An analysis of Banking Sector Stocks using Yahoo Finance (yfinance) in Python

Kauntey Acharya

December 11, 2025

This report provides an explanation of banking stock analysis using Python and the `yfinance` library. Each computational step is supported with financial theory, mathematical formulations, and interpretation of why the technique is used in real-world quantitative finance. This includes data normalization, return calculation, risk measures, volatility modelling, monthly resampling theory, and correlation analysis. Visualisation techniques are theoretically justified to show how they aid interpretation of high-dimensional financial time series.

1 Introduction

Financial markets produce time-series data that contain patterns of investor sentiment, macroeconomic conditions, and institutional behaviour. Stock price analysis is therefore rooted in extracting meaningful statistical patterns from noisy market data. The Python file `yfinance_banking_stock_analysis.ipynb` automates this process by collecting data directly from Yahoo Finance and performing a structured evaluation.

The banking sector is particularly sensitive to monetary policy, interest rate cycles, liquidity conditions, and economic growth. Thus, analysing banking stocks provides insight into financial dynamics.

The purpose of this report is to describe, in theoretical detail, every mathematical and computational step used in the notebook, equipping the reader with both conceptual understanding and applied skills.

2 Financial Market Data Structure

Stock market data consist of OHLCV attributes, each carrying unique information about market microstructure:

- **Open** represents the clearing price when the market opens. It reflects the absorption of overnight news.
- **High** indicates intraday upward volatility and price extremes.
- **Low** signals downside risk and bearish price pressure during the session.
- **Close** is the most widely used value because institutional trading typically clusters near the end of the day, making it a stable representation of final market consensus.

- **Adjusted Close** corrects for dividends, splits, and corporate actions. It is essential for accurate long-term return analysis because it ensures that changes in price reflect true economic gain.
- **Volume** measures trading activity. Higher volume indicates greater liquidity and reliability of price signals.

Financial models generally rely on closing prices because intraday volatility patterns introduce noise that complicates return calculations. Using close prices standardises analysis.

3 Data Normalization

Normalization is a transformation that allows two or more price series—often with vastly different magnitudes—to be compared directly. Raw price levels cannot be compared because:

- A 300 stock is not necessarily "more valuable" than a 50 stock.
- Growth rates matter more than absolute price.

Normalization is defined as:

$$P_t^{\text{norm}} = \frac{P_t}{P_0}.$$

This rescales every stock to start at 1. The theoretical interpretation is that normalization simulates an investor who invests 1 in each stock on day 0. The curves therefore represent growth trajectories, making differences in long-term performance visually meaningful.

Normalisation is essential for:

- Relative performance comparison
- Visual detection of outperformers and underperformers
- Portfolio backtesting

Without normalization, high-priced stocks would dominate graphs unfairly.

4 Return Calculations

Returns quantify how much a stock's value changes relative to its previous value. This is fundamental because investors care about percentage gain, not raw price difference.

4.1 Simple Daily Returns

The simple return formula:

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

measures proportional change. In financial theory, simple returns represent discrete compounding — the reality of most investments.

They are intuitive and easy to interpret, but not ideal for long-term mathematical modelling due to non-additivity.

4.2 Logarithmic Returns

Log returns are defined as:

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

They are widely used because:

- They are time-additive:
$$r_{1\to T} = \sum_{t=1}^{T} r_t.$$
- They assume continuously compounding returns, a natural assumption in stochastic calculus.
- They approximate simple returns when price changes are small.
- They align with the lognormal distribution assumption in the Black–Scholes model.

Log returns are foundational in quantitative modelling, risk analysis, and machine learning on financial data.

4.3 Cumulative Returns

Cumulative returns measure total growth:

$$CR_t = \prod_{i=1}^{t}(1 + R_i) - 1.$$

This assumes reinvestment of gains — a realistic assumption for portfolio growth. Cumulative return graphs show wealth trajectories and long-term performance divergence.

5 Monthly Resampling

Financial data often contains microstructure noise on a daily scale. Monthly resampling allows analysts to reveal macro trends.

The monthly close is:

$$P_m^{\text{monthly}} = P_{\text{last trading day of month}}^{\text{close}}.$$

Reasons for monthly analysis:

- Institutions evaluate portfolios monthly.
- Macroeconomic reports (GDP, inflation, policy rates) operate monthly.
- Smoother time-series reduces noise and highlights structural patterns.

