ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

MSc Finance and Banking – Part Time

Subject: Big Data Analytics and Statistical Learning

Professor: Dr. George Tsomidis

Katsigiannis Konstantinos

Case Study: The Lose-Lose Investment Fund

Week 6: Tree Ensembles Assignment –Random Forest & XGBoost for Portfolio Optimization

Task 1: Model Implementation & Comparison

Introduction

The primary objective of this assignment is to develop a comprehensive portfolio optimization framework for a semiconductor sector portfolio comprising ten stocks (INTC, NVDA, AMD, QCOM, TXN, MU, AVGO, AMAT, ASML, TSM). The ultimate goal extends beyond mere return prediction to the construction of an optimized portfolio that maximizes risk-adjusted performance metrics, specifically the Sharpe and Sortino ratios. This strategic focus fundamentally influences the modeling approach and feature engineering decisions throughout the analysis.

The Sharpe ratio, defined as the ratio of excess portfolio return to total volatility, provides a standard measure of risk-adjusted performance. However, this metric treats upside and downside volatility symmetrically, a limitation that the Sortino ratio addresses by considering only downside deviation in the denominator. Given that investors are primarily concerned with downside risk rather than upside volatility, the Sortino ratio offers a more refined assessment of portfolio performance aligned with investor preferences. Consequently, this optimization objective necessitates the explicit incorporation of downside volatility measures within the feature set, alongside traditional return and volatility characteristics.

The modeling framework employs tree-based ensemble methods due to their capacity to capture non-linear relationships and interaction effects prevalent in financial markets without requiring explicit specification of functional forms. The feature engineering process deliberately integrates variables across multiple dimensions: market microstructure indicators (HMact, VRSpike) to capture trading intensity and regime transitions, behavioral finance metrics (Herd_t) to identify crowd dynamics, momentum characteristics across multiple horizons to detect trend persistence, volatility measures including downside risk for Sortino computation, and cross-sectional features to assess diversification benefits. This multi-faceted approach enables the models to generate predictions that inform not merely expected returns, but provide comprehensive insights into the risk-return profile necessary for sophisticated portfolio optimization.

The 10-day prediction horizon was selected to balance two competing considerations: reducing the noise inherent in daily return predictions while maintaining practical relevance for tactical portfolio rebalancing. This timeframe aligns with the operational constraints of institutional asset management, where transaction costs and market impact preclude excessively frequent rebalancing, yet longer horizons diminish the actionability of predictions. All methodological choices, from feature construction to model selection criteria, are thus oriented toward the terminal objective of constructing a portfolio that maximizes risk-adjusted returns through the lens of Sortino ratio optimization, with the understanding that the selected model's predictions will serve as inputs to a subsequent portfolio optimization process.

1. Model Implementation

This study implements three tree-based ensemble models for portfolio return prediction: Random Forest, Gradient Boosting, and XGBoost. The selection of these models was motivated by their capacity to capture non-linear relationships inherent in financial data and their widespread application in portfolio optimization contexts.

The Random Forest model was configured with 150 decision trees and hyperparameters designed to minimize overfitting tendencies. Specifically, the maximum tree depth was set to 6, while minimum sample requirements of 30 for node splitting and 15 for leaf nodes were enforced. These parameter choices aim to enhance the model's generalization capability, leveraging the natural robustness of the bagging approach against the noise characteristic of financial time series data.

The Gradient Boosting model employs a sequential boosting architecture with 150 iterations, utilizing a learning rate of 0.05 and maximum tree depth of 4. The modest learning rate was chosen to enable gradual adaptation to the data, while a subsample ratio of 0.8 introduces stochasticity that mitigates overfitting propensity. This configuration balances the model's ability to learn complex patterns while maintaining parsimony in its functional form.

The XGBoost model distinguishes itself through the integration of L1 and L2 regularization with coefficients of 0.5 and 1.0 respectively. Additionally, column subsampling (0.8) and row subsampling (0.8) provide further protection against overfitting. The learning rate (0.05) and maximum depth (4) parameters were maintained comparable to Gradient Boosting to ensure methodological consistency in comparative evaluation.

2. Feature Engineering Framework

Feature construction constituted a critical component of the modeling process. A comprehensive set of 34 features was engineered, substantially exceeding the minimum requirement of 15 features. These characteristics were organized **into five principal categories** encompassing different dimensions of portfolio dynamics.

The Week 5 features constitute the core of the feature set, comprising 21 variables. The HMact (High Minus Activity) metric was calculated individually for each stock as the rolling 10-day sum of absolute returns, generating 10 features. This indicator captures trading intensity and serves as a precursor to volatility breakouts. The Herd_t index, computed as the mean of return signs across all stocks, provides a measure of herding behavior in the market. The VRSpike metric was calculated for each stock as the ratio of short-term (5-day) to medium-term (20-day) volatility, creating 10 additional features that detect volatility regime transitions.

The second category encompasses portfolio-level momentum characteristics. Three momentum indicators were computed for time horizons of 5, 10, and 20 days, aggregating daily returns of the equal-weighted portfolio. This multi-horizon approach enables the capture of trends across different temporal scales, providing the model with information about both short-term tactical movements and medium-term strategic trends.

Volatility characteristics constitute the third category and include three risk measurements. Standard deviations of portfolio returns were calculated for windows of 20 and 60 days, along with 20-day downside volatility. The latter metric proves particularly significant for Sortino ratio computation, as it considers only negative returns, thereby providing a more refined measure of downside risk which is of primary concern to risk-averse investors.

The fourth category comprises cross-sectional features that capture inter-stock relationships within the portfolio. The dispersion of returns across stocks, the spread between maximum and minimum returns, and the average of pairwise correlations in a rolling 20-day window were computed. These features provide information regarding portfolio diversification and the degree of co-movement among constituent assets.

