Do Tech Stocks Move Together?

Market Concentration Risk & Diversification Limits in Big Tech

Note: The HTML/notebook analysis was completed collaboratively; this final report was written independently by Maria George.

AAPL · MSFT · NVDA · AMZN | January 2015 - April 2026 | ECON 323

Training period: Jan 2015 - Dec 2025 | Test period: Jan 2026 - Apr 2026

Executive Summary

This report examines whether holding Apple (AAPL), Microsoft (MSFT), NVIDIA (NVDA), and Amazon (AMZN) together in a portfolio provides meaningful diversification, or whether they are so tightly linked that the portfolio effectively behaves as a single bet on the technology sector.

The short answer: they move together far more than most investors realise. Daily return correlations across all six stock pairs fall between 0.52 and 0.67; high enough that a simultaneous drawdown in all four is not a tail event; it is the norm during periods of market stress. The COVID crash of March 2020 and the 2022 rate-hike selloff both confirm this, with all four names declining sharply at the same time.

On the forecasting side, neither a statistical model (ARIMA) nor a machine learning model (XGBoost) was able to predict daily returns with any practical precision; both performed close to a naive "predict zero every day" baseline. This is consistent with how efficient these markets are: with thousands of analysts and algorithms trading AAPL and MSFT around the clock, there is little predictable signal left.

The key takeaway for a portfolio manager: genuine diversification alongside a tech allocation requires moving outside the sector into bonds, commodities, or non-tech equities.

1. The Question and Why It Matters

The four stocks in this study are Apple, Microsoft, NVIDIA, and Amazon, they collectively represent over $11 trillion in market capitalisation and account for roughly 20% of the S&P 500. A passive investor in any major index fund already has heavy, concentrated exposure to these names, often without realising it.

The standard argument for holding all four is diversification: "they're different businesses." Apple makes consumer hardware, Microsoft sells enterprise software, NVIDIA makes chips, and Amazon runs cloud infrastructure and e-commerce. But business model diversity does not automatically translate into return independence. If all four stocks respond to the same macro forces; interest rate expectations, risk appetite, sector-wide news; they will move together regardless of their underlying differences.

This study tests that assumption with data. Using daily log returns from January 2015 to December 2025, it measures how correlated these stocks are, how much of one stock's daily moves can be explained by the

others, and whether any model can forecast their short-term direction. The 2026 year-to-date data serves as an out-of-sample test.

2. Data & Approach

Adjusted daily closing prices were pulled from Yahoo Finance for all four tickers over January 2015 to April 2026. "Adjusted" prices account for stock splits and dividends, giving a clean total-return series. The raw prices were converted to daily log returns, the standard transformation for financial return analysis, which removes the price-level trend and allows like-for-like comparison across stocks with very different absolute prices.

Figure 1 shows the growth of $100 invested in each stock at the start of 2015, plotted on a log scale. The log scale is necessary because NVIDIA's ~375× appreciation would otherwise make the other three lines invisible near the x-axis.

