Machine Learning Approach

Kevin J. DiCiurcio, Boyu Wu, Fei Xu, Scott Rodemer, and Qian Wang

Kevin J. DiCiurcio
is the head of Vanguard Capital Markets Model Development within Vanguard's Investment Strategy Group at The Vanguard Group, Inc. in Malvern, PA.
kevin_j_diciurcio@vanguard.com

Boyu Wu
is a senior investment strategist within Vanguard's Investment Strategy Group at The Vanguard Group, Inc. in Malvern, PA.
boyu_wu@vanguard.com

Fei Xu
is a senior active equity portfolio manager within Vanguard's Quantitative Equity Group at The Vanguard Group, Inc. in Malvern, PA.
fei_xu@vanguard.com

Scott Rodemer
is the head of Factor-Based Strategies within Vanguard's Quantitative Equity Group at The Vanguard Group, Inc. in Malvern, PA.
scott_rodemer@vanguard.com

Qian Wang
is a principal, regional chief economist for Asia-Pacific and the global head of the Vanguard Capital Markets Model within Vanguard's Investment Strategy Group at The Vanguard Group, Inc. in Malvern, PA.
qian_wang@vanguard.com

KEY FINDINGS

- The two-stage approach significantly enhances factor prediction outcomes, underlining the importance of determining the market risk regime prior to evaluating factor performance. This finding highlights the necessity of incorporating market conditions in dynamic factor investment strategies.
- Market risk signals (financial turbulence) and financial conditions emerge as critical features for determining factor performance across different market risk regimes. This insight underscores the value of incorporating these features when developing factor prediction models to better navigate varying market conditions.
- The incorporation of both macroeconomics and market risk factors further improves the model's predictive capabilities. This finding suggests that integrating both general market features and factor-specific attributes can lead to more effective factor investment strategies, ultimately providing investors with a more comprehensive understanding of factor dynamics.

ABSTRACT

Equity factor investing has gained traction due to its ability to systematically capture premia for risk or behavioral reasons. However, developing a robust factor timing investment framework remains challenging. In this article, the authors propose a two-stage machine model for dynamic factor rotation, which adapts to varying market conditions. In the first stage, the authors employ both supervised and unsupervised machine learning techniques to identify dynamic market risk regimes, which reflect the prevailing economic environment. Subsequently, the second stage utilizes additional ensemble supervised machine learning methods, incorporating the features identified in the first stage, to predict factor performance within each regime. The authors' findings demonstrate that the proposed model delivers robust results across all regimes. Consequently, this hybrid machine learning approach offers an innovative alternative for dynamic factor investment strategies, providing investors with the tools to navigate diverse market conditions.

Factor investing has gained significant attention within the finance community over the past decade, revolutionizing traditional investment approaches. Fama and French's (1992) seminal study introduced the concept of factor-based investing, a multi-factor extension to Sharpe's Capital Asset Pricing model (CAPM) (Sharpe 1964), showing that investors could achieve enhanced returns by focusing on specific factors that influence asset pricing and performance. These factors include size, value, and momentum, and they have since become foundational components of modern portfolio construction.

As the field of factor investing evolved, academics and practitioners identified the set of factors that help characterize equity styles. These factors include size, which captures the impact of company size on returns; value, reflecting the potential undervaluation of certain assets; growth, indicating the growth potential of companies; quality, assessing the financial stability and profitability of firms; momentum, highlighting the price trend of securities; and leverage, gauging the impact of financial leverage on returns, among many others (Fama and French 2016). The systematic categorization of these factors allows investors to tailor their strategies according to specific investment goals and risk tolerance.

While a substantial body of research has examined the long-term return differentials among these factors (Banz 1981, Harris and Marston 1994), relatively fewer academic studies have delved into understanding the driving forces behind cyclical style return variations. This knowledge gap represents an exciting opportunity for further exploration and refinement of factor-based investment strategies. Additionally, tactical style allocation, which involves dynamically adjusting factor exposures based on market conditions, remains an area with untapped potential. Few studies have effectively explored whether and how tactical allocation strategies can improve risk-adjusted returns and enhance the overall performance of factor-based portfolios.

Although the ability to outperform a benchmark by accurately timing these dimensions remains debatable, understanding how factors perform under different market conditions and economic cycles can help investors design more robust and adaptive portfolios, which has led to a burgeoning area of research exploring the potential to optimize investment outcomes (enhance returns and reduce risk) through factor timing, or systematically shifting exposure to these various factors based on their expected performance.

As proposed by Ilmanen and Kizer (2012), accurate factor timing has the potential to significantly enhance portfolio returns and reduce risk. Their research provided evidence that asset class diversification only offers modest improvements to the portfolio's Sharpe ratio, whereas factor diversification proves highly effective due to the significantly lower average correlation among factors (nearly zero, around –0.02). Consequently, they concluded that factor diversification more effectively reduces portfolio volatility and market directionality. Despite the promising benefits of factor timing, accurately predicting shifts in factor performance remains a formidable challenge. Historically, many investors have relied on intuitive or ad hoc methods to guide their factor timing decisions. Our research aims to offer alternative solutions to this complex problem, seeking to provide valuable insights into more reliable and data-driven approaches for factor timing strategies.

Previous research has established a strong link between macroeconomics and the relative performance of different factors in the financial markets. Kao and Shumaker (1999) demonstrated that factors' outperformance can be attributed to specific economic fundamentals, business cycles, and trends in corporate earnings. Many factor timing models incorporate key macroeconomic variables to better understand and predict factor performance. Among the essential macro variables included in factor timing models are the yield-curve spread and real bond yield. These variables may significantly influence valuations given their implications toward corporate earnings, investment spending, and discount rates. Additionally, GDP growth plays a crucial role, as it reflects corporate profit cycles. During economic expansions, companies tend to experience higher profit growth, and value stocks, which benefit from operating leverage, are more likely to outperform during such periods (Choi 2013). Another vital factor is the equity earning yield gap, which represents the difference between bond yields and the market earnings yield. In environments with low earnings yield and high interest rates, value-related stocks tend to be favored. Furthermore, historical Consumer Price Index (CPI) data can significantly impact factor performance due to its effect on inflation expectations. Inflation can influence interest rates, affecting bond yields and the attractiveness of value-related stocks in comparison to other investments (Maio and Santa-Clara 2017). When the CPI shows signs of increasing, investors might favor inflation-hedging assets, which could impact the relative performance of different factors. Moreover, changes in macroeconomic indicators like employment data and consumer sentiment can also play a vital role in factor performance. For instance, during periods of robust job growth and high consumer confidence, growth-oriented factors may outperform, as investors have higher expectations for corporate earnings and economic expansion. Furthermore, changes in global economic conditions and geopolitical events can have far-reaching effects on financial markets, including factor performance. Geopolitical tensions, trade disputes, or shifts in international monetary policies can lead to fluctuations in currency valuations and asset prices, potentially impacting factors with significant exposure to international markets.

