The Effect of Market Concentration on Beta, Implied Volatility, and Long-Short Investment Strategy Using Deep Learning

Abstract

The S&P 500 index has become increasingly concentrated in a handful of stocks known as the Magnificent 7 (NVDA, META, TSLA, AMZN, GOOG, MSFT, AAPL), which accounted for 31.9% of the index as of February 14, 2025. This growing dominance raises critical questions about beta, a fundamental measure of systematic risk in portfolio management. Our research finds that Magnificent 7's contribution to the index's total beta has increased from 10% in 2010 to nearly 45% today, making market risk more reflective of their individual risk profiles rather than the broader economy. We show that beta estimates for non-dominant stocks are significantly affected by the choice of market proxy—when excluding the Magnificent 7 from the index, betas for other stocks tend to increase, suggesting that traditional beta estimates may underestimate systematic risk. Additionally, we find that major price movements in one of these dominant stocks can create cascading effects, amplifying overall market volatility and distorting risk assessments. Our analysis highlights the need for alternative benchmarking methods and risk measures in an increasingly concentrated equity market.

Dataset Description

The dataset comprises daily data for all S&P 500 constituents from January 6, 2010, to February 14, 2025, sourced from Bloomberg. It includes stock-level information on index weights, betas relative to the S&P 500, daily returns, and 30-day implied volatility. Spanning approximately 3,800 trading days, the dataset is updated daily to reflect market capitalization-based weight adjustments.

A key focus is the Magnificent 7—Apple (AAPL), Microsoft (MSFT), Alphabet (GOOGL), Amazon (AMZN), Nvidia (NVDA), Meta Platforms (META), and Tesla (TSLA)—which have grown to represent a significant share of the index. Their weights, betas, and returns are analyzed alongside other constituents, providing a detailed view of their outsized impact on index risk and performance.

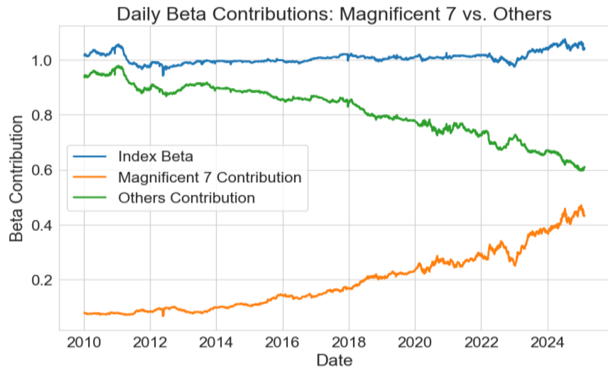
Additionally, a separate dataset was constructed using a selection of 10 correlated and 10 uncorrelated companies, representing both small and large enterprises, over the period from January 2, 2015, to February 14, 2025. This dataset was segmented into three categories—correlated, uncorrelated, and a mixed category containing both types—to evaluate the models' performance across different market structures.

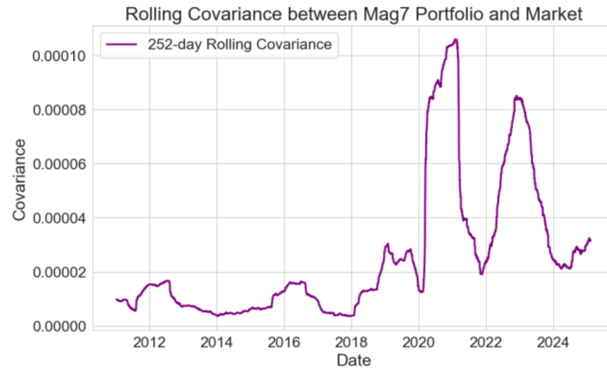
The Magnificent 7s Beta & Return Contribution Through Time

Since 2010, the rise of the Magnificent 7 has fundamentally altered the risk and return dynamics of the S&P 500. These companies, driven by successive waves of investment in technology, have grown to dominate the index in market capitalization, returns, and risk. As their influence expanded, so did their impact on the S&P 500's beta, a key measure of market risk. Since beta is weighted by market capitalization, the increasing concentration of these firms has made the index more reflective of their individual risk profiles rather than the broader market.