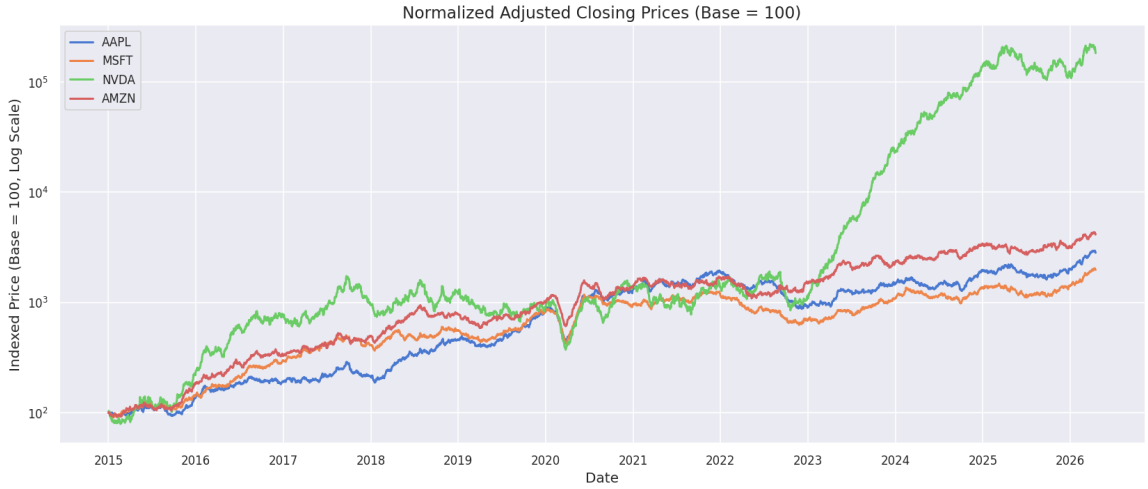

Figure 1: Indexed price performance (base = $100 at Jan 2015, log scale). All four stocks delivered strong returns, but NVIDIA's AI-driven surge from 2023 onward dwarfs the others.

Despite NVIDIA's outperformance in price terms, the daily return correlations are what matter for portfolio risk. Figure 2 shows those daily returns over time.