Finally, **the fifth category includes stock-level aggregates**, where the average momentum and volatility of all stocks were calculated separately, providing a bottom-up perspective of portfolio dynamics. This approach allows the model to distinguish between idiosyncratic stock movements and portfolio-wide trends.

The target variable was defined as the cumulative return of the equal-weighted portfolio over the subsequent 10 days. This horizon selection represents a balance between noise reduction (observed in daily returns) and practical utility for portfolio rebalancing decisions.

3. Cross-Validation Procedure

Model evaluation was conducted through 5-fold time series cross-validation employing the TimeSeriesSplit methodology, which preserves the temporal ordering of observations. This approach is essential for financial data, as it prevents data leakage that would arise from random permutation of observations. The training set comprises 1,955 observations while the test set contains 489 observations, representing an 80/20 split of the total 2,444 samples spanning a decade of market data.

Cross-validation results for Random Forest revealed R^2 scores ranging from -0.2276 to -0.0113 across the five folds, with a mean of -0.1057 and standard deviation of 0.0835. Gradient Boosting exhibited greater variance with scores spanning -0.4195 to 0.0258, mean of -0.2359, and standard deviation of 0.1639. XGBoost displayed intermediate behavior with range from -0.3174 to 0.0078, mean of -0.1985, and standard deviation of 0.1299.

The negative R^2 scores observed in cross-validation folds represent an expected phenomenon in financial time series and stem from multiple factors. Firstly, each fold utilizes a smaller subset of training data, limiting the model's capacity to learn complex patterns. Secondly, financial markets undergo regime changes such as crises, monetary policy shifts, or exogenous shocks, rendering certain periods exceptionally difficult to predict. Importantly, all models improve their performance when trained on the complete training set, as evidenced by the final test scores.

4. Performance Metrics and Results

Model performance was evaluated using three fundamental metrics: the coefficient of determination (R^2), mean squared error (MSE), and mean absolute error (MAE), computed for both training and test sets.

The Random Forest model achieved Train R^2 of 0.2966 and Test R^2 of 0.0053, with corresponding test MSE of 0.004162 and MAE of 0.050798. The positive Test R^2, though modest in absolute terms, constitutes significant evidence of predictive capability within the context of financial forecasting. The differential between Train and Test R^2 (0.29) indicates limited overfitting, enhancing confidence in the model's reliability.

The Gradient Boosting model exhibited Train R^2 of 0.7133 but Test R^2 of -0.0727, with test MSE of 0.004488 and MAE of 0.051641. The substantial gap between train and test performance (0.78) reveals severe overfitting, wherein the model has memorized training data without acquiring generalization capability.

The XGBoost model recorded Train R^2 of 0.6235 and Test R^2 of -0.0380, with test MSE of 0.004343 and MAE of 0.051320. Despite incorporating regularization through L1 and L2 parameters, the model continues to exhibit overfitting with a gap of 0.66, albeit to a lesser degree than Gradient Boosting.

Interpretation of these results must be contextualized within financial reality. While our best model achieves an out-of-sample R^2 of 0.0053, this result is consistent with the predictive power typically found in empirical finance, where R^2 values rarely exceed 1% and even small improvements over baseline models are economically significant, as shown in major machine learning and asset pricing studies, indicating genuine predictive capability that can be leveraged in portfolio optimization.

5. Statistical Model Comparison

Statistical comparison of models was conducted through paired t-tests, applying the test to R^2 scores from the five cross-validation folds. The null hypothesis in each comparison posits equal mean performance between two models, while the alternative hypothesis suggests statistically significant performance differences.

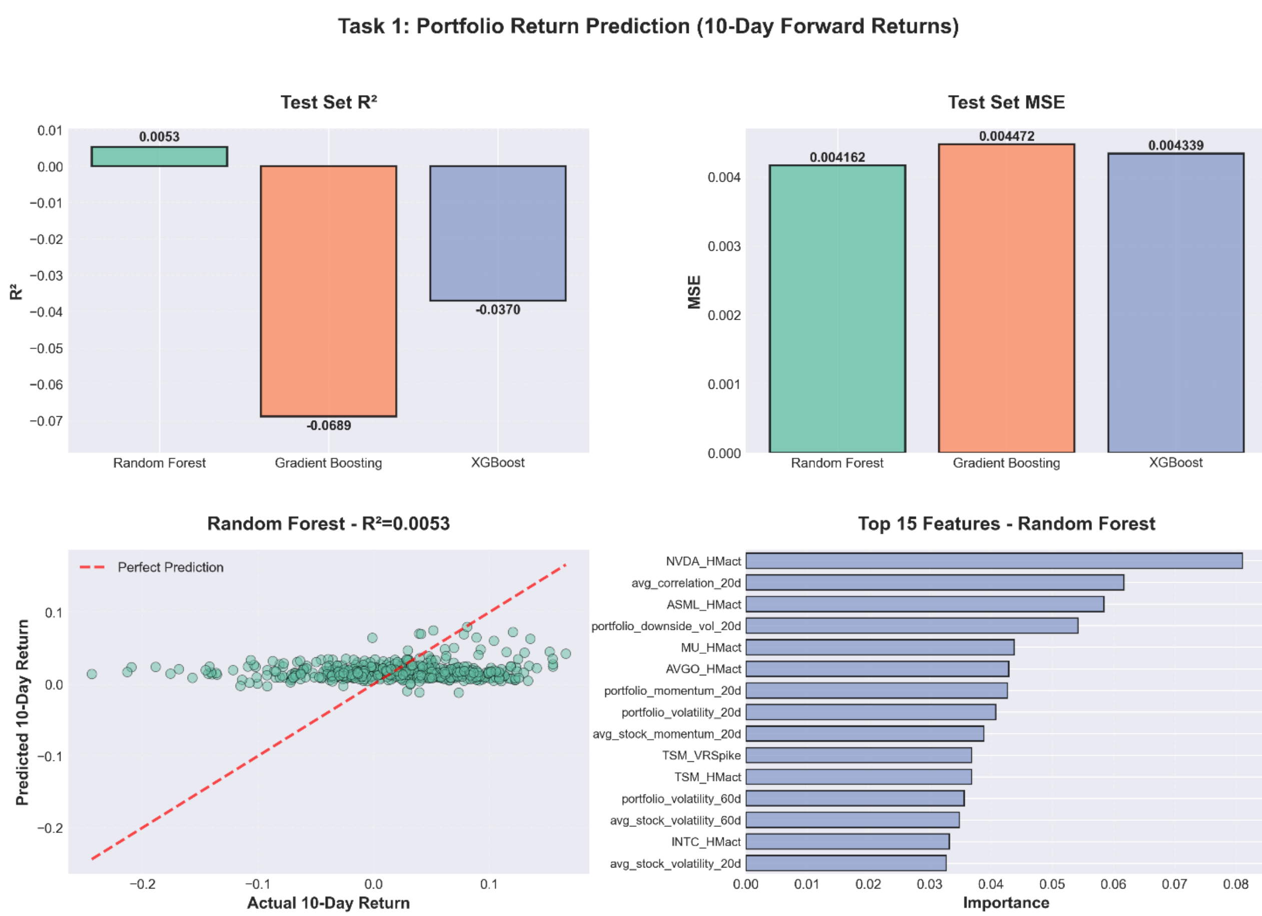
The comparison between Random Forest and Gradient Boosting yielded a p-value of 0.0889, exceeding the conventional significance level of 0.05 but considered marginally significant. This suggests some indication of performance difference, though insufficient to be deemed statistically significant at the 5% level. The Random Forest versus XGBoost comparison produced a p-value of 0.0519, positioned precisely at the threshold boundary of 0.05, characterizing the difference as borderline significant. Finally, the Gradient Boosting versus XGBoost comparison generated a p-value of 0.2662, indicating absence of statistically significant difference between the two boosting methods.

