Financial Risk Analysis in Stock Markets: A Data Lifecycle Approach

Fatma Braham, Apolline Hadjal & Elise Deyris

Due on 3rd December, 2025

https://github.com/FatmaZBh/Financial-risks-in-stock-markets

Introduction

This research uses a complete data lifecycle management approach to analyze financial risks in stock market operations. The main objective of this project is to **identify which stocks are the riskiest and to determine whether there are significant correlations between stocks across different sectors**. Through this dual analysis of risk and inter-sector correlation, we aim to better understand market behavior and diversification opportunities.

This report presents the data collection process, the key financial concepts related to the topic, and the methodology used to conduct the analysis. It also discusses the results obtained, as well as the limitations of the study and the difficulties encountered throughout the project.

Dataset

Data Source

The dataset was originally collected from Yahoo Finance, an open financial data provider. We chose to work with daily granularity, meaning we have one value per trading day for five years starting from January 1st, 2019 to the day of running the code. We selected five companies from five different economic sectors in order to ensure sectoral diversity in our analysis. Apple was selected to represent the Technology sector, Amazon the Consumer sector, JP Morgan the Finance sector, Johnson & Johnson the Healthcare sector, and TotalEnergies the Energy sector. This allowed us to compare risk and correlation patterns across sectors and to evaluate the potential benefits of diversification

Use of Messy Data

We initially planned to use yfinance, a built-in Python library that imports historical stock data directly from Yahoo Finance. However, these data were too clean and well-structured for a realistic project. Therefore, we downloaded messy datasets containing missing values, duplicates, incorrect data types, and other imperfections that better reflect real-world data challenges. This approach allowed us to demonstrate essential data cleaning and preprocessing skills necessary in practical financial analysis.

Key Financial Concepts and Definitions

Definitions of Terms

Volatility and Relation to Risk

Volatility measures the degree of variation of a stock's returns over time. It reflects how strongly the price of a stock fluctuates around its average value. A high volatility indicates that the stock price changes rapidly and unpredictably, which implies a higher level of risk, while a low volatility indicates more stable price movements and therefore lower risk.

Diversification

Diversification is an investment strategy that consists of spreading investments across different assets or sectors in order to reduce overall portfolio risk. The idea is that losses in one asset can be offset by gains in another. Diversification is most effective when the selected assets are weakly correlated or negatively correlated, meaning they do not move in the same direction at the same time.

Financial Formulas

Simple Returns

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}}$$

Used to measure daily percentage change in price.

Log Returns

$$r_t = \ln\left(\frac{P_t}{P_{t-1}}\right)$$

Log returns are preferred because:

- They are time-additive.
- They are more symmetric and statistically stable.
- They reduce the impact of extreme values.

Value at Risk (VaR)

VaR measures the maximum expected loss over a given time horizon for a given confidence level.

$$\text{VaR}_{95\%} = \text{5th percentile of returns}$$

$$\text{VaR}_{99\%} = \text{1st percentile of returns}$$

Interpretation:

- With 95% confidence, the stock will not lose more than VaR_{95} in one day.
- A more negative VaR implies higher downside risk.

Maximum Drawdown

$$\text{Drawdown}_t = \frac{P_t - \max(P)}{\max(P)}$$

Maximum drawdown measures the worst peak-to-trough loss during the period. It reflects the most severe crash an investor would have experienced.

Sharpe Ratio

$$S = \frac{R_p - R_f}{\sigma_p}$$

It measures return per unit of risk. A higher Sharpe ratio means better risk-adjusted performance. Note that while this metric is conceptually important, it was not computed in our analysis.

Diversification

Diversification consists of investing across different sectors to reduce portfolio risk. If sectors are weakly correlated, diversification is effective. If correlations are high, diversification loses its power.

Correlation

$$\rho = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

- $\rho \approx 1 \rightarrow$ Assets move together
- $\rho \approx 0 \rightarrow$ Assets move independently
- $\rho \approx -1 \rightarrow$ Assets move oppositely

Results and Insights

Riskiest Stocks by Volatility

Our analysis of annualized volatility across the five sectors revealed significant disparities in risk profiles. Table 1 presents the top three stocks ranked by volatility.

Table 1: Top Three Stocks by Annualized Volatility

Rank	Stock	Sector	Annualized Volatility
1	TotalEnergies	Energy	45.2%
2	JP Morgan	Finance	38.7%
3	Amazon	Consumer	36.4%

TotalEnergies emerged as the most volatile stock in our dataset, exhibiting substantial price fluctuations throughout the analysis period. The prominence of Energy and Finance sectors in the top rankings suggests these industries are particularly susceptible to market turbulence, likely due to their sensitivity to macroeconomic factors such as commodity prices, interest rate changes, and regulatory developments.

Largest Drawdowns

Maximum drawdown analysis provided further insight into downside risk during market stress periods. TotalEnergies not only demonstrated the highest volatility but also experienced the largest peak-to-trough decline, reinforcing its position as the riskiest asset in our portfolio. Consumer and Finance sector stocks also exhibited substantial drawdowns during crisis periods, reflecting their cyclical nature and exposure to economic downturns. In contrast, Healthcare stocks demonstrated more defensive characteristics, with markedly smaller maximum drawdowns, consistent with the sector's reputation for stability during market turbulence.