Combining macroeconomic and cyclic views, a common approach among practitioners and academics involves leveraging the business cycle to time factors. In their comprehensive study, Polk, Haghbin, and de Longis (2020) offer valuable insights into factor cyclicality. The central concept revolves around interpreting signals that anticipate changes in the business cycle through a factor lens. If a signal suggests positive future market fundamentals, strategies with relatively high cash-flow betas become appealing. Conversely, if the signal indicates negative future market fundamentals, strategies with relatively low cash-flow betas become more attractive. The researchers considered five common factors: Low volatility, size, value, momentum, and quality in their analysis. Their research findings allowed them to implement a business cycle-based factor timing approach with strategic overweighting of factors during different stages of the business cycle. During the recovery stage, they favored size and value factors, while momentum was prioritized during the expansion stage. In the slowdown stage, low volatility and quality factors were preferred, and during the contraction stage, the focus was on overweighting low volatility and quality factors. The application of this business cycle-based factor timing approach resulted in outperforming their benchmark, the Russell 1000 R1 comprehensive factor index. This demonstrates the practical significance of integrating macro views and cyclic perspectives when constructing factor-based investment strategies.

Relying solely on the business cycle for factor timing may oversimplify the complexities of market dynamics. Factors beyond the usual pattern can disrupt the business cycle signal, leading to suboptimal timing decisions. Moreover, the slow-moving nature of macroeconomic variables can introduce lag in the strategy. To address these limitations, our research endeavors to refine factor timing by incorporating market risk regimes and additional risk factors into the model. Notably, we consider the financial turbulence ratio, a crucial determinant of systematic risk identified by Kritzman and Li (2010). Additionally, our model integrates other macroeconomic, fundamental, and state variables, such as financial conditions, business cycle labels, equity market valuations, and Federal Reserve policy rate expectations. By incorporating these factors, we aim to enhance the precision and adaptability of our factor timing approach, enabling investors to make more informed decisions and navigate market fluctuations with greater confidence.

The preliminary results of our research are encouraging. By considering market risk regimes, financial turbulence ratios, and other macro variables, our model significantly improves the accuracy of factor timing decisions compared to equally weighted and pure business cycle-based approaches. This suggests that our approach could provide a valuable tool for investors seeking to enhance their factor timing strategies.

However, further research is required to validate and refine our model, and to explore the potential for additional enhancements.

This article is structured as follows: The methodology section details the construction of the two-step approach hybrid learning framework. This involves utilizing clustering for market risk regime creation and prediction, as well as supervised learning on each sub-market risk regime to predict the probability of factor relative performance. The feature construction and selection section present financial turbulence calculation methodology and the rationale for selecting macroeconomic variables. The model results and feature analysis section detail the model performance and overall feature importance analysis. The portfolio simulation section offers out-of-sample portfolio results, using equally weighted factors and a business cycle-only approach as benchmarks to demonstrate how our model compares to other common approaches. Finally, the conclusion section presents our research findings along with conclusions and recommendations.

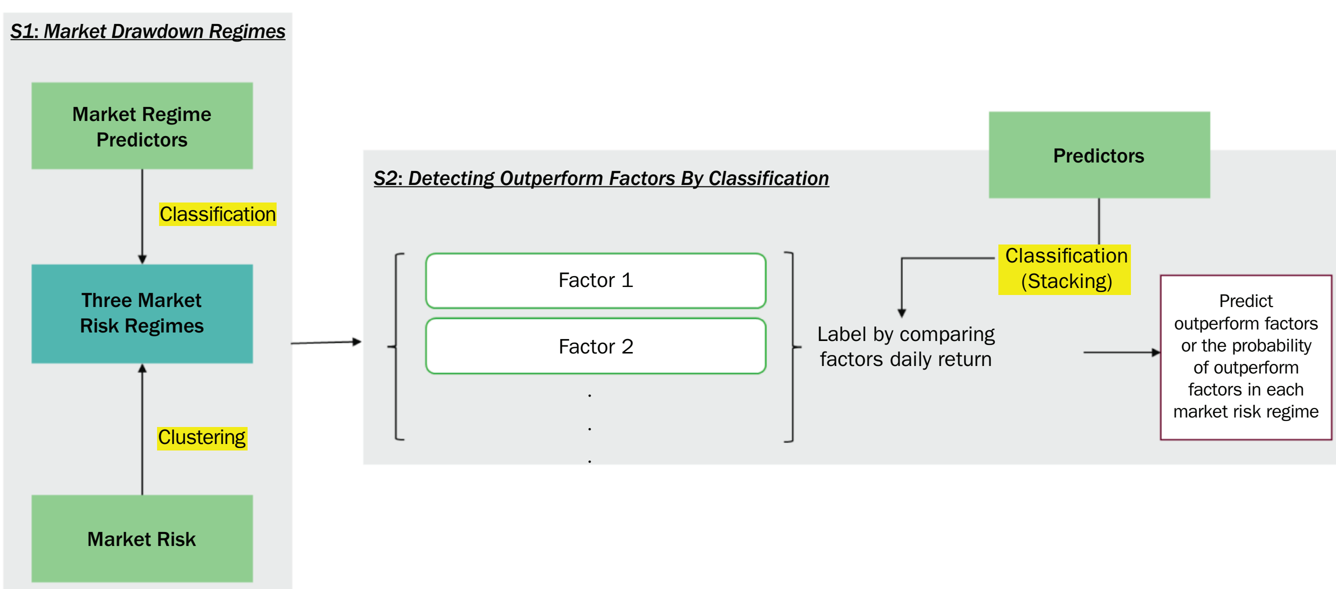
METHODOLOGY

The Overview of Two-Stage Machine Learning Framework and Business Cycle Benchmark

Exhibit 1 showcases our dual-stage factor timing model. In the initial stage, we use K-means clustering to categorize market risk regimes, focusing on S&P 500 maximum drawdown data. This results in distinct clusters reflective of varying market conditions. We then employ a variety of classification techniques to predict these regimes, using the labels we've generated. We also utilize decision-tree-based feature importance functions to highlight key drivers within each market risk regime. To address the issue of an imbalanced dataset, we implement oversampling techniques, including SMOTE, to create a balanced dataset that minimizes model bias.

EXHIBIT 1

The Two-Stage Machine Learning Framework



In the second stage, we utilize a relative return factor classification model. The entire dataset is segmented into sub-dataframes, each corresponding to the market risk regimes identified in the first stage. We then evaluate the monthly performance of each factor, designating the best-performing factor as '1' and the others as '0'. These labels serve as the groundwork for the classification model in this stage. Our models are trained on both macroeconomic variables and financial turbulence indicators. Once trained, these models enable us to forecast the likelihood of each factor outperforming, thus supporting a data-driven factor timing strategy. This refines our decision-making and optimizes factor allocations according to market risk, thereby boosting portfolio performance.

For our out-of-sample testing purposes, we employ a rolling window model training approach to enhance the model's dynamism, rather than using extending windows. As a benchmark for business cycle factor timing, we rely on leading business cycle data from The Conference Board. Factors are then overweighted or underweighted based on their full sample historical average performance within each business cycle stage using point-in-time economic data.