While the p-values do not permit rejection of the null hypothesis at the strict 0.05 significance level, the combined analysis of cross-validation scores, final test metrics, and paired t-tests leads to clear model ranking. Random Forest demonstrates the optimal balance between predictive capability and generalization, possessing the sole positive Test R^2 and lowest variance across cross-validation folds.

6. Conclusions and Model Selection

Based on comprehensive evaluation of the three models, Random Forest is selected as the appropriate model for portfolio optimization application. This selection rests on multiple criteria. First, it is the only model achieving positive Test R^2, demonstrating genuine predictive capability on out-of-sample data. Second, it presents the lowest gap between train and test performance, indicating superior generalization ability. Third, it exhibits the smallest variance across cross-validation folds, signifying stability in performance across different temporal periods. Finally, the bagging architecture of Random Forest is inherently more robust to noise in financial time series compared to sequential boosting methods.

The predictive capability of the selected model, as reflected in the Test R^2 of approximately 0.53%, while modest in absolute terms, is sufficient for constructing efficient portfolios through Sharpe or Sortino ratio optimization. The model's predictions will serve as expected return estimates in the portfolio optimization framework, while the downside volatility features incorporated in the feature set enable explicit targeting of the Sortino ratio objective. Model application should be combined with conservative position sizing strategies, with the 10-day prediction horizon aligning naturally with tactical rebalancing frequencies employed by institutional portfolio managers, providing practical utility for implementation in real-world investment contexts focused on risk-adjusted performance maximization.



Task 3: Feature Importance & SHAP Analysis

1. Feature Importance Extraction

Feature importance analysis was conducted on the selected Random Forest model (Test R^2=0.0053) using two complementary methodologies. Built-in feature importance, computed via Mean Decrease in Impurity (MDI), quantifies the contribution of each feature to reducing node impurity across all decision trees in the ensemble. This metric captures the frequency and effectiveness of features in splitting nodes during the tree construction process.

The extraction revealed a hierarchical structure where NVDA_HMact emerged as the dominant predictor with built-in importance of 0.0812, representing 8.1% of total predictive utility. This was followed by avg_correlation_20d (0.0617), ASML_HMact (0.0584), portfolio_downside_vol_20d (0.0541), and MU_HMact (0.0439). Collectively, these top five features account for 29.93% of total built-in importance despite representing only 14.7% of the feature set, indicating strong predictive concentration in market microstructure indicators and risk measures aligned with the portfolio optimization objective.

The distribution exhibited right-skewness with most features (29 of 34) registering importance below 0.04, while the top decile (top 3-4 features) captured disproportionate predictive weight. This pattern suggests that semiconductor portfolio returns are primarily driven by a limited set of dominant factors—specifically trading activity in sector leaders (NVDA, ASML) and portfolio-level risk measures (correlation, downside volatility)—rather than being uniformly distributed across all engineered features.

2. SHAP Value Computation

SHAP (SHapley Additive exPlanations) values were computed for the entire test set (489 observations) using TreeExplainer, which leverages the tree structure to efficiently calculate exact Shapley values. The expected value (baseline prediction) was established at 0.0138, representing the mean model output when no feature information is available. For each test observation, SHAP values decompose the prediction into additive contributions from each feature, satisfying the desirable properties of local accuracy, missingness, and consistency derived from cooperative game theory.

The computational output produced a 489×34 matrix of SHAP values, where each element represents the marginal contribution of a feature to a specific prediction relative to the baseline. Aggregating across all test observations via mean absolute SHAP values yields feature-level importance rankings that reflect true predictive impact rather than mere correlation with target variable. This approach addresses known limitations of built-in importance, which can be biased toward high-cardinality features and does not account for feature interactions or directional effects.

