Stevens Institute of Technology

Financial Engineering

Market Regimes in QWIM

Authors:
Masumeh Babaei
Mark Wong
Mahender Teegala

Supervisor:
Dr. Khaldoun Khashanah
Dr. Christian Homescu

November 26, 2024

Abstract

Market regimes have been a main-stay topic of discussion amongst industry professionals and academics. Most traditionally, the market can be separated into a 'bear' and 'bull' market. However, as research into market regimes continues, the method of detection and the amount of these regimes have been suggested to be much more nuanced than pointing out when the market goes up or down. In this paper, we explore three distinct methodologies to both identify regimes within financial markets and utilize these regimes to generate a profitable portfolio. Firstly, we introduce a variety of machine learning models, notably the Gradient Boosted Decision Tree. Second, we examine a method using Principal Component Analysis (PCA). PCA is commonly used for dimensionality reduction and was applied to economic data publicly available from the Federal Reserve Economic Data (FRED) database. Following the application of PCA, we utilized K-means clustering to identify distinct regimes within the transformed principal components. Finally, we introduce an improved method of K-means clustering that uses Wasserstein distance as opposed to Euclidean distance. Results demonstrate that the regimes produced capture much more of the intended market regimes than traditional methods and generate portfolios with outsized risk adjusted returns relative to SPY.
Keywords: Market Regimes, Principal Component Analysis, K-means, Wasserstein Distance, Garch, GJR-Garch, Gradient Boosting Decision Trees

1 Introduction

Investors with long-term horizons often face potential losses due to mean reversion, where asset prices and historical returns gradually revert to their long-term average. This phenomenon can undermine the time value of money and diminish investment returns over time. To mitigate such risks, identifying distinct market regimes can be instrumental. These regimes, which reflect different phases such as growth or recession, and high or low volatility, allow investors to tailor their strategies to prevailing economic conditions.

K-means clustering is a natural choice in clustering market regimes as it is a very straightforward and easy to implement method for generic clustering. However, as pointed out in Blanka Horvath (2021), the traditional Euclidean measure is insufficient in capturing entire regimes since it does not function on the probability space. While K-means produces clear clustering of positive vs negative returns, these do not consistently capture the behavior of log-returns within a market regime. By lifting the problem of market regime clustering to a space of streams of log-returns, we can look at the probability distribution of returns within each regime. Similar to Gaussian mixture methods like that in Aigner (2023) and David Hallac (2019) we define regimes by an empirical measure of their probability distribution. This allows us to both identify intermittent market regimes and extended market regimes.

There exists some classical choices to perform the clustering on the space of probability distributions such as the Kullback-Leibler divergence or the Kolmogorov-Smirnov statistic. However, both methods are limited by their interpretability. The former is not a symmetric statics and the latter tends to be more sensitive to the center of distribution than at the tails so it is difficult to interpret the latent behavior when there are behaviors within the probability space that do not follow intuition. Wasserstein distance becomes a natural choice as it is tractable in the 1-dimensional data series we are working with and is capable of producing symmetric distances between two probability distributions.

As another method investigated our analysis, we utilized Principal Component Analysis (PCA) to process a variety of economic indicators, including Gross Domestic Product (GDP), Unemployment Rate, Consumer Price Index (CPI), Money Supply, Federal Funds Rate, Personal Income, Inflation Rate, Total Public Debt, and metrics relating to Housing and Urban Development, Gross Private Domestic Investment, Monetary and Fiscal Policies, and the Producer Price Index (PPI). We first standardized these indicators to neutralize unit disparities, ensuring each variable contributed equally to the analysis. PCA was then applied to reduce the dimensionality of this data, with the aim to retain components that collectively accounted for up to 90% of the variance, thus focusing on the most significant features.

To enhance the robustness of our PCA, we selected three principal components from the initial analysis, standardized this subset of data again, and performed another round of PCA. This method helped to distill the information further and concentrate on the most impactful features. Subsequently, K-means clustering was applied to this reduced data set. The number of clusters was determined using techniques such as the Elbow Method, which identifies the optimal count by balancing explained variance against the number of clusters.

The final step involved visualizing these market regimes through a plot that categorizes the clustered data points based on their principal components. This visualization not only aids in recognizing the transitions between different market conditions but also serves as a practical tool for investors to adjust their investment strategies dynamically, thereby potentially enhancing returns and minimizing the adverse effects of mean reversion. Through the combined use of PCA and K-means clustering, we effectively segmented the financial market into discernible regimes, guiding strategic investment decisions.

In the rest of the report we explain our methodology and clustering results in Section 3. We then follow with portfolio construction using the methods in Section 4. Finally we conclude in Section 5.

2 Literature review

2.1 Method 1

The literature on forecasting regime changes in financial markets predominantly focuses on leveraging advanced statistical and machine learning techniques to predict market turns effectively. These methods aim to capture signals from complex market data that may indicate potential transitions from bull to bear markets or vice versa. This review highlights significant contributions using volatility models like GARCH, as well as machine learning approaches including Gradient Boosted Decision Trees (GBDT) and Neural Networks. Volatility models, especially the Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model, have been extensively utilized to capture the dynamic nature of market volatility and its implications on regime changes. The application of such models provides a nuanced understanding of market conditions by assessing how past volatilities influence future values, which is crucial for predicting volatile shifts that characterize regime changes. Recent advancements in machine learning offer promising techniques for enhancing the prediction of financial market regimes. Specifically, methods like GBDT and Neural Networks have shown considerable success due to their ability to handle large datasets and extract complex patterns that are not immediately apparent. In the study by Zhang et al. (2023) "Volatility Forecasting with Machine Learning and Intraday Commonality" machine learning models were applied to forecast intraday realized volatility by pooling stock data and incorporating market volatility proxies. The study found that neural networks outperformed linear regressions and tree-based models because of their superior ability to identify complex patterns in the data. GBDT models are lauded for their precision in handling various types of data anomalies and imbalanced datasets, which are common in financial markets. These models operate by building an ensemble of trees incrementally and adjusting at each step, focusing on correcting the errors made by previous trees.

A study by Benhamou et al. (2014), on exploring the GBDT focusing on using a mix of technical, fundamental, and macroeconomic features to plan regime changes on the S&P 500 shows that GBDT models are highly accurate in predicting substantial market downturns. They highlighted the application of Shapley values to robustly identify and explain the features most crucial for predicting stock market crises, especially useful in the analysis of the March 2020 financial meltdown.

In the study that was conducted by Chiara Lattanzi (2019) on the tail distribution, they demonstrate that the inclusion of different extreme regimes outperforms both static and dynamic competing approaches in financial applications. Neural networks, particularly deep learning architectures like MLPs (Multi-Layer Perceptron's), have been utilized for their capacity to model non-linear relationships in data.

Lingfei Li and Bo Wu Lingfei Li (2023) had examined this model and found that Neural Network can learn from vast amounts of unstructured data, making them ideal for capturing the complex interactions in financial markets that traditional models might miss.

The literature suggests a growing consensus on the efficacy of combining traditional volatility models with advanced machine learning techniques to forecast financial market regimes. GBDT and neural networks, in particular, stand out due to their robustness and adaptability, which are crucial for navigating the complexities of financial markets. Future research might explore further integration of these models with real-time data and alternative machine learning approaches to enhance predictive power and reliability.

2.2 Method 2 - Wasserstein K-means

As of today, a final consensus on the number of regimes within a market is not yet decided. This is readily clear by approaching the market from different lenses. With different contextual backgrounds and different detection methodologies and regime definitions it can be difficult to exactly pin-point the exact number of regimes.

A large portion of literature tends to presume the existence of only a 'bull' and 'bear' market and builds their models around that assumption. Many of such papers are based on Hamilton (1989) Markov switching model which usually assumes a priori that there are only two market states. Even recent papers like that by Pier Francesco Procacci (2019) begin their studies under the assumption of the number of clusters being 2.It is unclear whether these models perform very well as the clusters increase.

Some papers like Blanka Horvath (2021) suggest that the choice of these parameters may very well be more of an art than a science in that the results of the parameter selection may affect the outcome of the model but is more based on intuition rather than there being an objective answer. On the other hand Mathieu Gatumel (2014) suggested that the traditional binary regime model of 'bull' vs 'bear' was wholly insufficient in capturing the different behaviors of asset returns and that there may require between 2 and 5 regimes to fully capture all features of the asset's distribution. The core conclusion here is that models that cannot correctly identify the number of regimes or are not robust to new regimes tend to fail in actual performance.

While the results of their paper is based on manual inspection, these implications are followed by several papers that seek to automatically determine the number of regimes based on data computation. For example, David Hallac (2019) produces a Gaussian Segmentation model that is able to automatically determine the number of regimes.

In this paper we explore many of the more recent methodologies to classify and identify regimes and analyze the potential of these models to forecast regimes into the future and generate portfolios that can be profitable in a variety of market states.

2.3 Method 3 - PCA and K-means

In Peter Akioyamen (2021), researchers used method principal component analysis to reduce the dimension of dataset and regime detection was done by k-means clustering. I picked my method from this paper but the novel about our paper is we selected three different methods and compared them with regime detection for best earnings.

The paper Chan et al. (2021) helps in portfolio construction by comparing different methods of portfolio construction in regimes that gives best outcome. CPO methods were compared with other portfolio optimization approaches such as Equal Weights, Risk Parity, Markowitz, and Minimum Variance. This paper helps me for portfolio construction once I have identified the regimes.

The paper Carlos Heitor Campani (2021) helped me learn how to change investment decisions according to regime change. Investments like bond, gold, stock again choosing mid or large cap also use other methods to predict regimes. The objective is to generate the most effective wealth allocation strategy among these asset classes on a monthly basis, ensuring swift optimization.

3 Methodology

3.1 Method 1: Regime Change Prediction using Score Model, ML Models, Advanced volatility Model

This paper aims to forecast the regime change which is an integrated approach using a scoring system Called Topology 1, machine learning models and advanced volatility models . The methodology outlined in this study offers a sophisticated fusion of machine learning techniques, including Gradient Boosting Decision Trees (GBDT) , Neural Networks and logistic regression with traditional volatility forecasting methods such as GARCH models and analysis of realized volatility. This integrated approach employs a multifaceted predictive framework that not only seeks to enhance the accuracy of regime change forecasts but also provides a deeper comprehension of market dynamics. By comparing three distinct forecasting approaches and examining the unique signals each produces, the study delivers a comprehensive evaluation of their predictive capabilities, thereby enriching financial decision-making with a more informed understanding of market volatility and its impact on regime shifts.

Model Development for Regime Change Prediction

Using the features described below, five models have been developed to predict regime changes:

1: Topology 1, Integrative Score, and Machine Learning System.

2: Integration of VAR Residuals, GARCH Parameters, and Moving Averages.

3: Advanced GARCH and Machine Learning Ensemble.

4: Realized Volatility Analysis for Regime Detection.

5: Synthesized Realized Volatility (RV) and GARCH Strategies with Machine Learning Enhancement.

