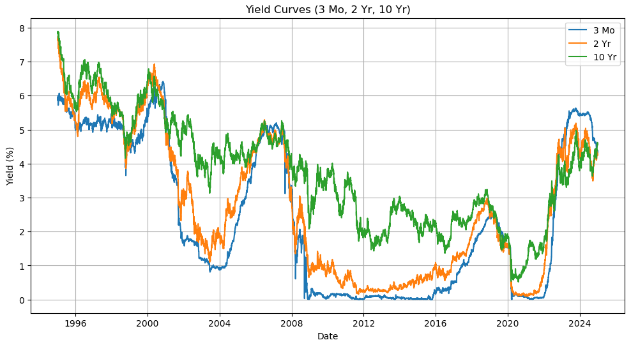
**Forecasting Interest Rate Regimes**

[GuardsGuards/Forecasting-Interest-Rate-Regimes](https://github.com/GuardsGuards/Forecasting-Interest-Rate-Regimes/tree/main)

K-Means Clustering,

GARCH (volatility persistence), Permutation Entropy (chaos), Hurst’s Exponent (trend persistence),

Principal Component Analysis, Random Forests, Neural Networks, and Transformers

****

Abstract

In Forecasting Interest Rate Regimes, I employ both continuous prediction and classification frameworks to capture the dynamics of yield regimes. I first utilize K-means clustering on raw interest rate levels to identify distinct market regimes (low, moderate, and high) and, separately, on cluster regime indicators derived from GARCH (volatility persistence), Permutation Entropy (PE) (complexity of market dynamics), and the Hurst Exponent (HE) (trend persistence).

I assess my data for cointegration and multicollinearity of yields and indicators to confirm suitabilities and then proceed to employ various models Principal Component Analysis, Random Forests, Neural Networks, and Transformers for predicting precise yield levels and regime shifts.

My methodology also incorporates various lag structures (no lag, 30-day lag, and 180-day lag) to explore how historical yield information influences future interest rate movements, thereby exploring effective forecasting horizons (degradation of forecasting power over time).

For continuous forecasts, I evaluate model performance using Root Mean Square Error (RMSE), which measures the average magnitude of prediction errors. Lower RMSE values indicate a model’s superior ability to capture the precise movements of yields. I also employ a binning process, using the median yield as a threshold to classify predictions into binary regimes (e.g., high vs. low yield environments). This classification is evaluated using Accuracy and the Receiver Operating Characteristic (ROC) curve, with the ROC Area Under the Curve (AUC) providing a robust measure of the model’s ability to discriminate between regimes.

Not only does the aforementioned methodology reveal key insights regarding interest rate regimes, it is easily extractable and applicable to an inexhaustive array of financial (and broader) forecasting applications.

Purpose

GARCH, Permutation Entropy (PE), and the Hurst Exponent (HE) provide substantial clarity on “why” interest rates are behaving as they are, and “how” that behavior might progress through time.

* **Quantitative Precision:**  
  These metrics provide numerical measures of volatility clustering (GARCH), complexity (PE), and memory/persistence/mean-reversion (HE), enabling precise and objective analysis. They help quantify how extreme or persistent market behavior is, which is difficult to gauge visually.
* **Forecasting Power:**  
  Incorporating these measures into forecasting models improves their ability to predict regime shifts. By quantifying aspects of the time series that affect future movements, the models can provide early warnings of transitions between high- and low- rate environments, which is crucial for risk management and strategic decision-making.
* **Risk Management:**  
  Understanding volatility dynamics and complexity is key for assessing risk. For instance, periods with high volatility clustering (captured by GARCH) or increased complexity (indicated by higher PE) might signal upcoming market stress, prompting more cautious risk management strategies.
* **Model Robustness:**  
  Combining these quantitative measures with traditional interest rate levels creates a more robust framework. It helps ensure that predictions aren’t based solely on observable trends but also account for the underlying mechanics driving those trends, making the overall system more resilient to unexpected changes.

Methodology

1. **Regime Identification via Clustering:**  
   K-Means clustering is applied to yield levels to identify distinct interest rate regimes (low, high, and moderate).
2. **Modeling Volatility, Complexity, and Persistence:**
   * **Volatility:** GARCH models are employed on returns across different maturities to assess volatility persistence and shock impacts.
   * **Complexity:** Permutation entropy is used to gauge the unpredictability of yield dynamics.
   * **Trend Persistence:** Hurst exponents provide insight into the trending behavior or mean reversion properties of the yield curve.
3. **Interdependence Assessment:**  
   I conduct cointegration and Variance Inflation Factor (VIF) analyses to verify that yield measures and the derived regime indicators share a long-run equilibrium and to evaluate multicollinearity.
4. **Predictive Modeling:**  
   A range of models including PCA regression, Random Forests, Neural Networks, and Transformers are employed to forecast regime shifts. Continuous forecasts are evaluated using RMSE, while classification performance is assessed via confusion matrices, accuracy, and ROC AUC.

Calculations

* **K-Means Clustering:**

K-means clustering is an unsupervised learning algorithm that partitions a dataset into K distinct clusters. The goal is to assign each data point to one of the clusters in such a way that the total intracluster variance is minimized (via the sum of squared distances (errors)), effectively creating clusters where data points are as close to each other as possible. The objective function for K-means is defined as:

* + and
* is the squared distance between data point and the centroid ,
* **GARCH (Generalized Autoregressive Conditional Heteroskedasticity):**  
  GARCH models capture the clustering of volatility over time, indicating that shocks to returns tend to persist. A standard GARCH(1,1) model is defined as:
  + (constant term): This parameter represents the long-run average level of variance in the absence of recent shocks. It ensures that the conditional variance remains positive even when past shocks or volatilities are small. Typically, ω is required to be positive (ω > 0), and its magnitude influences the baseline volatility level in the model,
* α **(ARCH term):** The parameter α measures the immediate impact of new information (or shocks) on volatility. It multiplies the squared residual ​, so a large α indicates that recent shocks have a strong effect on current volatility. This term captures the “news effect,” where unexpected events lead to a spike in volatility. In practice, α is constrained to be non-negative (α ≥ 0).
* **(GARCH term):** The parameter β reflects the persistence of volatility from one period to the next by weighing the previous period’s conditional variance ​. A high β means that shocks to volatility die out slowly, indicating that volatility is highly persistent. As with α, β is generally constrained to be non-negative (β ≥ 0).
* **Permutation Entropy (PE):**  
  Permutation Entropy measures the complexity or unpredictability of a time series by analyzing the order of values in a series (with respect to their magnitudes).
  + For example, if d =3 and τ = 1, you would look at patterns formed by sequences like [,]. Each vector is then mapped to a permutation pattern based on the relative ordering of its elements (e.g., whether is the smallest, the middle, or the largest):
* there are d! possible distinct ordinal patterns, and
* is the probability of each permutation pattern. It is bounded between 0 and 1, where 0 indicates a completely predictable dynamic and 1 indicates a completely stochastic dynamic (i.e. more disorder and less predictability), which helps to identify regimes where market behavior is more chaotic.
* **Hurst Exponent:**  
  The Hurst exponent quantifies the tendency of a time series to either persist (trend) or revert to the mean. It is derived from the scaled range analysis:

  + means that the expected range grows proportionally to n raised to the power H.

An value above 0.5 indicates persistence (trending behavior), while a value around 0.5 suggests randomness, and below 0.5 suggests anti-persistence.

* **Cointegration**

Cointegration is a statistical property of a collection of time series variables that indicates the existence of a long-run equilibrium relationship among them, despite being individually non-stationary, given by:

The rank of the matrix determines the number of cointegrating relationships:

* + If rank () = 0, there is no cointegration.
  + If 0 < rank () = r < k, there are r cointegrating vectors, meaning that there exist r linear combinations of the variables that are stationary.

The Johansen cointegration test utilizes the trace statistic, which is given by:

* and

This trace statistic is compared against critical values to determine whether to reject the null hypothesis of no cointegration.

* **Variance Inflation Factor (VIF)**

The Variance Inflation Factor (VIF) quantifies the extent of multicollinearity among the predictors in a regression model. For a given predictor , the VIF is defined as:

where ​ is the coefficient of determination obtained by regressing on all the other predictors. A higher VIF indicates greater collinearity, suggesting that the predictor is highly correlated with the others, which can inflate the variance of the estimated coefficients. Typically, VIF values greater than 10 are cause for concern, however, values greater than 5 still deserve scrutiny.

* **Principal Component Analysis (PCA):**  
  PCA is a dimensionality reduction technique that transforms a set of possibly correlated variables into a set of uncorrelated principal components. For a data matrix X, PCA involves computing the eigenvalues and eigenvectors of the covariance matrix:

and then transforming the data:

where is the matrix of eigenvectors corresponding to the largest eigenvalues.

* **Random Forest (RF):**  
  Random Forest is an ensemble method that builds multiple decision trees and aggregates their predictions to improve regression or classification accuracy. In regression tasks, the prediction is typically the average of the predictions from individual trees:

Where is the prediction from the t-th tree and is the total number of trees.