Sector Risk Ranking

Based on our comprehensive risk assessment incorporating volatility, drawdown analysis, and value-at-risk calculations, we established the following sector risk hierarchy:

1. **Energy** – Highest risk profile, characterized by extreme volatility and severe drawdowns
2. **Finance** – High risk, with significant sensitivity to macroeconomic conditions
3. **Consumer** – Moderate-high risk, demonstrating cyclical behavior

4. **Technology** – Moderate risk, balancing growth potential with volatility
5. **Healthcare** – Lowest risk, offering defensive characteristics

This ranking provides valuable guidance for portfolio construction and risk management decisions.

Correlation and Diversification Insights

Correlation analysis revealed important relationships for diversification strategies. Technology and Consumer sectors exhibited strong positive correlation ($\rho = 0.72$), suggesting limited diversification benefits when holding both sectors simultaneously. This relationship likely stems from shared exposure to consumer spending patterns and economic growth drivers.

Healthcare demonstrated notably lower correlations with other sectors, with an average cross-sector correlation of just 0.35, establishing it as an effective diversification tool within a multi-sector portfolio. The sector's defensive nature and independence from cyclical economic forces make it particularly valuable for risk reduction.

Energy sector stocks displayed more independent behavior relative to other sectors, though this independence came at the cost of elevated volatility. While Energy can provide diversification benefits, investors must carefully weigh these against the sector's substantial standalone risk.

Implications

Based on all of these results, we can trace guidelines for investors. When looking to invest, individuals should carefully consider sector allocation when constructing portfolios. Those seeking stability should prioritize Healthcare stocks, which demonstrated defensive characteristics and low correlation with other sectors, making them ideal for risk mitigation. On the contrary, Energy sector investments require caution due to their extreme volatility and severe drawdowns, and should only constitute a small portion of well-diversified portfolios. Investors should avoid overconcentration in Technology and Consumer sectors simultaneously, given their strong positive correlation, as this provides limited diversification benefits. For balanced portfolios, combining Healthcare's defensive properties with moderate exposure to Technology and Finance can optimize the risk-return tradeoff while maintaining adequate diversification.

Limitations

However, there are limits to our code, as it does not take into account some factors. For instance, we only analyzed five companies, so we cannot conclude whether our findings generalize to the broader sectors they represent. We also ignored transaction costs such as taxes and brokerage fees, even though they can have a significant impact on real investment performance. In addition, we did not include macroeconomic variables such as GDP growth or inflation, which, like period crises and Covid-19, could influence stock returns and correlations. Finally, the Value at Risk (VaR) was computed only using the historical method, whereas other approaches such as parametric or Monte Carlo methods could provide complementary insights.

Difficulties Encountered

Data Quality and Cleaning

Initially, we extracted data from Yahoo Finance, which provided surprisingly clean datasets. However, real-world financial projects rarely encounter such pristine data. To better simulate practical conditions,

we deliberately switched to messier datasets with missing values, duplicates, and incorrect data types that required substantial cleaning and demonstrated essential data wrangling techniques.

Selection of Risk Indicators

The field of financial risk analysis offers an extensive array of metrics, including Beta, Sortino ratio, Sharpe ratio, and CAPM-derived measures. We focused on three core measures: annualized volatility, Value-at-Risk (VaR), and maximum drawdown, selected for their intuitive economic interpretation and widespread acceptance in both academic and industry practice.

Visualization Strategy

With multiple stocks, sectors, and risk metrics, we faced the potential for producing dozens of visualizations. However, excessive graphs risk overwhelming readers and obscuring key insights. We limited our visual presentation to three essential chart types: volatility comparison plots, maximum drawdown bar charts, and sector correlation heatmaps.

Financial Domain Knowledge

As students without extensive financial backgrounds, we invested considerable time developing the necessary theoretical foundation. Key concepts requiring substantial study included the proper interpretation of Value-at-Risk, the rationale for using log-returns rather than simple returns, drawdown calculation mechanics, and the economic significance of volatility. This learning curve proved essential for conducting meaningful analysis and avoiding common pitfalls in financial risk assessment.

What we learned

This project helpes us gained a more gloabl understanding of metrics and concepts of financial analysis. It demonstrated that effective financial analysis requires multiple complementary metrics. For instance, volatility alone is not sufficient to measure financial risk, we also need to consider drawdown, and correlation. Financial indicators must be carefully chosen and well interpreted in order to generate useful investment insights. This project also showed us the importance of cleaning the dataset.

Conclusion

This study applied a comprehensive data lifecycle approach to assess financial risk across five major economic sectors. Our analysis identified Energy and Finance as the highest-risk sectors, with TotalEnergies exhibiting the greatest volatility and maximum drawdown, while Healthcare demonstrated defensive characteristics ideal for portfolio stabilization. The correlation analysis revealed that Healthcare's low correlation with other sectors makes it an effective diversification tool, whereas the strong positive correlation between Technology and Consumer sectors limits diversification benefits.