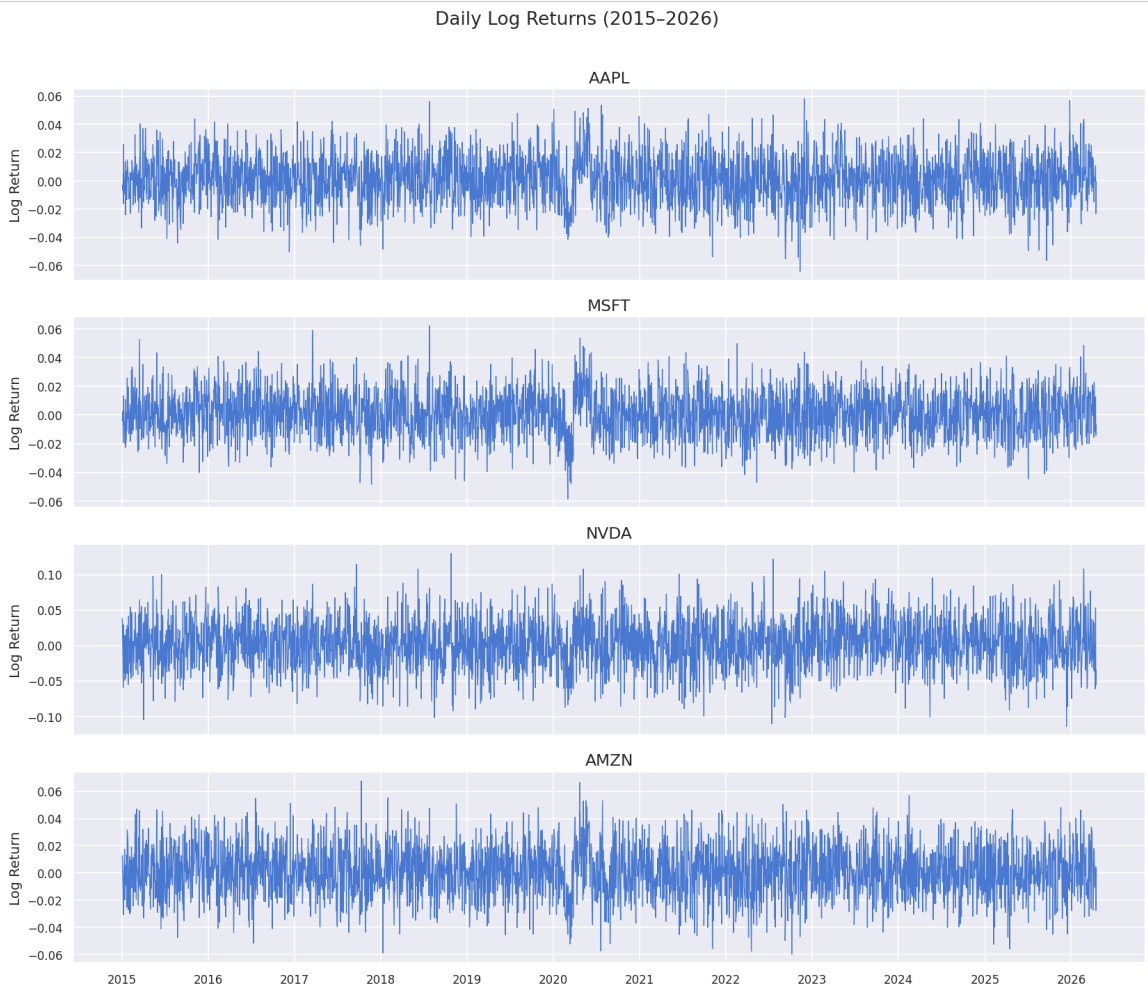

Figure 2: Daily log returns (2015–2026). The spike clusters around March 2020 (COVID crash) and 2022 (rate hikes) are clearly visible across all four stocks simultaneously, the defining characteristic of correlated assets.

A few things stand out immediately. First, the large swings happen at the same time across all four stocks. Second, NVIDIA is notably more volatile than the others, with daily return swings frequently exceeding ±10% during the AI boom. Third, all four series hover around zero on average, which is what you would expect from an efficient market where prices already incorporate all available information.

Figure 3 shows the distribution of daily returns more formally, with a smoothed density curve overlaid on each histogram.

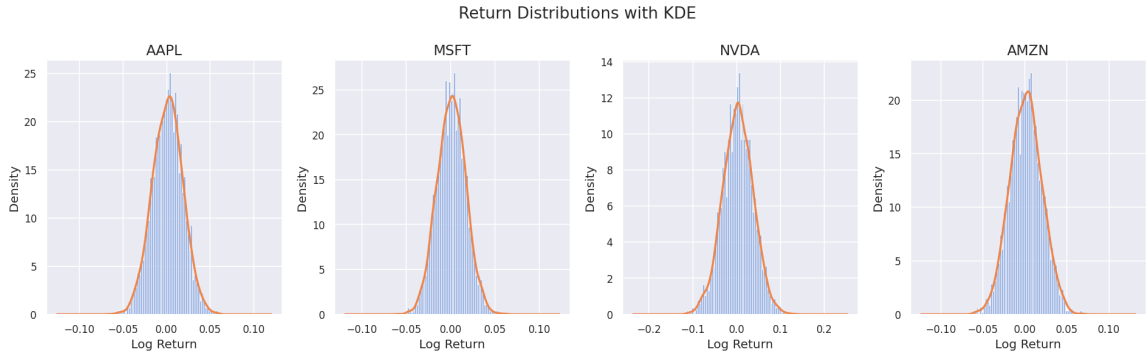

Figure 3: Return distributions with kernel density overlay (training period). NVDA's wider spread reflects its ~2× higher daily volatility versus MSFT. All distributions show slightly fatter tails than a normal bell curve - extreme days happen more often than theory would predict.

Stock	Avg. Daily Return	Daily Volatility	Annualised Vol
AAPL	0.10%	1.74%	27.6%
MSFT	0.09%	1.60%	25.4%
NVDA	0.25%	3.47%	55.1%
AMZN	0.12%	1.90%	30.1%

Table 1: Summary return statistics (training period 2015–2025).

3. How Correlated Are These Stocks?

Correlation measures how much two assets move together on a scale from −1 (perfectly opposite) to +1 (perfectly in sync). For diversification to work, you want low or negative correlation. Here is what the data shows:

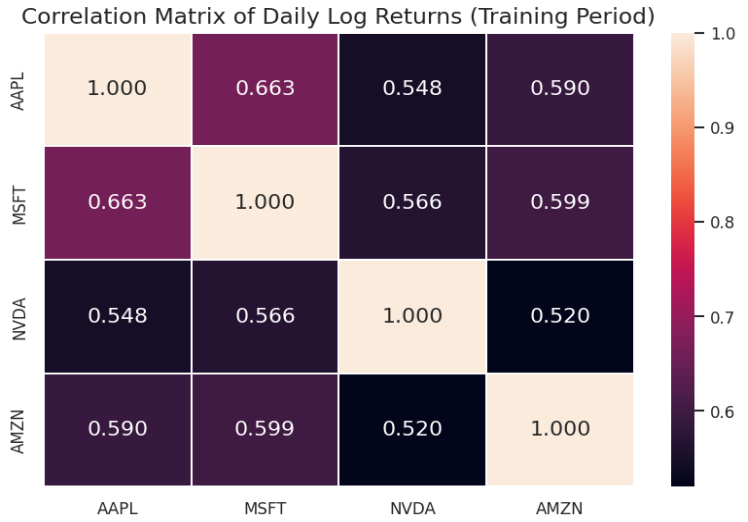

Figure 4: Pearson correlation matrix of daily log returns (2015–2025). Every pair exceeds 0.52 - there are no low-correlation pairs in this group.

