451 Feature Engineering: Week 4 Term Project Checkpoint A: Research Report

Shruti Kalaskar

October 12, 2025

Prospectus & Research Plan for an Actively Managed Tech–Macro ETF

Introduction

This research aims to design an **actively managed exchange-traded fund (ETF)** that focuses on large-cap technology companies, **Apple (AAPL)**, **Microsoft (MSFT)**, and **NVIDIA (NVDA)**, whose valuations are influenced by global innovation cycles, monetary policy, and investor sentiment toward artificial intelligence and cloud infrastructure.

The goal is to create a **data-driven, rule-based investment model** that dynamically rebalances positions based on evolving market signals. The ETF will serve investors seeking systematic exposure to innovation leaders, while maintaining risk control through macroeconomic awareness.

Drawing on **Grinold and Kahn's (2000)** principle of "superior information," this research integrates market, macroeconomic, and sentiment indicators to enhance predictive performance. The project bridges **quantitative finance and AI-driven analytics** to define a transparent, automated investment process.

Literature Review

Active portfolio management theory, developed by **Grinold and Kahn (2000, 2023)**, defines alpha as a product of information quality and forecast breadth. This philosophy supports the use of multiple, independent signals, such as momentum, quality, and sentiment, to improve predictive accuracy.

The foundational concepts of **mean–variance optimization (Markowitz, 1952, 1956)** and **risk–return trade-offs (Sharpe, 1963, 1994)** guide this ETF's portfolio construction and performance metrics. Extensions by **Hudson and Thames (2020)** provide systematic approaches to algorithmic execution.

Trend-following and risk-parity frameworks (**Covel, 2009; Greyserman & Kaminski, 2014; Clenow, 2023**) inform dynamic volatility targeting and adaptive rebalancing. Meanwhile, **Graham (2006)** and **Siegel (2022)** emphasize disciplined, long-term compounding—principles relevant for sustainable ETF design.

Recent literature highlights the predictive utility of **alternative data**. Studies by **Trivedi & Kyal (2021)** and **Zakamulin & Giner (2024)** demonstrate that integrating news sentiment and public attention (e.g., Google Trends) improves short-term forecasts. These findings motivate combining traditional financial indicators with real-time, behavior-driven features for this ETF.

Methods

Data Sources and APIs

Type	**Source**	**Access**	**Example Code**
Market Data	Yahoo Finance	yfinance API	yf.download(['AAPL','MSFT','NVDA'], start='2019-01-01', end='2024-12-31')
Macroeconomics	FRED (St. Louis Fed)	fredapi	fred.get_series('FEDFUNDS')
News Sentiment	NewsAPI / Google News RSS	REST API	Python requests

Retail Sentiment	Reddit & StockTwits APIs	OAuth	praw / HTTP queries
Public Interest	Google Trends (pytrends)	Python API	pt.build_payload(['AI chips','cloud computing'])

All data collection and preprocessing scripts are maintained in a public GitHub repository to ensure reproducibility and transparency.

Signal Design

1. **Momentum:** 12-month trailing returns (1-month skip)
2. **Quality:** Profitability proxies such as gross margin and ROIC
3. **Macro Regime:** Binary indicator for tightening vs. easing cycles (from FRED data)
4. **Sentiment:** Headline polarity and Reddit discussion frequency
5. **Public Attention:** Google Trends search intensity for "AI chips" and "cloud computing"

Signals will be standardized and combined through weighted averages or machine learning models (e.g. gradient boosting) to estimate expected returns (μ_{it}). Portfolio optimization will follow the **mean–variance** framework.

Trading and Risk Management

- **Rebalancing Frequency:** Weekly, with volatility targeting (12–15 % annualized)
- **Hedging:** Dynamic QQQ short overlay or cash position during risk-off regimes

- **Stop-Loss Policy:** −8 % hard stop per position or 10 % trailing stop
- **Evaluation Metrics:** CAGR, annualized volatility, Sharpe ratio, Sortino ratio, and maximum drawdown from 2019–2024 backtests

Results

Exploratory Data Analysis (EDA) was conducted on historical prices for AAPL, MSFT, and NVDA from January 2019 to December 2024 using the **Yahoo Finance API**.