Data Selection

For the analysis of regime changes in financial markets, my study uses historical data from two significant financial instruments: GLD and SPY. The dataset spans from January 2000 to December 2023, providing a comprehensive view across different market conditions, including various economic cycles and financial crises.

Gold (GLD): Gold is traditionally considered a "safe haven" asset. During financial uncertainty or economic downturns, investors often flock to gold, increasing its price. Analyzing GLD provides insights into periods of market fear and risk-averse behavior, indicative of potential regime shifts.

S&P 500 ETF (SPY): SPY reflects a broad spectrum of the U.S. equity market. Its performance is closely tied to the economic and corporate health of the United States, making it a valuable proxy for assessing the equity market's stance within various market regimes.

Modeling Approach

To capture and analyze the market dynamics associated with these assets, the study employs the scoring system "Topology 1," GARCH models (Generalized Autoregressive Conditional Heteroskedasticity and Jump-GARCH models), and Realized volatility. These models are adept at modeling financial time series data that exhibit time-varying volatility, a common characteristic of asset returns.

GARCH Model: Estimates the conditional volatility of asset returns. It provides insights into how past volatilities and shocks affect current volatility, which is crucial during regime changes.

Jump-GARCH Model (JGR-GARCH): This model accounts for sudden, significant changes in volatility (jumps) that are often observed during market crashes or geopolitical events. It offers a nuanced analysis of abrupt market shifts, key indicators of regime change. Technical Indicators

In addition to volatility models, technical indicators were incorporated to identify potential regime shifts. Moving averages, particularly the 50-day and 200-day moving averages (MA50 and MA200), help determine market trend and momentum.

Dataset Significance

The use of GLD and SPY data over an extensive period (2000-2023) encompasses stable growth, recessions, and recoveries, providing a robust foundation for model testing and validation across diverse market scenarios. This strategic choice of gold and a major equity index ETF captures a wide array of market behaviors, offering a balanced perspective on potential triggers and characteristics of regime changes.

Combining historical data analysis, statistical modeling, and technical analysis creates a detailed framework for identifying and understanding regime changes in financial markets. This integrative methodology comprehensively assesses how varying market conditions influence investor behavior and asset price dynamics, essential for predicting regime shifts.

3.1.1 Model1: Topology1, Integrative Score and Machine Learning System

Topology 1 is a scoring system which is designed by Masumeh Babaei to synthesize complex financial data into a quantifiable metric that reflects the likelihood of a regime change. Financial markets are influenced by multiple factors. The purpose of this model is to combine diverse indicators of market behavior such as volatility and asymmetry effect into a comprehensive score, offering a holistic view of market dynamics where each parameter contributes uniquely to the score, enhancing its predictive power regarding future market states, particularly in identifying shifts from bull to bear markets or vice versa.

Development and Application of the Scoring Model

The scoring system is constructed through a weighted linear combination of selected model parameters. scores are calculated using a weighted sum of various parameters from the GARCH and GJR-GARCH models. These scores encapsulate the combined impact of several volatility indicators, including baseline volatility (omega), shock responses (alpha coefficients), volatility persistence (beta), and asymmetry in shock responses (gamma). The score for each trading day is calculated by summing these weighted parameters, effectively aggregating all relevant information into a single predictive metric. The formula is as follow:

$$S = \omega \cdot w_\omega + \sum_{i=1}^{15} \alpha_i \cdot w_{\alpha_i} + \beta \cdot w_\beta + \gamma \cdot w_\gamma \tag{1}$$

Where:

- ω is the baseline volatility parameter from the GARCH model.
- α_i represents the coefficients corresponding to the reaction of volatility to past squared residuals for each lag i, up to 15 lags.
- β is the coefficient indicating the persistence of the volatility over time.
- γ is the coefficient for the asymmetry in the volatility response, applicable in the GJR-GARCH model.
- w_i are the weights assigned to each parameter, reflecting their relative importance and statistical significance in the model.

The score integrates both the immediate and persistent effects of market shocks, as well as the differential impact of positive and negative shocks, providing a comprehensive measure of potential market volatility and regime shifts.

Topology 1 Features Engineering

To better elucidate the relationship between the two assets in question, a Vector Autoregression (VAR) model was employed. This strategic choice was aimed at capturing the intricate dynamics of interrelationships between these assets, thereby providing essential insights for the strategic development of the Topology 1 scoring system. The VAR model's results indicated that 15 lags were optimal, as determined by the Akaike Information Criterion (AIC), highlighting the significant interactions within this time frame. This finding supports the rationale that incorporating these characteristics into a scoring system designed to quantify regime changes is a prudent decision. Implementing this consistent lag structure across the scoring system not only enhances its predictive accuracy but also ensures analytical coherence. Consequently, the 15 lags from the GARCH model were integrated as parameters within the scoring system, leveraging their demonstrated predictive capabilities to enhance the system's overall effectiveness.

Transforming Scores into Probabilities

After computations of the scores using the formula 1, the transformation of scores into probabilities is accomplished using the logistic sigmoid function, which is particularly useful for converting any real-valued number into a probability format. The logistic function has an S-shaped curve (sigmoid curve), which is ideal for binary classification tasks such as predicting two market regimes (bullish & bearish).

The logistic sigmoid function used for the transformation can be mathematically expressed as:

$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{2}$$

If P is close to 1, it suggests a high probability of a bullish market regime, implying that market conditions are favorable, and volatility characteristics align with upward market trends. If P is close to 0, it suggests a high probability of a bearish market regime, indicating potentially unfavorable market conditions with characteristics typical of downward trends or increased market instability.

As a result, a threshold has been defined to classify the observations:

- If $P \geq 0.5$, the market is classified as bullish.
- If $P < 0.5$, the market is classified as bearish.

The derived probabilities can serve as powerful predictors in machine learning models alongside other financial indicators such as price levels, moving averages, technical indicators. This features has been used in three machine learning models, GBDT, Neural Networks and Logistic Regression. the best result was for GBDT model where shows 95 percent accuracy of the predictions for the bull and bear market regime.

Precision is the ratio of correctly predicted positive observations to the total predicted positives. High precision relates to the low false positive rate. For the bearish class (0), 88 percent of predicted bearish results were actually bearish. For the bullish class (1), 100 percent of predicted bullish results were actually bullish.

Table 1: Classification Report

Class	Precision	Recall	F1-score	Support
0	0.88	1.00	0.93	7
1	1.00	0.92	0.96	13
Overall Metrics				
Accuracy	0.95 (20 instances)			
Macro Avg	0.94	0.96	0.95	20
Weighted Avg	0.96	0.95	0.95	20

Recall (also known as sensitivity) is the ratio of correctly predicted positive observations to all observations in the actual class. For the bearish class (0), the model correctly identified all bearish instances.For the bullish class (1), the model correctly identified 92 percent of all actual bullish instances.

The F1 Score is the weighted average of Precision and Recall. Therefore, this score takes both false positives and false negatives into account. It is a good way to show that a class has a good recall and precision balance. The bearish class (0) had an F1-score of 0.93, indicating a very good balance between precision and recall. The bullish class (1) had an F1-score of 0.96, showing an excellent balance.

3.1.2 Integration of VAR Residuals and Tail Distribution

In the pursuit of a robust model for predicting market regime changes, Model 2 incorporates a strategic combination of VAR residuals, GARCH parameters, and moving averages, each selected for their unique contributions.

Residuals from the VAR model provide a refined measure of the market's behavior by isolating unpredictable movements after accounting for known linear interdependencies between SPY and gold returns. The residuals are crucial as they potentially indicate the influence of external, unmodeled shocks or events on market dynamics, making them valuable for detecting regime shifts. In Model 2, VAR residuals are employed as predictive features to enhance the forecasting of regime changes. This is crucial because VAR residuals capture the component of asset returns that is not explained by the linear interdependencies between multiple financial time series, such as SPY and gold returns.

In model 2, Initially, the VAR model is fitted to the time series data of SPY and gold returns. This model helps to quantify the interdependencies between these assets. The number of lags used in the VAR model is selected based on the lowest AIC value, ensuring that the model is neither underfit nor overfit. The model incorporated 15 lags to capture the complex, long-term interactions between these two assets. The inclusion of multiple lags allows us to examine the extended influence past values of each series exert on the present, a critical factor in financial markets where past trends can inform future movements. The VAR model itself is primarily designed to capture linear relationships between multiple time series; however, it can be adapted or extended to study tail risks and distributions through several approaches. The vector autoregressive model for order 1, denoted as VAR (1), is as follow:

$$
\begin{aligned}
Y_{1t} &= c_1 + \alpha_{11}Y_{1,t-1} + \alpha_{12}Y_{2,t-1} + \varepsilon_{1t} \\
Y_{2t} &= c_2 + \alpha_{21}Y_{1,t-1} + \alpha_{22}Y_{2,t-1} + \varepsilon_{2t}
\end{aligned} \tag{3}
$$

Where y_{1t} and y_{2t} are the vectors of observations for both assets at time t. C_i represents the intercept terms for each equation. α_{ij} are the coefficients which represent the effect of asset i on asset j at time $t-1$. ϵ_1 and ϵ_2 are the error terms.

The VAR residuals has been stored and incorporated into the dataset as new features. This is done because the residuals may contain information about potential regime shifts that are not captured by typical price movements or volatility indicators alone. The rationale is that these residuals, indicating deviations from expected behavior based on historical correlations, might be early indicators of a shift in market dynamics.

The choice of a VAR model is particularly apt for financial time series data where the behavior of one asset can be influenced by the performances of others. By using a multivariate approach, we gain a richer, more nuanced understanding of market dynamics, essential for accurate regime prediction, furthermore using a Vector Autoregression (VAR) model to investigate the tail distribution of assets involves examining the behavior of the assets during extreme market conditions, such as financial crises or other tail events.

The output of fitting the VAR model on SPY and GOLD provides this information: The Vector Auto-regressive (VAR) Model used the sample size (N_{obs}) of 5103 observations, indicating the number of daily returns used in the model after accounting for lags. The number of equations is 2, corresponding to the two-time series (SPY returns and gold returns). The result of AIC values helps in determining the optimal number of lags for the model. The model has chosen 15 lags based on the AIC criterion, which suggests a relatively complex interrelationship with significant past values influencing current values.

With 15 lags, the model is designed to capture complex, long-term inter dependencies between the assets. This might indicate that the effects of changes in one asset on the other can be delayed or spread out over an extended period that can be employed for deriving signals for regime change. After using the VAR residuals accompanied with volatility results and moving averages result into the GBDT, Neural Networks and Logistic regression for predicting the regime change, The result of the machine learning models showed perfect accuracy and precision which indicates overfitting problem which is an indication that the model requires tuning parameters which requires advanced technologies both hardware and software systems as a result we postpone this section to further analysis for future.