This shift is evident in the data. In 2010, Magnificent 7 accounted for roughly 10% of the index's total beta; today, they represent nearly 45%. Historically, risk was more evenly distributed across sectors, preventing any single group from dictating overall volatility. Now, a narrow set of technology firms wields disproportionate influence, making index movements more dependent on sector-specific trends.

The concentration of returns further reinforces this trend. The growing imbalance raises important considerations for investors who track the index passively. While the S&P 500 is designed to represent a broad cross-section of the market, its performance is increasingly tied to a small group of stocks. Investors may assume they are diversifying across industries, but their returns are heavily influenced by the fortunes of a few dominant firms.





Implications on Wider Portfolio Decisions

The increasing dominance of the Magnificent 7 in the S&P 500 has important consequences for how beta is used in forecasting returns and valuing companies. Using CAPM framework:

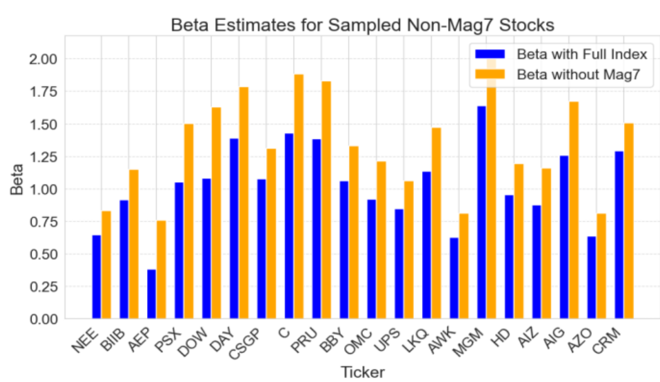
$$E(R_i) = R_f + \beta_i(E(R_m) - R_f) \text{ where: } \beta_i = \frac{cov(R_i, R_m)}{var(R_m)}$$

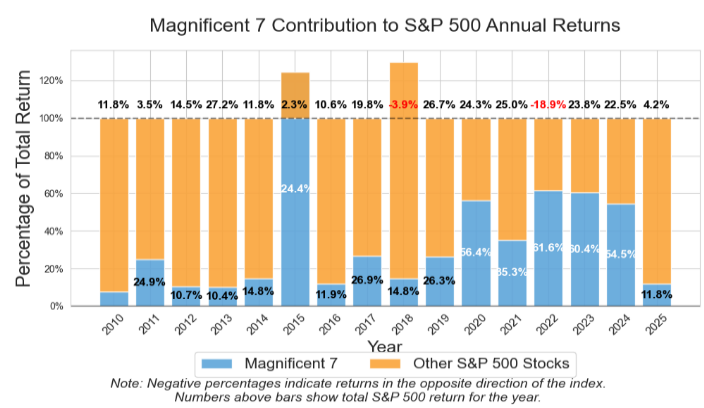
beta measures a company's systematic risk relative to the overall market, which is the critical input for measuring the discount rate for company valuation purposes. However, when the market proxy (i.e. the full index) is now largely driven by a small, concentrated group of technology stocks, the beta computed for many non-dominant companies can be biased downward.

To illustrate, we compared the beta of a sample of non-Mag7 companies using two market proxies, which are the full index and an alternative benchmark which excludes the Mag7 stocks. Specifically, we randomly selected 50 non-dominant stocks from the index and computed their respective betas.

Our analysis shows that non-Mag7 companies tend to exhibit a higher beta when the market proxy excludes the Mag7. In other words, using the full index as a benchmark "dilutes" the sensitivity of these companies' returns because the high volatility and unique behavior of the Mag7 increase the overall variance of the index. When the Mag7 stocks are removed, the resulting benchmark is less volatile, so the same covariance translates into a higher beta.

This result is critical for investment decisions. If beta is used to forecast expected returns (via the CAPM, for example), underestimating beta means that the risk premium may be understated. In practice, this could lead to mispricing of risk and suboptimal asset allocation. Investors and analysts need to be aware of this concentration effect—and possibly adjust their benchmarks or use alternative risk measures – to ensure that the true systematic risk of individual companies is accurately captured.





