA-APARCH vs. APARCH:** The ARMA-APARCH model has a lower AIC value (16262.017) compared to the simple APARCH model (16265.024). This suggests that the ARMA component in the ARMA-APARCH model improves the model's fit to the data.
- **ARMA-FIGARCH vs. FIGARCH:** Similarly, the ARMA-FIGARCH model has a lower AIC value (16249.454) compared to the simple FIGARCH model (16252.088). This indicates that the ARMA component in the ARMA-FIGARCH model also improves its fit.
- **Overall Comparison:** Among all four models, the ARMA-FIGARCH model has the lowest AIC value (16249.454), suggesting that it is the best-fitting model based on the AIC criterion.

The small difference in AIC values between the ARMA-FIGARCH and ARMA-APARCH models does suggest that both models provide a reasonably good fit to the data. However, the ARMA-FIGARCH model has a slightly lower AIC value, indicating that it is a slightly better fit.

While the difference may seem small, it is important to consider the context of the analysis. In some cases, even small differences in AIC can be meaningful, especially when dealing with complex time series models.

1.34.8 Ljung-Box Test ARMA-APARCH and simple APARCH

Model	Ljung-Box Statistic (lb_stat)	p-value (lb_pvalue)
ARMA-APARCH (10 lags)	29.672629	0.000969
APARCH (10 lags)	11.459125	0.322883

Table 28: Ljung-Box test results for ARMA-APARCH and APARCH models

1.34.9 Ljung-Box Test ARMA-FIGARCH and simple FIGARCH

Model	Ljung-Box Statistic (lb_stat)	p-value (lb_pvalue)
ARMA-FIGARCH (10 lags)	33.302323	0.000242
FIGARCH (10 lags)	12.162653	0.274318

Table 29: Ljung-Box test results for ARMA-FIGARCH and FIGARCH models

The provided tables present the results of the Ljung-Box test for different GARCH models: ARMA-APARCH, APARCH, ARMA-FIGARCH, and FIGARCH. The Ljung-Box test is used to test for autocorrelation in the residuals of a time series model.

- **ARMA-APARCH Model:** The ARMA-APARCH model exhibits significant autocorrelation in the residuals, as evidenced by the p-value of 0.000969. This suggests that the model may not be capturing all the relevant dependencies in the data.
- **APARCH Model:** The simple APARCH model also shows some evidence of autocorrelation, with a p-value of 0.322883. However, the autocorrelation is not as significant as in the ARMA-APARCH model.
- **ARMA-FIGARCH Model:** The ARMA-FIGARCH model also exhibits significant autocorrelation in the residuals, with a p-value of 0.000242. This indicates that the model may not be capturing all the relevant dependencies, particularly those related to long memory.
- **FIGARCH Model:** The simple FIGARCH model shows some evidence of autocorrelation, but the p-value of 0.274318 is not as significant as in the ARMA-FIGARCH model.

1.34.10 Jarque-Bera Test ARMA-APARCH and ARMA-FIGARCH

Model	JB Statistic	p-value
ARMA-APARCH	515.364250	1.2305×10^{-112}
ARMA-FIGARCH	548.868975	6.5255×10^{-120}

Table 30: Jarque-Bera test results for ARMA-APARCH and ARMA-FIGARCH models

The provided table presents the results of the Jarque-Bera test for two GARCH models: ARMA-APARCH and ARMA-FIGARCH. The Jarque-Bera test is used to assess the normality of the residuals in a time series model.

- **ARMA-APARCH Model:** The ARMA-APARCH model exhibits significant non-normality in the residuals, as evidenced by the p-value of 1.2305×10^{-112}. This indicates that the residuals deviate significantly from a normal distribution.
- **ARMA-FIGARCH Model:** The ARMA-FIGARCH model also exhibits significant non-normality, with a p-value of 6.5255×10^{-120}. This suggests that the residuals from this model are even more non-normal than those from the ARMA-APARCH model.

1.35 RMSE and MAE on FIGARCH and APARCH Models to ARMA Residuals

The provided table presents the Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) for different GARCH models: ARMA-APARCH, APARCH, ARMA-FIGARCH, and FIGARCH. These metrics are used to assess the overall accuracy of the models in forecasting conditional volatility.

RMSE:

- The ARMA-APARCH model has the lowest RMSE (1.409789), indicating that it has the smallest average squared error in predicting volatility.
- The ARMA-FIGARCH model has a slightly higher RMSE (1.409838) compared to ARMA-APARCH, but the difference is minimal.
- The simple APARCH and FIGARCH models have slightly higher RMSE values (1.410056 and 1.409993, respectively).

MAE:

- The ARMA-APARCH model also has the lowest MAE (0.975960), indicating that it has the smallest average absolute error in predicting volatility.

- The ARMA-FIGARCH model has a slightly higher MAE (0.976189) compared to ARMA-APARCH, but the difference is minimal.
- The simple APARCH and FIGARCH models have slightly lower MAE values (0.973196 and 0.973259, respectively).

1.36 Standardized Residuals for ARMA-GARCH

1.36.1 Standardized Residuals for ARMA-APARCH and simple APARCH

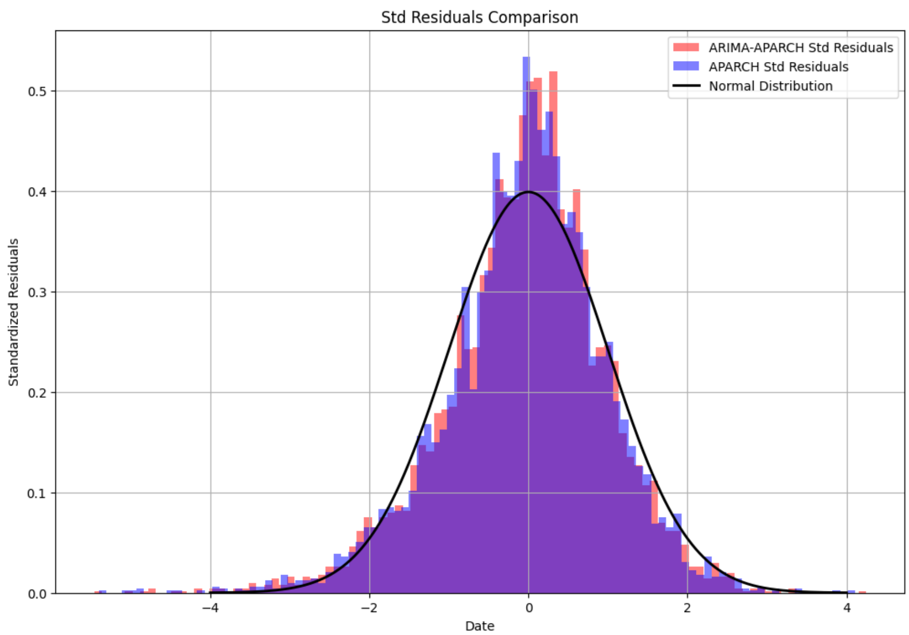


Figure 79: Standardized Residuals for ARMA-APARCH and simple APARCH

The provided image compares the standardized residuals from two GARCH models: ARMA-APARCH and APARCH. Standardized residuals are the residuals from a model divided by their estimated standard deviation, allowing for a direct comparison of the residuals across different models.

Distribution:

- Both ARMA-APARCH and APARCH standardized residuals appear to be centered around zero, indicating that the models are capturing the overall mean of the data.

Kurtosis:

- The ARMA-APARCH residuals exhibit heavier tails than the APARCH residuals, suggesting that the ARMA-APARCH model may be better at capturing extreme events or outliers in the data.

Skewness:

- Both distributions appear to be roughly symmetric, with a slight leftward skew for the ARMA-APARCH residuals and a slight rightward skew for the APARCH residuals.

Comparison to Normal Distribution:

- The black line represents the normal distribution. Both sets of residuals deviate from the normal distribution, particularly in the tails, suggesting that the data may have non-normal characteristics, such as heavy tails or skewness.

1.36.2 Standardized Residuals for ARMA-FIGARCH and simple FIGARCH

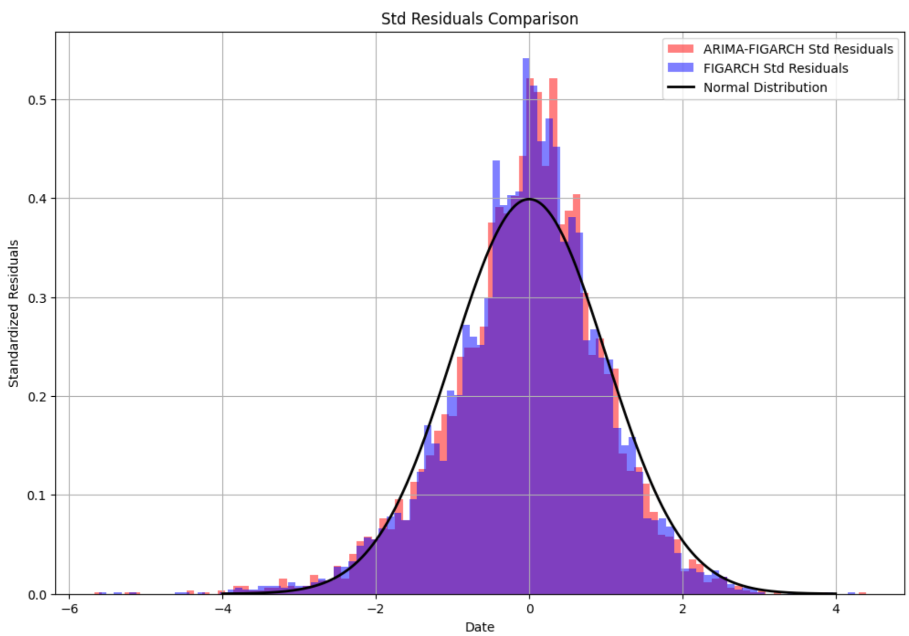


Figure 80: Standardized Residuals for ARMA-FIGARCH and simple FIGARCH

The provided image compares the standardized residuals from two GARCH models: ARMA-FIGARCH and FIGARCH. Standardized residuals are the residuals from a model divided by their estimated standard deviation, allowing for a direct comparison of the residuals across different models.