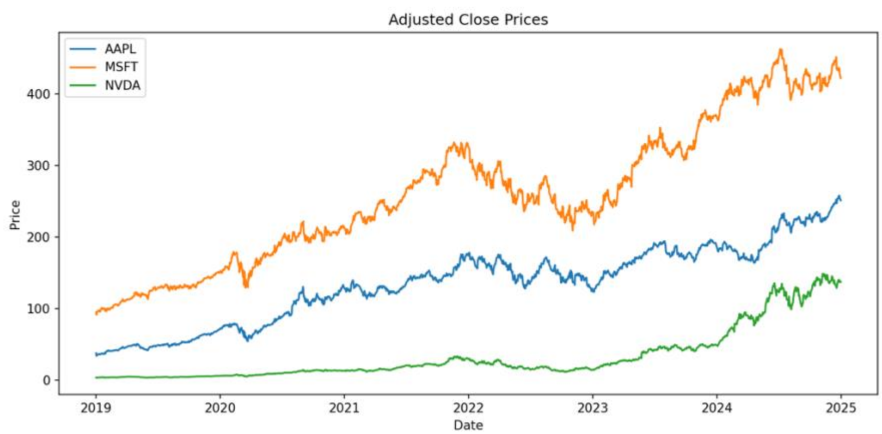

Figure 1. Adjusted Close Prices (2019–2024)

The **Adjusted Close Prices** (Figure 1) show consistent upward trends across all three stocks, punctuated by drawdowns during early 2020 (COVID-19) and mid-2022 (rate-hike volatility). **Normalized Price Indices** (Figure 2) reveal that NVDA outperformed dramatically, increasing over 40× since 2019, demonstrating strong momentum and sector leadership in AI hardware.

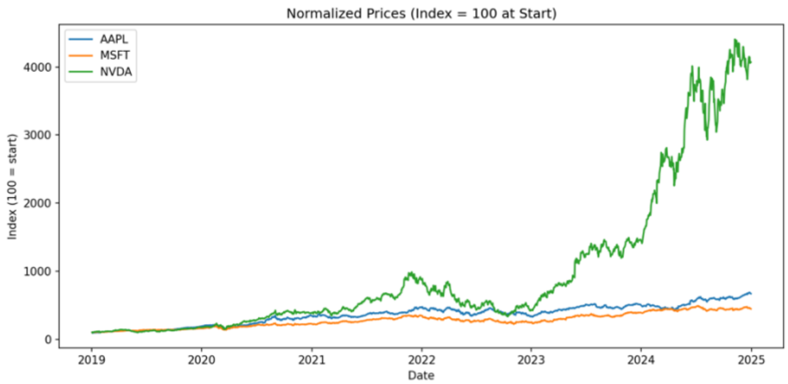

Figure 2. Normalized Prices (Index = 100 at Start)

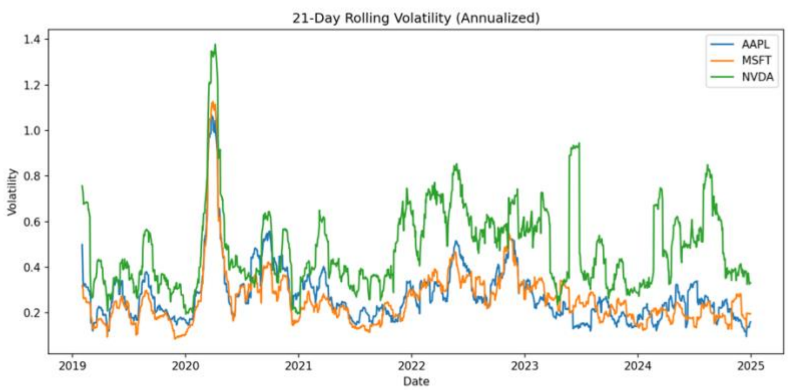

Figure 3. 21-Day Rolling Volatility (Annualized)