Tesla 2024 Drop

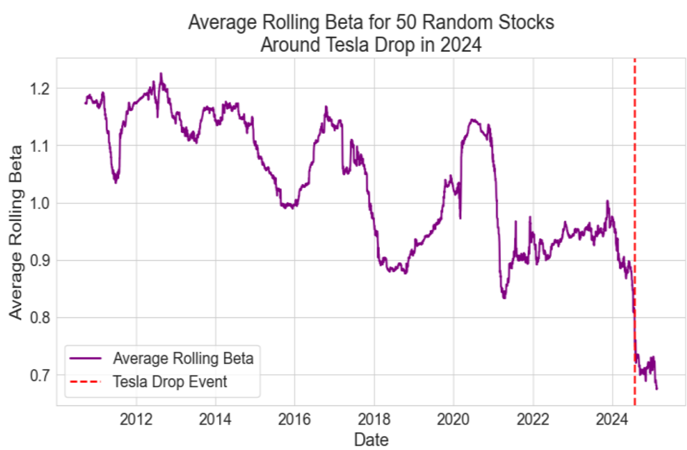
With nearly half of the index's beta driven by just seven stocks, the overall risk measure becomes highly sensitive to any large moves in one of these dominant names. For instance, in 2024 a significant drop in Tesla's share price not only impacted its own beta, but also altered the weighted composition of the index. When Tesla experiences a sharp decline, its individual beta and market capitalization may change abruptly. Since index beta is computed as a weighted average of individual betas, even if most stocks remain stable, a dramatic Tesla move can cause a sudden jump (or drop) in the index beta.

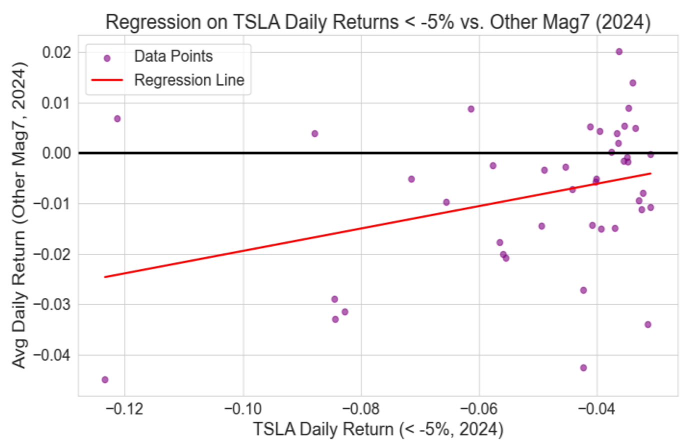
This matters for investors who use beta to forecast expected returns (for example, via CAPM) or to hedge systematic risk. If the market proxy becomes more volatile due to a shock in a single, dominant stock, then a company's beta estimated relative to that index may be misleadingly low or high. In our analysis, we can see that on the day of Tesla's dramatic drop—and in the days immediately following—the overall index beta shifts noticeably. This suggests that using the full index as a market proxy can mask the true systematic risk for companies, potentially leading to underestimation or overestimation of risk in portfolio construction.

Our regression analysis on days when TSLA experienced a significant negative return (below negative 3%) reveals a statistically significant relationship between TSLA's negative performance and the average daily returns of the other Magnificent 7 stocks (coefficient ≈ 0.2222, p=0.026). This finding indicates that when TSLA drops, the other stocks in the group tend to follow suit, evidencing an increased correlation among these dominant firms during market stress.

Since the S&P 500's beta is calculated as a weighted average of individual stock betas, and the Magnificent 7 now accounts for nearly half of the index's risk, this synchronized movement has profound implications. When one of these stocks—such as Tesla—suffers a significant drop, it can cause a disproportionate shift in the overall market return. As a result, the beta estimates for individual companies become highly sensitive to these isolated shocks in the dominant group. In practice, this means that the beta of a company, which is a critical input in models such as the CAPM, becomes volatile, unpredictable, and less reliable as a measure of systematic risk.

In summary, the increased concentration of the Magnificent 7 has led to a scenario where negative returns in one dominant stock trigger a cascade of similar movements among its peers. This co-movement inflates the correlation during downturns, thereby distorting beta estimates. Investors and analysts who rely on beta for forecasting returns or making hedging decisions must recognize that in such a concentrated market environment, beta may no longer serve as a stable or accurate indicator of risk. Alternative risk metrics or adjusted benchmarks may be necessary to capture the true systematic risk in this evolving market landscape.