Table 2: Model Performance Comparison

Metric	GBDT	Logistic Regression	Neural Network
Precision	1.00	1.00	1.00
Recall	1.00	1.00	1.00
F1 Score	1.00	1.00	1.00

3.1.3 Advanced GARCH and Machine Learning Ensemble

Garch and JGR GARCH Volatility

The primary objective of Model 3 in this financial analysis is to enhance the estimation of market regimes through dynamic strategies, particularly by employing advanced GARCH models to evaluate the volatility of returns on two key assets: SPY (S&P 500 ETF) and GLD (Gold ETF). The standard GARCH model is further extended to include JGR-GARCH modeling to capture leverage effects, which play a pivotal role in identifying market regimes. This extension is essential as it adds depth to the volatility assessment, providing a more nuanced understanding of market dynamics under varying conditions.

GARCH model, Generalized Arch model by Bollerslev (1986) define the conditional volatility of asset returns with respect to the information from previous days. Given the assumption that asset returns are normally distributed with mean zero and conditional variance:

$$(y_t|Y_{t-1}) \sim N(0, h_t), \tag{4}$$

$$h_t = \mathrm{Var}(y_t|Y_{t-1}) = \alpha_0 + \alpha_1 y_{t-1}^2 + \beta_1 h_{t-1} \tag{5}$$

Imposed conditions are:

$$\alpha_0 > 0, \quad \alpha_1 \geq 0, \quad \beta_1 < 1 \tag{6}$$

y^2 is the asset return and h_t is the conditional volatility.

The JGR GARCH model, also known as Glosten, Jagannathan, and Runkle's GARCH model, is designed to accurately reflect the dynamics of financial time series that exhibit asymmetric volatility. This model is particularly effective in capturing the "leverage effect," a phenomenon where negative asset returns increase future volatility more significantly than positive returns. By incorporating an additional term that adjusts volatility based on the sign of the previous return, the JGR GARCH model enhances the standard GARCH model's ability to predict conditional volatility under conditions of market stress or financial turbulence. This makes it an invaluable tool for financial modeling and risk management, providing a more nuanced understanding of how news and events differently impact volatility depending on their nature (positive or negative).

The JGR GARCH model utilizes an indicator function to capture asymmetry in the volatility response to returns. Specifically, the indicator function $I(y_{t-1} < 0)$ is employed within the model's variance equation. This function is defined as follows:

$$I(y_{t-1} < 0) = \begin{cases} 1 & \text{if } y_{t-1} < 0 \\ 0 & \text{otherwise} \end{cases}$$

This function allows the model to differentiate the impact of negative returns from positive returns on future volatility. When y_{t-1} is negative, indicating a decline in asset prices, the function takes the value of 1, which activates the leverage term α_1' in the model's equation. This term adjusts the volatility prediction upward, reflecting the typical observed behavior that negative shocks increase volatility more than positive shocks of the same magnitude. The equations incorporating this function are centered and labeled as follows:

$$y_t = \sqrt{h_t}\epsilon_t, \quad \epsilon_t \sim NID(0,1), \tag{7}$$

$$h_t = \alpha_0 + (\alpha_1 + \alpha_1' I(y_{t-1} < 0))y_{t-1}^2 + \beta_1 h_{t-1} \tag{8}$$

To aid in regime change detection, the analysis incorporates moving averages (MA50 and MA200) as features, which are commonly used indicators for trend analysis. These features, along with the volatilities predicted by the GARCH models, form the basis for the feature set used in machine learning models, including Neural Networks, Gradient Boosting Decision Trees (GBDT), and Logistic Regression. Among these, Neural Networks demonstrated superior predictive accuracy, guiding the dynamic adjustment of asset allocation between SPY and GLD. For consistency with the overall investment strategy presented in the paper, a tactical asset allocation of 70 percent to 30 percent between the assets, depending on the predicted regime, was adopted. The performance of this dynamic strategy was then quantitatively assessed, measuring the effectiveness of the regime prediction and subsequent asset allocation adjustments in optimizing portfolio returns.

3.1.4 Realized Volatility, Bipower Variation Analysis for Regime Detection

The primary aim of Model 4 in this analysis was to improve the prediction of market regime changes by incorporating advanced volatility measures—specifically realized volatility and bipower variation—into the modeling framework. Realized volatility is derived from daily squared log returns, which provides a more granular and accurate depiction of the market's actual volatility dynamics than traditional historical volatility metrics.

The formula for realized volatility (RV) is given by:

$$RV = (\log(\frac{P_t}{P_{t-1}}))^2$$

where P_t and P_{t-1} are the asset prices at times t and $t-1$ respectively.

Bipower variation (BV), calculated using the product of the absolute values of consecutive daily log returns scaled by a factor of $\pi/2$, complements realized volatility by estimating daily volatility while mitigating the influence of outliers and market jumps. The formula for bipower variation is:

$$BV = \frac{\pi}{2}\sum_{i=1}^{n-1} |r_i||r_{i+1}|$$

where $r_i = \log(\frac{P_i}{P_{i-1}})$ are the log returns.

By integrating these sophisticated volatility metrics with moving averages (MA50 and MA200), which serve as indicators of longer-term market trends by smoothing out price data over specific periods, Model 4 leverages a comprehensive set of features. These features combine short-term volatility signals with longer-term trend information, creating a robust framework for detecting shifts in market regimes. The hypothesis is that this combination of immediate and extended data insights enables more timely and informed adjustments to investment strategies.

Within this enhanced feature set, Model 4 employs three distinct machine learning models—Neural Networks, Gradient Boosting Decision Trees (GBDT), and Logistic Regression—to evaluate and compare the predictive capabilities of each approach under the enriched data environment. The utilization of realized volatility and bipower variation, in particular, aims to equip these models with the nuanced data necessary to effectively anticipate regime shifts, thereby optimizing dynamic asset allocation and potentially enhancing portfolio returns through informed, data-driven decision-making.

The Gradient Boosting Decision Tree (GBDT) model has demonstrated exceptional performance in Model 4, characterized by high precision in both bear (97%) and bull (95%) market conditions. This has significantly influenced the portfolio's overall metrics, leading to notable achievements in terms of returns and risk management.

3.1.5 Synthesized RV and GARCH Strategies with ML Enhancement

The purpose of Model 5 was to develop a sophisticated dynamic asset allocation strategy by integrating the combined strengths and features of Models 1, 3, and 4. This comprehensive model aimed to enhance portfolio management by dynamically adjusting allocations between SPY (S&P 500 ETF) and GLD (Gold ETF) based on predictions of market regimes. The integration of various features from previous models allowed Model 5 to leverage a broad spectrum of predictive insights, maximizing potential returns during bullish market conditions and minimizing losses during bearish phases.

The features for Model 5 were meticulously constructed from an extensive set of variables that included all the volatility models and moving averages developed in the earlier models. This included realized volatility, GARCH and JGR-GARCH model outputs, and moving averages like MA50 and MA200. By amalgamating these features, Model 5 utilized a rich dataset that encompassed both short-term volatility signals and long-term trend data, providing a robust framework for the machine learning algorithms to predict market regimes. This integration enabled the model to effectively decipher complex patterns in market behavior, thus optimizing the decision-making process for asset allocation in response to predicted market changes.

Model 5 further refines its asset allocation strategy by employing a specific distribution of assets based on the regime predictions provided by its sophisticated predictive models. In a bullish market regime, the strategy allocates 70 percent of the portfolio to SPY (S&P 500 ETF) and 30 percent to GLD , capitalizing on the upward momentum of the stock market. Conversely, in bearish conditions, the allocation shifts to enhance risk management by favoring a safer asset, allocating 30 percent to SPY and 70 percent to GLD.

The performance metrics of Model 5 along with model 1, 3, 4 are summarized in section 4 Portfolio construction.

3.2 Method 2 - Wasserstein K-Means

We begin by defining and setting up the problem. Since we are looking to cluster the regimes using the prices of an asset we define the time series of data as a stream of data $\mathcal{S}$ over $\mathcal{X}$ where $\mathcal{X}$ is the daily prices.

$$\mathcal{S}(\mathcal{X}) = \{\mathbf{x} = (x_1, ..., x_n) : x_i \in \mathcal{X}\} \tag{7}$$

Therefore, $\mathcal{S}$ defines the price path of our asset across time. Intuitively, clustering based on prices is not very useful for the performance of a stock within a given regime and so using $\mathcal{S}$ directly is not valid. Therefore, we calculate a vector of log returns associated to r_i^S for every price x_i in $\mathcal{X}$.

$$r_i^2 = log(s_{i+1}) - log(s_i) \quad \text{for} \quad 0 \leq i \leq N-1 \tag{8}$$

The most crucial part of the Wasserstein K-means algorithm is the transition from Euclidean space to the space of distributions. To generate a measure then requires us to produce a set of segments of data, such that each segment can then produce for us an empirical measure of the probability distribution so that we may perform the clustering technique. We define the lifting function:

$$l = (l^1, ..., l^v) : \mathcal{S}(\mathcal{X}) \to \mathcal{S}(\mathcal{S}(\mathcal{V})) \tag{9}$$

Intuitively, each l is some smaller segment of the entire time series. Additionally, we define parameters h_1 and h_2. The former is the length of the segment while the latter is the amount that each segment overlaps. It is these lifting segments of data that we will cluster. To make the nomenclature and equations clear when we cluster we define the following:

$$Q^j : \mathcal{S}(\mathbb{R}) \to \mathbb{R} \tag{10}$$

Q is the j-th order statistic. We define also the μ, an empirical measure, to represent each segment of data where $\mathcal{X}$ is the indicator functon.

$$\mu^{\mathbf{x}}((-\inf, x]) = \frac{1}{N}\sum_{i=1}^{N} \mathcal{X}_{\{Q^i(\mathbf{x}) \leq x\}}(x) \tag{11}$$

For each μ associated with a segment of data l, we develop the family of measures $\mathcal{K}$ which acts as the set of data whose members μ_i we will be clustering.

$$\mathcal{K} = \{(\mu_1, ..., \mu_M) : \mu_i \in \mathcal{P}_p(\mathbb{R}) \text{ for } i = 1, ..., M\} \tag{12}$$

With the clustering entities μ defined, we can beging the the Wasserstein K-means algorithm summarized in Algorithm 1. The goal is clear, which is to produce k-centroids to reflect the desired k clusters we intend to find within the data.

To begin the k-means clustering, we randomly sample some some of initial centroids equal to the intended number of clusters k. Next, for every empirical measure μ we must assign it to its nearest centroid by calculating the Wasserstein distance from it to the initial centroid. For some empirical measures μ and ν, we can calculate this distance in equation (13). Due to being 1-dimensional we can simplify the original 1-Wasserstein equation as follows:

$$\mathcal{W}^1_{\mu,\nu} = \frac{1}{N}\sum_{i=1}^{N} |\alpha_i - \beta_i| \tag{13}$$

After we have assigned all μ to some cluster, we recalculate the centroid as the Wasserstein barycenter within each new cluster C_l.