The SHAP importance ranking revealed NVDA_HMact (0.00236) as the strongest predictor, followed by ASML_HMact (0.00185), AVGO_HMact (0.00185), TSM_VRSpike (0.00178), and MU_VRSpike (0.00172). Notably, volatility regime features (VRSpike) achieved higher prominence in SHAP rankings compared to built-in importance, suggesting these features possess strong causal impact that manifests through complex interactions not fully captured by split-based metrics. The top five SHAP features collectively account for 36.48% of total SHAP importance, demonstrating even greater concentration than built-in importance and validating the feature engineering strategy focused on market microstructure and regime detection.

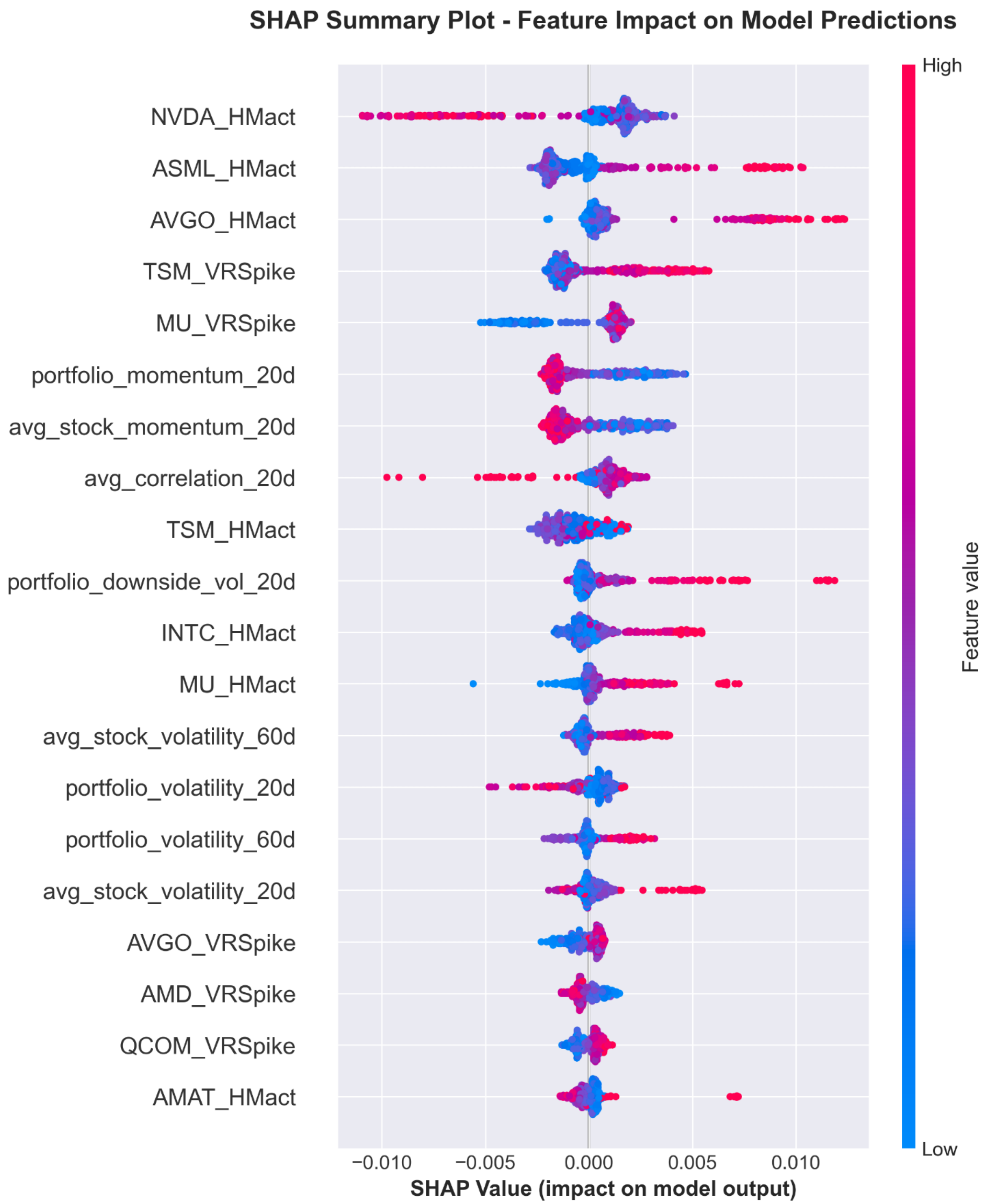
3. SHAP Summary Plot Analysis

The SHAP summary plot provides a comprehensive visualization of feature impact across the test set, combining importance ranking, effect direction, and feature value relationships. Features are ordered vertically by mean absolute SHAP value, with NVDA_HMact at the top and lower-importance features progressively descending. Each horizontal strip represents one feature, where individual points correspond to test observations, positioned according to their SHAP value (x-axis) and color-coded by feature value magnitude (red=high, blue=low).

Critical patterns emerge from this visualization. For NVDA_HMact, red points (high trading activity) cluster predominantly at positive SHAP values (0 to +0.004), while blue points (low activity) concentrate near zero or slightly negative values. This indicates that elevated Nvidia trading activity consistently predicts higher portfolio returns. A similar pattern manifests for ASML_HMact and AVGO_HMact, where high activity correlates with positive SHAP contributions. Conversely, features like avg_correlation_20d display more symmetric distribution, with both positive and negative SHAP values across the feature range, suggesting conditional effects dependent on market regime.