Volatility of Index Options

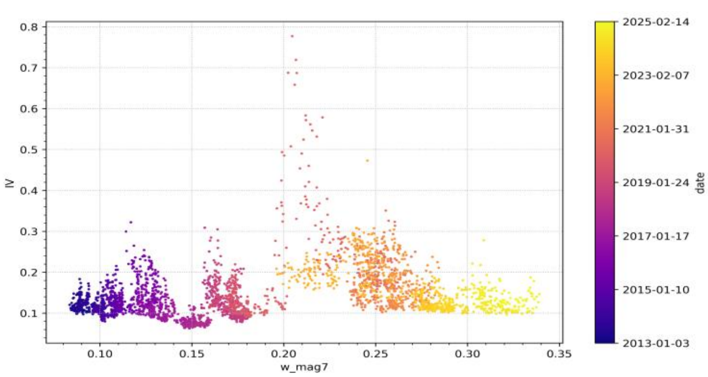
Options on the index are amongst the most actively traded. If the index is so highly concentrated, what are the implications for the pricing of the option (if any) or the implications for the volatility of the option (remember that options are basically a means for trading volatility)? What happens to the option is there is a significant increase in the volatility of one or several of the underlying securities?

The impact of concentration in option pricing can be evaluated studying the relationship between the latter and implied volatility. The surface, rather than individual prices, is the ubiquitous object of interest of market participants. This is because implied volatility is uniquely determined for both puts and calls at the same strike and maturity, and it provides critical insights into the market's expectations for the distribution of future asset prices. The analysis relies on the historical 30-day at-the-money forward (ATMF) implied volatility time series from 01-02-2013 to 02-14-2025. Ideally, historical VIX futures data with the same expiry would be the preferred metric, as the VIX index is designed to incorporate information across multiple strikes. However, obtaining this dataset is prohibitively expensive. Plotting the implied volatility against the mag7 weight, with dates encoded via a colormap, immediately reveals that a regression model based solely on the weight level is inadequate (note: when the colormap variable is placed on the x-axis instead, it suggests that the relationship might be dynamic rather than static).

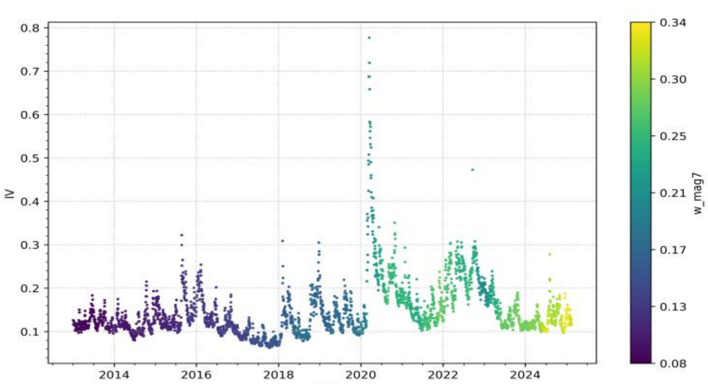
Concentration is inherently a relative measure, reflecting the performance of certain assets relative to the rest of the index, in this case that of the magnificent seven. Economic intuition suggests that the change in concentration has greater explanatory power than its absolute level. For instance, an increase in concentration, driven by large caps outperforming small caps, might be associated with higher or lower volatility than would otherwise be observed, vice versa for a decrease. Based on these insights, it may be advantageous to incorporate lagged or smoothed measures of index performance as control variables. Given that implied volatility is path-dependent (see Guyon, 2022), employing exponentially smoothed index returns can better capture key features of the volatility process, such as the leverage effect and volatility clustering, providing a benchmark to evaluate whether market participants attribute importance to whether or not a market move is driven by the magnificent seven stocks. To test the significance of the changes in concentration on the level of implied volatility we propose a linear model of the following form:

$$IV_t = \beta_0 + \beta_1 R_{1,t} + \beta_2 \Sigma_t + \beta_3 R_{3,t} + \epsilon_t$$
$$R_{1,t} = \sum_{t_i \le t} K(t - t_i) r_{t_i}$$
$$\Sigma_t = \sqrt{R_{2,t}}$$
$$R_{2,t} = \sum_{t_i \le t} K(t - t_i) r_{t_i}^2$$
$$R_{3,t} = \sum_{t_i \le t} K(t - t_i) c_{t_i}.$$