Algorithm 1: Wasserstein K-means algorithm

Result: k centroids
calculate $l(r^S)$ given S
define family of empirical distributions $\mathcal{K} = \{\mu_j\}_{1 \le j \le M}$
initialise centroids $\overline{\mu}_i = 1, ..., k$ by sampling k times from $\mathcal{K}$
while *loss* > *tolerance* **do**
  foreach μ_i **do**
    calculate Wasserstein distance from μ_i to each centroid $\overline{\mu}_j$
    assign μ_i to closest cluster $C_l, l = 1, ...k$ by Wasserstein distance
  end
  calculate Wasserstein barycenter for each cluster C_l
  update centroids for each cluster C_l as the Wasserstein barycenters
  calculate loss
end

$$\overline{\mu} = \underset{\nu \in \mathcal{P}(X)}{\arg\min} \sum_{\mu_i \in \mathcal{K}} \mathcal{W}^1(\mu_i, \nu) \tag{14}$$

Finally, we calculate the loss function, which is defined as the distance from the old centroids to the new centroids. And until we reach a certain tolerance level, we continue the process.

$$l(\overline{\mu}^{n-1}, \overline{\mu}^n) = \sum_{i=1}^{k} \mathcal{W}^1(\overline{\mu}^{n-1}, \overline{\mu}^n) \tag{15}$$

In the end, the centroids define the center of the clusters while its membership is the resulting assignment of each empirical measure to its respective cluster.

3.2.1 Comparison to Original Paper

Compared to the paper results, while the paper uses the hourly time scale, even using the same parameters (h1, h2) = (35, 28) we see in Figure 1 and Figure 2 that results match up very closely to the expected results from the paper.

Similar to the paper, the regimes are well captured for all the expected periods of economic downturn as we'd expect. However we note that due to the increased time frame, our method does not react as quickly as the paper. This is expected since we are looking at more historical data (1.5 months) while the paper looks at 1 week.

3.2.2 Baseline Comparison - Vanilla K-means

When compared to the baseline method in Figure 3, we see that by looking at segments of data and performing the clustering on the probability space, we do see that we properly capture the performance of regimes rather than the intraday performance. While the vanilla K-means does segment the green and red days, there are still green days marked as a bull regime (such as in 2008) when clearly the market is in a bear state.

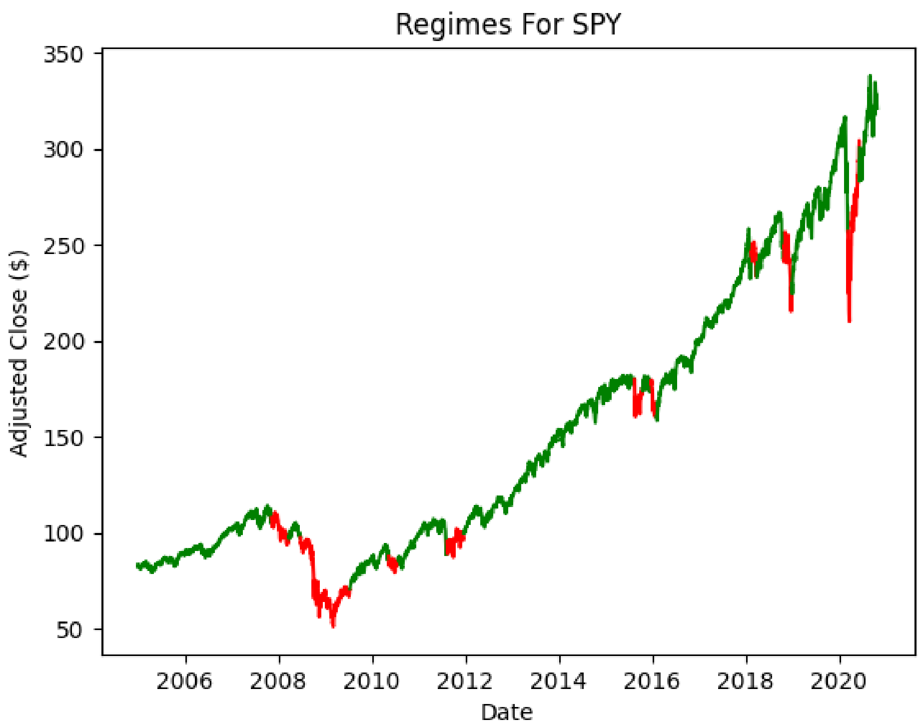


Figure 1: Clustered Regimes

Figure 2: Regimes from Horvath Paper

3.2.3 Hyperparameter Analysis

Analysis of h1 and h2

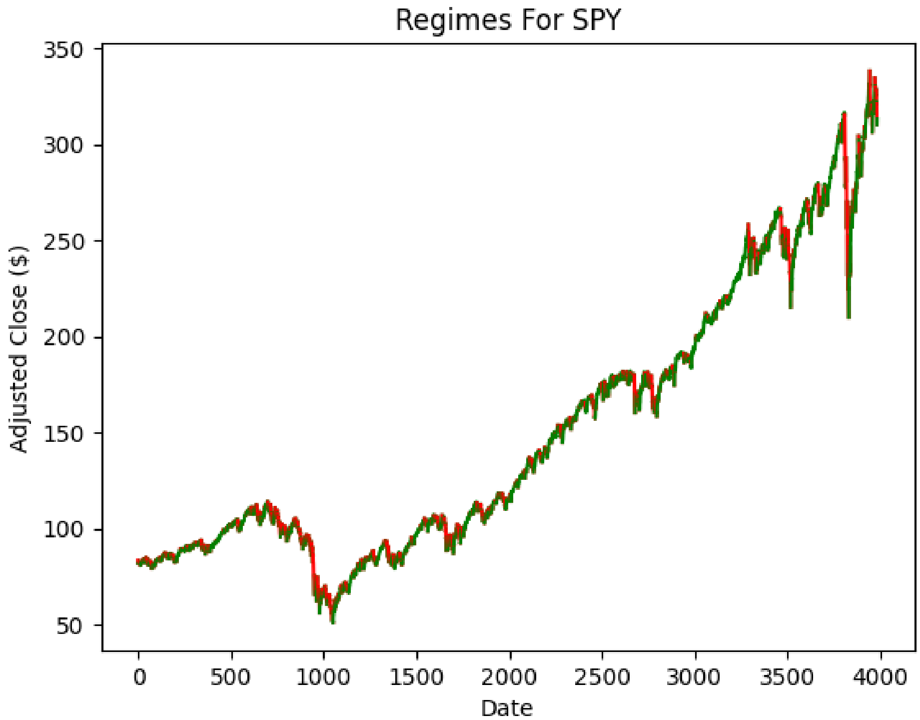


Figure 3: Regimes from Vanilla K-means

From the paper, it is stated that h2 does not appear to hold very much effect, but is quite useful in creating more data points when h1 is high since the higher h1 is the less segments you can perform. Therefore if you overlap segments, you can create as many data segments as days of data even if h1 is high. Nonetheless, h2 does not have a big impact on the clustering results.

The h1 parameter on the other hand does demonstrably have an impact as shown below. We see that with a shorter h1 (5), the regimes likewise follow the much shorter segmentation and are overeager in classifying results as bear market. This is expected behavior since each empirical measure only has 5 data points as its constituents so it is highly weighted towards present performance and does not have much memory of overall market trends.

As the paper suggests, h1 then becomes more of an “art rather than a science” to tune. We notice that for more volatile stocks, it may be better to have a lower h1 to be more updated while a less volatile stock, and those with prolonged downtrends are better serviced with higher h1.

For example, with TSLA, we see in Figures 5 and 6 that the clustering algorithm better captures the rising and falling trends of the volatile stock. Most notably in 2023, the higher h1=20 fails to capture the downtrends while h1=10 does.

Furthermore, we examine markets that have longer more protracted bear markets like that exhibited in the Chinese market (modeled by the index MCHI) as shown in Figures The longer h1=40 models the bear market much more effectively than the h1=20 which suggests a change into the bull market erroneously like in 2022.

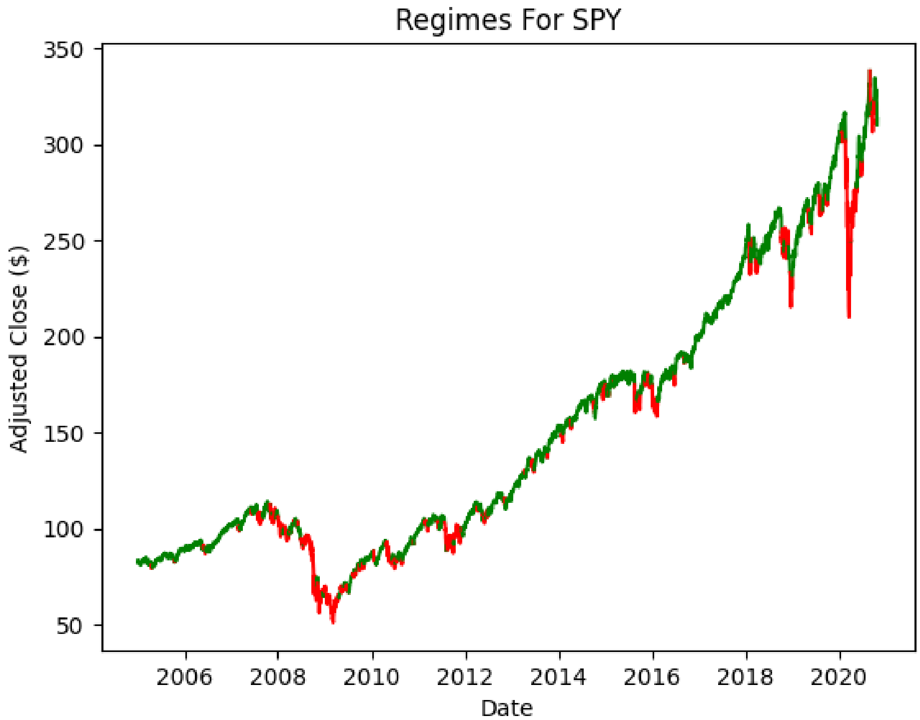


Figure 4: Regimes Cluster (h1 = 5)

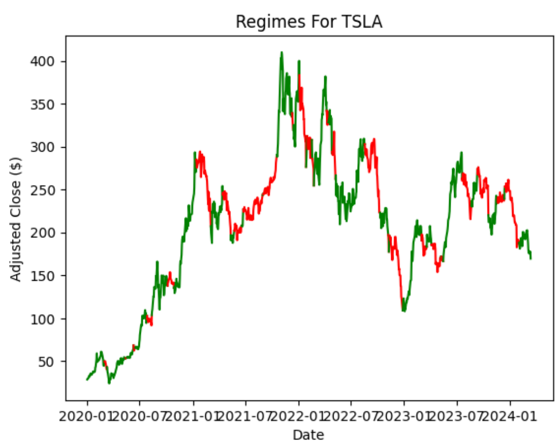


Figure 5: TSLA Regimes (h1 = 20)

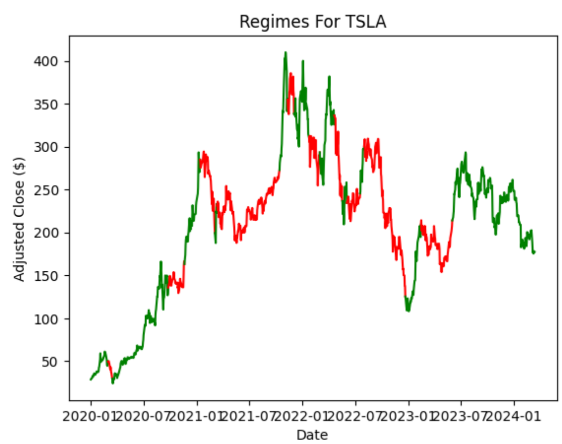


Figure 6: TSLA Regimes (h1 = 10)

3.2.4 Different Number of Clusters (k=3)

We experiment with 3 clusters and find that the results are not very clear like the bull vs bear regimes. The intuition is that there are bull, bear and flat regimes, but based on Figure 9, there is no clear difference among the 3 regimes that would translate into meaningful portfolio changes.