Distribution:

- Both ARMA-FIGARCH and FIGARCH standardized residuals appear to be centered around zero, indicating that the models are capturing the overall mean of the data.

Kurtosis:

- The ARMA-FIGARCH residuals exhibit slightly heavier tails than the FIGARCH residuals, suggesting that the ARMA-FIGARCH model may be better at capturing extreme events or outliers in the data.

Skewness:

- Both distributions appear to be roughly symmetric, with a slight leftward skew for both ARMA-FIGARCH and FIGARCH residuals.

Comparison to Normal Distribution:

- The black line represents the normal distribution. Both sets of residuals deviate from the normal distribution, particularly in the tails, suggesting that the data may have non-normal characteristics, such as heavy tails or skewness.

1.36.3 Standardized Residuals for ARMA-GARCH and simple GARCH models

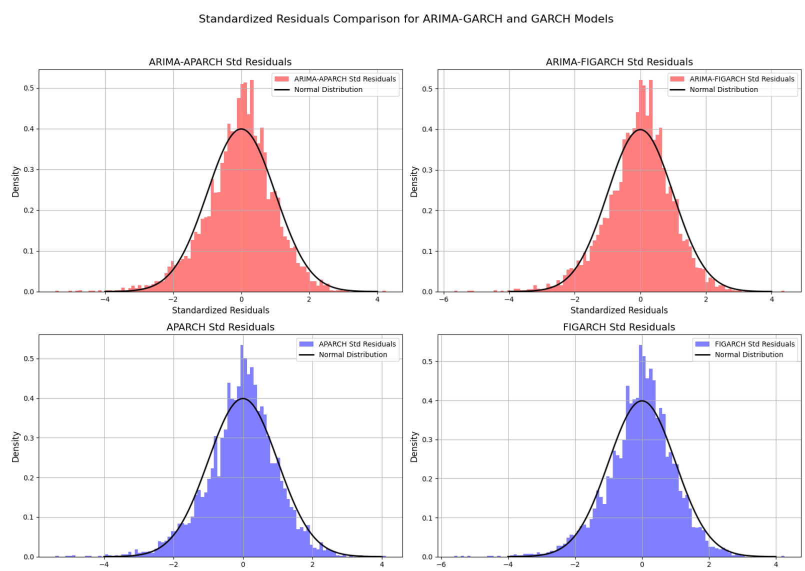


Figure 81: Standardized Residuals for ARMA-GARCH and simple GARCH models

The provided image compares the standardized residuals from four GARCH models: ARMA-APARCH, APARCH, ARMA-FIGARCH, and FIGARCH. Standardized residuals are the residuals from a model divided by their estimated standard deviation, allowing for a direct comparison of the residuals across different models.

Distribution:

- All four sets of standardized residuals appear to be centered around zero, indicating that the models are capturing the overall mean of the data.

Kurtosis:

- The ARMA-APARCH and ARMA-FIGARCH residuals exhibit slightly heavier tails than the APARCH and FIGARCH residuals, suggesting that these models may be better at capturing extreme events or outliers in the data.

Skewness:

- All distributions appear to be roughly symmetric, with a slight leftward skew for the ARMA-APARCH and ARMA-FIGARCH residuals and a slight rightward skew for the APARCH and FIGARCH residuals.

Comparison to Normal Distribution:

- The black line represents the normal distribution. All sets of residuals deviate from the normal distribution, particularly in the tails, suggesting that the data may have non-normal characteristics, such as heavy tails or skewness.

1.37 Second Part Conclusion

In this second part of the VGT index analysis, we focus on understanding the characteristics of returns, volatility, and distributions. Algorithms were developed to identify the optimal ARMA model, alongside residual analysis to assess model fit and performance.

Subsequently, an algorithm was created to identify the best GARCH model across various orders (p, q), GARCH types, and non-Gaussian distributions tailored to handle skewness and heavy tails. Tests and residual analysis were conducted to evaluate the models' ability to handle financial time series data effectively.

The objective is to gain insights into the specific behavior of VGT and, more broadly, to understand how to model financial series, particularly in terms of volatility and distribution characteristics.

Starting with basic metrics, VGT displays a slightly higher mean daily return compared to the S&P 500 (0.06% vs. 0.04%) but also a higher standard deviation (1.41% vs. 1.20%). VGT exhibits less pronounced negative skewness (-0.14 vs. -0.26) and about half the excess kurtosis of the S&P 500 (7.10 vs. 12.94). Both series exhibit expected features of financial data: negative skewness and excess kurtosis, indicating asymmetry and heavy tails.

We then calculated Sharpe ratios under both current rates (September 2024) and projected 2025 lower rates (based on Federal Reserve targets). In both scenarios, VGT shows stronger risk-adjusted performance, suggesting a robust outlook despite its sector-specific exposure.

In volatility analysis, VGT shows higher levels of both upside and downside volatility. Examining drawdowns, we find the S&P 500 experienced a marginally worse maximum drawdown, with an average drawdown of -8.76% compared to VGT's -8.21%.

Distribution analysis and simulations were conducted to determine which distributions best describe financial behaviors. The Normal distribution, while straightforward, fails to capture the extreme behaviors of SPX returns. The Skew-t distribution, in contrast, provides the best fit for capturing potential extreme events, making it valuable for risk management. The GED offers a middle ground, reflecting moderate extremes with manageable complexity.

Several statistical tests confirm key financial return characteristics, including non-normal distributions, significant autocorrelation, heavy tails, and negative skewness. An algorithm was then used to identify the optimal ARMA model, followed by residual analysis, revealing ARCH effects and autocorrelation in residuals.

This was followed by the selection of the optimal ARCH model, testing different GARCH types, orders, and distributions. The analysis focused on two models: the FIGARCH(1,1) with skew-t distribution and the APARCH(1,1) with skew-t distribution.

Through residual analysis, model accuracy in forecasting conditional volatility, information criteria, and performance when applied to ARMA residuals, we selected the APARCH(1,1) with skew-t distribution as the final model.

The APARCH (Asymmetric Power ARCH) model provides advantages in modeling financial time series by:

- Capturing asymmetric effects, such as positive and negative shocks,
- Allowing a power parameter that adjusts for different tail thickness in residuals, and
- Offering flexibility, making it suitable for a wide range of financial data.

The standardized skew Student's t (skew-t) distribution in the APARCH model enables:

- **Tail thickness:** Capturing heavier tails than the normal distribution, which is crucial for modeling extreme events,
- **Skewness:** Accounting for asymmetry in the distribution.

In the final section, we will concentrate on forecasting volatility and returns, as well as on applications in risk management. [1]

[1] data from 2004-02-01 to 2024-09-08 Code developed in Google Colab and reviewed with OpenAI This content was written with the assistance of Google Gemini and OpenAI

1.38 Conditional Volatility Forecasting

In the next chapter, we will focus on forecasting conditional volatility using the APARCH(1,1) model with a skew-t distribution

The APARCH (Asymmetric Power ARCH) model is a generalized version of the GARCH (Generalized Autoregressive Conditional Heteroskedasticity) model. It allows for:

- Asymmetric effects: Positive and negative shocks can have different impacts on volatility.
- Power parameter: The model incorporates a power parameter that can capture different levels of tail thickness in the distribution of residuals.
- Flexibility: APARCH is more flexible than standard GARCH models, making it suitable for modeling a wider range of financial time series.

The standardized skew Student's t (skewt) distribution in the APARCH model allows for:

- **Tail thickness:** The skewt distribution can capture heavier tails than the normal distribution, which is important for modeling extreme events in financial time series.
- **Skewness:** The skewt distribution can also capture skewness, which is the asymmetry of the distribution. This is important for modeling financial time series that have a tendency to have more large positive or negative returns.

1.38.1 One-step ahead forecast

Table 31: One-step ahead forecast of conditional volatility

Date	**h.1**
2024-09-06	1.845256

[1]

[1] This content was written with the assistance of Google Gemini and OpenAI

1.38.2 10-step ahead forecast of conditional volatility

To extend the forecast horizon beyond one step ahead, it was necessary to generate iterative forecasts, as the model is inherently designed for one-step-ahead predictions. This process involves forecasting one step ahead, extracting the predicted variance, and then taking the square root to obtain the volatility. The forecasted volatility is stored, and the model is updated with the new data by treating the forecast as an actual observation.

Table 32: 10-step ahead forecast of conditional volatility

Date	Forecasted Volatility
2024-09-09	1.845256
2024-09-10	1.914160
2024-09-11	1.805711
2024-09-12	2.017824
2024-09-13	2.100529
2024-09-16	1.991288
2024-09-17	2.117620
2024-09-18	2.012527
2024-09-19	1.958594
2024-09-20	2.139759

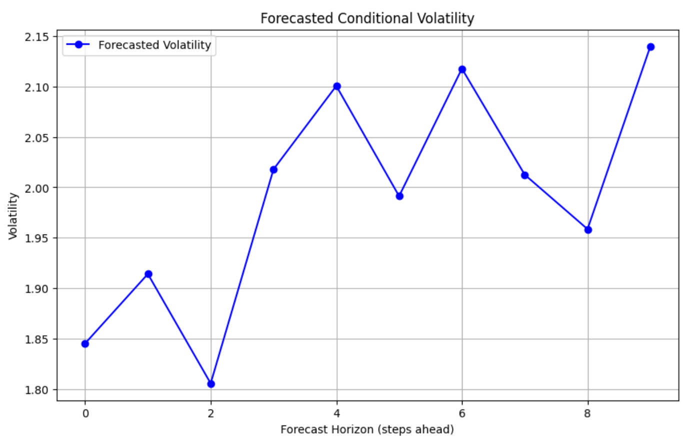


Figure 82: 10-step ahead forecast of conditional volatility

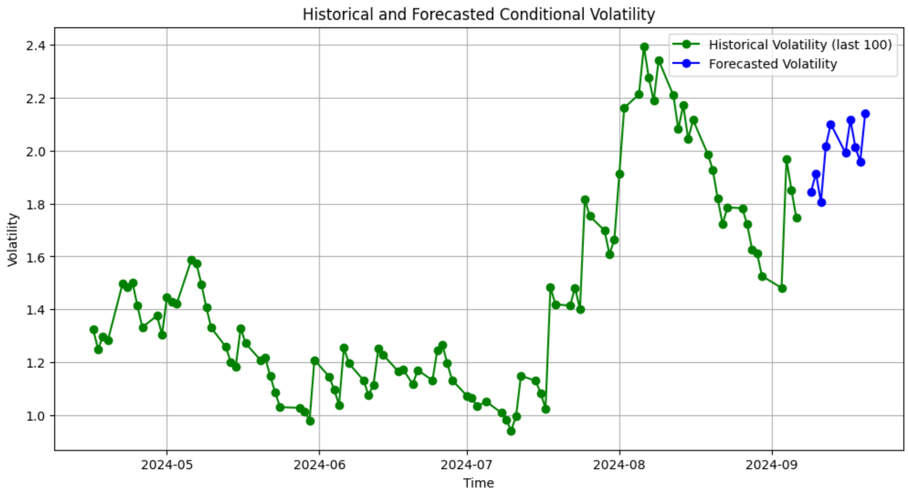


Figure 83: Conditional Volatility series with forecasts

1.39 Returns Forecasting

To forecast the returns, I first extracted the parameters from my APARCH(1,1) model using a skew-t distribution.