IV (%) vs. total weight (%) mag7

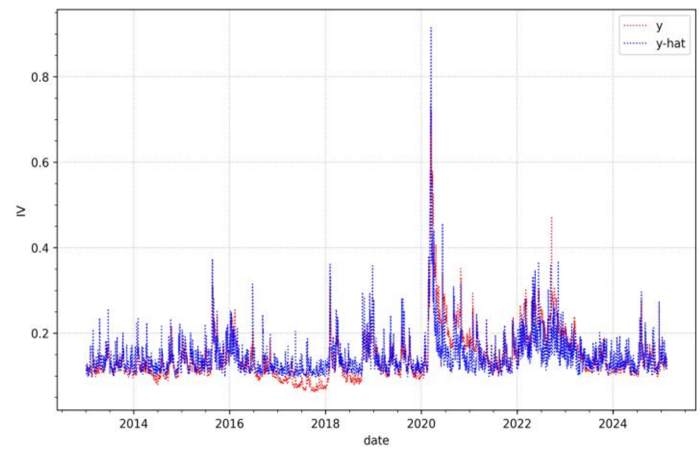


IV (%) time series



$K(\tau) = e^{-\tau}$ represents the exponential smoothing kernel, r_{t_i} and c_{t_i} are the percentage changes in the S&P500 and in the weight of the magnificent seven respectively. Empirical evidence indicates that the beta coefficient for the index returns control variable tends to be negative, while that for the volatility control variable is generally positive (see Guyon, 2022; Andrés et al., 2023). The model implicitly distinguishes between four extreme regimes: (1) a market upturn driven by the mag7; (2) a market upturn driven by the remaining stocks; (3) a market downturn driven by the mag7; and (4) a market downturn driven by the remaining stocks. Given the leverage effect, where high implied volatility is typically associated with market downturn and low implied volatility with market upturns, the hypothesis is that market participants assign an even higher or lower volatility premium when a market move coincides with changes in concentration. Parameter estimation is performed using ordinary least squares (OLS), with inference based on a heteroskedasticity and autocorrelation consistent (HAC) covariance matrix that employs eleven lags, as determined by the Newey & West automatic bandwidth selection procedure.

30d vs. estimate



Residuals time series

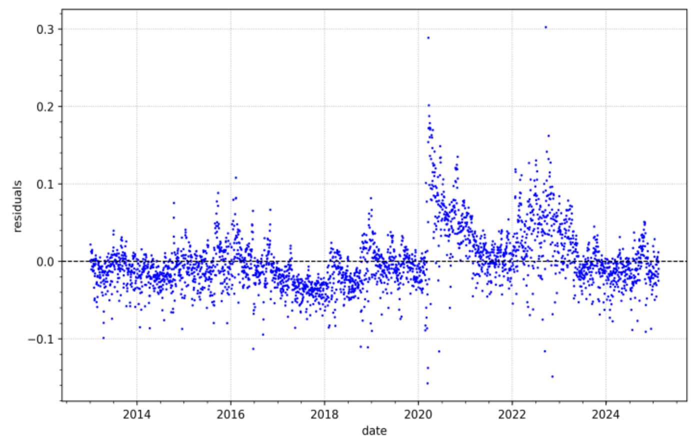


Figure 2 presents the model estimates along with their errors, while Figures 4 and 5 provide the OLS summary tables for the complete model and for the model excluding the concentration feature, respectively.