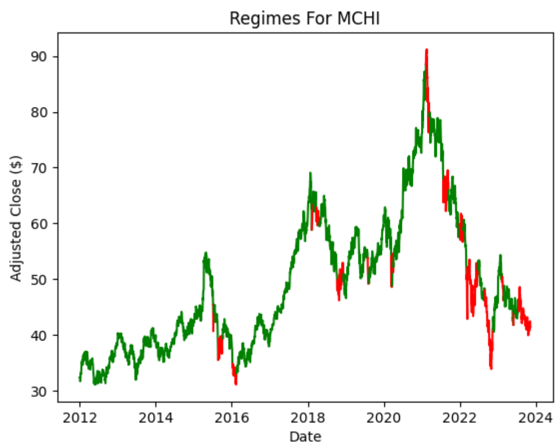


Figure 7: MCHI Regimes (h1 = 20)

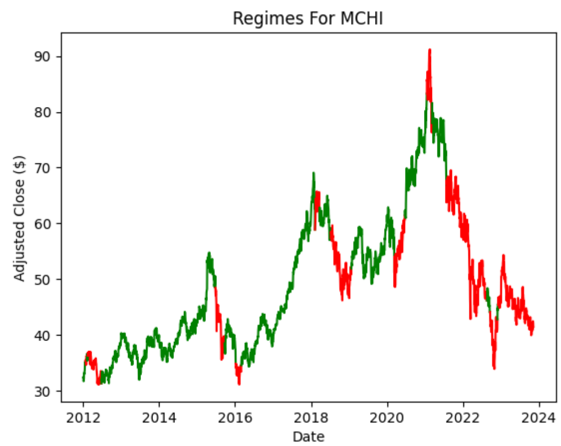


Figure 8: MCHI Regimes (h1 = 40)

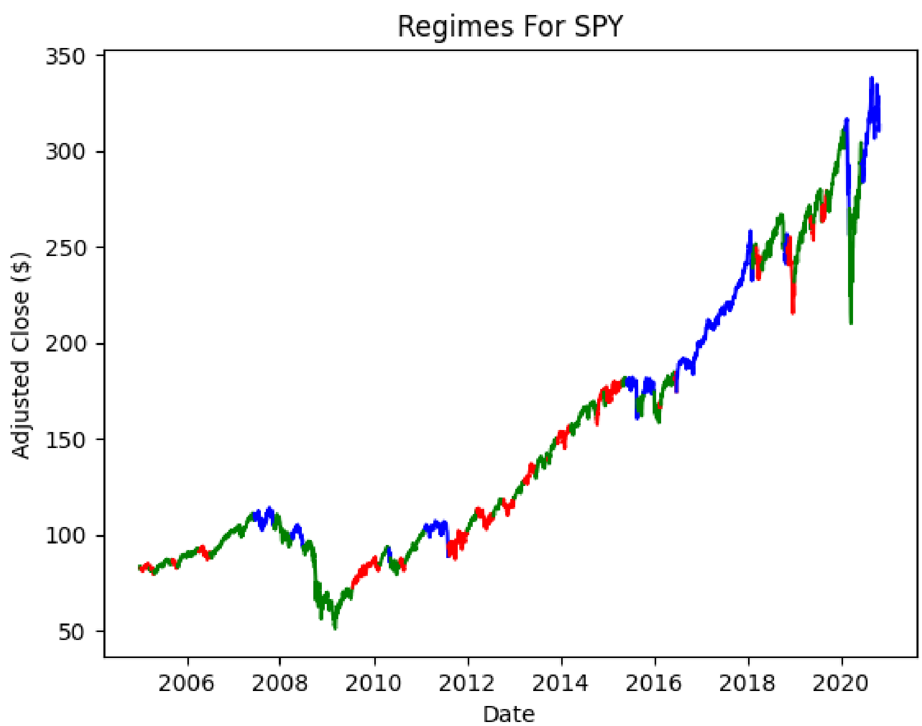


Figure 9: Regimes with K=3

3.3 Method 3 - PCA and K-means

In this study, we analyzed time series data spanning from the year 2003 to 2023, utilizing publicly available datasets from the Federal Reserve Economic Data (FRED). The data were recorded on a quarterly basis, reflecting a total of 94 observations per series over the 23-year period. These datasets were categorized into thirteen distinct groups, each representing different economic indicators and metrics. This classification facilitated a comprehensive analysis of various economic regimes and their dynamics over the specified period. The use of FRED data ensured that the study was grounded in reliable and widely recognized sources, allowing for meaningful interpretations of the economic trends and patterns.Divided data into two segments training data 2003 to 2018 and testing data from 2019 to 2023

Variable Symbol	Variable Name
GDP	Gross Domestic Product
UNRATE	Unemployment Rate
CPIAUCSL	Consumer Price Index
WM2NS	Money Supply
FEDFUNDS	Federal Fund Rate
PI	Personal Income
PCE	Inflation
GFDEBTN	Total Public Debt
HOUST	Housing and Urban Development
GPDI	Gross Private Domestic Investment
EPUMONETARY	Monetary Policy
EPUFISCAL	Fiscal Policy
PPIACO	Producer Price Index

The SPDR S&P 500 ETF (symbol: SPY) was selected for constructing the portfolio. This ETF was back-tested using the methodology described previously, which involved analyzing time series data from 2000 to 2023 with a quarterly frequency. The cumulative return of the SPY ETF was calculated to assess its performance over the specified period. For trading within the identified economic regimes, a principal component analysis augmented with K-means clustering was utilized to detect regime shifts. The trading strategy employed was tail hedging, which involves buying out-of-the-money put options to protect against potential downturns. Essentially, this strategy aims to safeguard profits by hedging against downside risks whenever a new regime is identified. By implementing this approach, the portfolio is designed to capitalize on the stability of the SPY ETF while minimizing losses during volatile market phases. The use of quarterly data provides a granular view of the market's movements, allowing for precise adjustments to the hedging positions as new economic regimes emerge. This strategy ensures that the portfolio is not only responsive to immediate market conditions but also well-prepared for potential future fluctuations.

Principal component Analysis

Principal Component Analysis is a linear dimensionality reduction technique. First, we must standardize our data to have a mean (μ) of 0 and a standard deviation (σ) of 1. Here, x represents our variable values, μ is the mean, and σ is the standard deviation.

$$Z = \frac{x - \mu}{\sigma}$$

After Standardization we need to find a covariance matrix. It would be a square matrix depending on the number of variables in data. So data in this method considered thirteen variables then matrix would be 13*13. After this step eigenvalues and eigenvectors are calculated.

$$\det(A - I) = 0 \quad \text{(Where } I \text{ is the Identity matrix)}$$

$$\det(A - I)\mathbf{V} = 0$$

Then PCA is applied $Y = XW$ where X is standardized data and W is the matrix of eigenvectors. In this method, in order to decide on the number of PCA components, we plot the variances considered up to 90% of the total variance.

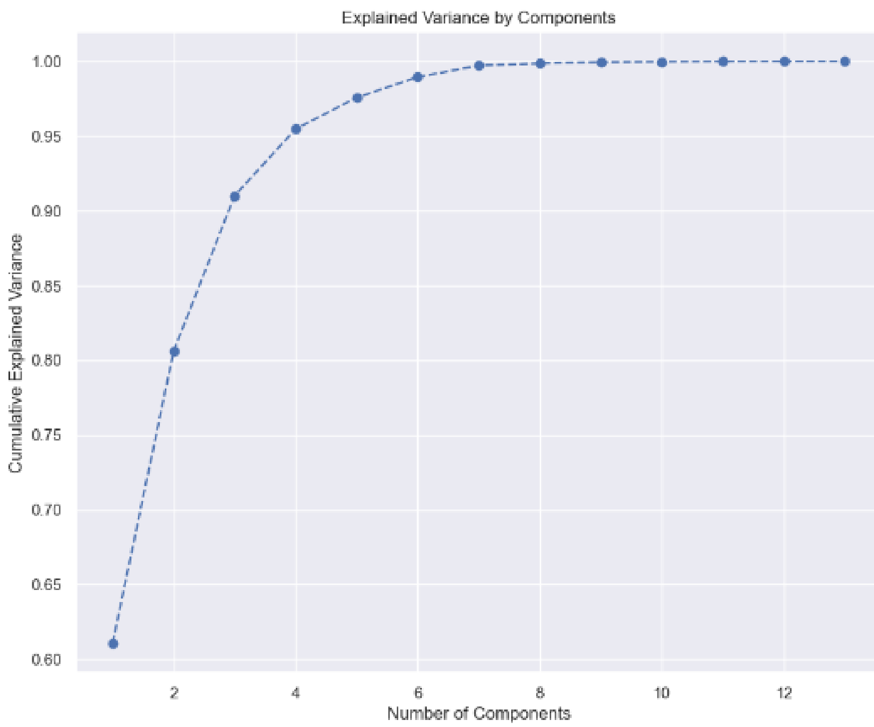


Figure 10: This plot helps in deciding number of principal components.

In the plot, up to 90% of the total variance, there are approximately three dots. Thus, we decided on three principal components for the analysis: Component 1, Component 2, and Component 3.

Regime Clustering

K- means clustering was applied. It is a method to cluster data points of the nearest centroid based on the number of clusters chosen.

$$J = \sum_{i=1}^{n} \sum_{k=1}^{K} w_{ik} \left\| x_i - \mu_k \right\|^2$$

Where n is the number of datapoints,k means the number of clusters,wik is binary indicator 1 if belongs to centroid 0 if doesn't belong to clusters. In order to decide upon the number of clusters used a method called elbow.