Daily Returns analysis confirms mean reversion around zero with frequent volatility clustering, while the **21-Day Rolling Volatility** chart (Figure 3) highlights that NVDA experiences higher and more variable volatility compared to AAPL and MSFT. The **Correlation Matrix of Daily Returns** (Figure 4) indicates moderately high correlations (≈0.65–0.85), suggesting exposure to common tech-sector risks but allowing partial diversification.

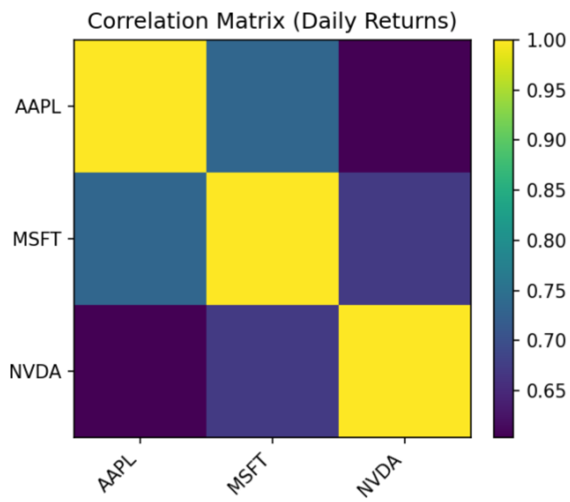

Figure 4. Correlation Matrix of Daily Returns

Table 1 summarizes annualized performance metrics estimated over this period. NVDA achieved the highest return and risk-adjusted performance, consistent with its strong market trajectory during the AI-driven cycle of 2023–2024.

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

Table 1. Estimated performance metrics for AAPL, MSFT, and NVDA (2019–2024), assuming a 2% risk-free rate.

These results validate the suitability of large-cap technology stocks for a momentum-plus-macro ETF strategy. The strong co-movement between AAPL, MSFT, and NVDA implies shared exposure to macroeconomic conditions, while NVDA's excess returns provide scope for alpha generation. Future stages will incorporate macroeconomic indicators and sentiment features to enhance signal robustness and improve drawdown management.

Conclusions

The exploratory findings demonstrate that AAPL, MSFT, and NVDA display **sustained growth, persistent momentum, and distinct volatility patterns**, aligning well with a systematic active-management approach. NVDA's dominant upward trajectory underscores the role of innovation cycles in driving excess returns.

The observed correlations and volatility patterns confirm that diversification within the large-cap tech sector remains feasible but limited, supporting the need for macro-aware hedging strategies.

Future work will extend this research to incorporate **FRED macroeconomic variables**, **sentiment indicators**, and **Google Trends attention metrics** into the ETF's decision rules. These enhancements aim to refine the ETF's risk-adjusted performance and ensure adaptability across market regimes.

References

- Grinold, R. C., & Kahn, R. N. (2000, 2023). *Active Portfolio Management.* McGraw-Hill.
- Markowitz, H. (1952, 1956). *Portfolio Selection. Journal of Finance.*
- Sharpe, W. F. (1963, 1994). *Capital Asset Pricing Model and Risk–Return Trade-Off.*
- Graham, B. (2006). *The Intelligent Investor.*
- Siegel, J. (2022). *Stocks for the Long Run.*
- Covel, M. (2009, 2011, 2017). *Trend Following.*
- Greyserman, A., & Kaminski, K. (2014). *Trend Following with Managed Futures.*
- Clenow, A. (2023). *Following the Trend.*
- Trivedi, A., & Kyal, N. (2021). *Quantitative Sentiment Analysis in Financial Markets.*
- Zakamulin, V., & Giner, G. (2024). *Practical Time-Series Momentum Strategies.*