Next, I selected the number of future returns to simulate, aligning it with the forecast horizon of the conditional volatility.

I then simulated standardized returns from the skew-t distribution and scaled these standardized returns by the forecasted conditional volatility to obtain the future returns.

(a) APARCH(1,1) Model Parameters for VGT Returns

Parameter	Description
μ	Location (mean)
ω	Scale
α_1	Skewness
β_1	Shape parameter (degrees of freedom)

(b) Estimated Values of APARCH(1,1) Parameters

Parameter	Value
Location (μ)	0.0946
Scale (ω)	0.0237
Skew (α_1)	0.1083
Shape (β_1)	0.8906

Table 33: Parameters and Estimated Values of APARCH(1,1) Model with Skew-t Distribution for VGT Returns

Table 34: Forecasted Returns

Date	Forecasted Return
2024-09-09	0.45139647
2024-09-10	-0.18776025
2024-09-11	0.48645572
2024-09-12	-3.44435975
2024-09-13	0.11336563
2024-09-16	0.03001757
2024-09-17	-0.05825534
2024-09-18	0.0355505
2024-09-19	-1.60200612
2024-09-20	0.52707187

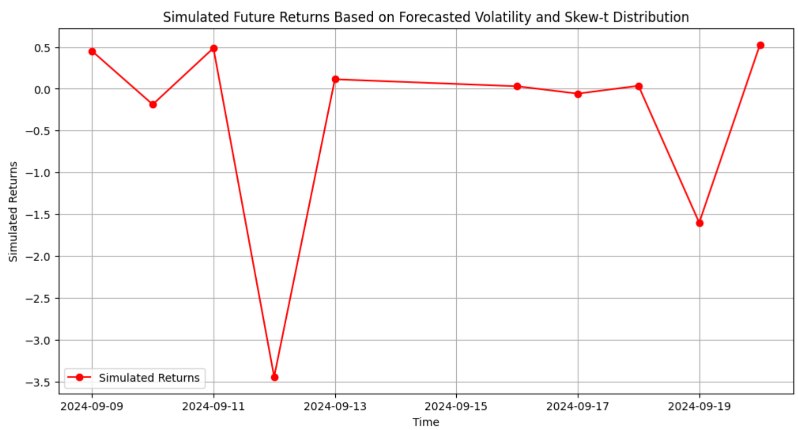


Figure 84: Returns Forecasting

[1]

[1]Code developed in Google Colab and reviewed with OpenAI

1.40 Risk Management Strategies: VaR and Expected Shortfall

I divided the returns data into a training set comprising 75% of the data and a test set consisting of the remaining 25%.

Using the training data, I calculated the Value at Risk (VaR) and the Expected Shortfall (ES).

Subsequently, I forecasted the VaR for the test data and computed the Expected Shortfall based on these forecasted VaR values.

1.40.1 VaR based on GED

A rolling Value at Risk (VaR) based on the Generalized Error Distribution (GED) is calculated for a dataset. A loop iterates through the dataset, creating a rolling window of data for each iteration. The GED is fitted to this window, and the 95th percentile VaR is calculated, representing the threshold for potential losses at a 5% significance level.

Value at Risk (VaR) is a widely used risk management tool that quantifies the potential loss in value of a portfolio over a defined period for a given confidence interval. When calculating VaR, the choice of distribution for asset returns is crucial, as it can significantly impact the risk estimates. Using the Generalized Error Distribution (GED) can offer several advantages over the normal distribution in certain contexts:

- **Flexibility in Tail Behavior:** The GED is more flexible than the normal distribution in modeling the tails of the distribution. It can accommodate various levels of kurtosis, meaning it can better capture the likelihood of extreme events (fat tails) or more peaked distributions compared to the normal distribution. This is particularly useful in finance, where asset returns often exhibit higher kurtosis due to the prevalence of extreme outcomes.

- **Empirical Fit to Financial Data:** Financial return data frequently show asymmetry and non-normality. The GED can model these characteristics more accurately than a normal distribution, leading to more reliable estimates of risk. Since the GED allows for both skewness and kurtosis adjustments, it can provide a better fit to actual return distributions observed in financial markets.

- **Improved Risk Assessment:** Using a GED-based VaR can lead to more conservative risk assessments. Since the GED accounts for extreme values more effectively, it may indicate higher potential losses in stressful market conditions compared to a normal distribution, which could underestimate risk.

- **Adaptability to Different Asset Classes:** Different asset classes (equities, bonds, commodities) exhibit varying return distributions. The GED's flexibility allows it to be adapted for different assets more easily than the normal distribution, making it suitable for multi-asset portfolios where returns might not be identically distributed.

Table 35: Value at Risk (VaR) based on GED at 5%

Date	**VaR_95**
2004-03-17	-2.485409
2004-03-18	-2.549134
2004-03-19	-2.219550
2004-03-22	-2.262838
2004-03-23	-2.349841

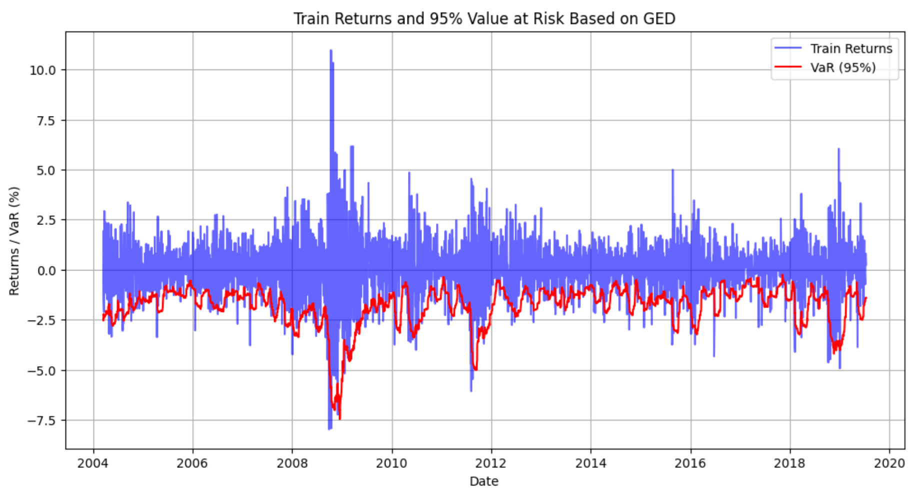


Figure 85: VaR 5% based on GED

VaR Failure Rate: 6.84%

A 6.84% failure rate suggests that in approximately 6.84% of the observed periods (or trading days), actual losses were greater than the VaR estimate. For example, if the VaR is calculated at a 95% confidence level, it means that in 5% of cases, the actual loss could exceed the predicted VaR.

In practical terms, if you assess the performance of your VaR model over a certain period (say, one year), and it records that actual losses exceeded the VaR estimate on about 6.84% of occasions, this indicates that your risk model might be underestimating risk or that the assumptions made in calculating VaR are not fully capturing the tail risk of the asset or portfolio.

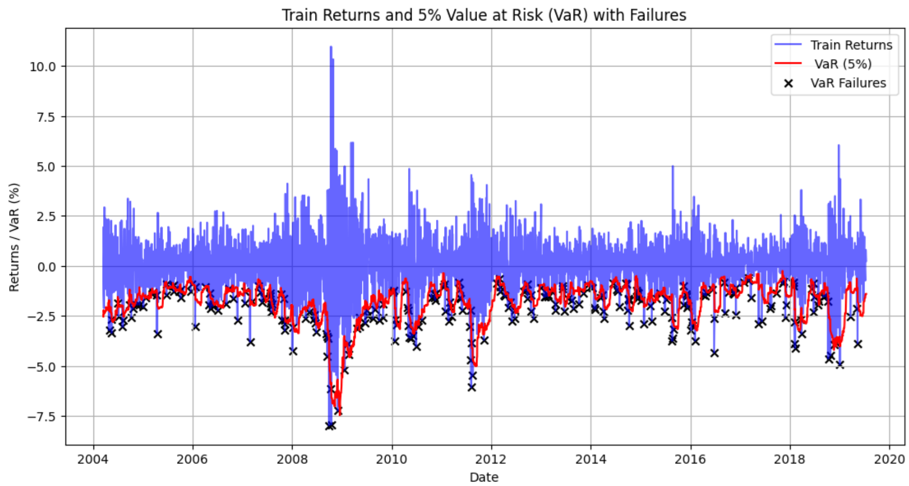


Figure 86: VaR 5% based on GED with Failures

To evaluate the effectiveness of different Value-at-Risk (VaR) models in financial series, we begin by considering the failure rate of the VaR at the 5% level based on the Generalized Error Distribution (GED), which is 6.84%.

As a benchmark, we use the VaR calculated with the Normal distribution, analyzing its corresponding failure rate for comparison.

Finally, we assess the VaR based on the skew-t distribution and its failure rate to determine which model provides the most accurate risk assessment for financial data.