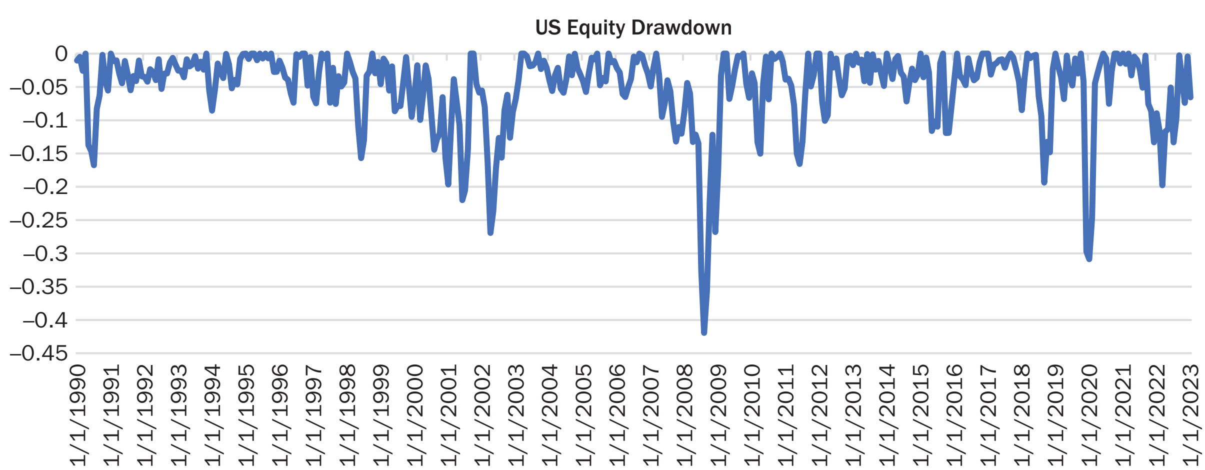
FEATURE SELECTION AND FEATURE CONSTRUCTION

Data and Preliminary Statistics

Our sample dataset starts on March 1987 and ends in March 2023, which yields a total of 433 monthly observations. We obtain equity and Treasury return related variables from Refinitiv Eikon Datastream and macroeconomic variables from the St. Louis Federal Reserve's database (FRED). The market drawdown of S&P 500 is shown in Exhibit 2. As mentioned previously, we settled on a combination of factors that contain macroeconomics information about inflation, real GDP growth, financial condition, and policy expectations. Details of the 25 macro drivers considered for

EXHIBIT 2

S&P 500 Three Months' Rolling Drawdown



NOTES: Past performance is no guarantee of future returns. The performance of an index is not an exact representation of any particular investment, as you cannot invest directly in an index.

SOURCES: Authors' calculations based on S&P 500 price data sourced from Factset. The S&P 500 is a product of S&P Dow Jones Indices LLC, a division of S&P Global, or its affiliates ("SPDJI").

EXHIBIT 3

Macroeconomic and Financial Market Factors Considered for Machine Learning Testing

Factors	Rationale	Variables Considered for Machine Learning Testing
Inflation	Inflation can significantly impact factor performance due to its effect on inflation expectations. Inflation can influence interest rates, affecting bond yields and the attractiveness of value-related stocks in comparison to other investments.	Trailing ten-year annualized changes in core CPI Trailing one-year annualized changes in core CPI Sources: FRED, Federal Reserve Economic Data
GDP Growth/ Business Cycle	GDP growth plays a crucial role, as it reflects corporate profit cycles. During economic expansions, companies tend to experience higher profit growth, and value stocks, which benefit from operating leverage, are more likely to outperform during such periods.	US Real GDP Growth Conference Board Business Cycle Labels Sources: FRED, Federal Reserve Economic Data
Financial Conditions	The impact of financial conditions on factor performance is a crucial aspect in the realm of financial markets. Fluctuations in financial conditions, such as interest rates, credit availability, and market volatility, can significantly influence the investment landscape.	Financial Condition – Risk Financial Condition – Credit Financial Condition – Leverage Sources: FRED, National Financial Conditions Index
Monetary Policy Expectations	Policy expectations play a pivotal role in influencing factor performance within financial markets. Anticipation of changes in monetary policy, fiscal measures, or regulatory decisions can significantly impact investor behavior and asset valuations. Factors such as growth, value, and momentum can experience shifts in performance based on policy expectations.	Effective Federal Funds Rate Monetary Policy Expectations Sources: FRED, Federal Reserve Economic Data
Equity Earning Yield	The equity earning yield significantly influence valuations based on companies' future earnings expectations.	Cyclically Adjusted PE Ratio (CAPE Ratio) Sources: Online Data Robert Shiller

EXHIBIT 4

Selected Equity Factors' Succinct Definition

Equity Factors	Succinct Definition	Selection Universe
Value	1/3 of stocks with the lowest price-to-book ratio	Russell 1000 Index
Growth	1/3 of stocks with the highest price-to-book ratio leverage are more likely to outperform during such periods	Russell 1000 Index
Momentum	1/3 of stocks with the highest 12-month trailing returns	Russell 1000 Index
Low Volatility	1/5 of stocks with the lowest annualized return volatility	Russell 1000 Index
Quality	1/3 of stocks with the highest quality score	Russell 1000 Index
Small-Cap	2/3 of stocks with the lowest market capitalization	Russell 3000 Index

SOURCE: Factset Data.

initial feature selection are presented in Exhibit 3 and the details of six selected equity factors are listed in Exhibit 4.

Feature Construction—The Financial Turbulence

Kritzman and Li (2010) present a pioneering approach for detecting market systemic risk. They introduce the concept of financial turbulence, which characterizes a condition wherein asset prices deviate from their typical historical patterns of behavior. This phenomenon encompasses extreme price movements, the decoupling of previously correlated assets, and the convergence of uncorrelated assets.

The identification of financial turbulence allows for a deeper understanding of market dynamics and the potential risks associated with market-wide disruptions.

They measure financial turbulence as:

$$d_t = (y_t - \mu)\sum^{-1}(y_t - \mu)'$$

where

d_t = turbulence for a particular time period t
y_t = vector of asset returns for period t
μ = sample average vector of historical returns
Σ = sample covariance matrix of historical returns

Following Kritzman and Li's approach, we leveraged monthly returns from ten S&P 500 sectors (info tech, energy, financials, healthcare, staples, discretionary, utilities, industrials, telecoms, and materials) and returns from various maturity Treasury instruments (12 months, 2 years, and 10 years maturity) to construct our financial turbulence series, as demonstrated in Exhibit 5. In our model, we recognize that market risk regime determination should consider the evolution of market conditions over time, not just under the prevailing conditions.