Pair	Correlation	Interpretation
AAPL – MSFT	0.663	Strongly linked; both respond to enterprise tech cycles
AAPL – AMZN	0.590	High; share macro and consumer sentiment sensitivity
MSFT – AMZN	0.599	High; both large cloud platforms with common risk drivers
MSFT – NVDA	0.566	High; AI spending links both tightly from 2023 onward
AAPL – NVDA	0.548	High; weakest AAPL pair but still substantial
NVDA – AMZN	0.520	Lowest pair; NVDA's chip cycle adds some independence

Table 2: All pairwise correlations (training period). None fall below 0.52.

To put these numbers in context: a correlation of 0.65 means that on any given day when MSFT rises sharply, AAPL has a strong tendency to rise sharply too. From a risk management standpoint, this means the "diversification discount" from holding both is much smaller than most investors assume.

This was most visible during the two major drawdowns in the sample. In March 2020, all four stocks fell 30–40% within three weeks. In 2022, all four dropped 25–65% as the Fed hiked rates. Diversification within the tech sector offered no protection in either episode. During a systemic selloff, high-correlation assets all go down together, the opposite of what diversification is supposed to achieve.

4. How Much Do They Drive Each Other?

Correlation tells you the direction and strength of a relationship. Regression quantifies it: specifically, what percentage of one stock's daily moves can be explained purely by another's movement? This R^2 figure is the key metric.

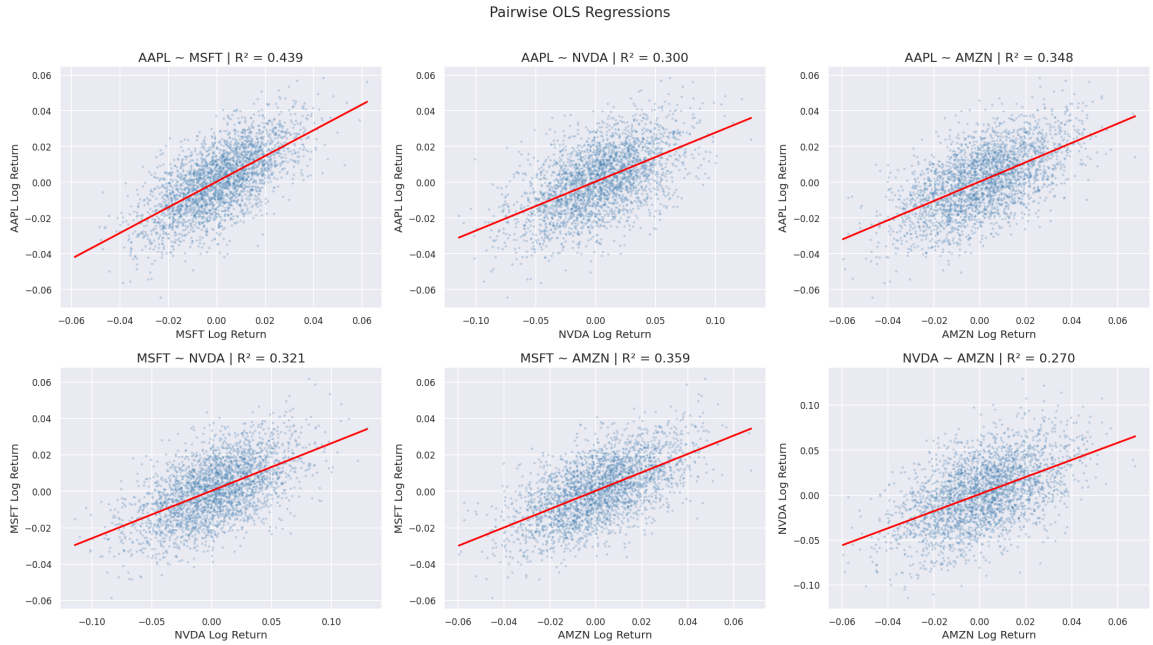

Figure 5: Pairwise regression scatter plots. Each panel fits a straight line through the return data. The tighter the cloud around the line, the higher the R^2. All six pairs show a clear positive slope.