1.40.2 ES based on GED

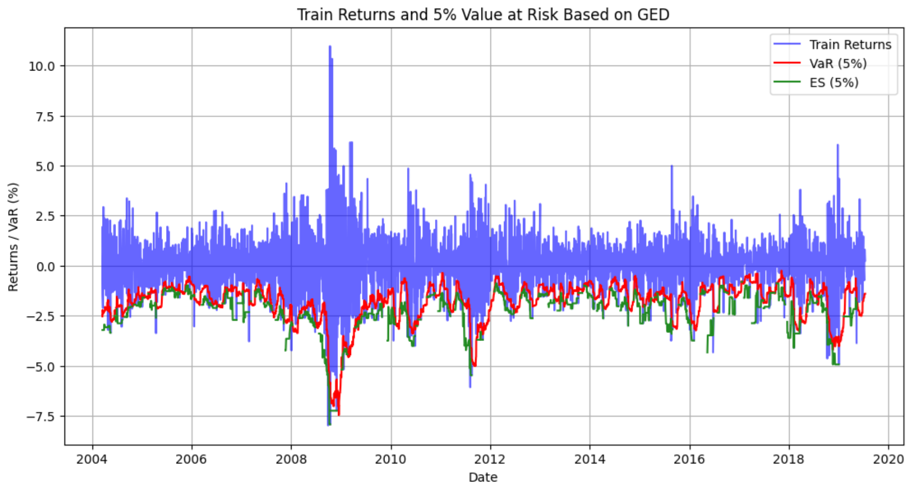


Figure 87: VaR and ES 5% based on GED

Expected Shortfall (ES) is considered a better risk measure than Value-at-Risk (VaR) because it addresses some key limitations of VaR, providing a more comprehensive view of tail risk:

- **Captures Tail Risk Better**: While VaR indicates the maximum potential loss at a specified confidence level (e.g., 5%), it does not account for the severity of losses beyond that threshold. In contrast, ES measures the average loss when the loss exceeds the VaR, thus offering insight into the "worst-case" scenarios and providing a more accurate picture of extreme risks.
- **Sub-additivity and Coherent Risk Measure**: ES satisfies the properties of a coherent risk measure, such as sub-additivity, which means that the risk of a combined portfolio is no greater than the sum of the individual risks. VaR can sometimes fail this property, potentially underestimating the risk of diversified portfolios. ES, being sub-additive, ensures more consistent risk assessments across portfolios.
- **Sensitivity to Tail Events**: ES is more sensitive to changes in the distribution of extreme losses. If the tail risk increases, ES will reflect this increase more accurately than VaR, which only considers the specific quantile without capturing the distribution beyond that point.
- **Regulatory Preference**: Financial regulators, such as the Basel Committee on Banking Supervision, have moved towards using ES over VaR for capital adequacy requirements because ES provides a more robust measure of potential losses in extreme market conditions.

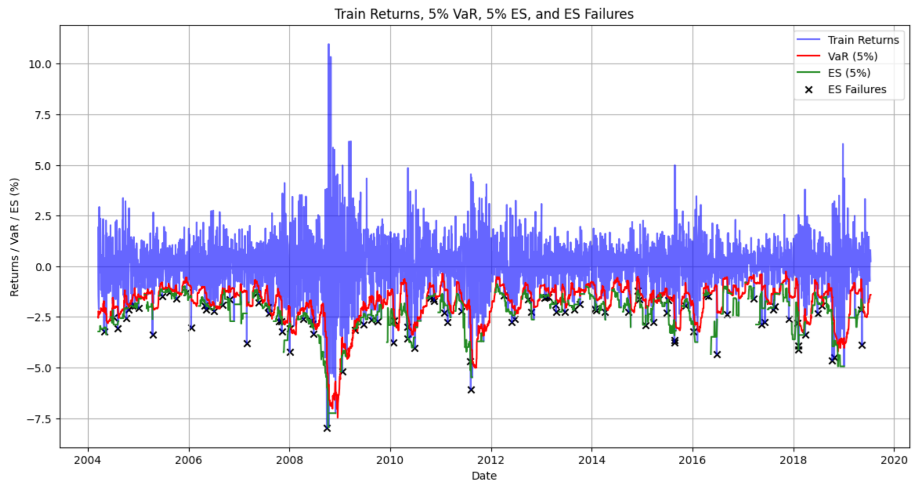


Figure 88: VaR and ES 5% based on GED with ES Failures

ES-Ged Failure Rate: 2.62%

A 2.62% failure rate means that in approximately 2.62% of the instances (such as trading days) observed, the actual losses were greater than the ES estimate. This implies that during those instances, the losses exceeded what was predicted by the ES model.

If the ES is calculated at a certain confidence level (for example, 95%), the failure rate indicates that there were times when losses in the worst-case scenarios were greater than what the model estimated.

Model Effectiveness: A failure rate of 2.62% may be seen as relatively low, especially if the ES was estimated at a high confidence level (like 95%). It suggests that the model is effectively capturing the tail risk of the portfolio to a reasonable extent.

1.40.3 VaR based on Normal distribution

Using a normal distribution to calculate VaR for financial data has several limitations:

- **Assumption of Symmetry**
  - **Symmetry in Returns:** The normal distribution assumes that the returns are symmetrically distributed around the mean, meaning that positive and negative returns of the same magnitude are equally likely.
  - **Real-World Data:** In reality, financial returns often exhibit skewness, where the distribution is not symmetric. For example, negative returns may be more frequent or severe than positive returns, particularly in turbulent market conditions. This skewness is not captured by a normal distribution.
- **Underestimation of Tail Risk**
  - **Thin Tails in Normal Distribution:** The normal distribution has "thin tails," meaning that extreme events (large gains or losses) are considered to be very unlikely.
  - **Fat Tails in Financial Data:** Financial returns often show "fat tails" or a higher probability of extreme outcomes compared to the normal distribution. As a result, using a normal distribution-based VaR may underestimate the likelihood of significant losses during periods of market stress, making it inadequate for capturing tail risk.
- **Inability to Account for Volatility Clustering**
  - **Constant Volatility Assumption:** VaR based on the normal distribution typically assumes that volatility is constant over time.
  - **Volatility Clustering in Markets:** Financial markets often experience periods of high volatility followed by high volatility, or low volatility followed by low volatility, known as volatility clustering. This dynamic nature of market volatility is not captured well by the constant volatility assumption inherent in a normal distribution-based VaR calculation.
- **Ignores Non-Normal Features like Kurtosis**
  - **Normal Distribution Kurtosis:** A normal distribution has a kurtosis of 3, which indicates moderate thickness of tails and height of the peak.
  - **Excess Kurtosis in Financial Data:** Financial returns usually exhibit excess kurtosis, meaning there is a higher probability of extreme values than predicted by a normal distribution. Using normal distribution-based VaR may therefore fail to account for the occurrence of such extreme events, leading to underestimation of risk.
- **Assumption of Independent Returns**
  - **Independent Returns Assumption:** The normal distribution-based VaR assumes that asset returns are independent and identically distributed over time.
  - **Serial Correlation in Financial Data:** In reality, returns may exhibit some degree of correlation across time (e.g., autocorrelation), and events in the market may impact returns for multiple consecutive periods. This correlation can lead to an underestimation of risk when using the normal distribution.

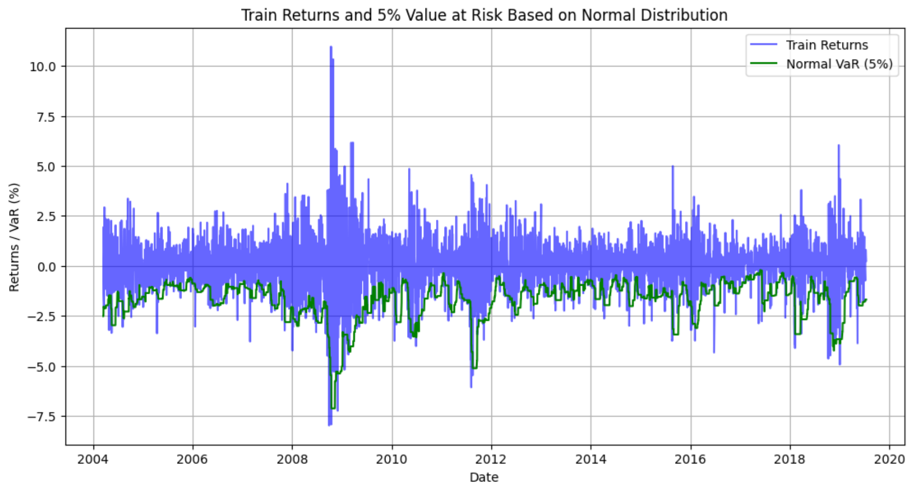


Figure 89: VaR based on Normal distribution

A rolling window approach to compute the VaR for each subset of data, sliding the window across the entire dataset has been used.

For each window, it has been calculated the 5th percentile of the returns, representing the 5% VaR.

Date	**VaR_Normal_95**
2004-03-17	-2.519824
2004-03-18	-2.519824
2004-03-19	-2.041589
2004-03-22	-2.041589
2004-03-23	-2.119646

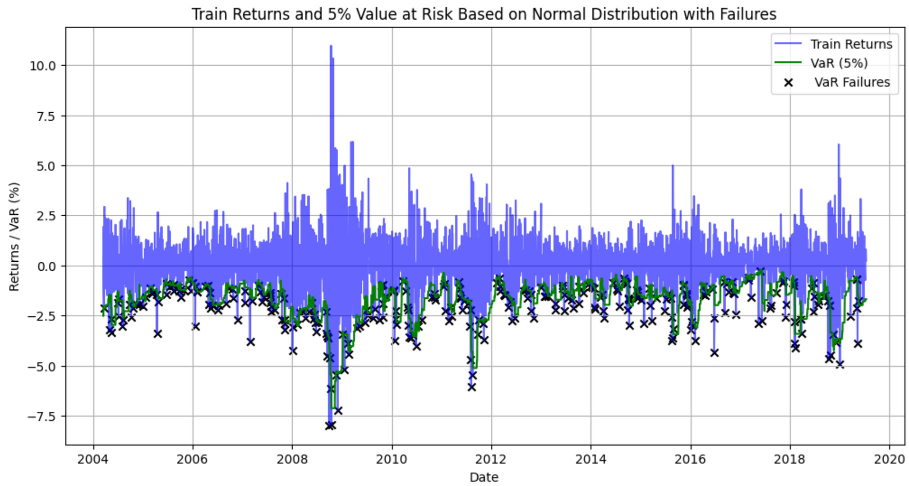


Figure 90: VaR based on Normal distribution with Failures

Normal VaR Failure Rate: 7.85%

- The VaR failure rate measures how often actual losses surpass the predicted VaR threshold. For example, if VaR is calculated at a 95% confidence level, we would expect losses to exceed the VaR level in about 5% of the cases.
- If the failure rate is higher than the expected rate (here, 5%), it suggests that the VaR model is underestimating risk.
- A 7.85% failure rate is higher than the expected 5% for a 95% VaR model. This indicates that losses are exceeding the VaR estimate more frequently than predicted.
- **Underestimation of Risk:** This higher failure rate suggests that the normal distribution-based VaR may not be adequately capturing the true risk, likely due to its limitations, such as assuming symmetric returns and thin tails.
- **Model Limitations:** The result may highlight that the assumption of normally distributed returns does not align well with actual financial return distributions, which often exhibit fat tails and skewness.