This is a simple method you see in the plot that there inconsistent drop of data points up to four. So the number of clusters for regime identification is four. Below

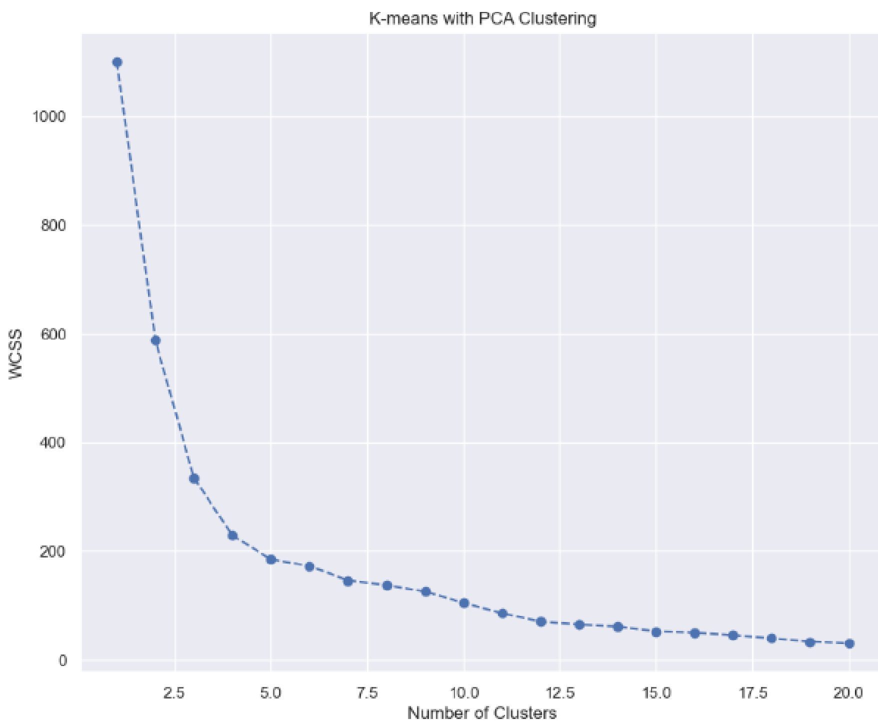


Figure 11: This plot helps in deciding number of clusters.

the plot there are four regimes: yellow represents regime1 ,green represents regime 2, blue indicates regime 3 & purple identifies as regime 4. Red line chart is index SPY.

Regime Detection

Regime detection as each data point allocated to the nearest centroid based on the number of centroid regimes are formed as we can see there are four regimes. We made regimes using economic data. Time frame of data was 2000 to 2023. For regime detection there are four centroids. Index was considered to backtest in the regimes. These regimes are more accurate if we also consider volatility of Stock combined with economic data.

4 Portfolio Construction

4.1 Method 1: Regime Change Prediction using Score Model, ML Models, Advanced volatility Model

Back testing and Portfolio Strategy Using the Score System

The investment strategy employed was straightforward yet strategic, focusing on asset allocation based on predictive market regimes. The allocation strategy involved directing 70 percent of the portfolio's capital towards the S&P 500 Index ETF (SPY) as the primary risky asset, and 30 percent towards Gold (GLD) during periods forecasted as bull markets. Conversely, in anticipated bear markets, the strategy shifted to

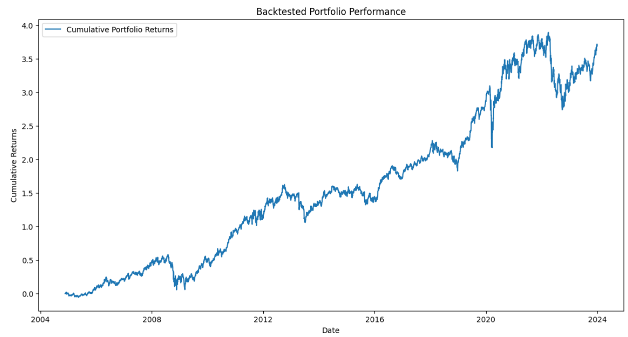


Figure 12: Model 1 Cumulative Returns of Dynamic Allocation Strategy

allocate 70 percent of the portfolio to Gold (GLD) for safety and 30 percent to SPY. This dynamic reallocation was predicated on the scores generated from our financial models, which aimed to predict market regimes effectively.

Portfolio Performance Metrics

A Sharpe ratio of 55 percent suggests that the portfolio provided a moderate return per unit of risk, which is a favorable outcome.

Table 3: Portfolio Performance Metrics Model 1 - Score system

Metric	Value
Annualized Portfolio Return	8.93%
Annualized Portfolio Volatility	14.38%
Annual Sharpe Ratio	0.5517 (55%)

Table 4: Portfolio Performance Metrics Model 3

Metric	Value
Annualized Return	9.88%
Annualized Volatility	12.79%
Sharpe Ratio	0.6949 (69.49%)

The graph shows the cumulative returns of a dynamic allocation strategy based on GBDT predicted market regimes from 2004 through 2024. This visual representation highlights a successful upward trajectory, indicating that the investment strategy effectively capitalizes on the GBDT ability to forecast different market conditions. The overall growth trend is punctuated by several significant dips and recoveries, which align with historical periods of market volatility. These fluctuations demonstrate the strategy's adaptability, as it dynamically adjusts asset allocations in response to predicted regime changes. Notably, the substantial increase in cumulative returns underscores the potential of using advanced predictive models to enhance portfolio performance over long periods.

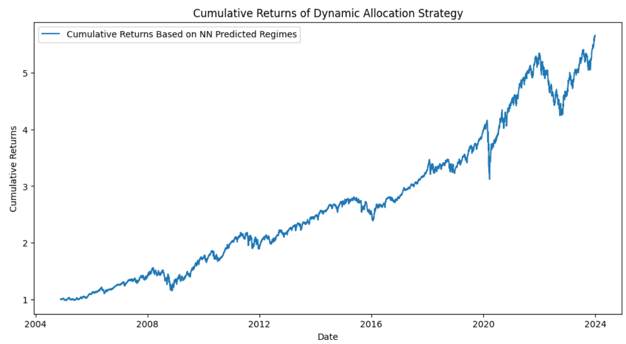


Figure 13: Model3 Cumulative Returns of Dynamic Allocation Strategy

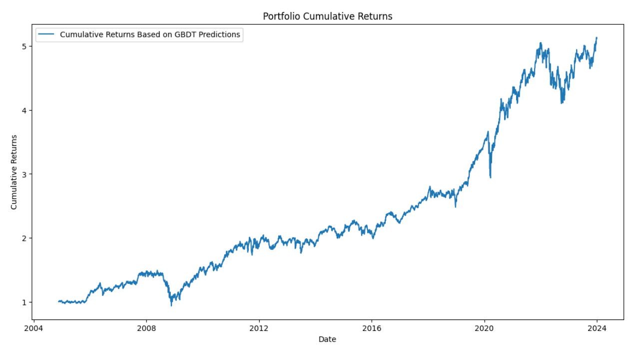


Figure 14: Model 4 Cumulative Returns of Dynamic Allocation Strategy

The GBDT model emerged as the top performer in predicting market regimes, demonstrating high precision rates—97 percent for bear markets and 95 percent for bull markets. This superior predictive ability contributed to impressive portfolio metrics, including a high annualized return and a reasonable Sharpe ratio, reflecting both strong returns and moderate volatility management.

Table 5: Portfolio Performance Metrics Model 4

Metric	Value
Annualized Return	99.70%
Annualized Volatility	14.37%
Sharpe Ratio	0.6057

Table 6: Portfolio Performance Metrics Model 5

Metric	Value
Annualized Return	11.02%
Annualized Volatility	13.70%
Sharpe Ratio	0.7315

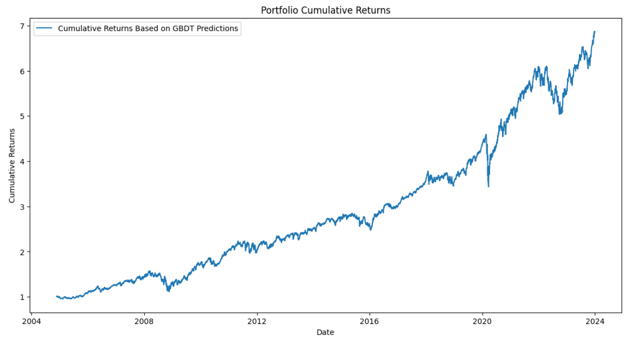


Figure 15: Model 5 Cumulative Returns of Dynamic Allocation Strategy

Model 5 showcases enhanced performance metrics, attributed to its comprehensive use of combined features from previous models. An annualized return of 11.02 percent signifies strong growth potential, while an annualized volatility of 13.70 percent reflects a well-managed risk profile. The Sharpe Ratio of 0.7315 indicates a favorable risk-adjusted return, suggesting that the model effectively capitalizes on the integrated volatility models and moving averages to optimize portfolio outcomes. These metrics collectively validate the robustness of Model 5 in achieving superior returns through informed, dynamic asset allocation.

These methods aimed to leverage a combination of advanced machine learning techniques and optimized feature engineering to predict market regime changes effectively. The most successful model emerged from an integration of various volatility models, which proved crucial in generating reliable signals for identifying regime changes. To better understand the practical implications of these predictions, a straightforward asset allocation strategy was implemented using the most accurate model predictions.

The results clearly demonstrate that engineered features derived from volatility models serve as robust indicators for regime changes. This is evident in the strategic allocation of the portfolio, which significantly benefited from the predictive power of the refined models. Furthermore, this study lays a strong foundation for future research, suggesting that there is considerable potential for further enhancement of feature engineering to improve predictive accuracy and investment outcomes even more.

In conclusion, model 5's performance highlights the advantage of using comprehensive data sets and complex algorithms to capture nuanced market dynamics, thereby facilitating informed investment decisions that optimize both returns and risk.

4.2 Wasserstein K-means

4.2.1 Signal Computation

The signal generation is straightforward. The prediction relies on the core assumption that regimes do not last for singular days and so if we have a change in the regime prediction then it should result in a period of different price behavior. Therefore, when get a single different signal then we can make a trade assuming the future regime will follow suit.

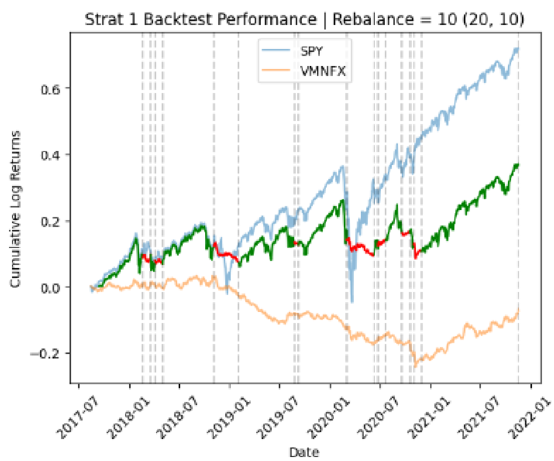
Using the algorithm we have already set up for the Wasserstein K-means, the regime change signal is calculated by taking the most updated stream of prices with sliding window parameter h1 and calculating its distance to pre-trained centroids on historical data.

It should be noted that the original paper by Blanka Horvath (2021) states that the Wasserstein method is well designed for "a posteriori" classification of regimes, so this section is mostly a foray of curiosity into a priori regime classification under the initial assumption.