Regression	R^2	What it means
AAPL explained by MSFT	44%	44% of Apple's daily moves are traceable to Microsoft
MSFT explained by AMZN	43%	43% of Microsoft's variance tracks Amazon
NVDA explained by AMZN	27%	Lowest pair; NVDA retains more idiosyncratic moves

Table 3: Selected pairwise R^2 values (training period).

Going further, each stock was regressed against the other three simultaneously. The results are striking: 42-60% of every stock's daily return variance is explained by the movement of its three peers alone. In other words, most of what moves Apple on any given day is the same force moving Microsoft, NVIDIA, and Amazon.

Stock	R^2 vs. Other Three	Residual (unexplained)
AAPL	49%	51%
MSFT	60%	40%
NVDA	42%	58%
AMZN	47%	53%

Table 4: Multiple regression R^2 (each stock vs. the other three ~ training period.

NVIDIA has the most independence (42% explained), reflecting that its chip cycle and AI-specific demand story gives it a distinct return driver. MSFT has the least independence (60% explained), consistent with its role as a broad-based technology platform tied closely to enterprise and cloud spending that touches all four names.

5. Can We Forecast Daily Returns?

Two forecasting models were tested on the 2026 out-of-sample data: ARIMA (a classical statistical time-series model) and XGBoost (a gradient-boosted machine learning model). Both were trained exclusively on 2015–2025 data and then asked to predict 2026 daily returns.

5.1 ARIMA: the Statistical Baseline

ARIMA models work by identifying repeating patterns in historical return sequences. For each stock, the model specification was chosen to best fit the training data (using the AIC criterion). The 2026 forecasts are shown in Figure 6.