1.40.4 VaR based on Skewt

- **Advantages of VaR Based on the Skew-t Distribution**
  - **Ability to Capture Skewness and Heavy Tails:** The Skew-t distribution is a flexible model that accounts for both skewness (asymmetry) and fat tails (kurtosis) in the data. This makes it particularly suitable for financial returns, which often exhibit these characteristics, especially during periods of market turmoil.
  - **Modeling Extreme Events:** The heavy tails of the Skew-t distribution allow it to better capture the likelihood of extreme losses or gains, which are often underestimated by the normal distribution. This results in more realistic risk assessments.
  - **Asymmetric Behavior:** Financial returns often show asymmetric behavior, where negative outcomes may be more severe or frequent than positive outcomes. The Skew-t distribution can adjust for this asymmetry, providing more accurate risk estimates for downside risk.
- **Why Skew-t VaR is Better than Normal Distribution-based VaR**
  - **Tail Risk Assessment:** The normal distribution assumes "thin tails," meaning that the probability of extreme outcomes is low. In contrast, the Skew-t distribution has "fat tails," making it more suitable for capturing tail risk in financial markets, where extreme events occur more frequently than what the normal distribution predicts.
  - **Non-Symmetric Return Distributions:** The normal distribution assumes symmetric returns around the mean, which is often not the case for financial data. The Skew-t distribution's ability to handle skewness helps better represent the actual distribution of returns.
  - **Adaptability to Market Conditions:** The Skew-t distribution can model periods of high market volatility more accurately by allowing for greater flexibility in the shape of the distribution, unlike the normal distribution, which may underestimate risk under such conditions.
- **Differences Between Skew-t VaR and GED VaR**
  - **Tail Flexibility:** While both the Skew-t and GED distributions can handle fat tails, the Skew-t distribution specifically incorporates skewness as a parameter. This gives it an advantage when modeling data with significant asymmetry, whereas GED primarily adjusts for kurtosis.
  - **Parameterization:** The Skew-t distribution has parameters for location, scale, skewness, and degrees of freedom (controlling tail thickness), allowing for more granular control over the distribution's shape. GED, on the other hand, mainly adjusts for shape via the tail parameter, which controls the tail thickness but does not explicitly account for skewness.
  - **Application in Stress Scenarios:** The Skew-t VaR may perform better when the data exhibits strong skewness and heavy tails simultaneously. In contrast, GED VaR might be more appropriate when excess kurtosis (fat tails) is present without significant skewness.

- **When to Use Skew-t VaR vs. GED VaR**
  - **Skew-t VaR:** It is more suitable when the financial return data show both skewness and heavy tails, such as during market crashes or sudden economic shocks where negative returns are much more pronounced.
  - **GED VaR:** It may be preferable when the primary feature of the data is excess kurtosis without substantial skewness, as it can effectively capture fat-tailed distributions.

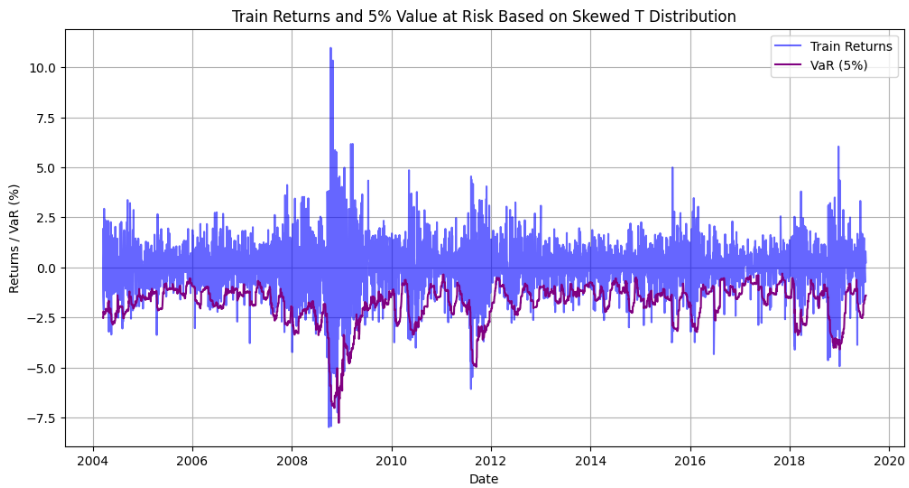


Figure 91: VaR based on Skew-t distribution

For each rolling window over the training data, a subset of data points was extracted. A Skewed t-distribution was then fitted to this subset to estimate the distribution's parameters. Using the fitted parameters, the 5% VaR was computed, representing the threshold below which 5% of the distribution falls. These rolling VaR values were calculated for each window and stored in a DataFrame, aligned with the corresponding dates.

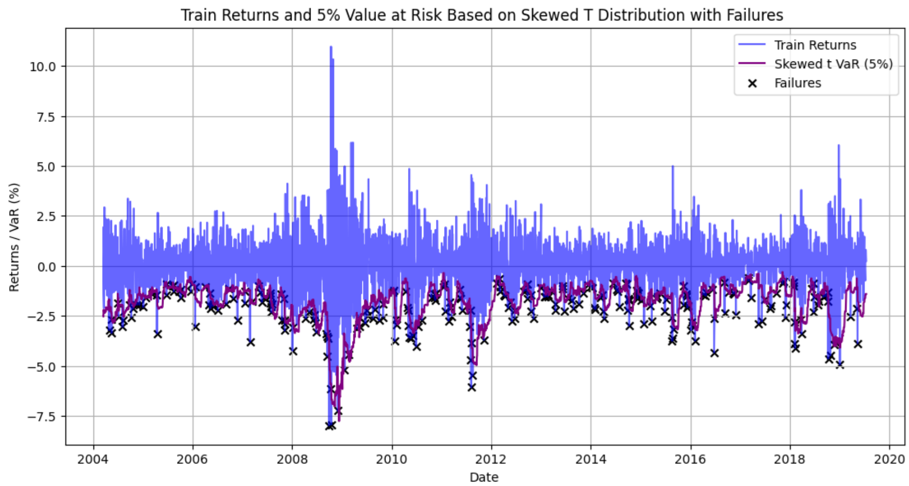


Figure 92: VaR based on Skew-t distribution with Failures

Skewed t VaR Failure Rate: 6.82%

A Skewed t VaR failure rate of 6.82% means that actual losses exceeded the estimated 5% VaR threshold more frequently than the expected 5% of the time. Since the VaR was calculated at a 95% confidence level, losses should ideally exceed the VaR estimate about 5% of the time under a well-calibrated model.

A higher failure rate, such as 6.82%, suggests that the Skewed t VaR may still be underestimating risk, although it may be performing better than models based on normal distribution assumptions. This underestimation could be due to limitations in capturing the actual behavior of the return distribution, even when accounting for skewness and heavy tails.

1.40.5 ES based on Skewt

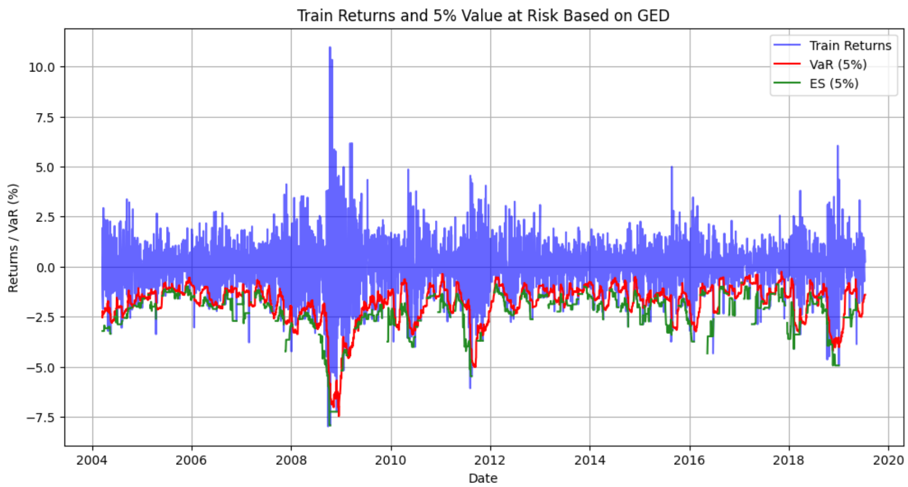


Figure 93: VaR and ES 5% based on Skewt

The analysis of failure rates will help identify the most suitable distribution for risk measurement

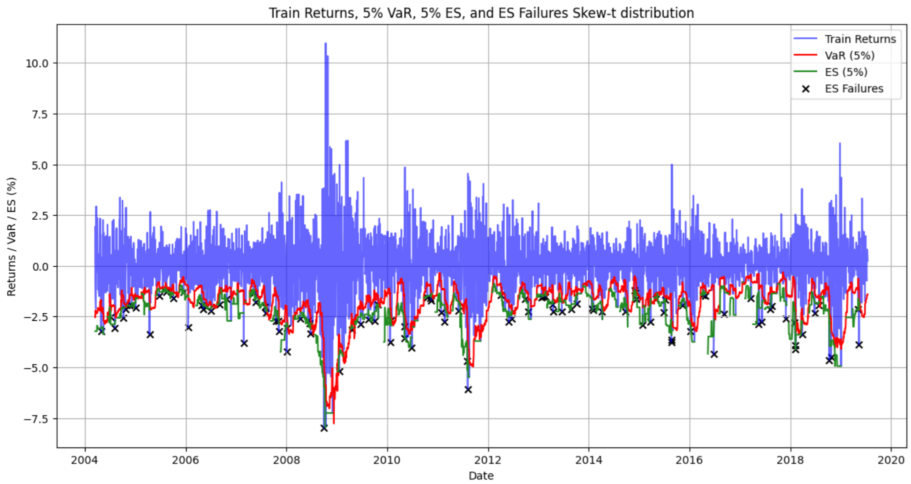


Figure 94: VaR and ES 5% based on Skewt with Failures

ES-Skewt Failure Rate: 2.67%

An ES-Skewt failure rate of 2.67% means that, when using the Expected Shortfall (ES) calculated based on the skew-t distribution, 2.67% of the observed losses exceeded the predicted risk threshold. In other words, in 2.67% of the cases, the actual losses were greater than what the ES model, using the skew-t distribution, anticipated for the given confidence level.