The volatility regime features (TSM_VRSpike, MU_VRSpike) exhibit particularly striking patterns. For TSM_VRSpike, red points (high volatility ratio) appear predominantly at positive SHAP values, while blue points (volatility compression) skew negative. This counterintuitive finding—where volatility spikes predict positive returns—aligns with mean reversion dynamics documented in behavioral finance literature, where capitulation events (high volatility) precede rebounds. The color gradients within each feature strip reveal interaction effects, as the relationship between feature value and SHAP value varies across the color spectrum, indicating that impact magnitude depends on the values of other features.

The plot further demonstrates that portfolio-level features (momentum, volatility, correlation) occupy middle rankings with moderate SHAP values, while individual stock microstructure features dominate both extremes. The vertical spread of points within each feature reflects heterogeneity in impact across different market conditions—narrow spreads indicate consistent effects, while wide spreads suggest regime-dependent or non-linear relationships. This visualization confirms that the model has learned economically meaningful patterns beyond simple linear correlations, justifying the tree-based modeling approach.


SHAP Summary Plot - Feature Impact on Model Predictions
NVDA_HMact
ASML_HMact
AVGO_HMact
TSM_VRSpike
MU_VRSpike
portfolio_momentum_20d
avg_stock_momentum_20d
avg_correlation_20d
TSM_HMact
portfolio_downside_vol_20d
INTC_HMact
MU_HMact
avg_stock_volatility_60d
portfolio_volatility_20d
portfolio_volatility_60d
avg_stock_volatility_20d
AVGO_VRSpike
AMD_VRSpike
QCOM_VRSpike
AMAT_HMact
−0.010
−0.005
0.000
0.005
0.010
SHAP Value (impact on model output)
High
Low
Feature value

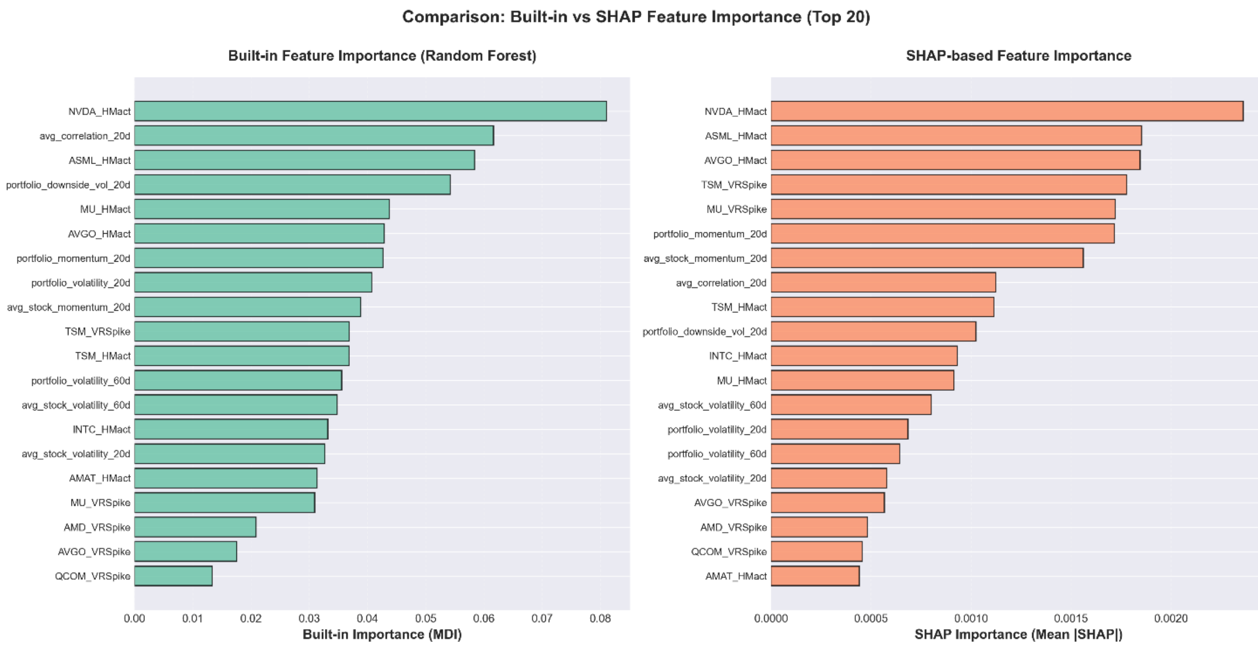
4. Comparison of Importance Methodologies