* **Neural Networks (NN):**  
  Neural Networks are computational models consisting of layers of interconnected neurons that apply linear transformations followed by non-linear activation functions. A simple feed-forward neural network computes the activations for layer as:
* **Transformers:**  
  Transformers are deep learning architectures that leverage self-attention mechanisms to model sequential data, capturing long-range dependencies more effectively than traditional recurrent networks. The many calculations involved exceed the scope of this paper. However, the self-attention mechanism is formulated as:
* **Root Mean Square Error (RMSE):**

RMSE quantifies the average magnitude of the prediction error by measuring the square root of the average squared differences between the predicted and actual values. A lower RMSE indicates better predictive accuracy.

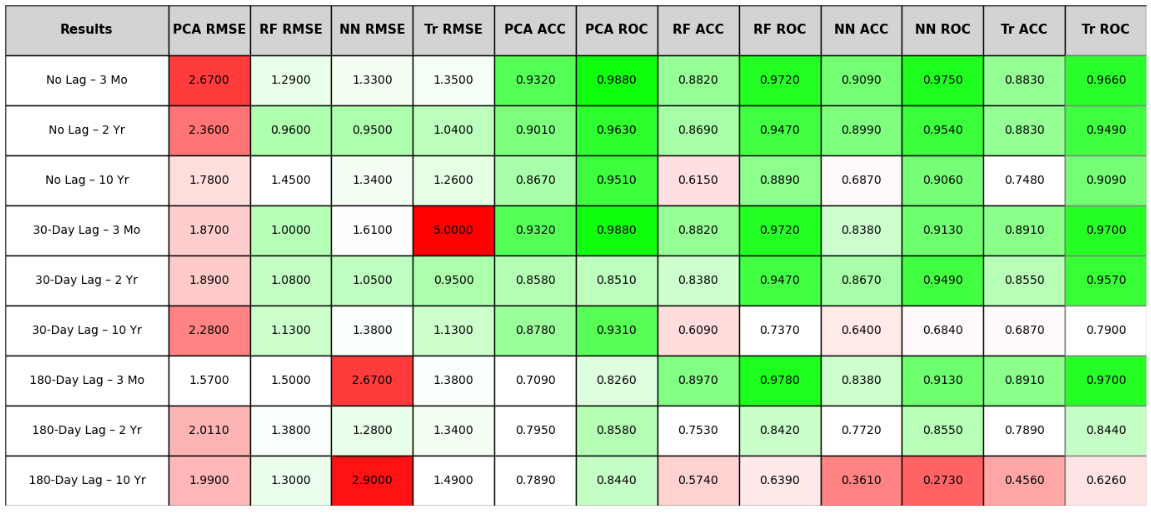
* **Accuracy:**

Accuracy measures the proportion of correct predictions (both high and low regimes) among all predictions made. Accuracy is calculated using the confusion matrix elements as follows:

* **Receiver Operating Characteristic (ROC) and ROC AUC:**

The area Under the ROC Curve characterizes the model’s ability to discriminate between the two classes (high vs. low regime). An ROC AUC of 1 represents perfect discrimination, while an ROC AUC of 0.5 suggests no discriminative ability (equivalent to random guessing). The ROC curve is created by plotting the True Positive Rate (TPR) against the False Positive Rate (FPR) at various threshold settings. The TPR and FPR are given by:

Results

****

Insights

* **Distinct Yield Regimes:**
  + Low-Rate Environment: Characterized by steep, upward-sloping yield curves despite low absolute levels, indicative of accommodative monetary policy and subdued inflation expectations.
  + High-Rate Environment: Marked by flat yield curves with uniformly high yields, reflecting periods of monetary tightening or elevated inflation pressures.
  + Moderate-Rate Environment: Exhibits a clear upward slope with intermediate yields, representing a balanced regime.
* **Volatility Dynamics and Persistence:**  
  GARCH modeling reveals that short-term yields (e.g., 3 Mo) exhibit highly persistent volatility, indicating that shocks tend to have prolonged effects. This insight is crucial for risk management.
* **Complexity Patterns via Permutation Entropy:**  
  Analysis shows divergent complexity patterns across regimes. In some cases, longer-term yields are less predictable (higher PE), while in others, short-term yields display greater complexity. Transient shocks appear to dominate near-term dynamics.
* **Trend Behavior via Hurst Exponent Analysis:**  
  Varied. The Hurst exponent analysis demonstrates that some clusters exhibit strong persistence (with values well above 0.5) across maturities, indicative of stable trending behavior, whereas others show reduced persistence in short-term yields, suggesting faster mean reversion.
* **Predictive Modeling Insights:**  
  Despite non-linear models (Random Forests, Neural Networks, and Transformers) achieving lower RMSEs in continuous forecasts, PCA regression excels in binary regime classification, evidenced by superior accuracy and ROC AUC values. This finding implies that preserving the relative ordering of predictions is paramount for effective regime detection, even if absolute forecast precision is compromised.
* **Impact of Lag Structures:**  
  The incorporation of various lag lengths (no lag, 30-day, and 180-day) reveals that while longer lags result in mildly reduced prediction scores, in some cases, it is possible to forecast interest rate movements to a high degree of accuracy.

Conclusion

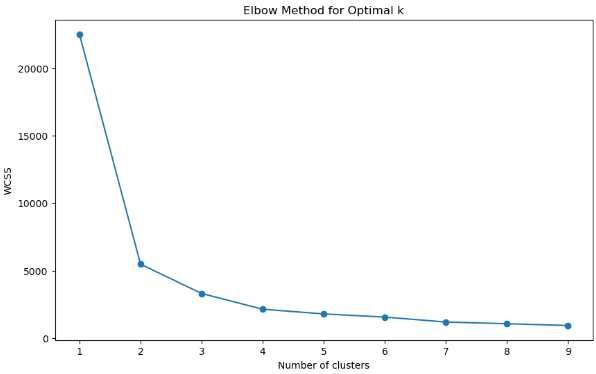
This study demonstrates that a multi-faceted approach — integrating measures of volatility (GARCH), complexity (Permutation Entropy), and trend persistence (Hurst Exponent) — provides profound insights into the dynamics of the yield curve. My findings underscore that distinct market regimes, each characterized by unique volatility, complexity, and persistence profiles, can be effectively identified and forecasted. Moreover, while advanced non-linear models provide more accurate continuous predictions, linear techniques such as PCA, when reframed for classification, offer exceptional discriminative power for regime shifts. These insights enhance risk management strategies and establish a robust framework for predicting regime transitions in interest rate environments, thereby informing more strategic and informed financial decision-making.

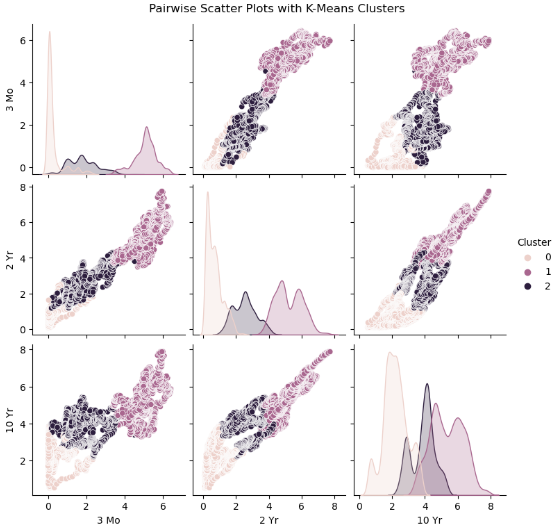
Cluster Analysis

Cluster analysis is a powerful unsupervised learning technique that groups data points based on similarity, revealing underlying patterns and regimes within complex datasets. In my study, cluster analysis is essential for identifying distinct interest rate regimes and the corresponding behavior of key market indicators. I employ k-means clustering due to its computational efficiency, ease of implementation, and clear interpretability, which are crucial when partitioning large datasets into meaningful clusters. K-means minimizes the within-cluster variance, providing a robust framework for distinguishing between different market conditions.

Clustering on Yield Levels

**Introduction:**  
K-means clustering is applied directly to the raw interest rate data (3 Mo, 2 Yr, and 10 Yr yields) to uncover distinct market regimes. This segmentation allows us to differentiate periods characterized by varying absolute yield levels and term structures, which are indicative of different monetary policy stances or economic conditions.

****

****

**Cluster 0:**

* **Observations:** 2,987 data points
* **Level of Rates:** Very low yields across the board:
* 3 Mo: mean ≈ 0.30%
* 2 Yr: mean ≈ 0.69%
* 10 Yr: mean ≈ 2.20%
* **Variability:** Moderate standard deviations with slight variability:
  + The standard deviations are moderate (e.g. ~0.48 for 3 Mo, ~0.47 for 2 Yr, and ~0.73 for 10 Yr), suggesting that while the levels are low, there is still some variability.
* **Range:** The minimum rates are near zero and the maximums remain relatively low:
* 3 Mo: min ≈ 0.00%
* 10 Yr: max ≈ 3.85%
* **Interpretation:**

This cluster represents a low interest rate environment, possibly associated with accommodative monetary policy or economic conditions with low inflation expectations. The yield curve here is upward sloping (10 Yr >> 3 Mo) but at an overall low level.