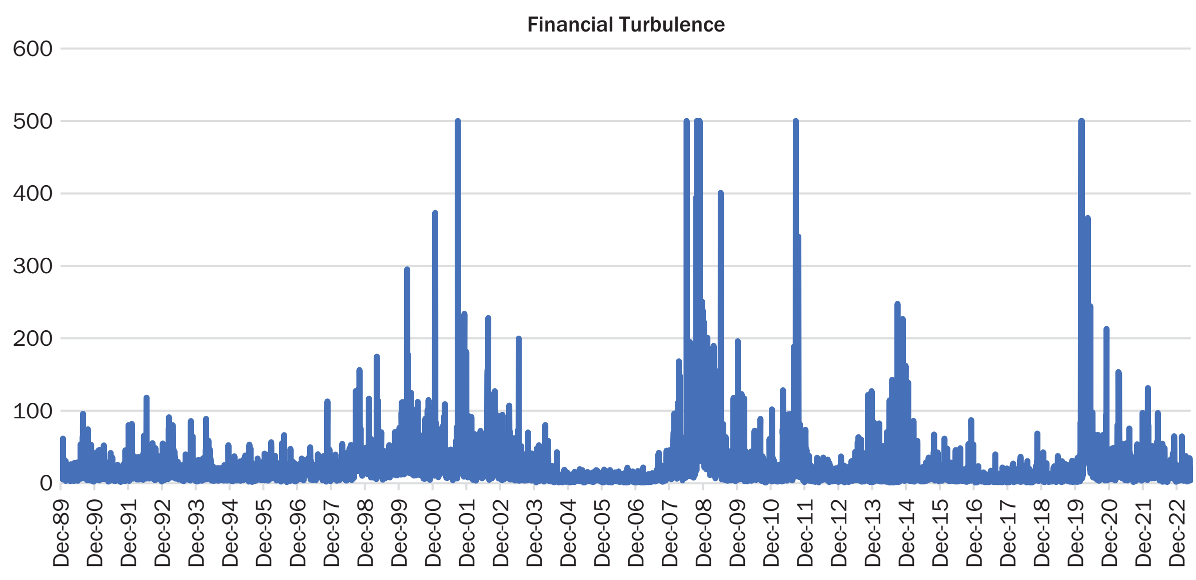
MODEL RESULTS AND FEATURE ANALYSIS

Stage One Model Results and Feature Analysis

As previously mentioned, our initial step involved utilizing K-Means clustering to group US Equity (S&P 500) drawdown data into distinct clusters. To determine the

EXHIBIT 5

Financial Turbulence from Ten S&P 500 Sectors and Various Maturity Treasuries

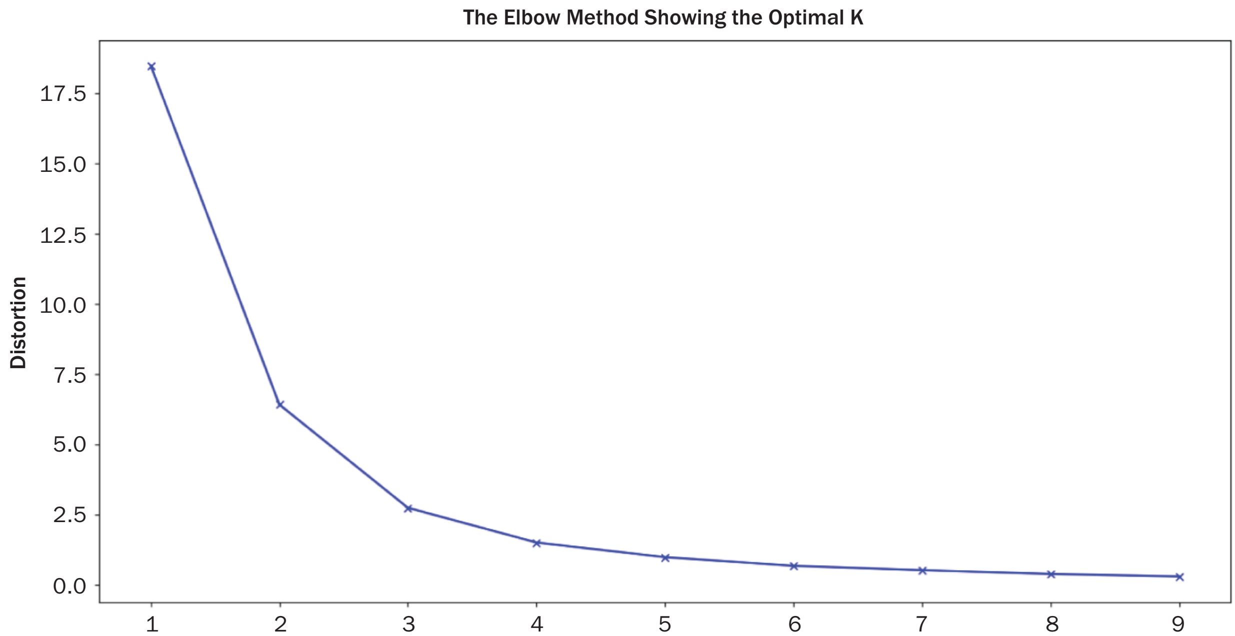


NOTES: **Past performance is no guarantee of future returns. The performance of an index is not an exact representation of any particular investment, as you cannot invest directly in an index.**

SOURCES: Authors' calculations based on data from FRED database and Factset.

EXHIBIT 6

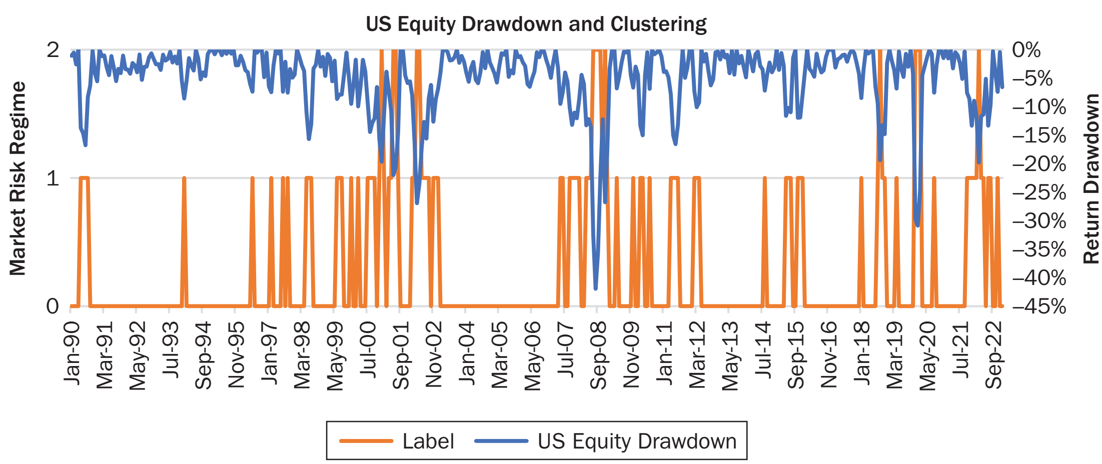
The Elbow Method of K-Means Clustering



SOURCES: Authors' calculations based on data from FRED database and Factset.

EXHIBIT 7

The S&P 500 Three Months' Rolling Drawdown and Clustering



NOTES: Cluster 0 shows "normal market risk regime," cluster 1 shows "market correction regime,, and cluster 2 shows the "bear market regime." **Past performance is no guarantee of future returns. The performance of an index is not an exact representation of any particular investment, as you cannot invest directly in an index.**

SOURCES: Authors' calculations based on S&P 500 price data sourced from FRED database and Factset. The S&P 500 is a product of S&P Dow Jones Indices LLC, a division of S&P Global, or its affiliates ("SPDJI").