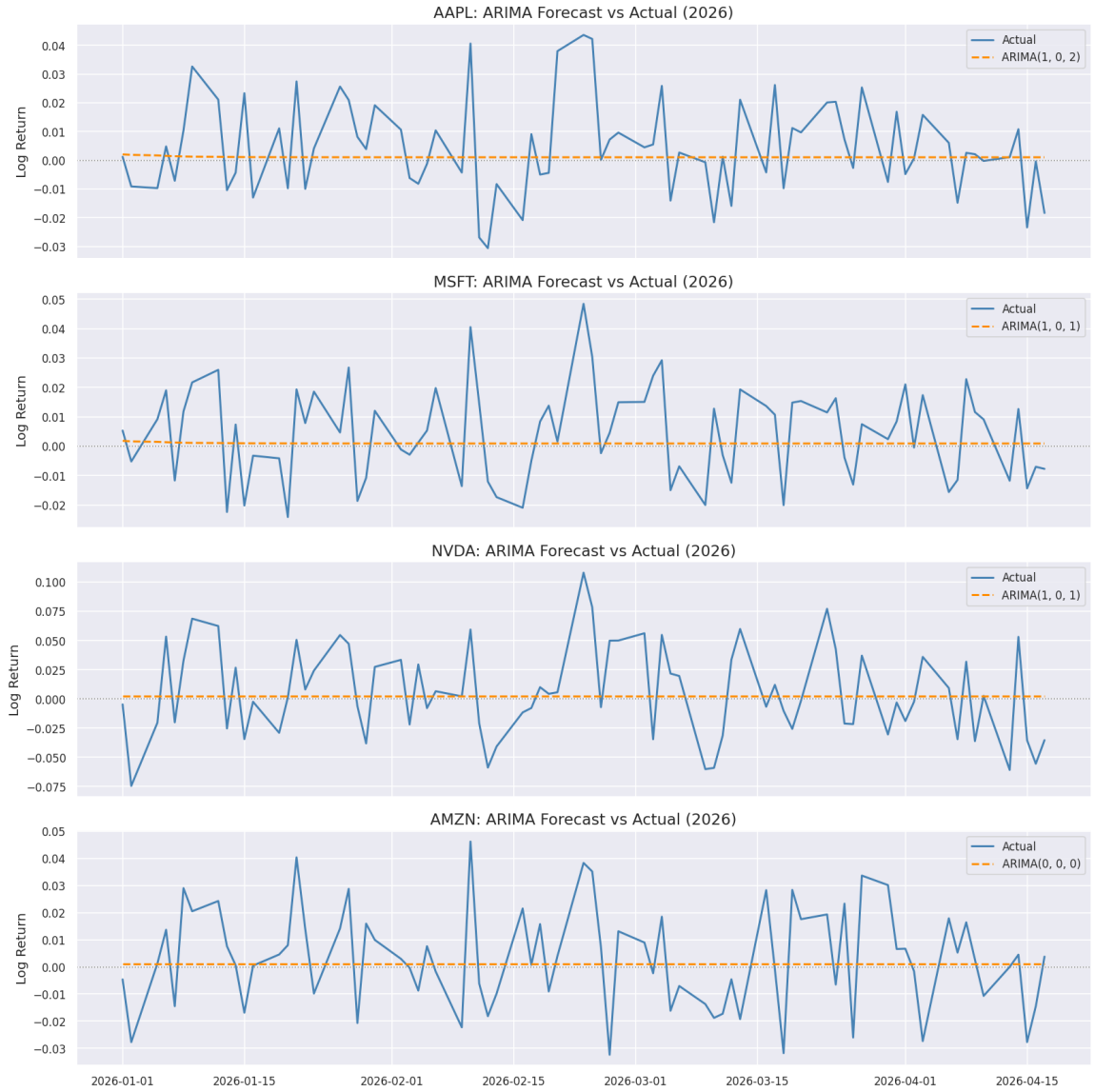

Figure 6: ARIMA forecasts vs actual returns (2026). Orange dashed = model prediction; blue = actual return. The model produces a near-flat line close to zero, while actual returns exhibit sharp unpredicted spikes.

The flat, near-zero predictions are not a failure of calibration, they reflect what the data implies. Amazon's best-fit ARIMA model was ARIMA (0,0,0), which simply predicts the mean (approximately zero) every day. That is the mathematically correct answer when a series has no detectable autocorrelation. The same logic applies to all four stocks: there is no consistent pattern in past returns that reliably predicts tomorrow's direction.

5.2 XGBoost: the Machine Learning Challenger

XGBoost can capture non-linear relationships between lagged returns and future returns that ARIMA cannot. Each model was given the five most recent daily returns as inputs and trained to predict the next day's return.