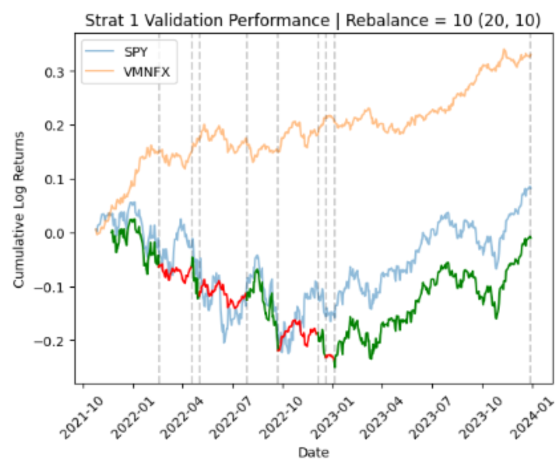
4.2.2 Portfolio Strategy - Long only SPY VMNFX

This strategy involves longing SPY (S&P 500) when we are in a bull market and then switching to long VMNFX (Vanguard Market Neutral Fund) in a bear market. The idea here is to avoid the market and move funds into a market neutral fund when the SPY is suspected to be in a bear market.

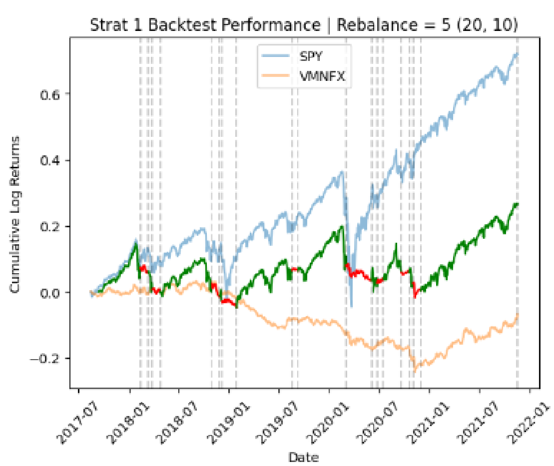
Predictions utilize the most up to date segment of length h1 to compute the distances to existing centroids and determine membership and thus regime labels. We introduce an additional parameter to our testing, which is the re-balancing period R.



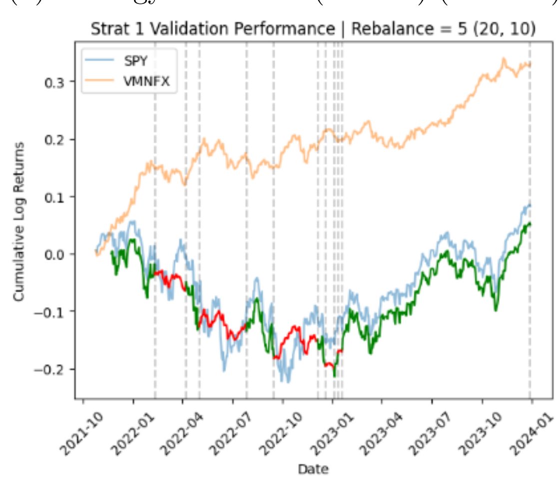
(a) Strategy (R = 10) (h1 = 20)



(b) Strategy Validation (R = 10) (h1 = 20)



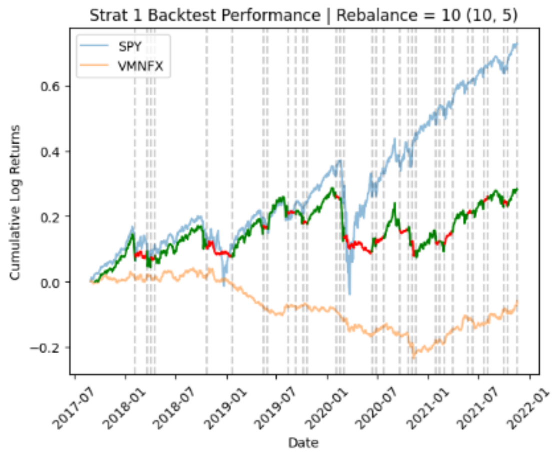
(c) Strategy (R = 5) (h1 = 20)



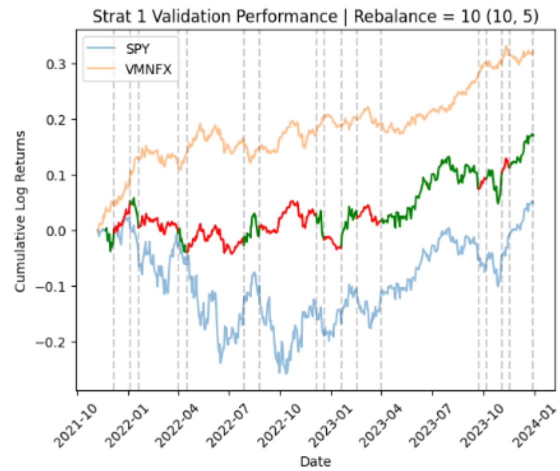
(d) Strategy Validation (R = 5) (h1 = 20)

Figure 16: Testing and Validation Performance with Various Re-balance Periods (h1 = 20)

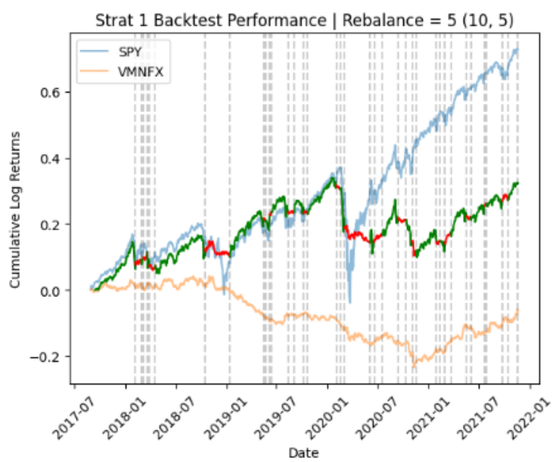
We see from Figures 16a which is a monthly look back, the stock appeared to perform decently for the testing periods with no significant difference between biweekly or weekly re-balance (R = 10 or 5). However, we notice in Figure 16b that it performs very poorly in the validation set, so this may not be a valid choice. It seems that the longer period is too slow to pick up on changes in regime and therefore is too delayed in selecting the correct stocks.



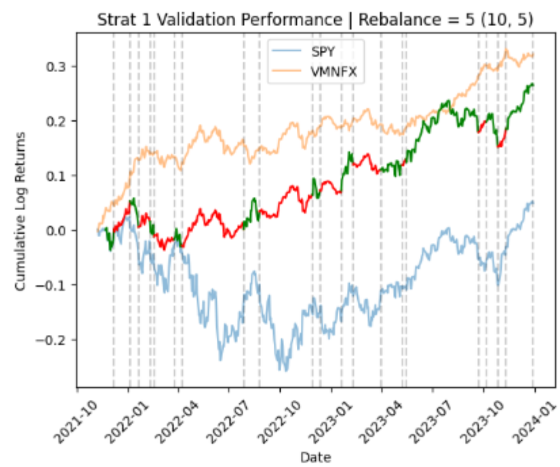
(a) Strategy (R = 10) (h1 = 10)



(b) Strategy Validation (R = 10) (h1 = 10)



(c) Strategy (R = 5) (h1 = 10)



(d) Strategy Validation (R = 5) (h1 = 10)

Figure 17: Testing and Validation Performance with Various Re-balance Periods (h1 = 10)

Next we reduce the parameter h1 from 20 to 10, which is 2 week look back. We see in Figure 17 that the testing performance is relatively similar even as we adjust re-balancing dates. However, we do see a significant improvement of validation set performance. This is likely because we are minimizing the amount of lagged data we need to perform the distance computation, which allows us to keep mostly on pace with developing market conditions.

Overall, it seems that the success of the model is in prolonged bear markets, where the algorithm is able to switch to VMNFX and stick with it for a long period of time to avoid the losses incurred in SPY. These longer periods reduce the impact of our delayed reaction and therefore are able to capitalize on the more steady returns of VMNFX.

4.2.3 Performance Metrics and Analysis

Table 7 shows the portfolio performance of each strategy in validation set. From overall comparison we can see that as we decrease both the re-balance period and the sliding window parameter, we produce better volatility adjusted performance. Across all metrics, we obtain better returns, better worse-case scenarios, and better volatility. This strongly reinforces the need to be up to date with the changing regimes, since recent days have more weight as the re-balance period and sliding window parameter decreases. Sharpe Ratio is substantially better for (R, h1) = (5, 10) than even the close second at 1.1201 compared to 0.6518.

We also notice that the sliding window parameter appears to have a greater effect than the re-balancing parameter. A 50% decrease in the sliding window parameter results in a 0.6518 Sharpe Ratio versus just 0.1693 after a 50% decrease in the re-balancing parameter parameter. This is likely because

Table 7: Validation Performance Statistics

(R, h1)	**(10, 20)**	**(10, 10)**	**(5, 20)**	**(5, 10)**
Mean Ret.	-0.0000	0.0003	0.0001	0.0005
Mean Ret. Ann.	-0.0049	0.0744	0.0239	0.1259
Min Return	-0.0445	-0.0344	-0.0445	-0.0200
Min Return Ann.	-11.2030	-8.6776	-11.2030	-5.6840
10-Day Drawdown	-0.1200	-0.0731	-0.0866	-0.0500
Volatility	0.0091	0.0072	0.0089	0.0071
Volatility Ann.	0.1441	0.1142	0.1412	0.1124
Sharpe Ratio	-0.0021	0.0411	0.0107	0.0706
Sharpe Ratio Ann.	-0.0341	0.6518	0.1693	1.1201
Skewness	-0.5768	-0.3198	-0.5884	-0.0973
Kurtosis	1.8728	1.4372	2.1457	1.2344

When analyzing the strategy across both Bull and Bear regimes as shown in Figure 18, the performance appears to be in between VMNFX and SPY. While it may not make as much as SPY in the longterm, the strategy ends up not limiting losses significantly in bear regimes. In fact, the Sharpe Ratio is on average better than both SPY and VMNFX when taking into account the different market regimes as shown in Table 8. We see that while the mean returns are not as high as SPY, its minimum returns and max 10-day drawdown are dramatically better than SPY itself. This lends to a lower volatility and thus better volatility adjusted performance.

From looking at Figure 18, it's evident that the strategy has some levels of loss due to delay on smaller short term bear markets, but is very good at protecting the portfolio from extended bear markets as shown in 2020. In addition, it seems to catch back up with the rising bull market following the long term bear market. This ability to automatically "stop loss" does well to reduce the stress on investors and despite a slightly lower return performs much more steadily across all market conditions.