---

optimal number of clusters, we employed the Elbow method, as illustrated in Exhibit 6, which led us to select three clusters as the most suitable solution for clustering the data. As shown in Exhibit 7, this three-cluster approach (cluster 0 shows "normal market risk regime," cluster 1 shows "market correction regime," and cluster 2 shows the "bear market regime") effectively differentiates the various market risk regimes

EXHIBIT 8

Accuracy Rate of Stacking Classification Mode

Accuracy (cross-validation five folds)	71%
Precision	0.753
Recall	0.721

SOURCES: Authors' calculations based on data from FRED database and Factset.

EXHIBIT 9

Feature Importance of Selected Variables

	Feature Importance
Equity Valuation	19.7%
Financial Turbulence	26.9%
Financial Conditions	34.1%
Monetary Policy Expectation	19.3%

SOURCES: Authors' calculations based on data from FRED database and Factset.

based on the market drawdown data. The key advantage of leveraging unsupervised learning over manual threshold determination and risk categorization lies in the dynamic adaptability it offers. By using unsupervised learning, we can dynamically choose strategy signals based on window length and data frequency, allowing for more flexible and data-driven decision-making processes. This enables the model to efficiently and effectively respond to evolving market conditions and tailor risk management strategies accordingly.

Given that most our data (77.5%) falls under the normal market risk regime, with 18.7% categorized as market correction, and only 3.8% classified as bear market regime, we adopted the synthetic minority oversampling technique (SMOTE) to address the class imbalance. By generating synthetic samples for the minority classes, we aimed to balance the dataset. Leveraging these engineered features, our model achieved an impressive 71% cross-validation accuracy rate in predicting significant shadow rate changes, as demonstrated in Exhibit 8. (In order to avoid data leakage in the time series data, we applied the time series split approach during cross-validation.)

In our effort to determine the importance ranking of the four factors in explaining market risk regimes, we employed the decision-tree feature importance function within the context of random forest. Details of this procedure are described in the Appendix. The outcome of this analysis is depicted in Exhibit 9. Our findings indicate that financial conditions and financial turbulence emerge as the primary factors influencing market risk regimes. However, we also observed that equity valuation (equity earning yield) and policy expectations play crucial roles, contributing valuable insights to the final decision-making process.

Stage Two Model Results and Feature Analysis

For our factor timing strategy, we carefully selected six common equity factors, including value, growth, momentum, low volatility, quality, and small size. In Exhibit 10, we present the full-sample return correlation among these six factors, revealing a notable level of correlation between them. This inherent correlation poses challenges when attempting to employ a linear approach based on macro or business cycles to distinguish the winning factors. To streamline the analysis and enhance convenience, we combined the bear market regime, representing only 3.8% of the time period, into the market correction regime. As displayed in Exhibit 11, the cross-correlation in the normal market and correction market regimes becomes relatively lower (for example, value with growth, small with quality etc.), offering empirical evidence to support the efficacy of using the market risk regime as stage one to discern the outperforming factors.

During the model training process, we assign labels to each month within each market regime's sub-dataframe based on the winning factor. For instance, if "value" is the winning factor for a particular month, we label it as 1 and label the remaining factors as 0. Similarly, if "growth" is the winning factor, we label it as 1, and so on for each factor. This labeling process is essential for training our random forest model to distinguish the outperforming factor accurately. This procedure is performed on a rolling basis to impart dynamism and to avoid introducing look-ahead bias. As a result

EXHIBIT 10

Cross Correlation throughout the Entire Sample Period

	Value	Growth	Momentum	Low Vol	Quality	Small
Value	1					
Growth	0.66	1				
Momentum	0.68	0.93	1			
Low Vol	0.76	0.70	0.71	1		
Quality	0.85	0.86	0.83	0.87	1	
Small	0.86	0.76	0.77	0.64	0.80	1

SOURCES: Authors' calculations based on data from FRED database and Factset.

EXHIBIT 11

Cross Correlation throughout Each Market Risk Regime

(Normal Risk)	Value	Growth	Momentum	Low Vol	Quality	Small
Value	1					
Growth	0.46	1				
Momentum	0.46	0.89	1			
Low Vol	0.63	0.56	0.55	1		
Quality	0.73	0.79	0.75	0.80	1	
Small	0.77	0.63	0.64	0.43	0.66	1

(Correction)	Value	Growth	Momentum	Low Vol	Quality	Small
Value	1					
Growth	0.54	1				
Momentum	0.61	0.86	1			
Low Vol	0.76	0.57	0.60	1		
Quality	0.86	0.70	0.68	0.88	1	
Small	0.88	0.64	0.69	0.67	0.81	1

SOURCES: Authors' calculations based on data from FRED database and Factset.

EXHIBIT 12

Historical Frequency Counting on Outperforming

Value	19.9%
Growth	15.9%
Momentum	14.4%
Low Volatility	21.3%
Quality	4.0%
Small Size	24.5%

NOTE: Past performance is no guarantee of future returns.

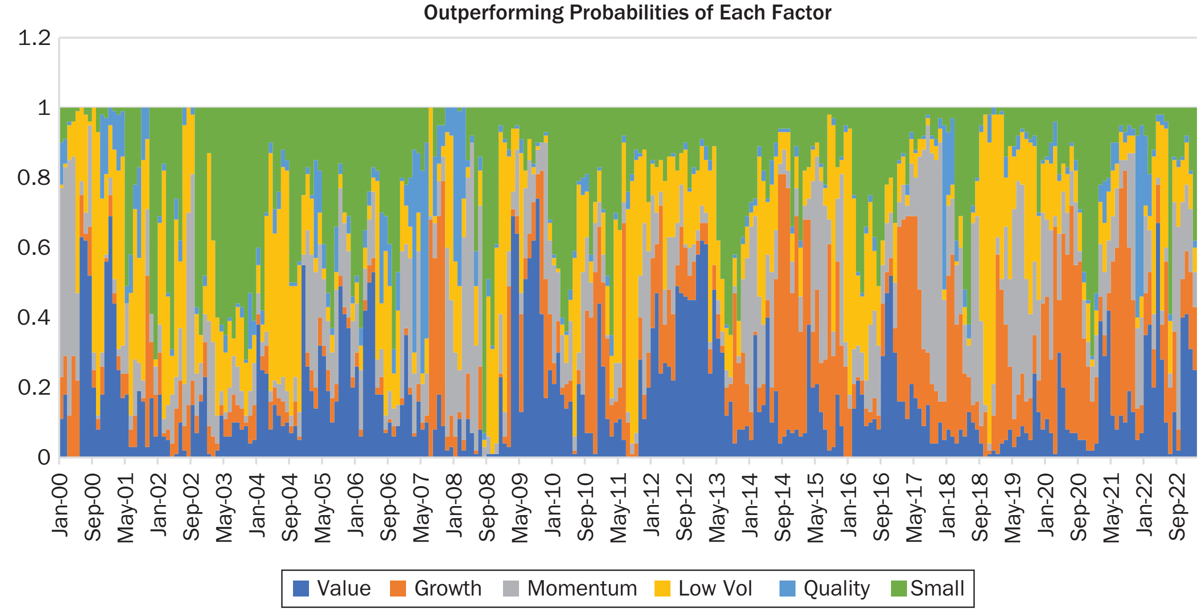
SOURCES: Authors' calculations based on data from FRED database and Factset.

of this training, we can then predict the probability of each factor to outperform other factors in the test set. This predictive capability empowers us to make informed decisions and tailor our factor timing strategy based on the probabilities, optimizing our approach for better performance in various market conditions.