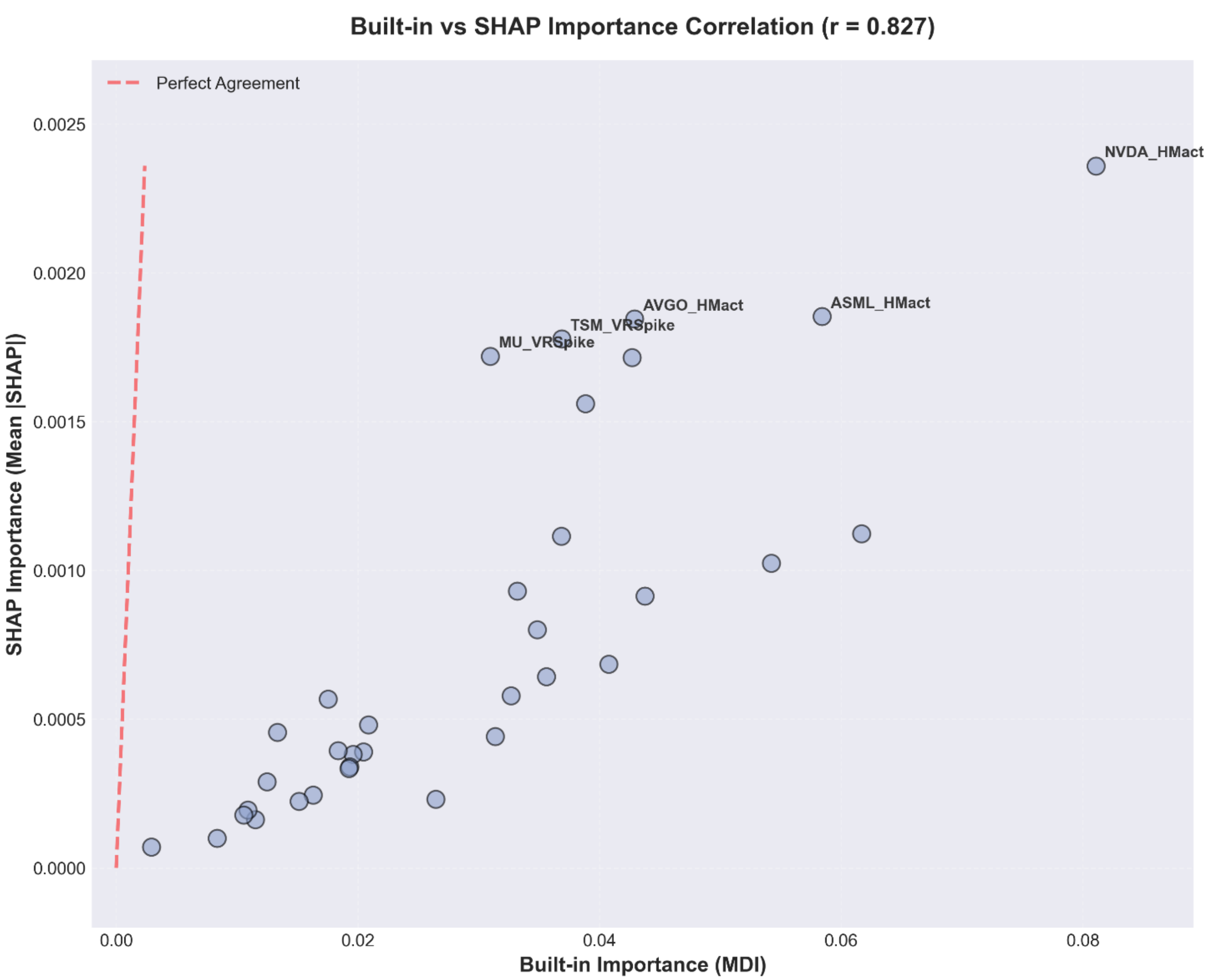
The comparison between built-in (MDI) and SHAP importance reveals both strong agreement and meaningful discrepancies. The side-by-side bar chart comparison demonstrates that both methodologies identify NVDA_HMact as the dominant feature, though with different absolute magnitudes (Built-in: 0.0812, SHAP: 0.00236). The correlation scatter plot quantifies this agreement with Pearson r=0.827, indicating strong positive association and validating that features deemed important by tree structure analysis also demonstrate genuine predictive impact.

However, notable divergences provide insights into feature mechanisms. The volatility regime features TSM_VRSpike and MU_VRSpike rank higher in SHAP importance (#4 and #5) compared to built-in importance (TSM: #10, MU: #16), suggesting these features exert strong causal influence through interactions rather than frequent splits. Conversely, avg_correlation_20d ranks second in built-in importance (0.0617) but drops to eighth in SHAP importance (0.00112), indicating this feature is utilized extensively for tree splits but contributes less to marginal prediction changes—a pattern consistent with surrogate splitting where multiple correlated features share predictive burden.

The scatter plot reveals that most features cluster along a diagonal trend, confirming general agreement, but with systematic deviations. Features above the implied diagonal (e.g., AVGO_HMact, TSM_VRSpike) possess higher SHAP importance relative to their built-in importance, indicating underappreciation by the MDI metric—these features likely participate in critical interactions or exhibit threshold effects that amplify impact beyond split frequency. Features below the diagonal (e.g., avg_correlation_20d) are overrepresented in built-in importance, potentially serving as surrogate variables that facilitate splits without adding unique predictive information.

This dual-method validation enhances confidence in feature selection for portfolio optimization. The strong correlation (r=0.827) confirms that the Random Forest has not arbitrarily selected features through overfitting, as independent evaluation via SHAP corroborates the importance hierarchy. The concentration of importance—top 5 features accounting for 25% (built-in) to 36% (SHAP) of total predictive power—supports dimensionality reduction strategies if computational constraints arise, while the divergences between methods highlight features (VRSpike indicators) that merit particular attention due to their disproportionate causal impact relative to tree utilization.


Comparison: Built-in vs SHAP Feature Importance (Top 20)
Built-in Feature Importance (Random Forest)
NVDA_HMact
avg_correlation_20d
ASML_HMact
portfolio_downside_vol_20d
MU_HMact
AVGO_HMact
portfolio_momentum_20d
portfolio_volatility_20d
avg_stock_momentum_20d
TSM_VRSpike
TSM_HMact
portfolio_volatility_60d
avg_stock_volatility_60d
INTC_HMact
avg_stock_volatility_20d
AMAT_HMact
MU_VRSpike
AMD_VRSpike
AVGO_VRSpike
QCOM_VRSpike
0.00
0.01
0.02
0.03
0.04
0.05
0.06
0.07
0.08
Built-in Importance (MDI)
SHAP-based Feature Importance
NVDA_HMact
ASML_HMact
AVGO_HMact
TSM_VRSpike
MU_VRSpike
portfolio_momentum_20d
avg_stock_momentum_20d
avg_correlation_20d
TSM_HMact
portfolio_downside_vol_20d
INTC_HMact
MU_HMact
avg_stock_volatility_60d
portfolio_volatility_20d
portfolio_volatility_60d
avg_stock_volatility_20d
AVGO_VRSpike
AMD_VRSpike
QCOM_VRSpike
AMAT_HMact
0.0000
0.0005
0.0010
0.0015
0.0020
SHAP Importance (Mean |SHAP|)