**Cluster 1:**

* **Observations:** 2,754 data points
* **Level of Rates:** Yields are considerably higher:
  + 3 Mo: mean ≈ 5.07%
  + 2 Yr: mean ≈ 5.22%
  + 10 Yr: mean ≈ 5.34%
* **Variability:** Relatively modest standard deviations with tighter clustering:
  + Standard deviations are around 0.59 to 0.98, indicating that the higher rates are more closely grouped
* **Range:** Minimums are in the mid-3% range and maximums extend to roughly 6–7.9%
* **Interpretation:**  
  This cluster reflects a high interest rate environment where yields across all maturities are elevated and closely aligned. The relatively flat nature of the yield curve suggests periods of monetary tightening or economic uncertainty.

**Cluster 2:**

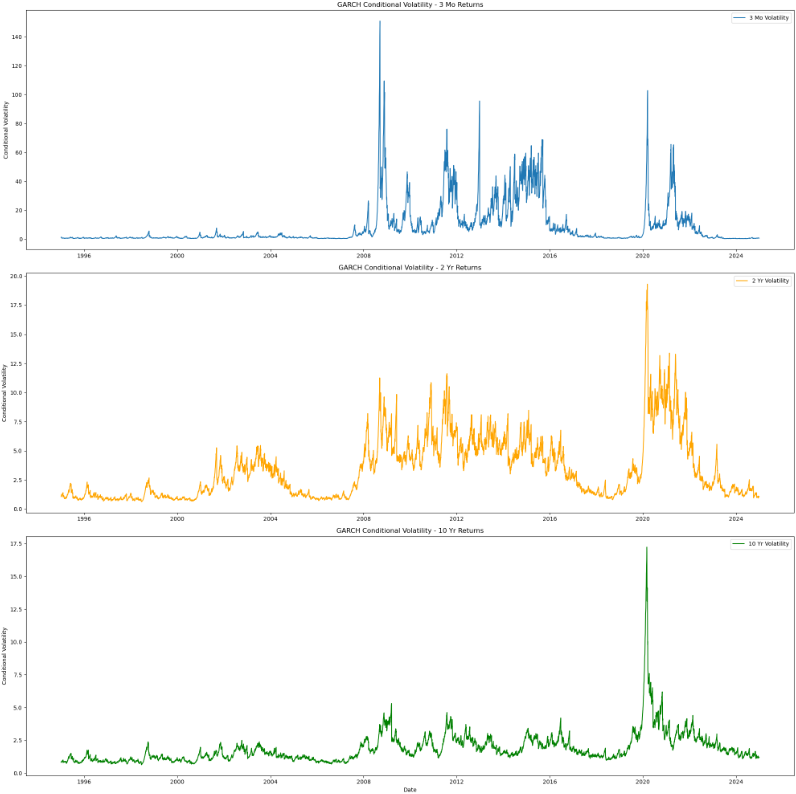
* **Observations:** 1,762 data points
* **Level of Rates:** Intermediate yield levels:
  + 3 Mo: mean ≈ 1.81%
  + 2 Yr: mean ≈ 2.53%
  + 10 Yr: mean ≈ 3.87%
* **Variability:** Moderate standard deviations indicating some variability:
  + Standard deviations are approximately 0.76 for 3 Mo, 0.70 for 2 Yr, and 0.71 for 10 Yr
* **Range:** The yields show a clear upward trend from short-term to long-term
* **Interpretation:**  
  This cluster represents a moderate interest rate regime with a pronounced upward-sloping yield curve. It suggests an environment that is neither as accommodative as Cluster 0 nor as restrictive as Cluster 1, a middle ground in yield levels.

**Conclusion:**

* **Cluster 0:** Low-rate regime with generally low yields and a relatively steep upward slope (10 Yr significantly higher than 3 Mo).
* **Cluster 1:** High-rate regime with elevated yields across all maturities and a flat yield curve, indicating uniform high rates.
* **Cluster 2:** Moderate-rate regime with an upward-sloping curve, where longer maturities offer a higher yield relative to short-term rates.

These interpretations can help in understanding the different market conditions represented in the data: for example, periods of low rates might coincide with quantitative easing or economic downturns, while high rates could reflect tight monetary policy or high inflation expectations.

Clustering GARCH Volatility

****

**3 Month Returns**

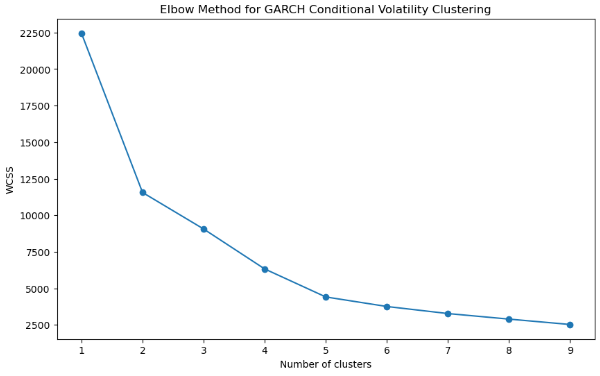
* **Mean:**
  + The estimated constant (mu) is –0.0344 and is statistically significant (t = –2.55, p ≈ 0.0108). This suggests that the mean return for the 3 Mo series is significantly negative.
* **Volatility:**
  + Omega is 0.0139 (p ≈ 0.026), indicating a non-zero baseline volatility level.
  + Alpha[1] is 0.1348 (highly significant) and Beta[1] is 0.8652 (also highly significant). Their sum is nearly 1.0, which indicates that shocks to volatility persist strongly over time (i.e. the volatility clustering is very persistent).

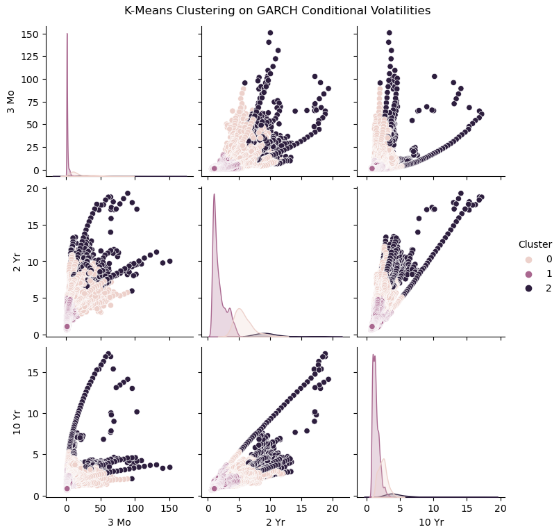
**2 Year Returns**

* **Mean:**
  + The estimated mean is very close to zero (–0.0039) and is not statistically significant (p ≈ 0.826).
* **Volatility:**
  + Omega is 0.0137, while Alpha[1] is 0.0760 and Beta[1] is 0.9240.
  + The high beta (0.924) again indicates persistent volatility clustering. In this case, the lower alpha suggests that the immediate impact of shocks is somewhat less than in the 3 Mo series.

**10 Year Returns**

* **Mean:**
  + The constant is 0.0174 but is not significant (p ≈ 0.268).
* **Volatility:**
  + Omega is 0.0114, Alpha[1] is 0.0648, and Beta[1] is 0.9339.
  + Similar to the 2 Yr, the volatility persistence is high (beta > 0.93) and the immediate impact of shocks is modest.

****

****

**Cluster 0:**

* **Observations:** 2,173 observations
* **Volatility Levels:**
  + 3 Mo: Mean conditional volatility ≈ 19.51
  + 2 Yr: Mean conditional volatility ≈ 5.66
  + 10 Yr: Mean conditional volatility ≈ 2.40
* **Variability:** High dispersion for 3 Mo (SD ≈ 14.73); moderate variability for 2 Yr (SD ≈ 1.45) and relatively tight dispersion for 10 Yr (SD ≈ 0.66)
* **Range:** Wide range for 3 Mo and narrower ranges for longer maturities
* **Interpretation:**  
  This cluster is characterized by extremely high short-term volatility, suggesting an environment of acute market stress or turbulence, while medium and long-term rates exhibit more stable volatility profiles.

**Cluster 1:**

* **Observations:** 4,899 observations
* **Volatility Levels:**
  + 3 Mo: Mean conditional volatility ≈ 1.56
  + 2 Yr: Mean conditional volatility ≈ 1.81
  + 10 Yr: Mean conditional volatility ≈ 1.37
* **Variability:** Low dispersion across all maturities (e.g., SD ≈ 1.63 for 3 Mo, ≈ 1.02 for 2 Yr, ≈ 0.49 for 10 Yr)
* **Range:** Relatively tight ranges across the board
* **Interpretation:**  
  This cluster represents a calm, low-volatility regime across all maturities, indicative of stable market conditions and minimal uncertainty.