OLS summary table, complete model

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                 30 day   R-squared:                       0.625
Model:                            OLS   Adj. R-squared:                  0.625
Method:                 Least Squares   F-statistic:                     158.5
Date:                Thu, 27 Feb 2025   Prob (F-statistic):           1.95e-95
Time:                        22:24:15   Log-Likelihood:                 5611.1
No. Observations:                3049   AIC:                        -1.121e+04
Df Residuals:                    3045   BIC:                        -1.119e+04
Df Model:                           3
Covariance Type:                  HAC
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0915      0.003     30.633      0.000       0.086       0.097
R_1           -0.7044      0.096     -7.357      0.000      -0.892      -0.517
Sigma          5.4651      0.309     17.671      0.000       4.859       6.071
R_3           -0.0613      0.108     -0.570      0.569      -0.272       0.149
==============================================================================
Omnibus:                      838.604   Durbin-Watson:                   0.601
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             3553.831
Skew:                           1.279   Prob(JB):                         0.00
Kurtosis:                       7.630   Cond. No.                         118.
==============================================================================
```

OLS summary table, model excluding concentration

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                 30 day   R-squared:                       0.625
Model:                            OLS   Adj. R-squared:                  0.625
Method:                 Least Squares   F-statistic:                     224.7
Date:                Fri, 28 Feb 2025   Prob (F-statistic):           9.50e-92
Time:                        19:01:07   Log-Likelihood:                 5610.8
No. Observations:                3049   AIC:                        -1.122e+04
Df Residuals:                    3046   BIC:                        -1.120e+04
Df Model:                           2
Covariance Type:                  HAC
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0914      0.003     30.869      0.000       0.086       0.097
R_1           -0.7030      0.095     -7.362      0.000      -0.890      -0.516
Sigma          5.4674      0.308     17.763      0.000       4.864       6.071
==============================================================================
Omnibus:                      836.100   Durbin-Watson:                   0.600
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             3534.832
Skew:                           1.275   Prob(JB):                         0.00
Kurtosis:                       7.617   Cond. No.                         117.
==============================================================================
```

Traditional Fama - French Model

The traditional approach taken is the Fama-French three-factor model —specifically, size (SMB) and value (HML)—affect a portfolio's returns. After deducting the risk-free rate (RF) to determine the excess returns, it performs a regression to determine the extent to which these factors account for the portfolio's performance. Additionally, the model calculates alpha, which indicates if the portfolio is producing higher returns than the factors predicted. The code also annualizes returns to provide a more comprehensive view of long-term performance. In the end, this aids in identifying whether the portfolio has a small-cap or value bias and whether any modifications are required. We calculated a custom SMB and HML for the execution and compared our results on the Sharpe Ratio and the Annualized Returns.

$$r = r_f + \beta_1(r_m - r_f) + \beta_2(SMB) + \beta_3(HML) + \varepsilon$$

Transformer Model

Input and Feature Engineering

In this study, the raw data is refined into a consistent format for model ingestion through a systematic series of preprocessing steps. Initially, all time-series features are normalized to a standard scale to stabilize gradients and mitigate the influence of outliers. Subsequently, key technical indicators—including the Relative Strength Index (RSI) and the Stochastic Oscillator – are computed to capture momentum and identify potential turning points in the price series. Finally, missing values are addressed using robust imputation techniques or forward/backward filling methods, thereby minimizing data discontinuities. Collectively, these transformation steps ensure that the final input tensor, which spans multiple assets and features, is coherent, uniform, and optimized for processing by the Transformer architecture.

Model

The proposed model employs a **deep Transformer architecture** to forecast multi-dimensional time series with multiple outputs (e.g., returns for various asset classes). Each input sample is organized as a sequence of observations over a fixed temporal window, where every timestep contains a feature vector representing the variables of interest. Inside each Transformer layer, **multi-head self-attention** captures long-range dependencies across time steps, enabling the model to learn complex temporal correlations without relying on recurrent operations.

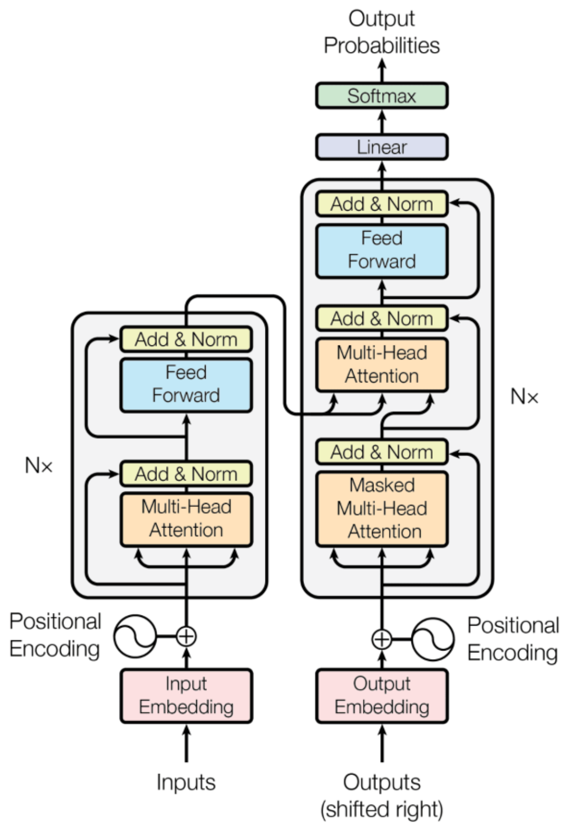
Formally, let Q, K, and V be the query, key, and value matrices derived from the layer input. The scaled dot-product attention for one head is computed as:

$$\mathrm{Attn}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \mathrm{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^{\top}}{\sqrt{d_k}}\right)\mathbf{V},$$

where d is the dimensionality of the keys, in multi-head attention, multiple such heads operate in parallel, each projecting the input into distinct subspaces before concatenating their outputs to form a richer representation. The result then undergoes a **feed-forward network** (two dense layers with a non-linear activation in between) to transform and refine the latent representation further. Residual connections and layer normalization are applied around the attention and feed-forward sub-layers to stabilize training and preserve gradient flow.

By stacking several Transformer blocks, the model can learn hierarchical patterns in the data. Finally, a **pooling step** produces a fixed-length embedding passed to a **dense output layer**. This output layer predicts the target values simultaneously, such as the returns for different assets. Compared to recurrent approaches, this deep Transformer design more effectively captures long-range dependencies, handles parallel computation over time steps, and often yields superior performance on complex time series tasks.

Transformer Model Architecture



In our proposed model, a transformer-based neural network optimizes long–short portfolio weights under realistic constraints. The network ingests asset returns R-long and R-short, and outputs weights w-long and w-short via an embedding layer and multi-head self-attention blocks. The daily portfolio return is computed as

$$R_p = w_{\text{long}} R_{\text{long}} - w_{\text{short}} R_{\text{short}},$$

with performance measured by the adjusted Sharpe ratio:

$$S = \frac{\mu\,(R_p - r_f)}{\sigma\,(R_p)},$$

where rf is the risk-free rate. Constraints are imposed to ensure:

$$\begin{aligned} &w_{\text{long}} + w_{\text{short}} = 1, \\ &R_{\text{long}} - R_{\text{short}} > 0, \quad \text{(positive credit spread)} \end{aligned}$$

Our model is trained using a custom loss that combines the negative Sharpe ratio and penalty terms for constraint violations. This end-to-end approach yields robust, risk-adjusted strategies that outperform risk-free benchmarks while satisfying practical allocation rules. By incorporating the risk-free rate, our Sharpe calculation reflects true excess returns, ensuring that the portfolio's performance is benchmarked against safe investments. Extensive experimentation demonstrates the method's effectiveness in achieving superior risk-adjusted returns under realistic market constraints.

Results

Category	Transformer		Traditional	
	Return (%)	Sharpe Ratio	Return (%)	Sharpe Ratio
Uncorrelated	11.26	0.74	0.19	0.25
Correlated	17.73	1.32	14.76	0.91
Mixed	13.22	1.06	10.18	0.71

The empirical results consistently demonstrate that the Transformer-based methodology outperforms conventional statistical approaches across all examined market conditions—Uncorrelated, Correlated, and Mixed. In the Uncorrelated scenario, the Transformer model achieves an annualized return of 11.26% with a Sharpe ratio of 0.74, substantially higher than the 0.19% return (Sharpe ratio: 0.25) realized by the traditional technique. This pronounced gap indicates that the self-attention mechanism effectively captures non-linear interactions and long-range temporal dependencies, even when data exhibit minimal or no explicit correlations—an area where standard models tend to underperform.

Under Correlated conditions, both models show improved outcomes owing to clearer structural patterns. Nonetheless, the Transformer retains a clear advantage by delivering a 17.73% annualized return (Sharpe ratio: 1.32), compared to the traditional method's 14.76% (Sharpe ratio: 0.91). The superior Sharpe ratio further underscores the Transformer's ability to balance return generation with risk mitigation, particularly in environments characterized by pronounced inter-asset correlations.