Monthly returns:

$$R_m^{\text{monthly}} = \frac{P_m - P_{m-1}}{P_{m-1}}.$$

Absolute monthly price change:

$$\Delta P_m = P_m - P_{m-1}.$$

These metrics help in studying:

- Market cycles
- Policy impacts
- Sector performance trends

Monthly data is especially useful in long-term investment strategies and strategic asset allocation.

6 Volatility and Risk Metrics

Volatility is the central quantity in risk measurement. It represents uncertainty in future returns, not loss itself.

6.1 Daily Volatility

Defined as the standard deviation of daily returns:

$$\sigma_{\text{daily}} = \sqrt{\frac{1}{N-1}\sum_{t=1}^{N}(R_t - \bar{R})^2}.$$

From a theoretical perspective:

- Volatility is a proxy for risk in Modern Portfolio Theory (Markowitz).
- Higher volatility means greater unpredictability.
- Many derivative pricing models (e.g., Black–Scholes) use volatility as the key input.

Volatility is not inherently bad — high-volatility stocks may offer higher returns.

6.2 Annualized Volatility

To convert daily fluctuation into yearly risk:

$$\sigma_{\text{annual}} = \sigma_{\text{daily}}\sqrt{252}.$$

This assumes returns are independent and identically distributed — a simplification but widely used in practice.

6.3 Sharpe Ratio

Sharpe ratio measures return per unit of risk:

$$S = \frac{\bar{R} - R_f}{\sigma_{\text{annual}}}.$$

The theory behind Sharpe ratio states:

- A higher Sharpe ratio means the asset compensates better for risk.
- A negative Sharpe ratio means risk-free assets outperform the stock.

It is the dominant metric for comparing investment strategies.

6.4 Maximum Drawdown

Drawdown measures how much wealth falls from peak before recovery:

$$DD_t = \frac{P_t - \max(P_1, ..., P_t)}{\max(P_1, ..., P_t)}.$$

Maximum drawdown is the worst historical loss:

$$\text{MDD} = \min(DD_t).$$

Theoretically important because:

- Investors strongly dislike large losses (asymmetric risk aversion).
- Drawdown reflects market crashes better than volatility.
- Many hedge funds use drawdown limits in risk management.

7 Correlation Analysis

Correlation measures co-movement between assets:

$$\rho_{ij} = \frac{\text{Cov}(R_i, R_j)}{\sigma_i \sigma_j}.$$

In theory:

- High correlation indicates systemic sector behaviour.

- Low correlation enables diversification (Markowitz portfolio optimisation).
- Banking stocks tend to move together due to policy-driven dynamics.

A correlation heatmap visually reveals clusters of similar behaviour and risk linkages in the sector.

8 Visualization Techniques

Visualization transforms complex financial data into interpretable forms. Each visual tool has theoretical justification:

- **Line charts of closing prices** reveal long-term trends, momentum, and growth.
- **Normalized performance plots** show relative outperformers — essential for portfolio allocation.
- **Histograms of daily returns** approximate return distributions, useful for checking normality assumptions.
- **Rolling volatility plots** reveal periods of market stress, uncertainty, or stability.
- **Correlation heatmaps** help identify diversification opportunities and systemic sector risks.
- **Monthly bar charts** display cyclical patterns and policy-driven shocks.

These visualisations combine statistical insight with intuitive financial interpretation, making them indispensable for quantitative analysis.

9 Python Workflow Summary

The Python notebook follows a logical structure consistent with financial data science pipelines:

1. Fetch data using `yfinance`.
2. Extract relevant columns (primarily closing prices).
3. Normalize price series for comparison.
4. Compute simple and log returns.
5. Resample stock prices to monthly frequency.
6. Compute risk and volatility metrics.
7. Generate all required plots.
8. Summarize outcomes.

Each step mirrors standard procedures in academic finance, quantitative research, and professional asset management.

10 Outcomes of the Analysis

The theoretical and computational findings indicate that:

- Banking stocks exhibit significant correlation due to regulatory and monetary policy influences.
- Volatility patterns identify which banks are riskier to hold.
- Monthly performance reveals macroeconomic cycles.
- Higher cumulative returns highlight long-term outperformers.
- Drawdowns expose vulnerabilities during market stress.

These outcomes align with empirical market behaviour and support deeper quantitative modelling.

11 Conclusion

This report presents an expanded theoretical foundation behind each computation in the Python file. The methodology, combined with real market data, provides a strong basis for empirical analysis, portfolio design, and further research in financial data science.