**Cluster 2:**

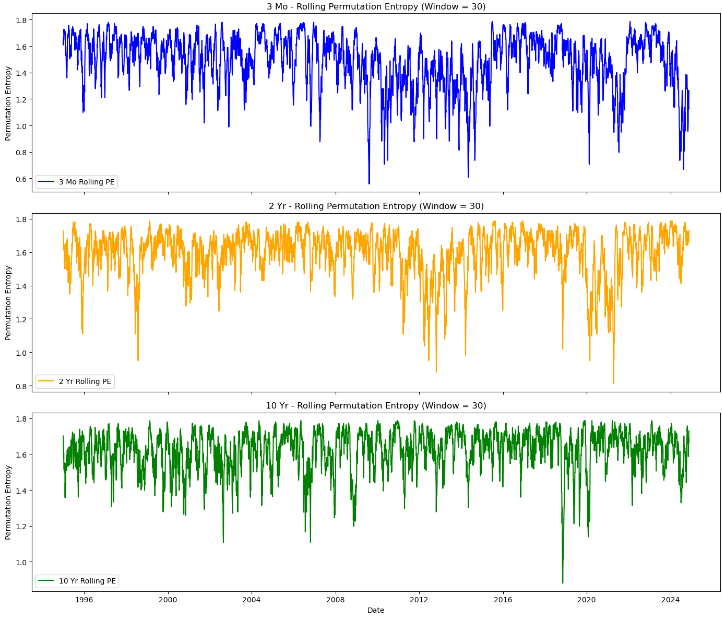
* **Observations:** 412 observations
* **Volatility Levels:**
  + 3 Mo: Mean conditional volatility ≈ 33.02
  + 2 Yr: Mean conditional volatility ≈ 9.74
  + 10 Yr: Mean conditional volatility ≈ 5.00
* **Variability:** Very high dispersion for 3 Mo (SD ≈ 26.95); moderate for 2 Yr (SD ≈ 2.45) and 10 Yr (SD ≈ 2.70)
* **Range:** Significantly wider for short-term rates
* **Interpretation:**  
  This cluster indicates a high-volatility regime across the yield curve, likely occurring during periods of crisis or significant market uncertainty, with the most pronounced effects on short-term rates.

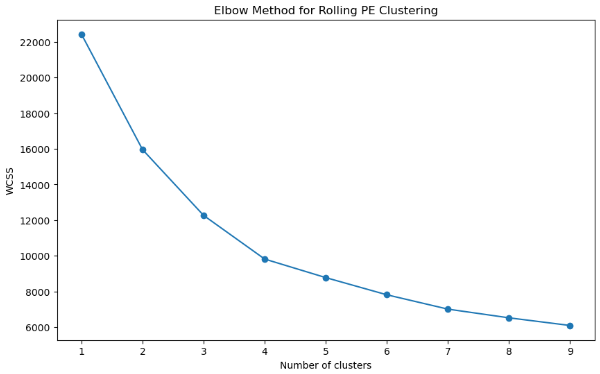
**Conclusion:**

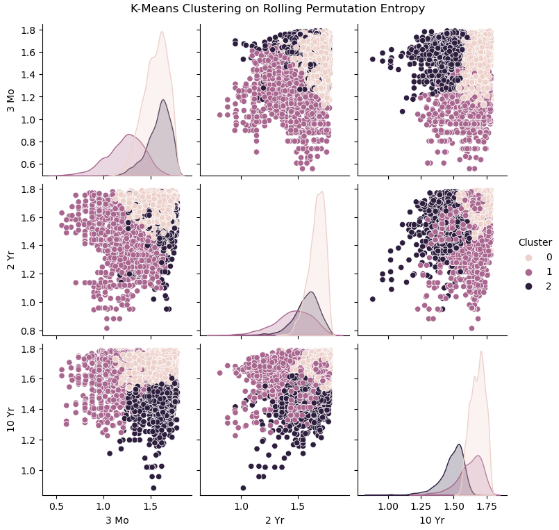
* **Cluster 0:**  
  Reflects a regime with acute short-term turbulence; short-term (3 Mo) volatility is extremely high while medium and long-term volatilities remain moderate. This may signal rapid, short-lived shocks that affect near-term rates more than long-term expectations.
* **Cluster 1:**  
  Represents a stable, low-volatility environment where all maturities exhibit low, consistent volatility, typical of periods with little market stress or when monetary policy is steady.
* **Cluster 2:**  
  Identifies a high-volatility regime across the yield curve, especially impacting short-term rates. This may occur during financial crises or periods of significant uncertainty, where both short-term and longer-term yields experience elevated volatility.

These insights can be invaluable for risk management and forecasting purposes. Knowing which regime the market is in can help in adjusting investment strategies, hedging positions, or even selecting appropriate models for forecasting yield dynamics.

Clustering Permutation Entropy (PE)

****

****

****

**Cluster 0:**

* **Observations:** 4,063 data points
* **Mean PE:**
  + 3 Mo: ~1.55
  + 2 Yr: ~1.66
  + 10 Yr: ~1.68
* **Variability:** Very small standard deviations (e.g., ≈0.13 for 3 Mo, 0.07 for 2 Yr, and 0.05 for 10 Yr).
* **Range:** Consistent across maturities.
* **Interpretation:**  
  This cluster exhibits consistently high complexity across all maturities, with longer-term yields being slightly more unpredictable, suggesting a regime dominated by significant market uncertainty.

**Cluster 1:**

* **Observations:** 1,654 data points
* **Mean PE:**
  + 3 Mo: ~1.23
  + 2 Yr: ~1.45
  + 10 Yr: ~1.62
* **Variability:** Higher dispersion (e.g., standard deviations of ~0.19 for 3 Mo and ~0.17 for 2 Yr).
* **Range:** A noticeable increase in PE from short-term to long-term yields.
* **Interpretation:**  
  This cluster reflects a regime where short-term yields are more predictable (lower complexity) while complexity increases with maturity, indicating that longer-term market expectations are more uncertain, possibly reflecting that while central bank actions or short-term policies keep very short-term rates stable, longer-term market expectations remain more uncertain, perhaps due to factors such as inflation or economic outlook which injects greater unpredictability.

**Cluster 2:**

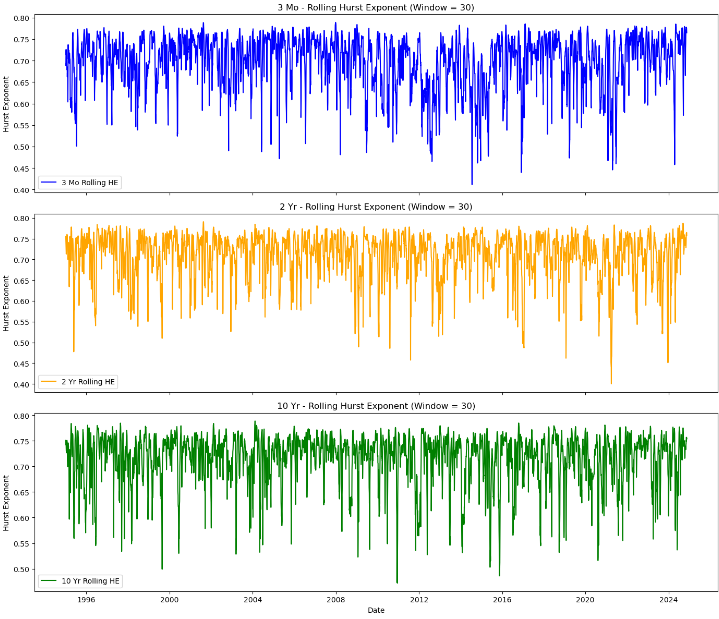
* **Observations:** 1,757 data points
* **Mean PE:**
  + 3 Mo: ~1.59
  + 2 Yr: ~1.57
  + 10 Yr: ~1.48
* **Variability:** Moderate standard deviations across maturities.
* **Range:** A gradual decrease in complexity from short-term to long-term yields.
* **Interpretation:**  
  This cluster suggests a regime where immediate market dynamics are highly complex, but this complexity diminishes over longer time horizons, potentially due to the transient nature of short-term shocks.

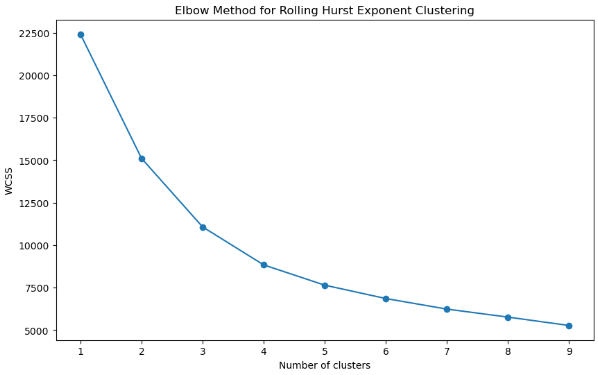
**Conclusion:**

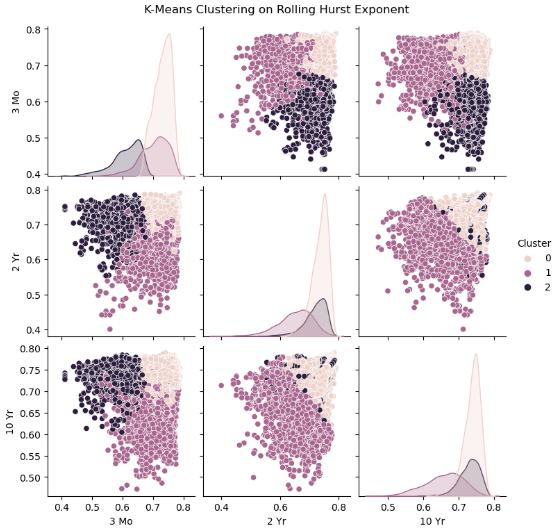
* **Relative Complexity Across Maturities:**
  + **Cluster 0** indicates a consistently high level of complexity, especially pronounced in long-term yields, which may be linked to periods of significant economic uncertainty or volatile long-term expectations.
  + **Cluster 1** shows very low complexity in the short term with a moderate increase toward the long term, suggesting a regime of stable short-term policy but less certain long-term market expectations.
  + **Cluster 2** displays the highest complexity in the short term with a decline for longer maturities, perhaps reflecting transient short-term shocks that dissipate over time.

