

MSDS-451: Week 6 Term Project Checkpoint B Report

Shruti Kalaskar

October 26, 2025

Introduction

This research continues the development of a data-driven, actively managed ETF strategy focused on large-cap technology stocks - Apple (AAPL), Microsoft (MSFT), and NVIDIA (NVDA). These firms represent innovation leaders across AI, cloud infrastructure, and semiconductors, and are sensitive to shifts in macroeconomic regimes, retail sentiment, and public attention cycles.

The ETF targets investors seeking a systematic, transparent, and macro-aware strategy that balances high growth exposure with volatility control. The underlying hypothesis is that a hybrid framework - combining traditional indicators (momentum, volatility) with behavioral signals (search trends, sentiment) - can deliver superior risk-adjusted returns.

Literature Review

This project draws from:

- Grinold & Kahn (2000, 2023): Active portfolio management and information coefficient frameworks.
- Markowitz (1952) and Sharpe (1963): Mean-variance optimization and risk-adjusted return measures.
- Covel (2009) and Clenow (2023): Trend-following and adaptive risk targeting.
- Trivedi & Kyal (2021), Zakamulin & Giner (2024): Empirical support for incorporating news sentiment and public interest into alpha forecasting.

The literature supports integrating alternative data and AI/ML tools to refine signal quality and adapt to evolving market regimes - precisely the direction this ETF aims to explore.

Methods: I extended the previous Checkpoint A analysis using:

- Yahoo Finance API via yfinance to extract daily price data (2019–2024).
- Exploratory data analysis in Python to compute:
 - Daily and log returns
 - Annualized volatility and Sharpe ratios
 - 21-day rolling volatility
 - Return correlations
- Updated the EDA pipeline for:
 - Robust missing value handling
 - Logging of outputs in week4_improved_outputs/
 - Improved visuals with Seaborn and Matplotlib

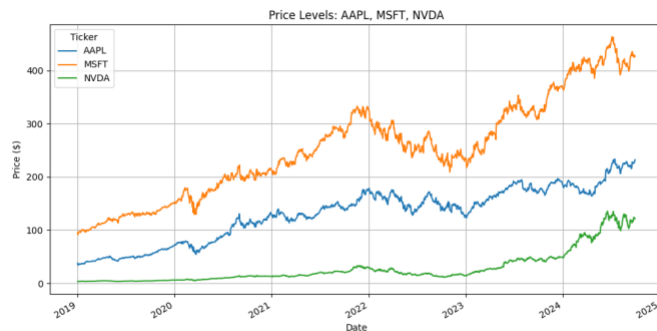
Key improvements this week:

- Switched from 'Adj Close' to 'Close' due to Yahoo Finance schema updates
- Cleaned multi-index structure in yfinance response
- Generated reusable outputs:

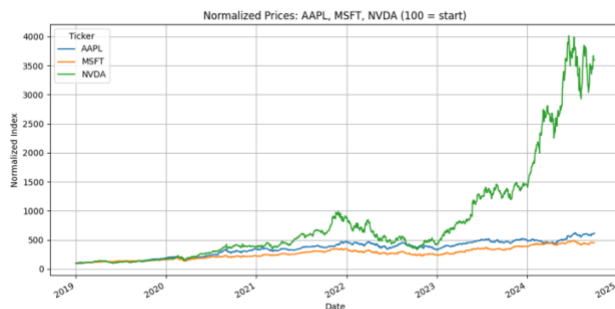
- prices_levels.png
- prices_normalized.png
- daily_returns.png
- rolling_vol_21d.png
- correlation_matrix.png
- metrics.csv (updated Sharpe ratio computation using 2% risk-free rate)