Exhibit 12 shows the historical "winning" frequency of each of six factors and Exhibit 13 showcases the dynamic moving winning probabilities of each factor since the year 2000. Our analysis began by using an initial training window of 84 months, and we subsequently rolling this training period in the post-2000 prediction data set. The exhibit illustrates how the probabilities of each factor outperforming others change over the years, providing valuable insights into their relative strengths in different market

EXHIBIT 13

Outperforming Probabilities of Each Factor from Out-of-Sample Prediction



NOTE: **Past performance is no guarantee of future returns.**

SOURCES: Authors' calculations based on data from FRED database and Factset.

EXHIBIT 14

Feature Importance within Each Market Risk Regime

	Normal Risk Regime	Market Correction Regime
Equity Valuation	36.4%	42.5%
Financial Turbulence	12.4%	13.1%
Inflation	16.3%	17.9%
Financial Conditions	22.4%	15.3%
Monetary Policy Expectations	9.3%	9.9%
GDP Growth	3.2%	1.3%

SOURCES: Authors' calculations based on data from FRED database and Factset.

conditions. By monitoring these moving winning probabilities, we gain a comprehensive understanding of the factors' performance trends, enabling us to make well-informed and adaptive factor timing decisions as market dynamics evolve. Exhibit 14 shows the feature importance of the classification models in both market risk regimes.

PORTFOLIO SIMULATION SECTION

To demonstrate the effectiveness of our dynamic "winning" probabilities, we developed a robust historical out-of-sample testing framework, adhering to our two-stage framework design. Initially, we trained the model using the specified training data between 1987 and 2000. Next, we divided the training sample into sub-samples corresponding to different market risk regimes. Within each risk regime, we trained separate models to predict relative factor performance. As the next step, during each out-of-sample time period, we first predict the current market risk regime. Based on this prediction, we then estimate the probabilities of outperforming for each selected factor in that particular risk regime.

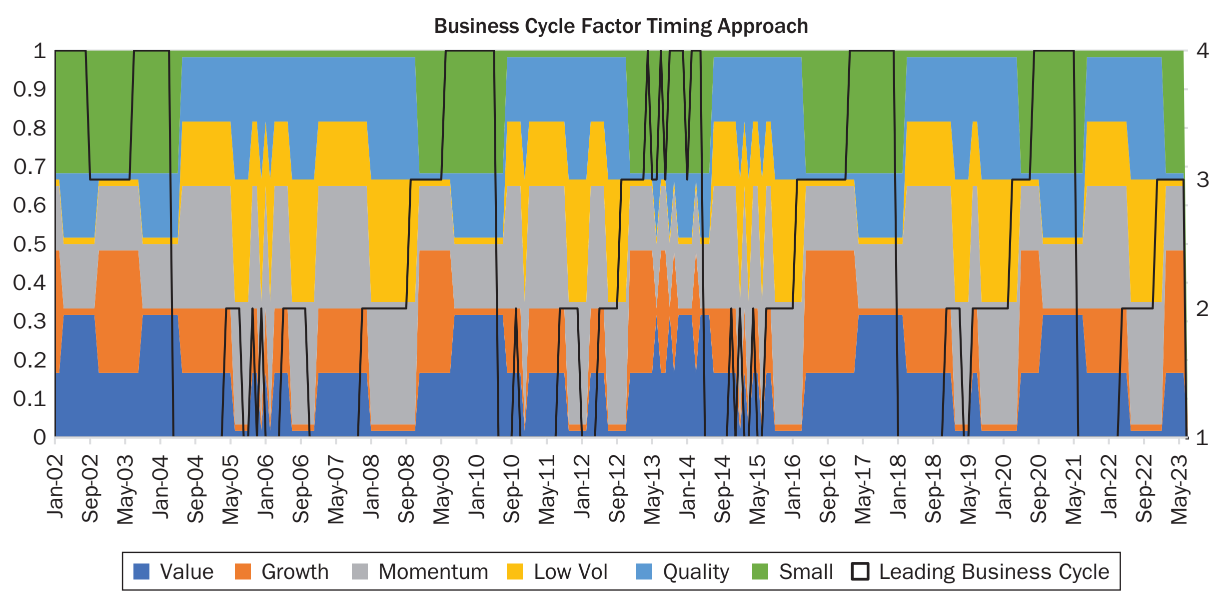
The primary benchmark commonly used in factor investing is the "1/N" approach, where an equal allocation is made to each factor (Khang, Picca, Zhang, and Zhu 2023). The second benchmark is the business cycle approach we mentioned previously. To simulate an optimal strategy using the business cycle approach, we first measured the average monthly performance for each factor in each stage of the business cycle across all cycles over the entire sample using perfect hindsight of the macro-economic data. We then used these average performance statistics to modify 1/N weights based on the point in time macro-economic data. Therefore, the optimal allocations are determined from full-sample averages with final macroeconomic data, but identification of business cycle stage during the out-of-sample test is based only on information available at the time. The resulting weights for each factor within specific business cycle regimes are illustrated in Exhibit 15. In Exhibit 16, we present the labels for leading state business cycles using data from The Conference Board. Subsequently, we adjust the risk budget, employing a 15% allocation, to modify the weights assigned to individual factors through a "1/N" approach.

In contrast, for the machine learning approach, we employed the probability of outperforming to weight each factor. While practitioners often incorporate a risk budget in real-world training scenarios, we directly utilized the "winning" probability of each factor as its weighting for the purpose of demonstrating signal effectiveness. By comparing the compound returns since 2002 and the excess return for each month, as depicted in Exhibit 17, we can observe the impact of our strategy relative to the broad market, the 1/N benchmark, and the business cycle approach. Notably, during most of the out-of-sample period, our strategy consistently delivers positive results. These metrics underscore the robustness and efficacy of our approach, showcasing its ability to generate favorable excess returns while demonstrating a strong risk-adjusted performance.