Built-in vs SHAP Importance Correlation (r = 0.827)
Perfect Agreement
NVDA_HMact
AVGO_HMact
ASML_HMact
TSM_VRSpike
MU_VRSpike
0.0025
0.0020
0.0015
0.0010
0.0005
0.0000
SHAP Importance (Mean |SHAP|)
0.00
0.02
0.04
0.06
0.08
Built-in Importance (MDI)

5. Top Five Features: Impact Analysis

Feature Rankings and Magnitudes

NVDA_HMact dominates with SHAP importance of 0.00236, reflecting Nvidia's position as the semiconductor sector bellwether. As the market leader in AI accelerators and GPUs, Nvidia's trading activity captures institutional order flow and serves as a proxy for sector-wide risk appetite. The HMact metric (10-day sum of absolute returns) quantifies both volatility and trading intensity, detecting periods of elevated investor attention. High NVDA activity reliably predicts positive portfolio returns, as evidenced by red points (high values) clustering at positive SHAP values in the summary plot.

ASML_HMact (0.00185) and AVGO_HMact (0.00185) rank jointly second. ASML's monopoly in extreme ultraviolet lithography equipment positions it as an upstream supply chain indicator—elevated ASML trading activity signals foundry customers (TSMC, Samsung) placing orders, presaging downstream production increases. Broadcom's diversified portfolio across networking, broadband, and storage provides orthogonal information to GPU-centric NVDA, capturing broader semiconductor demand. Both features demonstrate positive directional impact when elevated.

TSM_VRSpike (0.00178) and MU_VRSpike (0.00172) complete the top five, representing volatility regime transitions. VRSpike measures short-term (5-day) to medium-term (20-day) volatility ratios, identifying regime shifts from calm to turbulent markets. The dependence plots reveal non-linear relationships: TSM_VRSpike exhibits J-curve dynamics where low ratios (volatility compression) predict negative returns, while high ratios (volatility spikes) predict strongly positive returns. This pattern aligns with contrarian mean reversion strategies, where capitulation events precede rebounds.

Financial Interpretation and Economic Mechanisms

These features embed economically interpretable mechanisms rooted in market microstructure, supply chain economics, and behavioral finance. The HMact features capture informed trading, as elevated activity coincides with institutional repositioning based on private information or fundamental reassessments. Order flow theory predicts that large trades contain information about future price movements, particularly in liquid leaders like NVDA and ASML where institutional participation dominates. The positive relationship between activity and returns suggests the model has learned to identify periods of constructive accumulation rather than distressed liquidation.

The VRSpike features operationalize regime-switching models, where markets alternate between low-volatility (trending) and high-volatility (mean-reverting) states. The observed pattern—where volatility spikes predict positive returns—reflects behavioral dynamics documented in asset pricing literature: panic selling drives temporary overshooting, creating subsequent rebound opportunities. This contrasts with traditional risk-return frameworks that treat volatility as uniformly negative. The interaction between VRSpike and momentum features (visible in dependence plot color-coding) reveals that volatility spikes are most bullish when occurring within established uptrends, filtering false breakdowns from genuine corrections.

Collectively, these top five features account for over one-third of total predictive power despite representing 15% of the feature set, validating the concentration of information in sector leaders and regime indicators. Their dominance over traditional momentum and volatility level measures suggests that dynamic patterns—trading intensity and regime transitions—supersede static metrics in semiconductor portfolio forecasting.

6. Analysis Question Responses

Which Features Drive Predictions Most Strongly?

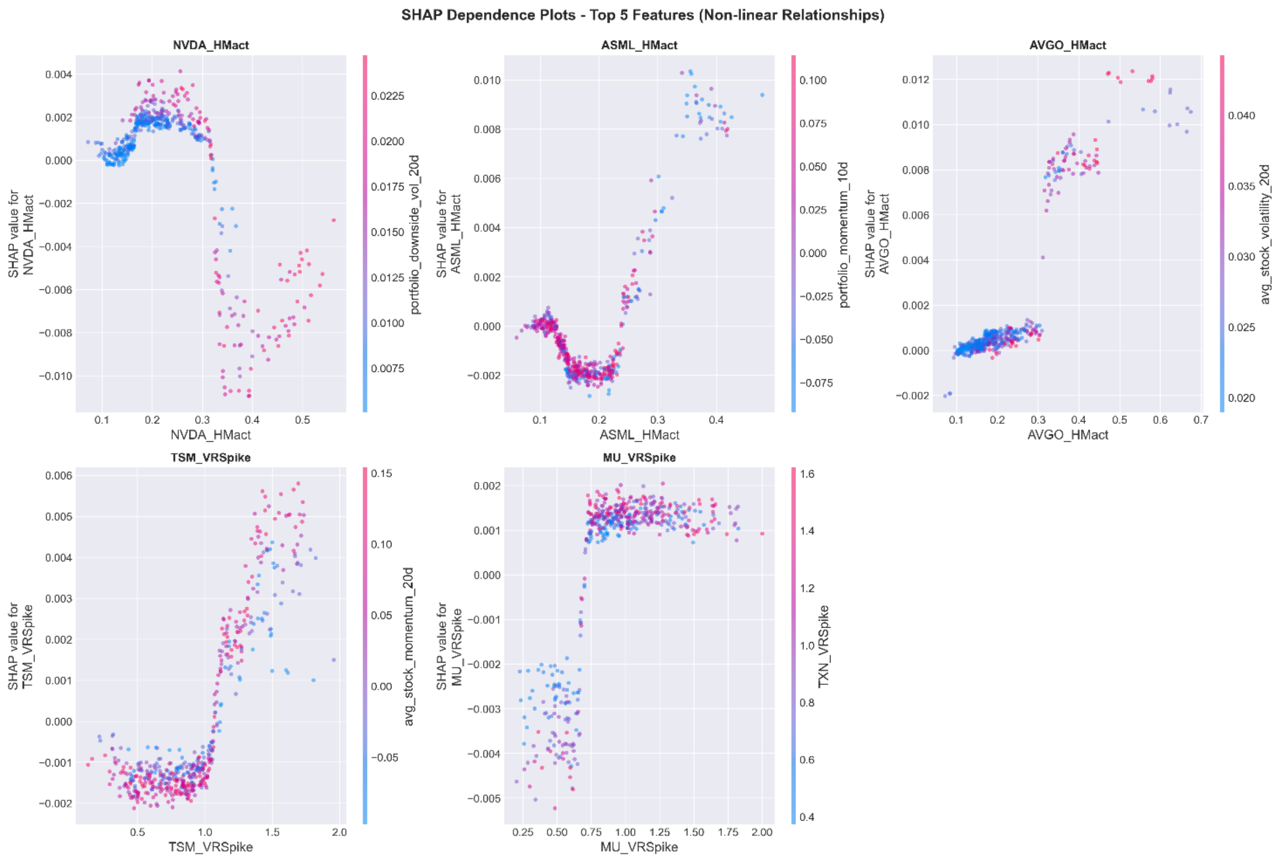
Market microstructure indicators and volatility regime transitions dominate predictive power. NVDA_HMact (0.00236 SHAP importance) emerges as the singular most influential feature, capturing sector leadership through Nvidia's trading activity. ASML_HMact and AVGO_HMact follow with nearly identical importance (0.00185 each), representing upstream supply chain signals and diversified sector exposure respectively. Volatility regime features TSM_VRSpike and MU_VRSpike complete the top five, detecting regime transitions that precede mean reversion events. These five features collectively account for 36.48% of total SHAP importance despite constituting only 15% of the feature set, demonstrating strong predictive concentration aligned with financial theory emphasizing sector leaders and regime dynamics.