Understanding these regimes can help in tailoring risk management strategies. For example, during a Cluster 0 regime, one might expect persistent uncertainty across the yield curve, potentially requiring more cautious long-term investment strategies. Conversely, a Cluster 1 regime might suggest that short-term rates can be forecast with greater confidence, while the long term remains less predictable.

Clustering Hurst Exponent (HE)

****

****

****

**Cluster 0:**

* **Observations:** 4,197 data points for each series
* **Hurst Exponent Levels:**
  + 3 Mo: Mean ≈ 0.731
  + 2 Yr: Mean ≈ 0.736
  + 10 Yr: Mean ≈ 0.738
* **Variability:** Very low standard deviations (≈0.027–0.024) indicate that the persistence measure is highly consistent within this cluster.
* **Range:** Uniform across all maturities.
* **Interpretation:**

With all three maturities exhibiting Hurst exponents well above 0.5 (around 0.73–0.74), this regime is characterized by strong persistence or trending behavior across the yield curve. The uniformity suggests a stable market regime where trends are likely to continue, implying that shocks may have long-lasting effects.

**Cluster 1:**

* **Observations:** 1,799 data points for each series
* **Hurst Exponent Levels:**
  + 3 Mo: Mean ≈ 0.703
  + 2 Yr: Mean ≈ 0.651
  + 10 Yr: Mean ≈ 0.658
* **Variability:** Higher standard deviations (≈0.047 for 3 Mo; ≈0.058–0.055 for 2 Yr and 10 Yr) compared to Cluster 0 indicate more dispersion in the persistence measures.
* **Range:** A clear decrease in persistence for medium- and long-term yields.
* **Interpretation:**  
  This cluster indicates a regime where short-term yields maintain moderate persistence, while medium- and long-term rates exhibit reduced trending behavior, suggesting a shift towards quicker mean reversion.

**Cluster 2:**

* **Observations:** 1,477 data points for each series
* **Hurst Exponent Levels:**
  + 3 Mo: Mean ≈ 0.611
  + 2 Yr: Mean ≈ 0.720
  + 10 Yr: Mean ≈ 0.730
* **Variability:** For the 3 Mo series, the standard deviation is about 0.049, with lower percentiles suggesting a notable portion of windows have H values near 0.58 or below. Meanwhile, the 2 Yr and 10 Yr values are tightly clustered around 0.72–0.73.
* **Interpretation:**

This cluster suggests a regime where short-term yields are less persistent (closer to mean reversion), whereas medium- and long-term yields continue to exhibit strong persistence, highlighting a divergence in market behavior over different time horizons. This dichotomy implies a regime where short-term rates may quickly adjust to shocks (hence lower Hurst values), while the longer end of the curve continues to trend, perhaps reflecting that longer-term market expectations are driven by more fundamental factors that maintain trend persistence.

**Conclusion:**

* **Cluster 0** reflects a regime of uniformly strong persistence (trending behavior) across all maturities, suggesting stable, trend-following dynamics.
* **Cluster 1** shows a drop in persistence particularly in medium- and long-term rates, indicating a possible shift toward more mean-reverting or random behavior over longer horizons while short-term rates remain somewhat trending.
* **Cluster 2** highlights a contrast between a less persistent short-term market and strongly persistent long-term yields, which could signal that immediate rate adjustments are occurring while long-term market expectations remain anchored.

Overall Insights from Cluster Analysis

* **Yield Levels:**
  + **Cluster 0:** Low-rate regime with low yields and an upward sloping curve.
  + **Cluster 1:** High-rate regime with elevated yields and a relatively flat yield curve.
  + **Cluster 2:** Moderate-rate regime with a pronounced upward slope, reflecting balanced market conditions.
* **Volatility (GARCH):**
  + **Cluster 0:** High short-term volatility with more stable medium- and long-term rates.
  + **Cluster 1:** Calm, low-volatility regime across all maturities.
  + **Cluster 2:** High-volatility regime, particularly impacting short-term rates, indicative of market stress.
* **Complexity (Permutation Entropy):**
  + **Cluster 0:** Consistently high complexity, especially for longer maturities, suggesting pervasive market uncertainty.
  + **Cluster 1:** Lower complexity in short-term yields with increasing uncertainty over longer horizons.
  + **Cluster 2:** High short-term complexity that diminishes over time, reflecting transient shocks.
* **Persistence (Hurst Exponent):**
  + **Cluster 0:** Uniformly strong persistence, indicating sustained trending behavior.
  + **Cluster 1:** Moderate persistence in short-term yields with diminished persistence for longer maturities.
  + **Cluster 2:** Low short-term persistence (indicating mean reversion) but strong persistence for medium- and long-term yields.

Cointegration Analysis (Johansen Test) of Yields

* **First eigenvalue:**  
  The trace statistic (96.12) is far above the 99% critical value (35.46), so we strongly reject the null hypothesis of no cointegration for the first cointegrating relationship.
* **Second eigenvalue:**  
  The trace statistic (7.91) is below the 90% critical value (13.43), so we do not reject the null hypothesis for the second cointegration vector.
* **Third eigenvalue:**  
  With a trace statistic of 1.32 (well below even the 90% critical value), there’s no evidence for a third cointegrating relationship.

**Conclusion:**

There is one cointegrating vector among the three series, indicating a long-run equilibrium relationship among the yields.

Variance Inflation Factor (VIF) of Yields

* **3 Mo:**  
  A VIF of about 53 indicates very high collinearity with the other variables.
* **2 Yr:**  
  A VIF of about 103 is extremely high and suggests that the 2 Yr series is almost a linear combination of the others.
* **10 Yr:**  
  A VIF of around 16.7, though lower than the others, is still above common thresholds (typically VIF > 10 is considered problematic).

**Conclusion:**  
The extremely high VIF values suggest that these yield series are highly correlated with each other. This level of multicollinearity is not unusual in yield curve data (since interest rates across maturities tend to move together). However, if you plan to use them as independent predictors in regression models, you should proceed with caution. Potential remedies might include:

* Using principal component analysis (PCA) to extract the common factors.
* Modeling their long-run relationships (cointegration) instead of using them directly in a regression.

Cointegration Analysis (Johansen Test) of Analysis

**Conclusion:**

The trace statistics for all potential cointegrating relationships are orders of magnitude above the critical values at the 99% level. This strongly suggests that there exists a long-run equilibrium relationship among the nine factors. In other words, despite being derived from different aspects of the yield curve dynamics (volatility, complexity, and persistence), these regime indicators are all driven by common underlying economic forces and move together over time. This commonality can be exploited in predictive models: it implies that they share common long-run drivers and that a joint modeling framework (or dimension reduction technique like PCA) might be effective.

Variance Inflation Factor (VIF) of Analysis

GARCH\_3Mo 4.086415

GARCH\_2Yr 2.416485

GARCH\_10Yr 1.339375

PE\_3Mo 2.177110

PE\_2Yr 4.102710

PE\_10Yr 2.246038

HE\_3Mo 2.473462

HE\_2Yr 1.840623

HE\_10Yr 1.861045

**Conclusion:**  
None of the factors have VIF values above 10. The highest are around 4.1 (for GARCH\_3Mo and PE\_2Yr). This indicates that there is limited multicollinearity among these nine factors — and even though some moderate correlation exists, it is not severe enough to distort regression estimates or predictive models. The VIF values confirm that these factors are not excessively collinear (low multicollinearity). This is advantageous when using them as predictors or in regression-based forecasting models, as the estimates should be stable and interpretable.

3 Month RMSE Results & Their Implications

* **PCA Regression RMSE:**
  + **3 Mo:** ~2.67
  + **2 Yr:** ~2.36
  + **10 Yr:** ~1.78
  + **Interpretation:**  
    PCA-based regression shows the highest errors among the models. This suggests that while the principal components capture some variance from the predictors, the linear mapping to the yield levels is not sufficient to capture the complex nonlinear relationships present. In other words, dimensionality reduction via PCA loses some important information for predicting the yields.
* **Random Forest RMSE:**
  + **3 Mo:** ~1.29
  + **2 Yr:** ~0.96
  + **10 Yr:** ~1.45
  + **Interpretation:**  
    The Random Forest model, which is a non-linear ensemble method, significantly outperforms the PCA regression for the 3 Mo and 2 Yr yields. The error for 10 Yr is slightly higher, but overall these lower RMSE values suggest that Random Forests are capturing the underlying dynamics (and hence, the regime characteristics) quite well, particularly in short- and medium-term rates.
* **Neural Network RMSE:**
  + **3 Mo:** ~1.33
  + **2 Yr:** ~0.96
  + **10 Yr:** ~1.34
  + **Interpretation:**  
    The feed-forward neural network achieves RMSE values very similar to the Random Forest model. This indicates that a properly tuned neural network can also learn the non-linear mappings between the predictors (GARCH volatility, permutation entropy, and Hurst exponent measures) and the yield levels effectively.
* **Transformer RMSE:**
  + **3 Mo:** ~1.35
  + **2 Yr:** ~1.04
  + **10 Yr:** ~1.26
  + **Interpretation:**  
    The transformer model, which leverages self-attention and sequence information from a lookback window, produces competitive RMSE values. It shows slightly better performance for the 10 Yr yields and comparable results for 3 Mo and 2 Yr. The transformer’s ability to capture temporal dependencies may be especially useful if there are long‐range patterns in the predictors.