EXHIBIT 15
Business Cycle Strategy

	Expansion	Slowdown	Contraction	Recovery
Value		–		+
Growth		–	+	–
Momentum	+	+		
Low Volatility		+	–	–
Quality		+	–	
Small-Cap	–	–	+	+

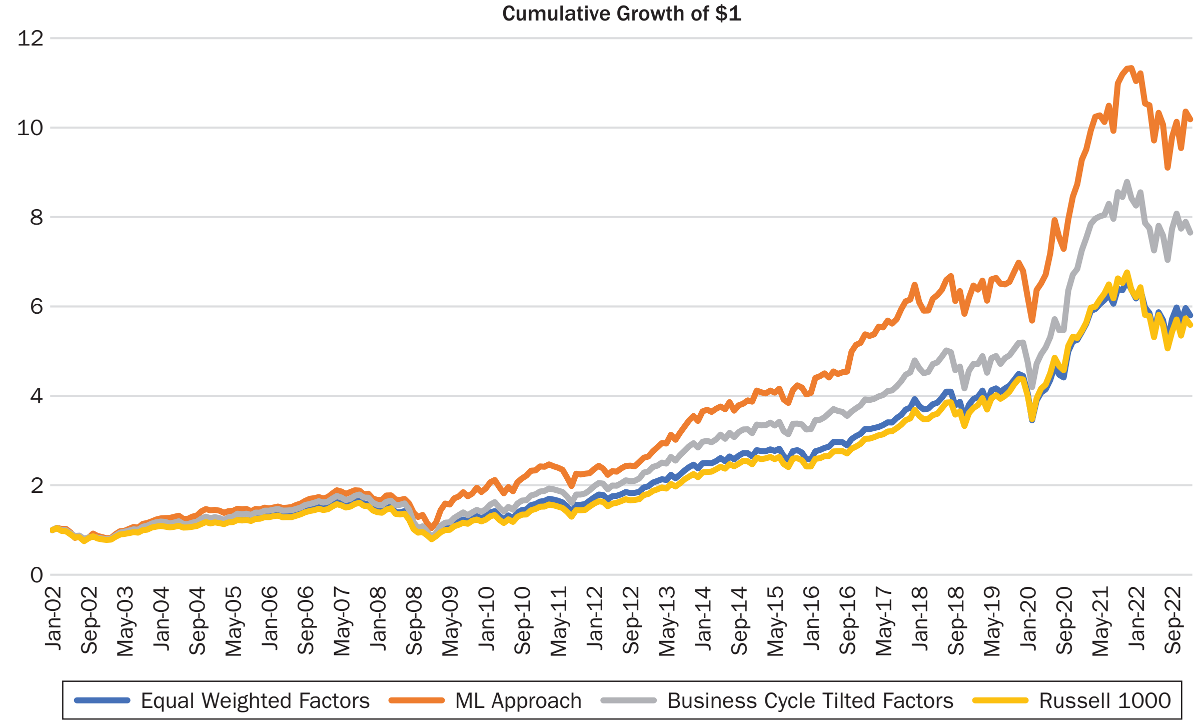
EXHIBIT 16
Business Cycle Labels



SOURCES: Authors' calculations based on data from FRED database and Factset.

EXHIBIT 17

Compound Return with Different Approaches



	Annualized Return	Annualized Volatility	Sharpe Ratio	Annualized Excess Return	Annualized Tracking Error	Information Ratio
Russell 1000	8.40%	15.49%	0.52			
Equal Weighted	8.63%	15.19%	0.54	0.18%	2.40%	0.07
Business Cycle Tilt	9.84%	16.06%	0.59	1.50%	3.95%	0.36
ML Approach	11.47%	16.81%	0.66	2.89%	6.68%	0.45

NOTE: Past performance is no guarantee of future returns. The performance of an index is not an exact representation of any particular investment, as you cannot invest directly in an index.

SOURCES: Authors' calculations based on data from FRED database and Factset.

CONCLUSION

Comprehending the dynamics of factors is vital to factor rotation strategies, and previous academic research has successfully established correlations between factor performance and macro variables. However, the true essence lies in effectively and systematically timing these factors to achieve tangible results. In this study, we harnessed the power of both macroeconomics and financial markets data, employing a two-stage machine learning framework to develop a more dynamic factor timing approach. The versatility of our framework allows for adoption across different time frequencies and facilitates testing with an expanded range of factors. As a result, we present a robust and alternative solution for factor timing, which holds the potential for wider application in asset pricing and investment strategies. The effectiveness of our approach in accurately predicting market risk regimes and determining outperforming factors showcases its practical relevance and contributes to advancing the field of factor investing. By harnessing the synergy between macroeconomic insights and machine learning techniques, our research opens avenues for optimizing portfolio allocation and enhancing investment decision-making in the ever-changing financial landscape.

APPENDIX

MACHINE LEARNING METHODOLOGIES

Unsupervised Learning Clustering—K-Means Clustering

We adopted the K-means clustering algorithm, a widely used unsupervised machine learning technique, to analyze the equity market drawdown data. K-means clustering aims to partition data points into K clusters (Hartigan and Wong 1979), where each data point belongs to the cluster with the nearest mean. Mathematically, given a set of N data points $\{x_1, x_2, \ldots, x_n\}$, K-means clustering involves finding K cluster centroids $\{\mu_1, \mu_2, \ldots, \mu_k\}$ to minimize the within-cluster sum of squares:

$$minimize\ \Sigma_i\ \Sigma_j\ ||x_i - \mu_j||^2$$

where $||x_i - \mu_j||$ *represents the Euclidean distance between data point* x_i *and cluster centroid* μ_j.

This approach allows us to identify underlying patterns and groupings within the market data, facilitating the creation of distinct market risk regimes. The implementation of K-means clustering in our research offers valuable insights into the dynamics of market behavior and enhances our understanding of the relationships between different market conditions.

Supervised Learning Classification—Random Forest Classification

Svetnik et al. (2003) introduced the random forest classification, which stands as one of the most renowned ensemble algorithms, employing multiple individual decision trees. The model utilizes bagging and randomness to construct underlying decision trees, ensuring their independence in the forest. By aggregating the outputs of these individual trees through voting or averaging, the random forest generates a final prediction that surpasses the accuracy achieved by any individual decision tree. This technique effectively leverages the strength of multiple decision trees to produce a more precise and robust classification outcome, making random forests an influential and widely adopted classification approach in various fields.

Assuming a random forest ensemble N trees underline, then we have $\{T_1(X), T_2(X), \ldots, T_n(X)\}$ where X is a multi-dimensional vector. Those decision trees produce N individual results, as $\{Y_1 = T_1(X), Y_2 = T_2(X), \ldots, Y_n = T_n(X)\}$ where Y_n is the prediction based on nth tree. Then the random forest final prediction result is based on averaging or majority voting of N individual results. We used this approach on both stage one to predict the market risk regime and stage two to predict the factor outperforming probabilities.

Decision Tree Model-Based Feature Importance

The feature importance function serves to assign a score to each input factor, reflecting its significance in predicting the dependent variable. This function, based on the decision tree model, proves valuable in both classification and regression algorithms. The decision tree-based feature importance function aims to identify the optimal split for the least impure node by employing the Gini index as a measure of impurity.

In our study, we utilized the random forest model, which, as previously mentioned, consists of multiple individual decision trees. For a single decision tree model, we obtain feature importance scores for each feature, quantifying their respective contributions to the model's predictive power. This analysis enables us to gain deeper insights into the key factors that influence the outcome and supports the use of random forest as an ensemble of decision trees to improve the accuracy and robustness of our classification results. For a single decision tree model, we have the importance for every feature i as

$$fi_i = \frac{\sum_{j:\text{node } j \text{ splits on feature } i} ni_j}{\sum_{k \in \text{all nodes}} ni_k}$$

where fi_i is the importance of feature i and ni_j is the importance of the node j. The latter is calculated as the decrease in the impurity of each child node of node j weighted by the probability of reaching that node. And to calculate the feature importance on ensemble level (random forest), we need to normalize the importance for each tree as:

$$normfi_i = \frac{fi_i}{\sum_{j \in \text{all features on decision tree}} fi_j}$$

Then sum normalized feature importance value from each tree to get the final feature importance on random forest level as:

$$RFfi_i = \frac{\sum_j normfi_{ij}}{\sum_{j,m} normfi_{jm}}$$

Where j belongs to all features in the data set, and m belongs to all the trees in the random forest. We apply this feature importance function to the whole training set to give us an overview of the features that are most important in predicting shadow rates.