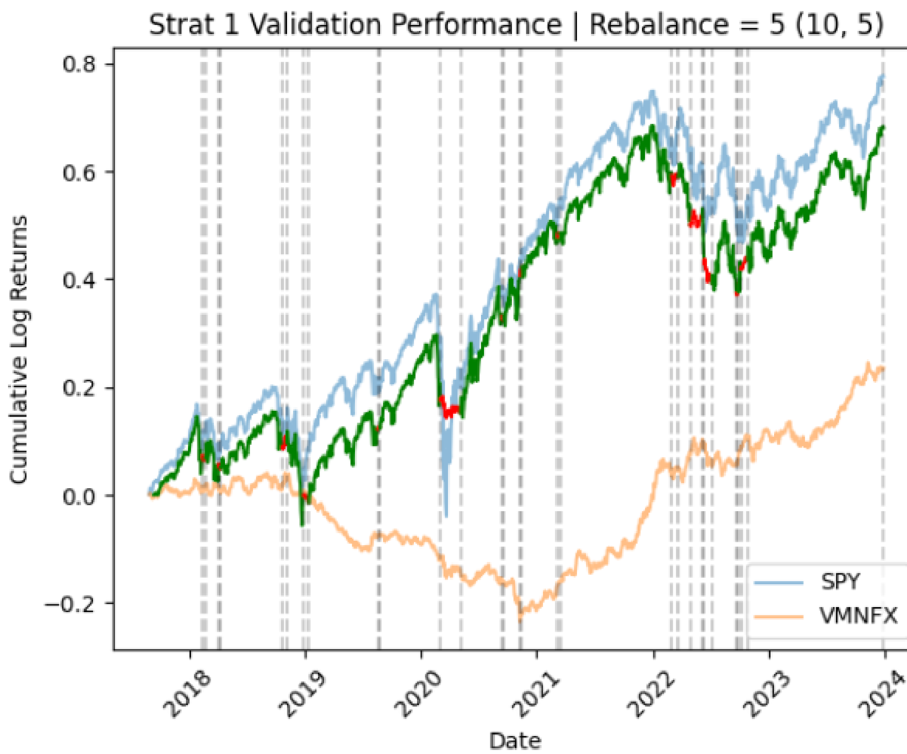


Figure 18: Performance of Strategy in Bull and Bear Regimes

Table 8: Portfolio Performance Benchmarked (R = 5) (h1 = 10)

	Strat.	**SPY**	**VMNFX**
Mean Ret.	0.0004	0.0005	0.0002
Mean Ret. Ann.	0.1082	0.1197	0.0379
Min Return	-0.0594	-0.1159	-0.0401
Min Return Ann.	-14.9631	-29.2034	-10.0926
Max 10-Day Drawdown	-0.1119	-0.2334	-0.0455
Volatility	0.0096	0.0126	0.0041
Volatility Ann.	0.1527	0.2000	0.0658
Sharpe Ratio	0.0446	0.0377	0.0363
Sharpe Ratio Ann.	0.7087	0.5985	0.5766
Skewness	-0.5536	-0.8057	-0.7502
Kurtosis	3.9351	12.5874	6.3140

Regardless, for future work, the biggest impact on performance is still, unfortunately, the reliance on historical data to calculate mu. For example, if current day k has progressed to a different regime, we do not detect this until a majority of days within the window of h1 are labeled with the new regime. Therefore, for larger h1 such as h1 $= 5$, we would need $k-1$, $k-2$ to all become a new regime before the new regime is classified. This is the core issue with this strategy and leads to some undue losses and missing out of some gains since we are sometimes too late to adapt to the new regime. This is especially the case when the initial change of regimes is a stark market crash or recovery.

Future explorations may involve taking smaller portions of the h1 window and then exploring what the possible obscured pieces are likely to be. This can be accomplished through deep learning or n-gram prediction while maintaining a bias towards minimizing the Wasserstein distance of the full h1 window to any given centroid. This will likely reduce the amount we have to wait before classification. Alternate and perhaps easier method may involve using hourly data, so that we can reduce the number of days to 1 or 2 and still produce a valid empirical measure. Either are possibilities to reduce the amount of lag and therefore produce a prediction close to real-time.

4.3 Method 3 - PCA and K-means

Regime Detection

Used a strategy called tail hedging to construct and backtest the results.it is simple strategy buy out of money put when new regimes starts. Calculated the Cumulative returns using this strategy and regimes. Below plot are regimes under training data.

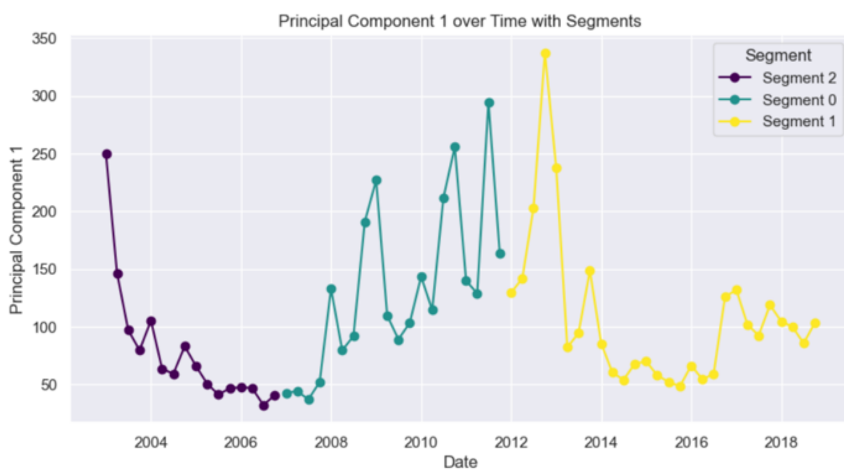


Figure 19: Regimes According to Component 1

These are regime according to testing regimes for principal component one. We can see that there are three regimes for principal component from year 2003 to 2018. Now after that we have predicted regimes using machine learning for testing data and these are results. Below table results for testing regime for spy using tail hedging strategy. In summary, the investment or strategy appears to perform well in terms of high annualized returns and an exceptionally high Sharpe ratio, suggesting efficient risk-adjusted returns. However, it exhibits significant potential risks, as indicated by high volatility, negative skewness, high kurtosis, and a sizable maximum drawdown. These risk factors should be carefully considered in the context of the investment's return profile.

Table 9: Summary of Portfolio Performance by PCA K-means

Index	**SPY**
Volatility	21%
Annualized return	19.83%
Sharpe Ratio	0.9442%
Skewness	-0.5529
Kurtosis	11.55
Drawdown	-0.330

5 Conclusion

Compiled Results of Methods

Table 10: Compiled Performance Metrics

	Annualized Ret	**Annualized Vol**	**Sharpe Ratio**	
SPY	11.97%	20.00%	0.5985	
VMNFX	3.79%	6.58%	0.5766	
Volatility Models, GBDT	11.02%	13.70%	0.7315	
Wasserstein	10.82%	15.27%	0.7087	
PCA	19.83%	21.00%	0.9443	

As shown in Table 10, GBDT and Wasserstein did not outperform the SPY (SPDR S&P 500 ETF) in terms of raw returns. However, these two strategies had better risk-adjusted performance with better Sharpe Ratios in comparison to SPY. Wasserstein resulted in better resturns at the cost of higher volatility.

In any case, PCA seemed to perform by far the best with a Sharpe Ratio of 0.9443. This comes significantly at the cost of a large 21% annualized volatility but in context, it's actually only 1% more than SPY itself. So while the volatility is high, it's annualized return easily make it more attractive than SPY.

It seems that the results of the K-means methods produce a much larger amount of annualized volatility. This is especially true for PCA since it simply holds SPY the entire time and hedges against a bear market with OTM puts. It could be that there is a degree of lag that creates an undesirable amount of volatility during regime changes.

In conclusion, this paper contributes to the ongoing discourse on market regimes by presenting three distinct methodologies for identifying and utilizing these regimes within financial markets. We demonstrate a nuanced approach to detecting market regimes beyond the traditional dichotomy of 'bull' and 'bear' markets. We show that by taking into account regimes in the construction of the portfolio we can produce better returns and better volatility adjusted returns relative to benchmarks.

6 Future Work

While explored some methods like Tail Hedging, other methods of hedging for the change of portfolios can be explored. Perhaps some methods will perform better at hedging for bear regimes and can potentially produce better outcomes. If we had employed more suitable methods of tail hedging rather than simply building a portfolio then the results of the methods can be either more comparable or competitive.

We also identify the crucial issue of lag within our K-means clustering results. GBDT method did not have this issue since it is an inherently predictive ML model. K-means by default is more suitable for a posterior classification than ex ante prediction. Therefore it may be more suitable to use the signals generated by the K-means methods as an input into a machine learning model as a feature to predict the momentum of a regime going forward. Perhaps if the lagged indicator was allowed to be more fuzzy within the confines of a ML model, it can produce value as a indicator of regime fluctuation.

Finally, there is likely interest in producing an ensemble method of sorts. If we can take the signals of all three models and average them out and see if 2 out of 3 agree or even all 3 agree, we could identify those moments as a much stronger signal and potentially produce a more reliable outcome.

Overall, we have just scratched the surface of the utility of these regime based models, and while we may have answered some questions we understand that the work of market regimes is far from fully encompassed.

References

Aigner, A. A. (2023). A gaussian mixture hidden markov model for the vix. "https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4389763".

Benhamou, E., J. Ohana, D. Saltiel, B. Guez, and S. Ohana (2014). Explainable ai (xai) models applied to planning in financial markets. "http://papers.ssrn.com/sol3/papers.cfm?abstract_id=3862437/".

Blanka Horvath, Zacharia Issa, A. M. (2021). Clustering market regimes using the wasserstein distance. "https://arxiv.org/abs/2110.11848".

Carlos Heitor Campani, Rene Garcia, M. L. (2021). Optimal portfolio strategies in the presence of regimes in asset returns. *Journal of Banking and Finance 123*, 1–2.

Chan, E., H. Fan, and S. Sawal (2021). Conditional portfolio optimization: Using machine learning to adapt capital allocations to market regimes. "https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4383184".

Chiara Lattanzi, M. L. (2019). A changepoint approach for the identification of financial extreme regimes. "http://https://arxiv.org/abs/1902.09205/".

David Hallac, Peter Nystrup, S. B. (2019). Greedy gaussian segmentation of multivariate time series. *Advances in Data Analysis and Classification 13*(3), 727–751.

Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica 57*(2), 357–384.

Lingfei Li, B. W. (2023). Reinforcement learning for continuous-time mean-variance portfolio selection in a regime-switching market. "http://papers.ssrn.com/sol3/papers.cfm?abstract_id=4396073/".

Mathieu Gatumel, F. I. (2014). The number of regimes across asset returns: Identification and economic value. *International Journal of Theoretical and Applied Finance 17*(6), 1450050+.

Peter Akioyamen, Yi Zhou Tang, H. H. (2021). A hybrid learning approach to detecting regime switches in financial markets. "https://arxiv.org/abs/2108.05801".

Pier Francesco Procacci, T. A. (2019). Forecasting market states. *Quantitative Finance 19*(9), 1491–1498.

Zhang, C., Y. Zhang, M. Cucuringu, and Z. Qian (2023). Volatility forecasting with machine learning and intraday commonality. "https://arxiv.org/abs/2202.08962".