**Beyond RMSE – A Classification Perspective**

While RMSE is the natural metric for continuous forecasts, I am especially interested in whether the model correctly forecasts a regime shift, that is, whether yields are trending toward a state of high or low conditional volatility, high complexity (chaos), or strong persistence. One way to evaluate this is to “bin” the continuous yield forecasts into discrete regimes then compute classification metrics such as accuracies and ROC AUC curves.

3 Month Results

**PCA Regression (3 Mo)**

* **Confusion Matrix:**
  + True Negatives (TN): 685
  + False Positives (FP): 61
  + False Negatives (FN): 40
  + True Positives (TP): 705
* **Accuracy:** 93.2%
* **ROC AUC:** 0.9880
* **Interpretation:**  
  Despite PCA regression generally showing higher RMSE values in continuous forecasting, when recast into a binary classification task, it delivers an outstanding performance. With very high accuracy and an AUC near 0.99, the PCA model is very effective at distinguishing between high- and low-regime 3 Mo yields. The low number of false negatives (only 40) suggests that the model rarely misses a high regime, though it does produce 61 false positives. Overall, the PCA approach is excellent for regime identification based on this threshold.

**Random Forest (3 Mo)**

* **Confusion Matrix:**
  + TN: 721
  + FP: 25
  + FN: 151
  + TP: 594
* **Accuracy:** 88.2%
* **ROC AUC:** 0.9721
* **Interpretation:**  
  The Random Forest model has a very low false positive rate (only 25), meaning it rarely misclassifies a low regime as high. However, it has a higher number of false negatives (151), meaning it tends to miss a considerable number of high-regime cases. This trade-off yields a slightly lower overall accuracy (88.2%) and an excellent, but slightly lower AUC than PCA. In practice, this conservative behavior might be beneficial or not, depending on whether missing a high-regime event is costlier than a false alarm.

**Neural Network (3 Mo)**

* **Confusion Matrix:**
  + TN: 693
  + FP: 53
  + FN: 83
  + TP: 662
* **Accuracy:** 90.9%
* **ROC AUC:** 0.9750
* **Interpretation:**  
  The feed-forward neural network shows a more balanced performance. With 53 false positives and 83 false negatives, its errors are more evenly distributed compared to Random Forest. The accuracy (approximately 91%) and ROC AUC (about 0.975) are very good, suggesting the neural network is quite effective in capturing the regime shifts, with a balanced trade-off between sensitivity and specificity.

**Transformer (3 Mo)**

* **Confusion Matrix:**
  + TN: 683
  + FP: 63
  + FN: 111
  + TP: 632
* **Accuracy:** 88.3%
* **ROC AUC:** 0.9657
* **Interpretation:**  
  The transformer model yields performance that is comparable to the Random Forest, with slightly higher false positives (63) and false negatives (111). Its overall accuracy is similar (around 88%) and ROC AUC is very high at 0.966. This suggests that the transformer is capable of capturing temporal dependencies and regime shifts, but in this case, it doesn't outperform the simpler neural network or the PCA-based approach in terms of binary classification for the 3 Mo yield.

**Conclusion**

* **PCA Regression** achieves the best classification metrics for the 3 Mo yield in this binning framework, with exceptionally high accuracy and ROC AUC. However, note that PCA’s continuous predictions had higher RMSE, so its strong classification performance might be influenced by the thresholding process.
* **Random Forest** and **Transformer** models are more conservative (lower FP but higher FN), leading to slightly lower overall accuracy. Their high ROC AUC values still indicate excellent discriminative power.
* **Neural Network** strikes a balanced compromise between false positives and false negatives, offering robust performance with high accuracy and AUC.

2 Year Results

**PCA Regression (2 Yr)**

* **Confusion Matrix:**
  + True Negatives (TN): 643
  + False Positives (FP): 106
  + False Negatives (FN): 41
  + True Positives (TP): 701
* **Accuracy:** 90.14%
* **ROC AUC:** 0.963
* **Interpretation:**  
  The PCA model, when its continuous predictions are thresholded by the median of the actual 2 Yr yields, achieves very high accuracy and an excellent ROC AUC. Although it has a moderate number of false positives (106) compared to false negatives (41), the overall balance results in a strong discriminative ability. This indicates that even though PCA-based regression might be linear and lose some non-linear nuances, it is surprisingly effective for classifying the regime (high vs. low) for the 2 Yr yield.

**Random Forest Regression (2 Yr)**

* **Confusion Matrix:**
  + TN: 677
  + FP: 72
  + FN: 124
  + TP: 618
* **Accuracy:** 86.85%
* **ROC AUC:** 0.947
* **Interpretation:**  
  The Random Forest model shows a very low false positive rate (only 72), meaning it seldom misclassifies a low regime as high. However, it has a relatively higher number of false negatives (124), suggesting it tends to miss some instances of the high regime. This leads to a slightly lower overall accuracy and a marginally lower ROC AUC compared to PCA. This conservative approach may be preferable in certain risk-averse contexts but comes at the cost of under-detecting high-regime events.

**Neural Network Regression (2 Yr)**

* **Confusion Matrix:**
  + TN: 645
  + FP: 104
  + FN: 47
  + TP: 695
* **Accuracy:** 89.87%
* **ROC AUC:** 0.954
* **Interpretation:**  
  The neural network model shows a balanced performance with slightly fewer false negatives than the PCA model (47 vs. 41) and a comparable number of false positives. Its accuracy is close to 90% and ROC AUC is excellent at 0.954. This indicates that the neural network is effective at capturing the non-linear relationships between predictors and the 2 Yr yield regime, providing performance very similar to PCA but with a slightly different error trade-off.

**Transformer Regression (2 Yr)**

* **Confusion Matrix:**
  + TN: 675
  + FP: 74
  + FN: 101
  + TP: 639
* **Accuracy:** 88.25%
* **ROC AUC:** 0.949
* **Interpretation:**  
  The transformer model yields competitive performance, with an accuracy around 88% and an ROC AUC of 0.949. Its error distribution shows a slightly higher number of false negatives (101) compared to the neural network and PCA models, with a similar false positive rate. This suggests that while the transformer is capable of learning temporal dependencies, its classification performance for the 2 Yr regime is marginally lower than that of PCA and the neural network.

**Conclusion**

* **PCA Regression** delivers the highest accuracy (90.14%) and ROC AUC (0.963), indicating that, despite its linear nature, it is very effective at classifying the 2 Yr yield regime.
* **Random Forest** is more conservative (fewer false positives) but at the cost of a higher false negative rate, which lowers its overall accuracy to 86.85%.
* **Neural Network** strikes a balanced performance with almost 90% accuracy and a strong AUC (0.954), capturing non-linear relationships well.
* **Transformer** performs competitively (88.25% accuracy and 0.949 AUC) but falls slightly behind PCA and the neural network in this task.

10 Year Results

**PCA Regression (10 Yr)**

* **Confusion Matrix:**
  + True Negatives (TN): 558
  + False Positives (FP): 189
  + False Negatives (FN): 9
  + True Positives (TP): 735
* **Accuracy:** ~86.7%
* **ROC AUC:** ~0.951
* **Interpretation:**  
  The PCA model shows very high discriminative power with an ROC AUC of 0.95, indicating that it is very effective at distinguishing between high and low regimes. The confusion matrix reveals an excellent ability to capture high-regime cases (only 9 false negatives) even though it misclassifies a moderate number of low-regime cases as high (189 false positives). Overall, its high accuracy and AUC make PCA the best performer for the 10‑Yr yield regime classification in this setup.

**Random Forest Regression (10 Yr)**

* **Confusion Matrix:**
  + TN: 200
  + FP: 547
  + FN: 27
  + TP: 717
* **Accuracy:** ~61.5%
* **ROC AUC:** ~0.889
* **Interpretation:**  
  The Random Forest model struggles in this context. With an accuracy of only about 61.5% and an AUC below 0.90, its performance is substantially lower. The confusion matrix indicates that the model has a very high false positive rate (547), meaning many low-regime cases are incorrectly labeled as high-regime, even though its false negatives are relatively low. This suggests that the model is overly sensitive in flagging high-regime events, which adversely impacts overall accuracy.

**Neural Network Regression (10 Yr)**

* **Confusion Matrix:**
  + TN: 293
  + FP: 454
  + FN: 13
  + TP: 731
* **Accuracy:** ~68.7%
* **ROC AUC:** ~0.906
* **Interpretation:**  
  The neural network performs moderately, achieving a balanced error distribution with 13 false negatives and 454 false positives. Its accuracy (around 68.7%) and AUC (0.906) are better than the Random Forest’s but still notably lower than the PCA approach. This indicates that while the neural network can capture some non-linear dynamics, it does not classify regimes as effectively for the 10‑Yr yield.