This indicates the level of accuracy and reliability of the ES model with the skew-t distribution in predicting extreme losses. A lower failure rate generally suggests a more accurate risk measure, as fewer instances exceed the expected loss threshold.

1.41 Forecasting Risk with Test data

Forecasting risk measures like Value-at-Risk (VaR) and Expected Shortfall (ES) in financial returns is crucial.

Financial institutions use VaR and ES to assess the potential losses in their portfolios under adverse market conditions. By forecasting these measures, firms can better understand their exposure to risk and take steps to mitigate potential losses, such as adjusting their positions or implementing hedging strategies.

1.41.1 Forecasting with GED distribution

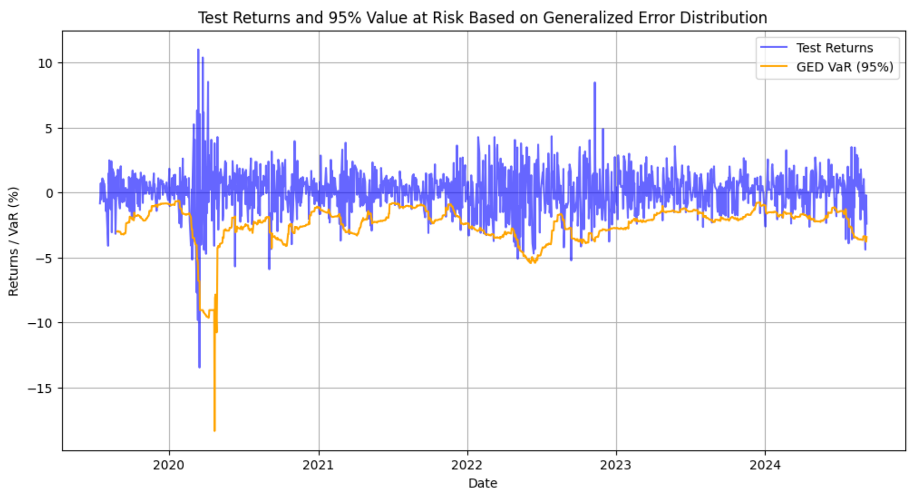


Figure 95: Forecasted VaR based on GED

It calculated the rolling Value at Risk (VaR) at a 95% confidence level as a forecast of potential future losses in a test dataset of financial returns. It initializes a list to store VaR values, then iterates through the dataset, applying a rolling window to estimate the parameters of the Generalized Error Distribution (GED). Using these parameters, it forecasts the VaR for each time point, which indicates the maximum expected loss over a specified period. The results are compiled into a DataFrame, providing a dynamic view of risk exposure over time.

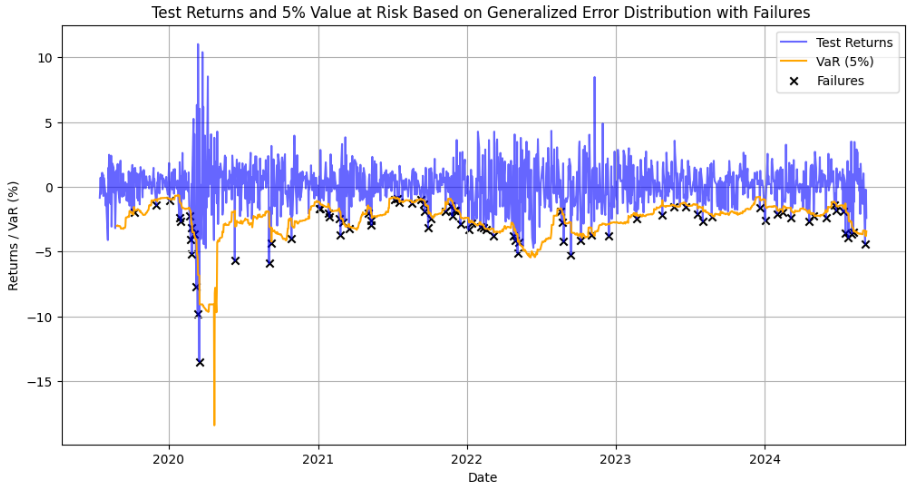


Figure 96: Forecasted VaR based on GED with Failures

GED Forecasted VaR Failure Rate on Test Data: 6.56%

The GED Forecasted VaR Failure Rate on Test Data of 6.56% means that the Value at Risk (VaR) model using a Generalized Error Distribution (GED) underestimated the actual losses in 6.56% of the cases during the out-of-sample testing period.

In other words, the model predicted that losses would not exceed a certain threshold (the VaR level) in 93.44% of cases, but in reality, losses exceeded this threshold in 6.56% of cases. This indicates that the model may be underestimating the potential for extreme negative returns.

A failure rate of 6.56% is relatively high, suggesting that the GED model may not be capturing the full extent of tail risk in the data.

It is generally difficult to achieve a failure rate significantly lower than 5% in financial data, especially during periods of high volatility or market stress

- Market Uncertainty: Financial markets are inherently unpredictable. Even the most sophisticated models can struggle to accurately forecast extreme events that occur infrequently but can have a significant impact.
- Data Limitations: Historical data may not adequately capture the full range of potential market conditions, especially during periods of crisis or economic shocks.
- Model Complexity: While more complex models can improve accuracy, they also introduce additional risks and uncertainties. There is a trade-off between model complexity and robustness.

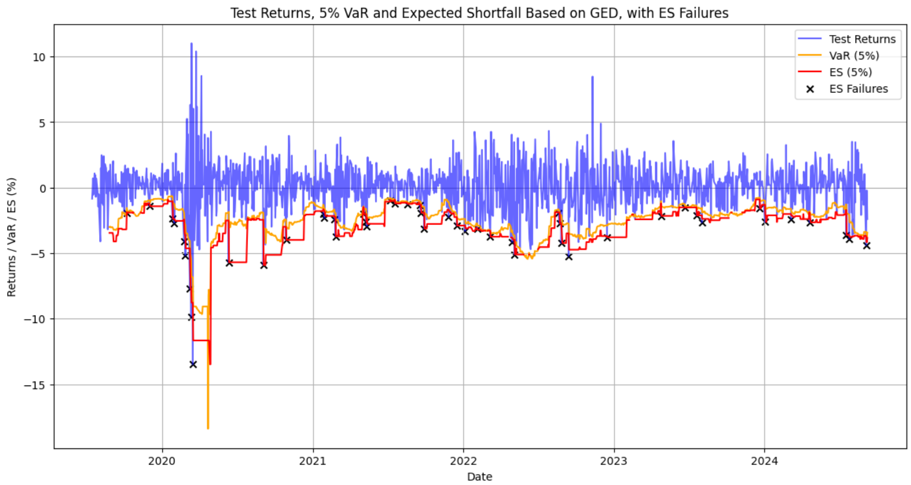


Figure 97: Forecasted VaR and ES based on GED with Failures

GED ES Failure Rate on Test Data: 3.79%

A forecasted Expected Shortfall (ES) failure rate of 3.79% indicates that, in about 3.79% of instances, the actual losses exceeded the predicted ES based on the Generalized Error Distribution (GED). This result suggests that the model's ability to predict extreme losses is relatively better than that of the Value at Risk (VaR) model, as the failure rate is below the typical 5% threshold often considered acceptable in risk management.

An ES failure rate of 3.79% implies that the model is effectively capturing potential tail risks, which is crucial for understanding the severity of losses in extreme scenarios.

1.41.2 Forecasting with Skewt distribution

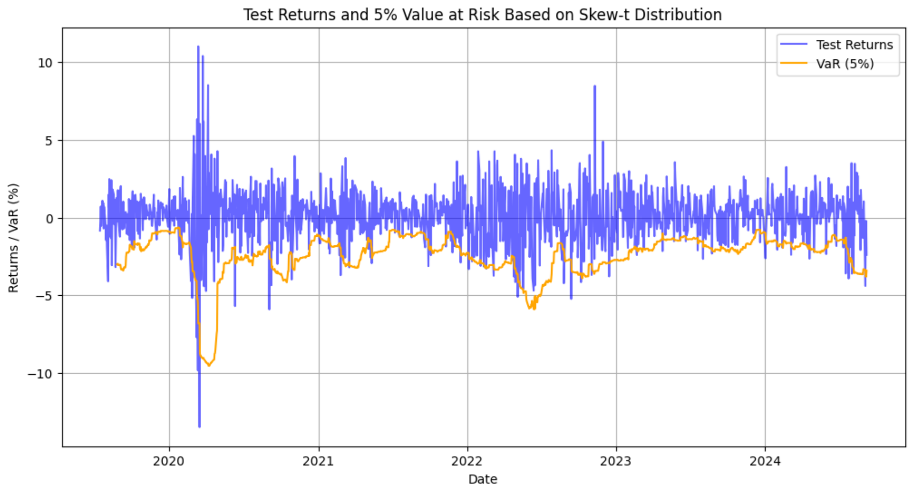


Figure 98: Forecasted VaR based on Skewt

In this procedure, the rolling Value at Risk (VaR) at a 95% confidence level is forecasted using a skewed t-distribution for both training and test datasets of financial returns.