The Stacking Ensemble Classification

In this study, we propose to employ the stacking methodology for classification tasks to improve predictive performance and enhance model generalization. We selected three powerful classifiers as our first layer: Random Forest (RF), Support Vector Classifier (SVC), and Gaussian Naive Bayes (GNB). These classifiers exhibit distinct strengths in handling different aspects of the data and have demonstrated their effectiveness in various machine learning tasks (Moguerza and Muñoz 2006, Xue and Titterington 2008). We denote the predictions of these base classifiers as follows: RF(x), SVC(x), and GNB(x), where x represents the input data.

To build our stacked ensemble, we combine the predictions of these base classifiers as inputs for our secondary model, Logistic Regression (LR). The logistic regression model is trained to learn the optimal weights, w1, w2, and w3, for each base classifier's output (RF(x), SVC(x), and GNB(x)) by minimizing the cross-entropy loss, given by:

$$L(w) = -\sum [y\_i * \log(y\_pred\_i) + (1 - y\_i) * \log(1 - y\_pred\_i)],$$

where y_i is the true label of the i-th sample, and y_pred_i is the predicted probability for the i-th sample obtained from the stacked ensemble. The final prediction y_pred, for a given input x, is calculated as follows:

$$y\_pred = \sigma(w1 * RF(x) + w2 * SVC(x) + w3 * GNB(x)),$$

where σ is the sigmoid function that maps the weighted sum of base classifiers' outputs to a probability between 0 and 1.

By allowing the Logistic Regression to learn the optimal weights, the ensemble can adapt to the complexity and uniqueness of the underlying data, leading to improved performance.

Synthetic Minority Over-Sampling Technique (SMOTE)

Imbalanced datasets pose significant challenges in various machine learning applications, as traditional algorithms tend to favor majority classes, leading to suboptimal performance for minority classes. To address this issue, researchers have developed various techniques for handling imbalanced data, such as Synthetic Minority Over-sampling Technique (SMOTE), which is developed by Chawla, Bowyer, Hall, and Kegelmeyer (2011). In this research, we leveraged SMOTE to tackle class imbalance and improving the performance of classification.

SMOTE is a popular approach that synthesizes new minority class instances by interpolating existing samples. The technique is based on the concept of generating synthetic data points along the line segments connecting nearest neighbors in the feature space. Mathematically, given a minority class instance x_i and its K nearest neighbors, SMOTE creates new synthetic instances x_new by randomly selecting a neighbor x_j and computing:

$$x\_new = x_i + \alpha * (x_j - x_i)$$

where α is a random value between 0 and 1. SMOTE effectively augments the minority class, providing the model with more diverse examples and mitigating the class imbalance problem.

ACKNOWLEDGMENTS

We wish to acknowledge Joseph H. Davis, PhD and John Ameriks, PhD for their useful comments. The views expressed in this article do not necessarily reflect the views of The Vanguard Group, Inc.

All investing is subject to risk, including the possible loss of the money you invest. There is no guarantee that any particular asset allocation or mix of funds will meet your investment objectives, or provide you with a given level of income. Diversification does not ensure a profit or protect against a loss. This material is provided for informational purposes only and is not intended to be investment advice or a recommendation to take any particular investment action.

REFERENCES

Banz, R. W. 1981. “The Relationship between Return and Market Value of Common Stocks.” *Journal of Financial Economics* 9 (1): 3–18. https://doi.org/10.1016/0304-405X(81)90018-0.

Chawla, N. V., K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer. 2011. “SMOTE: Synthetic Minority Over-Sampling Technique.” *The Journal of Artificial Intelligence Research* 16: 321–357. https://doi.org/10.1613/jair.953.

Choi, J. 2013. “What Drives the Value Premium?: The Role of Asset Risk and Leverage.” *The Review of Financial Studies* 26 (11): 2845–2875. https://doi.org/10.1093/rfs/hht040.

Fama, E. F., and K. R. French. 1992. “The Cross-Section of Expected Stock Returns.” *The Journal of Finance* 47 (2): 427–465. https://doi.org/10.1111/j.1540-6261.1992.tb04398.x.

——. 2016. “Dissecting Anomalies with a Five-Factor Model.” *The Review of Financial Studies* 29 (1): 69–103. https://doi.org/10.1093/rfs/hhv043.

Harris, R. S., and F. C. Marston. 1994. “Value versus Growth Stocks: Book-to-Market, Growth, and Beta.” *Financial Analysts Journal* 50 (5): 18–24. http://www.jstor.org/stable/4479770.

Hartigan, J. A., and M. A. Wong. 1979. “Algorithm AS 136: A K-Means Clustering Algorithm.” *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 28 (1): 100–108. https://doi.org/10.2307/2346830.

Ilmanen, A., and J. Kizer. 2012. "The Death of Diversification Has Been Greatly Exaggerated." *The Journal of Portfolio Management* 38 (3): 15–27. https://doi.org/10.3905/jpm.2012.38.3.015.

Kao, D.-L., and R. D. Shumaker. 1999. "Equity Style Timing." *Financial Analysts Journal* 55 (1): 37–48. http://www.jstor.org/stable/4480137.

Khang, K., A. Picca, S. Zhang, and M. Zhu. 2023. "How Inefficient Is the 1/N Strategy for a Factor Investor?" *Journal of Investment Management* 21 (1): 1.

Kritzman, M., and Y. Li. 2010. "Skulls, Financial Turbulence, and Risk Management." *Financial Analysts Journal* 66 (5): 30–41. http://www.jstor.org/stable/25741309.

Maio, P., and P. Santa-Clara. 2017. "Short-Term Interest Rates and Stock Market Anomalies." *Journal of Financial and Quantitative Analysis* 52 (3): 927–961. https://doi.org/10.1017/S002210901700028X.

Moguerza, J. M., and A. Muñoz. 2006. "Support Vector Machines with Applications." *Statistical Science* 21 (3): 322–336. http://www.jstor.org/stable/27645765.

Polk, C., M. Haghbin, and A. de Longis. 2020. "Time-Series Variation in Factor Premia: The Influence of the Business Cycle." *Journal of Investment Management* 18 (1): 69–89.

Sharpe, W. F. 1964. "Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk." *The Journal of Finance* 19 (3): 425–442. https://doi.org/10.2307/2977928.

Svetnik, V., A. Liaw, C. Tong, J. C. Culberson, R. P. Sheridan, and B. P. Feuston. 2003. "Random Forest: A Classification and Regression Tool for Compound Classification and QSAR Modeling." *Journal of Chemical Information and Computer Sciences* 43 (6): 1947–1958.

Xue, J., and D. M. Titterington. 2008. "Comment on "On Discriminative vs. Generative Classifiers: A Comparison of Logistic Regression and Naive Bayes." *Neural Processing Letters* 28 (3): 169–187. https://doi.org/10.1007/s11063-008-9088-7.