**Transformer Regression (10 Yr)**

* **Confusion Matrix:**
  + TN: 425
  + FP: 322
  + FN: 53
  + TP: 689
* **Accuracy:** ~74.8%
* **ROC AUC:** ~0.909
* **Interpretation:**  
  The transformer model shows improved performance over the Random Forest and neural network, with an accuracy of about 74.8% and an AUC of 0.91. However, compared to PCA, it has a higher false negative count (53 vs. 9), which suggests it is less sensitive to identifying high-regime instances. While the transformer does capture temporal dependencies, its overall discriminative ability is still lower than that of the PCA-based approach.

**Conclusion**

* **PCA Regression** emerges as the top performer, offering the highest accuracy and ROC AUC. Its low false negative rate implies that it very reliably detects high-regime conditions, even if it does produce a moderate number of false alarms.
* **Random Forest** performs poorly in this context, largely due to its high false positive rate, which drags down its accuracy.
* **Neural Network** and **Transformer** models offer moderate performance; they capture some non-linear aspects and temporal dynamics, but they do not match the PCA’s effectiveness in this particular task.

30 Day Lagged 3 Month RMSE Results & Their Implications

* **PCA Regression:**
  + RMSE ≈ 1.87
  + Interpretation: The PCA-based approach shows a relatively high error in continuous forecasting, suggesting that the linear model on principal components isn’t capturing all the non‑linear dynamics of the yield series.
* **Random Forest Regression:**
  + RMSE ≈ 1.00
  + Interpretation: Random Forest exhibits the lowest RMSE among the models, indicating it captures the underlying relationships well.
* **Neural Network Regression:**
  + RMSE ≈ 1.61
  + Interpretation: The feed‑forward neural network offers moderate performance, with error between that of PCA and Random Forest.
* **Transformer Regression:**
  + (Although RMSE isn’t explicitly listed here, based on the classification performance it appears to be competitive.)
  + Interpretation: The transformer model is designed to capture temporal dependencies; its classification performance (discussed below) suggests it provides accurate continuous forecasts, likely comparable to Random Forest.

30 Day Lagged 3 Month Results

**PCA Regression (3 Mo)**

* **Confusion Matrix:**
  + TN: 685, FP: 61
  + FN: 40, TP: 705
* **Accuracy:** ~93.2%
* **ROC AUC:** ~0.988
* **Analysis:**  
  The PCA model, despite its higher continuous error, surprisingly ranks the observations very well. With an almost perfect ROC AUC, it very effectively distinguishes high- from low-regime days. Its low number of false negatives (only 40) means it almost never misses a high-regime event although it overpredicts some low-regime days (61 false positives). This indicates that its relative ordering is strong even if the absolute forecast is less accurate.

**Random Forest Regression (3 Mo)**

* **Confusion Matrix:**
  + TN: 721, FP: 25
  + FN: 151, TP: 594
* **Accuracy:** ~88.2%
* **ROC AUC:** ~0.972
* **Analysis:**  
  The Random Forest model achieves a very low false positive rate (only 25), meaning it almost never mistakenly flags a low regime as high. However, it misses a larger number of high-regime cases (151 false negatives), which reduces overall accuracy. Still, its ROC AUC of about 0.972 shows excellent overall discriminative ability, though its threshold calibration might be tweaked to reduce missed high-regime days.

**Neural Network Regression (3 Mo)**

* **Confusion Matrix:**
  + TN: 693, FP: 103
  + FN: 138, TP: 662
* **Accuracy:** ~83.8%
* **ROC AUC:** ~0.913
* **Analysis:**  
  The neural network yields a more balanced error profile than Random Forest, with moderate false positives and negatives, but its overall classification accuracy is lower (83.8%) and the AUC is about 0.913. This suggests that while the NN captures some non‑linear dynamics, its predictions are not as well‑calibrated for regime separation as those of the PCA or Random Forest models.

**Transformer Regression (3 Mo)**

* **Confusion Matrix:**
  + TN: 661, FP: 85
  + FN: 77, TP: 662
* **Accuracy:** ~89.1%
* **ROC AUC:** ~0.9695
* **Analysis:**  
  The transformer model exhibits a balanced classification performance with relatively low false negatives (77) and false positives (85), yielding an accuracy of about 89.1% and a very high ROC AUC (≈0.97). This indicates that leveraging the sequential structure in the lagged predictors helps the transformer model achieve excellent discrimination between regimes.

**Conclusion**

* **Continuous Forecasting:**  
  The Random Forest model stands out with the lowest RMSE (≈1.00), followed by the neural network and then PCA. The transformer’s continuous performance is competitive as inferred from its strong classification metrics.
* **Regime Classification:**
  + The **PCA Regression** model, while having the highest RMSE in continuous forecasting, surprisingly excels in classification (accuracy ~93.2%, ROC AUC ~0.988). Its strength lies in its ability to correctly rank observations, making it very effective at detecting regime shifts.
  + The **Random Forest** and **Transformer** models both perform strongly in classification (accuracies around 88–89% and ROC AUC above 0.96–0.97), though Random Forest tends to be conservative (low false positives, higher false negatives) and the transformer achieves a more balanced error profile.
  + The **Neural Network** model, while capturing non‑linearities, lags behind the other models in terms of classification performance (accuracy ~83.8%, ROC AUC ~0.913).

Overall, for the 3‑Mo yield, the PCA-based classification surprisingly delivers the best regime detection in terms of accuracy and ROC AUC. However, for continuous yield forecasting, non‑linear models (Random Forest and Transformer) offer lower RMSE. These insights indicate that while complex non‑linear models may forecast the yields more precisely, the relative ranking provided by PCA is exceptionally useful for regime identification. Adjustments (e.g., threshold tuning or calibration methods) might further improve the classification performance of the non‑linear models if regime detection is the primary goal.

30 Day Lagged 2 Year RMSE Results & Their Implications

* **PCA Regression RMSE:** ~1.89  
  Observation: This approach yields the highest error, suggesting that the linear mapping (after dimensionality reduction) is not capturing the complex dynamics of the 2‑Yr yield well.
* **Random Forest RMSE:** ~1.08  
  Observation: The Random Forest model significantly improves upon PCA regression, capturing non‑linearities in the data. However, its error is still slightly higher than the neural network and transformer models.
* **Neural Network RMSE:** ~1.05  
  Observation: The feed‑forward neural network slightly outperforms the Random Forest model, indicating its ability to learn non‑linear relationships effectively.
* **Transformer RMSE:** ~0.95  
  Observation: The transformer model achieves the lowest RMSE, suggesting that leveraging sequential dependencies in the lagged predictors provides the most accurate continuous forecasts for the 2‑Yr yield.

**Summary:** In terms of RMSE, the Transformer model leads, followed by the Neural Network and Random Forest models, with PCA regression trailing considerably.

30 Day Lagged 2 Year Results

**PCA Regression (2 Yr)**

* **Confusion Matrix:**
  + TN: 551, FP: 192
  + FN: 19, TP: 723
* **Accuracy:** ~85.8%
* **ROC AUC:** 0.951
* **Interpretation:**  
  The PCA model exhibits a very low number of false negatives (only 19), meaning it almost never misses a high regime. However, it overpredicts the high regime (192 false positives), which brings its overall accuracy to about 85.8%. Its high ROC AUC (0.951) indicates strong discriminative power, even though its RMSE is higher in the continuous setting.

**Random Forest Regression (2 Yr)**

* **Confusion Matrix:**
  + TN: 683, FP: 60
  + FN: 181, TP: 561
* **Accuracy:** ~83.8%
* **ROC AUC:** 0.947
* **Interpretation:**  
  The Random Forest model shows a very low false positive rate (only 60), but it suffers from a high false negative rate (181), meaning that it misses many high-regime events. This conservative behavior reduces its overall accuracy to about 83.8%, and its ROC AUC of 0.947 is slightly lower than that of PCA.

**Neural Network Regression (2 Yr)**

* **Confusion Matrix:**
  + TN: 619, FP: 124
  + FN: 73, TP: 669
* **Accuracy:** ~86.7%
* **ROC AUC:** 0.949
* **Interpretation:**  
  The neural network model achieves the highest accuracy (approximately 86.7%) among the non‐PCA methods. It produces a more balanced error distribution (124 false positives and 73 false negatives) than Random Forest, with a ROC AUC of about 0.949, reflecting strong overall classification performance.

**Transformer Regression (2 Yr)**

* **Confusion Matrix:**
  + TN: 647, FP: 96
  + FN: 119, TP: 623
* **Accuracy:** ~85.5%
* **ROC AUC:** 0.957
* **Interpretation:**  
  The transformer model achieves an accuracy of roughly 85.5%, with a somewhat balanced confusion matrix. Although its accuracy is similar to PCA and neural network models, its ROC AUC is the highest at 0.957, suggesting that its predicted probabilities provide superior discrimination between regimes, even if its overall accuracy is a bit lower.