First, for the training data, the process involves iterating through the dataset using a rolling window. For each window, the parameters of the skewed t-distribution are estimated by fitting the model to the historical data. The 5% VaR is then calculated using the percent-point function (PPF) of the skewed t-distribution based on these fitted parameters. The computed VaR values are stored in a list and subsequently organized into a DataFrame.

Next, the focus shifts to forecasting the rolling VaR for the test dataset. The process begins by iterating through the test data, where NaN values are appended for the initial indices to account for insufficient data. Once enough data points are available, a rolling window is employed to fit the skewed t-distribution to the most recent observations in the test dataset.

This fitting allows for the calculation of the 5% VaR, which forecasts the potential maximum loss at a 95% confidence level for the current time point. The calculated VaR values are stored in a list and ultimately organized into a DataFrame.

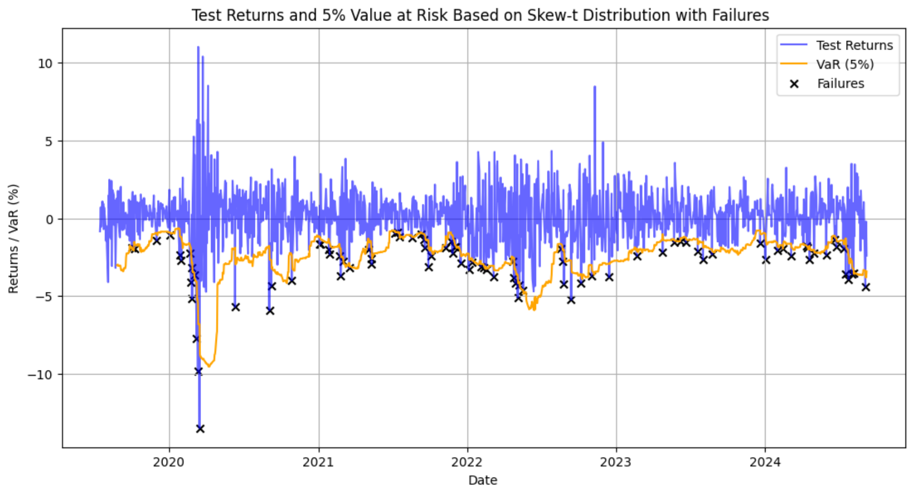


Figure 99: Forecasted VaR based on Skewt with failures

Skew-t VaR Failure Rate on Test Data: 7.03

A forecasted Skew-t Value at Risk (VaR) failure rate of 7.03% indicates that actual losses exceeded the predicted VaR in approximately 7.03% of instances for the test dataset. This failure rate is slightly above the commonly accepted threshold of 5%, suggesting that the model may be underestimating risk more frequently than expected.

This higher failure rate could imply that the skew-t distribution may not fully capture the underlying characteristics of the data, such as extreme events or tail risks.

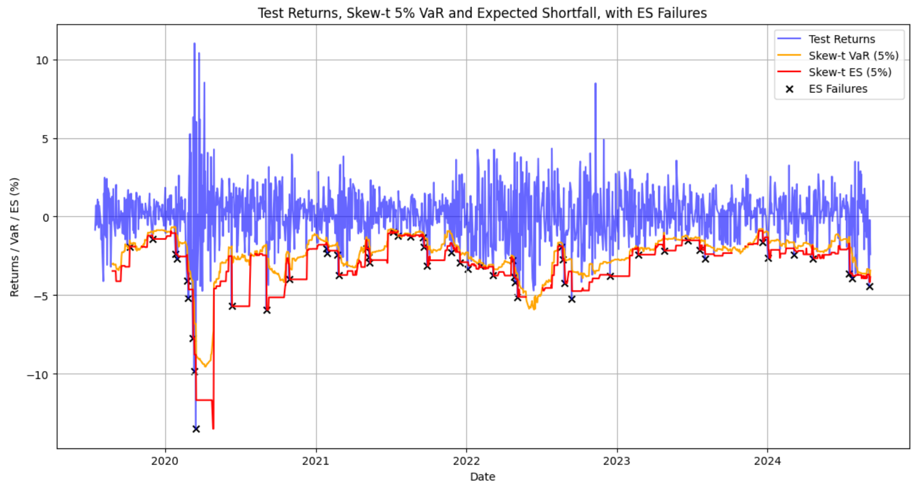


Figure 100: Forecasted VaR and ES based on Skewt with failures

Skew-t ES Failure Rate on Test Data: 4.03

A forecasted Skew-t Expected Shortfall (ES) failure rate of 4.03% suggests that actual losses exceeded the predicted ES in about 4.03% of cases within the test dataset. This result is below the typical threshold of 5%, indicating that the model is performing well in capturing potential tail risks.

The relatively low failure rate implies that the Skew-t ES model effectively forecasts extreme losses, making it a reliable tool for risk assessment in this context. A failure rate below 5% indicates that the model is appropriately capturing the risk of significant losses, which is crucial for effective risk management and capital allocation. [1]

[1]Code developed in Google Colab and reviewed with OpenAI This content was written with the assistance of Google Gemini and OpenAI

1.42 Summary of Risk Measures

1.42.1 Train Data Risk Measures

Table 36: VaR Failure Rates for Different Distributions

Distribution	**VaR Failure Rate (%)**
GED	6.84
Normal	7.85
Skewed t	6.82

The VaR (Value at Risk) failure rates for different models show some notable differences in how accurately each distribution captures the risk. The GED (Generalized Error Distribution) VaR has a failure rate of 6.84%, which is quite close to the Skewed t distribution's failure rate of 6.82%, indicating that both perform similarly in terms of capturing extreme losses. In contrast, the Normal distribution shows a higher failure rate at 7.85%, suggesting it underestimates risk more frequently compared to the GED and Skewed t distributions. This comparison indicates that the GED and Skewed t distributions may be more reliable for modeling risk, especially when considering the heavy tails and skewness present in financial returns.

Table 37: ES Failure Rates for Different Distributions

Distribution	**ES Failure Rate (%)**
ES-GED	2.62
ES-Skewed t	2.67

The Expected Shortfall (ES) failure rates for the GED and Skewed t distributions are very close, with ES-GED having a slightly lower failure rate at 2.62% compared to ES-Skewed t's 2.67%. This small difference suggests that both distributions perform similarly well in capturing extreme losses beyond the VaR threshold. The slight edge in the ES-GED failure rate indicates it may offer a marginally better fit for risk modeling in this context, though the difference is minimal.

1.42.2 Test Data Risk Measures

Table 38: Forecasted VaR Failure Rates on Test Data

Distribution	Forecasted VaR Failure Rate (%)
GED	6.56
Skew-t	7.03

The forecasted VaR failure rates on the test data reveal some differences between the GED and Skew-t distributions. The GED model shows a failure rate of 6.56%, which is lower than the Skew-t model's failure rate of 7.03%. This indicates that the GED distribution provides a slightly better fit for predicting extreme losses in the test data, as it captures the risk more accurately compared to the Skew-t distribution. The difference, although not large, suggests that the GED model may be more reliable for risk forecasting in this context.

Table 39: Forecasted ES Failure Rates on Test Data

Distribution	Forecasted ES Failure Rate (%)
GED	3.79
Skew-t	4.03

The forecasted ES failure rates on the test data show that the GED distribution has a lower failure rate at 3.79%, compared to the Skew-t distribution's failure rate of 4.03%. This suggests that the GED model is slightly better at capturing extreme tail risk in the test data than the Skew-t model. Although the difference is not large, the lower failure rate of the GED distribution implies it may offer a more accurate prediction of extreme losses, making it potentially more suitable for risk forecasting in this context.

1.43 Third Part Conclusion

In this final chapter, we focus on forecasting conditional volatility using the APARCH(1,1) model with a skew-t distribution. Forecasting both conditional volatility and returns is important for several reasons:

Why Forecast Conditional Volatility and Returns?

Risk Management

- **Volatility Forecasting**: Conditional volatility, modeled effectively by GARCH, reflects the time-varying risk of asset prices. Forecasting volatility helps analysts estimate potential future price swings, enabling institutions to manage risks associated with investments, credit, and market changes.
- **Return Forecasting**: Knowing expected returns is crucial for informed investment choices, helping investors weigh potential gains against risks and guiding strategic allocation and risk-reward trade-offs.

Understanding Financial Behavior

- **Volatility Forecasting**: Conditional volatility models, such as GARCH, capture market behaviors like volatility clustering, where periods of high or low volatility tend to persist.
- **Return Forecasting**: Examining return patterns reveals insights into market dynamics, investor sentiment, and price change drivers, aiding in explaining and anticipating market movements.

Market Timing and Strategic Planning

- **Volatility Forecasting**: Predicting high- or low-volatility periods can guide investment timing. During high volatility, investors may reduce exposure to manage downside risk, while others may increase it to capture potential gains.
- **Return Forecasting**: Return forecasts inform strategic entry and exit decisions, helping investors time the market for optimal performance.

Procedure

1. **Initial Forecasts**: The APARCH model provides a one-step-ahead forecast of conditional volatility. However, we extend this to a 10-step-ahead forecast by generating an iterative forecast, which is useful for longer-term risk assessment.

2. **Return Forecasting**: To forecast returns, we first extract parameters from the APARCH(1,1) model with a skew-t distribution. Then, we simulate standardized returns from the skew-t distribution and scale these by the forecasted conditional volatility to obtain future return estimates.

3. **Risk Measure Calculation**: After forecasting, we calculate Value at Risk (VaR) and Expected Shortfall (ES), aiming to improve upon the Normal-distributed VaR, which may not capture extreme risks effectively.

4. **Model Training and Testing**: We split VGT returns into training (75%) and testing (25%) sets. Using the training set, we fit and forecast VaRs. The failure rates are calculated to determine model effectiveness: for the GED VaR, a failure rate of 6.84%; for skew-t VaR, 6.82%; and for Normal VaR, 7.85%.

5. **Expected Shortfall Performance**: In Expected Shortfall, GED shows a failure rate of 2.62%, slightly better than the skew-t rate of 2.67%.