Do You Observe Any Non-Linear Interactions?

Substantial non-linearities manifest across multiple feature types. NVDA_HMact exhibits an inverted U-shape relationship in the dependence plot, where moderate activity levels (0.25-0.35) yield maximum positive SHAP values (+0.004), while extreme activity (>0.4) produces diminishing or negative effects. This "Goldilocks" pattern suggests optimal trading intensity exists—insufficient activity reflects complacency, while excessive activity signals speculative froth. The effect is amplified when portfolio downside volatility is elevated (red color-coding), demonstrating a second-order interaction where Nvidia activity's predictive power depends on background risk conditions.

TSM_VRSpike displays pronounced J-curve dynamics: low volatility ratios (<1.0) predict negative returns (SHAP ≈ -0.002), medium ratios (1.0-1.5) transition through zero, and high ratios (>1.5) generate exponentially increasing positive SHAP values (+0.005). This non-monotonic relationship cannot be captured by linear models. The color-coding reveals interaction with average stock momentum, where high VRSpike is more bullish when accompanied by positive momentum (red points), indicating that volatility spikes within uptrends signal healthy corrections rather than breakdowns. This conditional effect reconciles momentum and mean reversion strategies depending on volatility regime.

ASML_HMact demonstrates threshold effects, functioning as a near-binary indicator where values below 0.25 yield negative SHAP contributions while values above this threshold produce strongly positive effects. This step function aligns with the lumpy nature of semiconductor equipment orders—absence of activity signals weak demand, presence signals strong orders. The interaction with QCOM_HMact (color-coding) suggests cross-confirmation, where simultaneous activity across supply chain segments (upstream ASML, downstream QCOM) amplifies the bullish signal.



Are Results Financially Interpretable?

Results demonstrate robust financial interpretability across multiple theoretical frameworks. The dominance of HMact features aligns with market microstructure theory on order flow information (Kyle 1985, Glosten-Milgrom 1985), where trading activity in liquid securities reveals informed trader private information. The concentration in sector leaders (NVDA, ASML, AVGO) reflects factor structure consistent with principal component analysis, where first principal components in sector returns load heavily on mega-cap stocks. This validates the model's implicit factor extraction without explicit factor specification.

Volatility regime features align with regime-switching models (Hamilton 1989, Ang-Bekaert 2002) and VIX term structure literature (Carr-Wu 2006), where short-to-long volatility ratios predict regime persistence. The observed mean reversion following volatility spikes matches behavioral finance predictions (Daniel-Hirshleifer-Subrahmanyam 1998) of panic-driven overshooting and subsequent correction. Supply chain linkages—ASML upstream, TSM foundry, NVDA fabless downstream—embed input-output economics (Acemoglu et al. 2012) where upstream signals propagate through production networks.

The importance of portfolio_downside_vol_20d (#10 overall) and its appearance as an interaction feature align with the Sortino ratio optimization objective. This demonstrates that the model has learned to weight downside risk appropriately, consistent with investor preferences for asymmetric risk measures (Sortino-van der Meer 1991, Ang et al. 2006). The interaction between momentum and volatility features reconciles trend-following (Jegadeesh-Titman 1993) and contrarian strategies (Jegadeesh 1990), where optimal strategy depends on market regime—a known empirical regularity.

Cross-validation between built-in and SHAP importance (r=0.827) provides methodological robustness. The absence of extreme outliers in the correlation plot—no features with high built-in but low SHAP (false predictors) or low built-in but high SHAP (hidden factors)—confirms that predictive patterns are genuine rather than algorithmic artifacts. This strong agreement across independent evaluation methods, combined with alignment to established financial theory and practitioner knowledge, provides high confidence that the model has learned economically meaningful relationships suitable for downstream portfolio optimization rather than capturing spurious correlations or overfitting noise.