In Mixed regimes, defined by evolving or partial correlations, the Transformer continues to outperform its counterpart, achieving a 13.22% annualized return (Sharpe ratio: 1.06) versus the conventional model's 10.18% (Sharpe ratio: 0.71). These findings suggest that the architecture's attention mechanisms adaptively prioritize significant features amid changing market structures, thereby delivering robust risk-adjusted returns. Overall, the results indicate that Transformer-based frameworks are well-suited for capturing complex, non-stationary signals, largely due to their capacity to process sequential data without relying on strictly fixed temporal assumptions.

Future work

Our regression analysis shows that while the third parameter's estimated coefficient is negative its effect is not statistically significant at conventional thresholds. The 95% confidence interval for this parameter includes zero, indicating insufficient evidence to reject the null hypothesis that this parameter has no effect. Furthermore, removing this variable from the model does not materially change key performance indicators such as adjusted, log-likelihood, and information criteria (AIC/BIC). As a result, the simpler model without this parameter is equally supported by the data. These findings suggest that, at least within our current dataset, the third parameter does not contribute meaningfully to explaining or predicting the outcome variable.

The study demonstrates the efficacy of Transformer architectures across various correlation regimes, yet several promising avenues remain to be explored in equity markets. Future research could extend the analysis to encompass a broader array of equity indices—including small-cap, emerging markets, and sector-specific benchmarks—to test the generalizability of the Transformer-based approach under diverse market conditions. Such an investigation would determine if the observed performance persists during periods of heightened volatility or when sectors experience rotation. Additionally, integrating the Transformer model into long-short equity strategies holds significant promise. By identifying both undervalued and overvalued securities, the model can facilitate portfolio construction that leverages movements in both directions, potentially enhancing risk-adjusted returns through optimized pair trading or market-neutral strategies. It is also critical to assess how these architectures adapt dynamically to the distinct phases of equity market cycles, such as bull, bear, and sideways trends. Refinements in attention mechanisms may enable the model to better capture sudden regime shifts or transient momentum effects. Ultimately, by broadening equity coverage and fine-tuning long-short strategy designs, future studies can further substantiate the practical viability of Transformer-based models in real-world portfolio management.

References

1. *Reuters. (2025, January 27). China's DeepSeek sets off AI market rout. Retrieved from https://www.reuters.com/technology/chinas-deepseek-sets-off-ai-market-rout-2025-01-27/*
2. *Anderson, K., Brooks, C., & Katsaris, A. (2010). Speculative bubbles in the S&P 500: Was the tech bubble confined to the tech sector? Journal of Empirical Finance, 17(3), 345-361. https://doi.org/10.1016/j.jempfin.2009.11.002*
3. *Guyon, J., & Lekeufack Sopze, J. (2022, July 27). Volatility is (mostly) path-dependent. SSRN. Retrieved from https://ssrn.com/abstract=4174589 or http://dx.doi.org/10.2139/ssrn.4174589*
4. *Andrés, H., Boumezoued, A., & Jourdain, B. (2023). Implied volatility (also) is path-dependent. arXiv. Retrieved from https://arxiv.org/abs/2312.15950*
5. *Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. Journal of Financial Economics, 33(1), 3-56. Retrieved from https://www.bauer.uh.edu/rsusmel/phd/Fama-French_JFE93.pdf*
6. *Federal Reserve Economic Data (FRED). (n.d.). FRED API documentation. Federal Reserve Bank of St. Louis. Retrieved from https://fred.stlouisfed.org/docs/api/fred/*
7. *Wang, C., Chen, Y., Zhang, S., & Zhang, Q. (2022). Stock market index prediction using deep Transformer model. Expert Systems with Applications, 208, 118128. https://doi.org/10.1016/j.eswa.2022.118128*
8. *Yoo, J., Soun, Y., Park, Y., & Kang, U. (2021). Accurate multivariate stock movement prediction via data-axis Transformer with multi-level contexts. In Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining (KDD '21). https://doi.org/10.1145/3447548.3467265*
9. *Von Oswald, J., Niklasson, E., Randazzo, E., Sacramento, J., Mordvintsev, A., Zhmoginov, A., & Vladymyrov, M. (2023). Transformers learn in-context by gradient descent. Retrieved from https://arxiv.org/abs/2304.06969*
10. *TensorFlow. (n.d.). TensorFlow API documentation. Retrieved from https://www.tensorflow.org/api_docs/python/tf*