Why Expected Shortfall (ES) is Better than Value at Risk (VaR)

Captures Tail Risk More Effectively

- **VaR Limitation**: VaR shows potential maximum loss at a specific confidence level (e.g., 95%) but offers no insight into losses beyond this point. For instance, if VaR at 99% is $1 million, it indicates a 1% chance of exceeding this loss but does not specify by how much.

- **ES Advantage**: Expected Shortfall, however, calculates the average loss in the worst scenarios beyond the VaR threshold, providing a fuller picture of extreme loss risk.

Coherent Risk Measure

- **VaR Limitation**: VaR lacks subadditivity, meaning a combined portfolio might show a higher risk than the sum of individual assets, sometimes discouraging diversification.

- **ES Advantage**: ES is a coherent measure, satisfying subadditivity, monotonicity, and positive homogeneity. This coherence aligns with intuition, as the risk of a diversified portfolio should not exceed the sum of individual risks.

Better Insight for Stress Testing and Scenario Analysis

- **VaR Limitation**: VaR only addresses a single percentile, often missing insight into rare but extreme moves.

- **ES Advantage**: ES provides information about the magnitude of losses in extreme events beyond VaR, making it more suitable for stress testing and scenario analysis.

Results: Testing VaR performance on the test dataset shows the GED VaR has a failure rate of 6.56%, while skew-t VaR is slightly higher at 7.03%. In Expected Shortfall, the GED model again shows a slightly lower failure rate (3.79%) than skew-t (4.03%). Although these differences are modest, the GED distribution's lower failure rate suggests it may offer more accurate risk forecasts for extreme losses, making it a potentially better fit for financial risk assessment. [1]

[1] data from 2004-02-01 to 2024-09-08 Code developed in Google Colab and reviewed with OpenAI This content was written with the assistance of Google Gemini and OpenAI YahooFinance

Discussion and Conclusion

In this comprehensive analysis of the VGT ETF and its investment profile, we began by examining its long-term performance compared to the S&P 500 benchmark, represented by a compound annual growth rate (CAGR) of 13.21% versus 7.87% for the S&P 500. VGT's significant outperformance—total returns of 1187.69% over two decades compared to the S&P 500's 376.40%—can largely be attributed to its concentrated holdings in high-performing tech stocks like Apple, Microsoft, and Nvidia. This sector-specific exposure, focusing on Electronic Technology and Technology Services, has played a pivotal role, given these sectors' consistent growth and strong fundamentals. Additionally, key efficiency metrics, particularly ROIC, highlight VGT's superior profitability compared to the broader S&P 500, underscoring the competitive advantage of high-ROIC companies. Metrics such as EPS and FCF per share further emphasize that VGT's performance is underpinned by strong fundamentals in its top holdings, reinforcing a pattern of sustainable growth supported by higher operating margins and revenue growth rates than the S&P 500.

Moving beyond a retrospective analysis, the second part of this research delved into the characteristics of VGT's returns, volatility, and distributional behaviors. Through algorithmic identification of the optimal ARMA and GARCH models, we found that VGT exhibits higher daily mean returns and volatility compared to the S&P 500. Although VGT's return series shows typical financial data patterns, including negative skewness and excess kurtosis, its risk-adjusted performance, as captured by the Sharpe ratio, indicates that VGT holds a promising outlook, particularly under projected lower interest rates. To model financial series behavior effectively, simulations demonstrated that a skew-t distribution best describes the extreme events typical of the S&P 500, while the GED offers a moderate level of risk with less complexity. The APARCH(1,1) model, paired with a skew-t distribution, was ultimately chosen for its ability to capture asymmetric effects, such as differential responses to positive and negative shocks, which are crucial for accurately modeling VGT's volatility dynamics.

In the final stage, we extended the analysis to forecasting conditional volatility and returns, essential for both risk management and strategic investment planning. Forecasting volatility provides insights into potential price swings, enabling institutions to manage investment risks, while return forecasts allow investors to optimize market timing and allocation. Using the APARCH(1,1) model with skew-t distribution, we performed a 10-step-ahead forecast for conditional volatility and calculated risk measures such as Value at Risk (VaR) and Expected Shortfall (ES). The forecast models revealed that GED VaR holds a slightly lower failure rate than the skew-t VaR, though both were notably better than the Normal VaR. Importantly, ES demonstrated superior risk management potential by capturing tail risk beyond the VaR threshold, providing a more coherent and robust measure for stress testing and extreme scenarios.

In conclusion, our findings affirm that VGT's long-term outperformance is strongly rooted in the quality and resilience of its top holdings, especially in high-ROIC tech companies. The APARCH(1,1) model with a skew-t distribution proves an effective tool for capturing VGT's volatility dynamics, offering a nuanced approach to understanding and managing risk in tech-focused portfolios. Looking forward, the combination of robust fundamentals, strategic investments in AI, and favorable macroeconomic conditions suggests that VGT may continue to be well-positioned for sustained growth, particularly as the sector embraces a new growth phase backed by technological advancements. This holistic analysis thus underscores the value of sector-specific ETFs like VGT, offering enhanced risk-adjusted returns in alignment with evolving market dynamics. [1]

References

References

[1] Curvo. *Backtest by Curvo*. Available at: https://curvo.eu/backtest/

[2] FactSet. *Earnings Insight* by John Butters, VP, Senior Earnings Analyst. October 4, 2024. Available at: https://advantage.factset.com/hubfs/Website/Resources

[3] Economatica, a TC Company. *Value Reports - ROIC: SP 500 Performance*. Available at: https://valuereports.economatica.com/roic-sp-500-performance/

[4] FactSet. *Are the "Magnificent 7" the Top Contributors to Earnings Growth for the SP 500 for Q1?* by John Butters. April 22, 2024. Available at: https://insight.factset.com/are-the-magnificent-7-the-top-contributors-to-earnings-growth-for-the-sp-500-for-q1

[5] Polak, David. *Magnificent Seven: What do you need to believe?* Capital Ideas, January 9, 2024. Available at: https://www.capitalgroup.com/institutional/insights/articles/magnificent-seven-chart-diversify.html

[6] Goldman Sachs. *Global Investment Research*.

[7] MSCI. *MSCI US IMI Information Technology 25/50 Index (USD)*. Available at: https://msci.com

[8] Kennedy Capital Management. *ROIC – The Underappreciated Variable in Valuation*.

[9] CME Group. *FedWatch*.

[10] CNBC. *U.S. 10 Year Treasury*, September 21, 2024.

[11] Krauskopf, Lewis. *Echoes of Dotcom Bubble Haunt AI-Driven US Stock Market*. Reuters, July 2, 2024. Available at: https://www.reuters.com/markets/echoes-dotcom-bubble-haunt-ai-driven-us-stock-market-2024-07-02/

[12] LSEG Datastream.

[13] Lu, Marcus. *SP 500 vs. SP 500 Equal Weight Index*. Visual Capitalist. Available at: https://www.visualcapitalist.com/charted-sp-500-vs-sp-500-equal-weight-index/

[14] Lu, Marcus. *How the Top SP 500 Companies Have Changed Over Time*. Visual Capitalist. Available at: https://www.visualcapitalist.com/how-the-top-sp-500-companies-have-changed-over-time/

[1] Code developed in Google Colab and reviewed with OpenAI This content was written with the assistance of Google Gemini and OpenAI

[15] Bravo Research. *The Tech Sector: A Comparative Analysis of Valuations from the Dot Com Era to Today*. Available at: https://bravosresearch.com/blog/the-tech-sector-a-comparative-analysis-of-valuations-from-the-dot-com-era-to-today/

[16] Rao, Pallavi. *AIRanked: Tech Manufacturers by RD Investment Change in 2023*. Visual Capitalist. Available at: https://www.visualcapitalist.com/charted-tech-companies-r-d-change-2023/

[17] Zhu, Kayla. *Ranked: The Most Popular Generative AI Tools in 2024*. Visual Capitalist. Available at: https://www.visualcapitalist.com/ranked-the-most-popular-generative-ai-tools-in-2024/

[18] Jensen, Greg, Moriarty, Josh. *Are We on the Brink of an AI Investment Arms Race?* BridgeWater, May 20, 2024. Available at: https://www.bridgewater.com/research-and-insights/are-we-on-the-brink-of-an-ai-investment-arms-race

[19] Jensen, Greg, Narayan, Atul, Greene, Alex, Simon, Lauren. *Is an AI Bubble Ahead of Us or Behind Us?* BridgeWater, September 10, 2024. Available at: https://www.bridgewater.com/research-and-insights/is-an-ai-bubble-ahead-of-us-or-behind-us

[20] *Sectors of US Economy as Percent of GDP 1947-2009*. Wikipedia.

[21] Lu, Marcus. *US GDP by Industry 2023*. Visual Capitalist. Available at: https://www.visualcapitalist.com/visualizing-u-s-gdp-by-industry-in-2023/

[22] VettaFi. Available at: https://etfdb.com/etf/VGT/holdings

[23] Vanguard. Available at: https://investor.vanguard.com/investment-products/etfs/profile/vgt

[24] Yahoo Finance. Available at: https://finance.yahoo.com/quote/VGT/?guccounter=1guce$_r eferrer = aHR0cHM6Ly9maW5hbmNlLnlhaG9vLmNvbS8guce_r eferrer_s ig = AQAAAAnXvzMrw9XRKvpSUz5FGNRR53058n4h7GWughwgnR8nffRIsrPS35lfiPD7irltdPNwBwbup2R TCK-2OVjnmeAe7RKpvZ6CSucwBdx5PdP3BdHPnD0oRAadZ2Fxfw6UZdQP0fJ3HEEF$

[25] *Discounted Cash Flow*. Available at: https://www.discountedcashflow.com

[26] *ETF.com*. Available at: https://www.etf.com

[27] *Gurufocus*. Available at: https://www.gurufocus.com

[28] Lu, Marcus. *SP 500 Earnings Growth Forecasts vs. Big Tech*. Visual Capitalist. Available at: https://www.visualcapitalist.com/sp-500-earnings-growth-forecasts-vs-big-tech/

1.44 Disclaimer

Code developed in Google Colab and reviewed with OpenAI

This content was written with the assistance of Google Gemini and OpenAI

Data obtained from Yahoo Finance, Discounting Cash Flow, and Tickernomics; processed in Google Colab and Excel; and reviewed by OpenAI