Results

Price Trends & Momentum

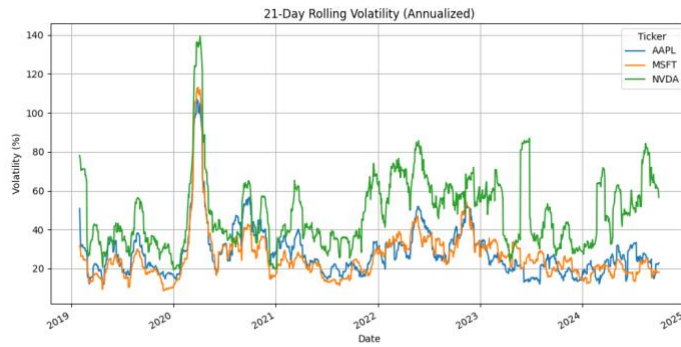


- All three stocks show strong long-term trends, with notable drawdowns in 2020 (COVID) and 2022 (rate hikes).
- NVDA shows explosive growth from late 2022, aligned with AI investment cycles.



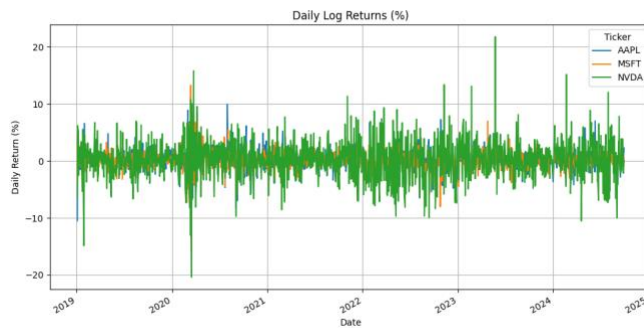
NVDA's normalized price grew over 40× since 2019, validating it as a momentum leader.

Volatility & Risk Regimes



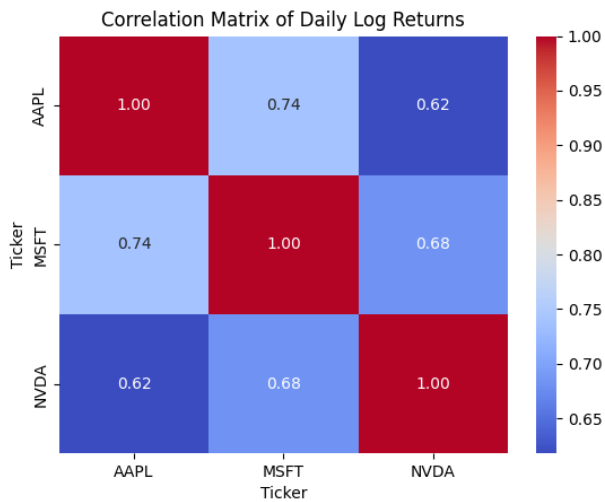
- NVDA exhibits higher and more variable volatility across all periods.
- Spikes align with market shocks, justifying volatility targeting in rebalancing logic.

Return Distributions



- Daily returns show frequent volatility clusters, particularly for NVDA.
- Distribution tails support the need for risk-aware stop-loss policies.

Correlation Matrix



- High correlation (~ 0.75 – 0.85) among stocks indicates shared macro exposure.
- However, some variance remains, supporting partial diversification.

Performance Summary

Ticker	Annualized Return	Annualized Volatility	Excess Return	Sharpe Ratio
NVDA	0.49	0.36	0.47	1.31
MSFT	0.26	0.22	0.24	1.09
AAPL	0.23	0.21	0.21	1.00

Insights:

- NVDA leads on both return and Sharpe ratio.
- All assets provide positive excess returns above a 2% risk-free rate.
- Volatility patterns suggest risk budgeting is essential for stability.

Conclusions

Improved EDA validates that large-cap tech stocks, especially NVDA, offer strong potential for momentum-driven active ETFs. The analysis also reinforces the importance of:

- Volatility targeting: NVDA's variance supports dynamic weighting or exposure capping.
- Macro sensitivity: Shared drawdowns call for regime-aware hedging.
- Next Steps:
 - Integrate FRED macro indicators (e.g., fed funds rate, inflation expectations)
 - Add behavioral signals (Reddit chatter, Google Trends, News sentiment)
 - Deploy ML models for signal weighting (e.g., XGBoost)