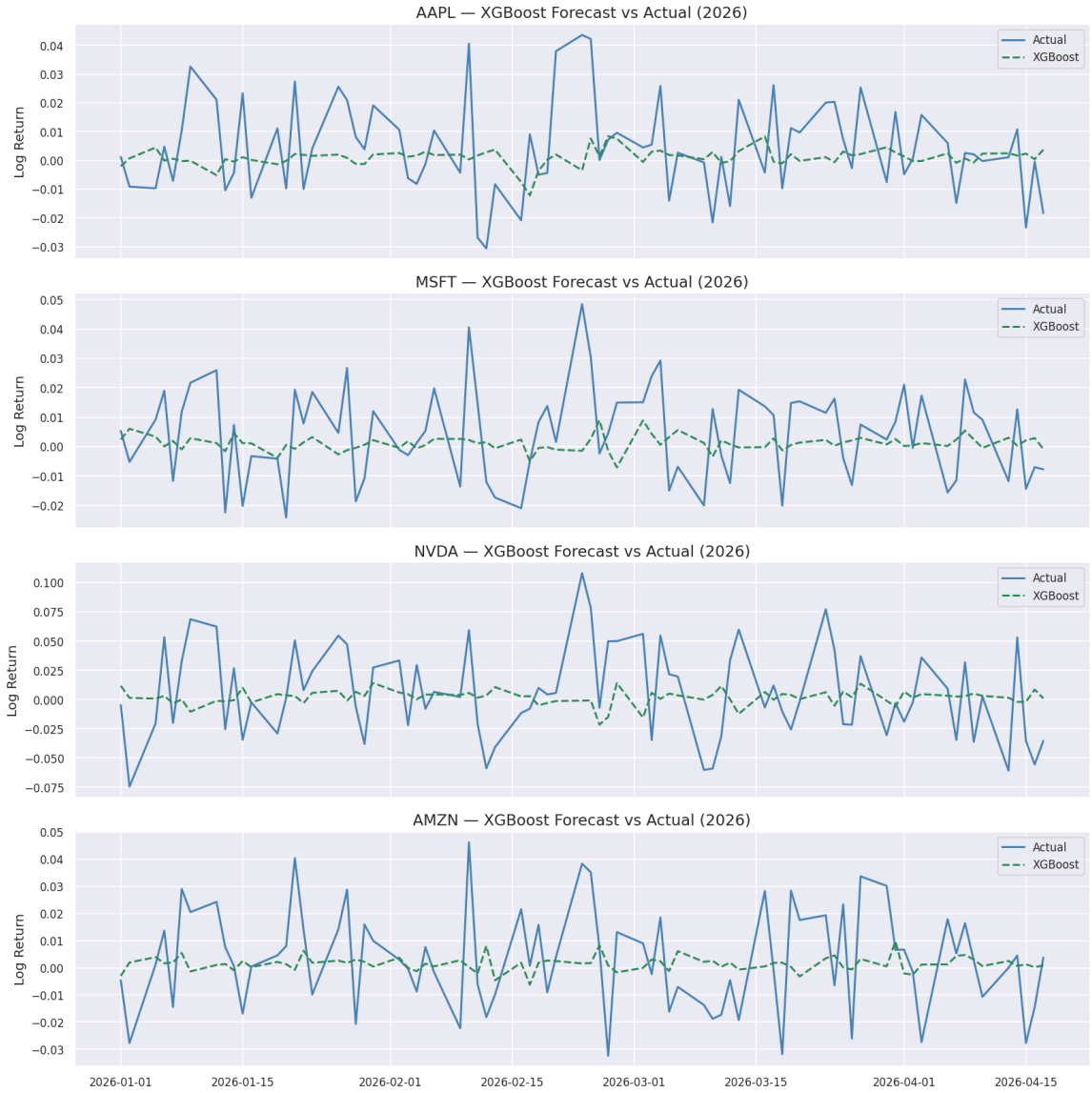

Figure 7: XGBoost forecasts vs actual returns (2026). Green dashed = XGBoost prediction; blue = actual. The model shows more variation than ARIMA but still fails to capture the direction of large return days.

XGBoost produces more varied predictions than ARIMA, which might look better at first glance. But the error metrics tell a different story.

5.3 Model Comparison

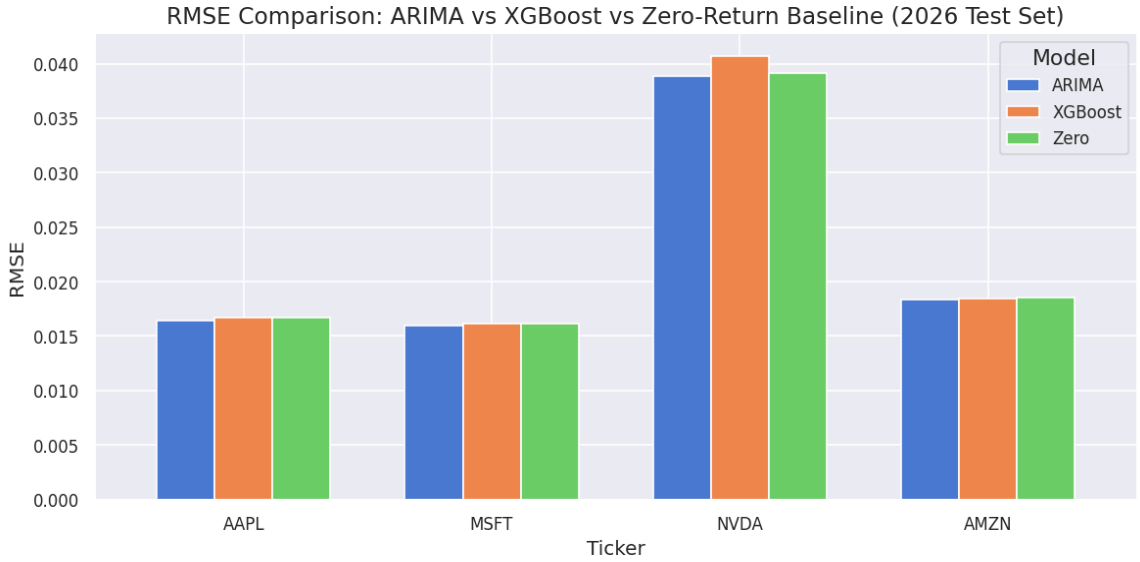

Figure 8: RMSE comparison across all three approaches. The three bars are nearly identical for each stock, confirming that none of the models adds material predictive value over simply predicting zero.

Stock	ARIMA RMSE	XGBoost RMSE	Predict Zero RMSE	Winner
AAPL	0.015617	0.015925	0.015535	ARIMA (marginally)
MSFT	0.021773	0.022400	0.021075	ARIMA (marginally)
NVDA	0.022242	0.022001	0.021998	ARIMA (marginally)
AMZN	0.020465	0.021020	0.020485	ARIMA (marginally)

Table 5: RMSE on 2026 test set. Lower is better. Differences between models are economically negligible.

ARIMA edges out XGBoost on all four stocks, but by margins so small they have no practical significance. More importantly, all three approaches, including simply predicting zero every day perform nearly identically. This is a strong signal: for large-cap, heavily traded technology stocks, there is essentially no predictable structure in daily returns that a model can exploit.

This is what market efficiency looks like in practice. Thousands of professional investors and algorithmic strategies are already trading on every piece of publicly available information about these companies. By the time any signal is visible in lagged return data, it has already been priced in.

6. Conclusions & Portfolio Implications

Finding 1: These stocks are highly correlated

All six pairwise correlations exceed 0.52, and multiple regression models explain 42-60% of each stock's daily return variance using only the other three as inputs. Holding all four together provides far less risk reduction than the diversity of their business models might suggest.

Finding 2: Concentration risk is real and asymmetric

During the two largest drawdowns in the sample, March 2020 and 2022, all four stocks fell simultaneously and sharply. This is the key point for risk management: correlation typically rises during market stress, which is exactly when diversification is most needed. A tech-only portfolio failed on this test in both episodes.

Finding 3: Daily returns are largely unpredictable

Neither ARIMA nor XGBoost meaningfully outperformed a naive zero-return prediction. This is not a modelling failure rather an evidence of market efficiency. For the most liquid equities in the world, all known information is already in the price. Predictability, if it exists, would need to be sought at different time horizons or in less-efficient corners of the market.

Practical Implications

- **True diversification requires leaving tech:** Adding more tech names does not solve the problem. Assets with genuinely low or negative correlation i.e. government bonds, commodities, international equities, or alternatives are required to reduce single-factor exposure.
- **Risk models should account for tail correlation:** Average correlations understate the risk. During drawdowns, these correlations rise further. Stress-test scenarios should assume all four stocks fall together.
- **Quantitative forecasting adds limited edge at daily frequency:** Resources spent on daily return prediction for large-cap tech may be better allocated to understanding the macro drivers (rates, earnings, sector flows) that explain the common factor these stocks share.

References

Bhowmik, M. (2023). Risk and return analysis using FAANG stocks. *Medium.* https://medium.com/@manisangsu/risk-and-return-analysis-using-faang-stocks-d0dc1026a416

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD Conference,* 785–794.

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *Journal of Finance, 25*(2), 383–417.

Markowitz, H. (1952). Portfolio selection. *Journal of Finance, 7*(1), 77–91.

Wang, Y., & Guo, Y. (2020). Forecasting method of stock market volatility in time series data based on mixed model of ARIMA and XGBoost. *China Communications, 17*(3), 205–221.