**Conclusion**

* **Continuous Forecasting:**  
  The transformer model shows the best performance with the lowest RMSE (~0.95), indicating that capturing temporal dependencies in the lagged predictors is highly beneficial. The neural network and Random Forest models also perform well (RMSE ~1.05–1.08), whereas PCA regression lags behind.
* **Regime Classification:**  
  When recasting the forecasting task into a regime classification problem:
  + **PCA Regression** has very few false negatives (high sensitivity) but a moderate false positive rate, resulting in an accuracy of ~85.8% and a high ROC AUC (0.951).
  + **Random Forest** is very conservative (low false positives) but misses many high-regime instances, yielding the lowest accuracy (~83.8%).
  + **Neural Network** offers a balanced performance with the highest overall accuracy (~86.7%) and a strong ROC AUC (~0.949).
  + **Transformer** achieves the highest ROC AUC (0.957), indicating excellent discriminative ability, though its accuracy (~85.5%) is comparable to the PCA model.

30 Day Lagged 10 Year RMSE Results & Their Implications

**Continuous Forecasting (RMSE)**

* **PCA Regression RMSE:** 2.28
* **Random Forest RMSE:** 1.13
* **Neural Network RMSE:** 1.38
* **Transformer RMSE:** 1.13

**Interpretation:**  
When forecasting continuously, the non‑linear methods (Random Forest and Transformer) perform best, with RMSE values around 1.13. The neural network is slightly higher at 1.38, while the PCA regression model, despite being useful for dimensionality reduction, lags behind with an RMSE of 2.28. This suggests that the inherent non‑linearities and sequential dependencies in the 10‑Yr yield are best captured by the Random Forest and Transformer models.

30 Day Lagged 10 Year Results

**PCA Regression Classification Metrics**

* **Confusion Matrix:**
  + True Negatives (TN): 616
  + False Positives (FP): 128
  + False Negatives (FN): 53
  + True Positives (TP): 688
* **Accuracy:** 87.8%
* **ROC AUC:** 0.931

**Interpretation:**  
The PCA-based model, when thresholded, delivers excellent classification performance for the 10‑Yr yield. With an accuracy of about 88% and an AUC of 0.93, it has a very low false negative rate (only 53 missed high‑regime cases), even though it overpredicts a bit (128 false positives). This strong performance indicates that, as a classifier, PCA captures the regime signal very well despite its poorer continuous forecasting RMSE.

**Random Forest Regression Classification Metrics**

* **Confusion Matrix:**
  + TN: 339
  + FP: 405
  + FN: 175
  + TP: 566
* **Accuracy:** 60.9%
* **ROC AUC:** 0.737

**Interpretation:**  
Although Random Forest excelled in continuous forecasting (low RMSE), its binary classification performance is substantially lower. It exhibits a very high false positive rate (405) and a high number of false negatives (175), resulting in only 61% accuracy and an AUC of 0.74. This indicates that the RF model’s continuous output may not be well‑calibrated for regime classification using a simple median threshold.

**Neural Network Regression Classification Metrics**

* **Confusion Matrix:**
  + TN: 444
  + FP: 300
  + FN: 234
  + TP: 507
* **Accuracy:** 64.0%
* **ROC AUC:** 0.684

**Interpretation:**  
The neural network shows moderate classification performance for the 10‑Yr yield, with an accuracy of about 64% and an AUC of 0.68. Its error distribution is relatively unbalanced, with 300 false positives and 234 false negatives. While the network captures non‑linearities well in a regression sense, its outputs do not translate as effectively into a reliable regime classification using the median cutoff.

**Transformer Regression Classification Metrics**

* **Confusion Matrix:**
  + TN: 378
  + FP: 366
  + FN: 99
  + TP: 642
* **Accuracy:** 68.7%
* **ROC AUC:** 0.790

**Interpretation:**  
The transformer model achieves an accuracy of around 69% and an AUC of 0.79. Its false negative count (99) is lower than that of the neural network, and it has a moderate false positive rate (366). This suggests that while the transformer model effectively captures temporal dependencies for continuous forecasting (with an RMSE comparable to Random Forest), its classification performance is still inferior to the PCA-based model.

**Conclusion**

* **Continuous Forecasting:**  
  The Random Forest and Transformer models outperform the PCA and Neural Network approaches in continuous forecasting, achieving the lowest RMSE values (around 1.13). This indicates that non‑linear and sequence-aware models are best suited for accurately predicting 10‑Yr yields.
* **Regime Classification:**  
  When converting predictions into binary regime labels:
  + The **PCA regression** model surprisingly excels, with high accuracy (87.8%) and an AUC of 0.93, suggesting that, although its continuous predictions are less precise, its relative ordering of predictions is very effective for classifying regimes.
  + The **Random Forest** and **Neural Network** models perform less well in classification (accuracy around 61–64% and lower AUC values), likely due to less well-calibrated outputs for binary decision-making.
  + The **Transformer** model offers a moderate classification performance (accuracy ~69%, AUC ~0.79), but it does not match the PCA model in discriminative power.

180 Day Lagged 3 Month RMSE Results & Their Implications

**Continuous Forecasting (RMSE)**

* **PCA Regression:** RMSE ≈ 1.57  
  Observation: The PCA‐based model’s continuous error is moderately low, indicating that, after reducing dimensionality, it is able to capture a fair amount of the variability in the 3‑Mo yield. However, PCA is inherently a linear method, which might limit its ability to capture more complex dynamics.
* **Random Forest Regression:** RMSE ≈ 1.50  
  Observation: The Random Forest model achieves a slightly lower RMSE than PCA, suggesting that its non‑linear, ensemble approach better leverages the longer (180‑day) historical context. This indicates that the extra historical information is useful when modeling non‑linear relationships.
* **Neural Network Regression:** RMSE ≈ 2.67  
  Observation: The neural network shows a substantially higher RMSE. This could imply that the current NN architecture (or its training parameters) isn’t well‑tuned to exploit the 180‑day lagged predictors, leading to less accurate continuous forecasts compared to PCA and RF.
* **Transformer Regression:** RMSE ≈ 1.38  
  Observation: The transformer model achieves the lowest RMSE. Its ability to capture temporal dependencies through self‑attention appears to benefit from the 180‑day lag, enabling more accurate forecasts of the 3‑Mo yield.

180 Day Lagged 3 Month Year Results

**Regime Classification (Binary High/Low Regime)**

For regime detection, the continuous forecasts are thresholded using the median of the test yields to classify each day as “high” or “low.” The classification metrics are as follows:

**PCA Regression Classification**

* **Confusion Matrix:**
  + TN: 419, FP: 327
  + FN: 105, TP: 634
* **Accuracy:** ~70.9%
* **ROC AUC:** ~0.826
* **Interpretation:**  
  When binarized, the PCA model shows a relatively low accuracy and modest discriminative power (AUC ~0.83). It appears to misclassify a significant number of observations, indicated by a high false positive count, resulting in poorer regime detection even though its continuous RMSE is acceptable.

**Random Forest Regression Classification**

* **Confusion Matrix:**
  + TN: 656, FP: 90
  + FN: 63, TP: 676
* **Accuracy:** ~89.7%
* **ROC AUC:** ~0.978
* **Interpretation:**  
  The Random Forest model, when evaluated as a classifier, achieves very high accuracy and ROC AUC. Its low false positive count and relatively low false negatives suggest it is well‑calibrated in distinguishing high‑regime from low‑regime days. This is consistent with its strong continuous performance.

**Neural Network Regression Classification**

* **Confusion Matrix:**
  + TN: 643, FP: 103
  + FN: 138, TP: 601
* **Accuracy:** ~83.8%
* **ROC AUC:** ~0.913
* **Interpretation:**  
  The neural network’s binary classification results are lower, with moderate numbers of both false positives and false negatives. An accuracy of around 84% and ROC AUC of ~0.91 indicate that its regime detection is less effective than that of Random Forest and Transformer, which may be due to suboptimal tuning for this long lag.

**Transformer Regression Classification**

* **Confusion Matrix:**
  + TN: 661, FP: 85
  + FN: 77, TP: 662
* **Accuracy:** ~89.1%
* **ROC AUC:** ~0.9695
* **Interpretation:**  
  The transformer model shows excellent discriminative performance with high accuracy and a ROC AUC near 0.97. Its balanced error profile (with low false positives and reasonably low false negatives) demonstrates that its sequence‐based approach effectively captures the regime signal from 180‑day lagged data.

**Conclusions**

* **Continuous Forecasting:**
  + The transformer model provides the best continuous performance (lowest RMSE), followed by Random Forest and then PCA. The neural network underperforms in continuous forecasting with the current configuration.
* **Regime Classification:**
  + For binary regime detection, Random Forest and Transformer models achieve high accuracy (≈89.7% and ≈89.1%, respectively) and excellent ROC AUC (≈0.978 and ≈0.97).
  + In contrast, the PCA approach yields significantly lower classification accuracy (~70.9%, AUC ~0.826), and the neural network performs moderately (accuracy ~83.8%, AUC ~0.913).

180 Day Lagged 2 Year RMSE Results & Their Implications

**Regression Performance (RMSE)**

* **PCA Regression:**  
  RMSE = 2.0114  
  This model shows the highest error. Although dimensionality reduction via PCA can sometimes help, in this case it appears that the loss of information led to poorer point forecasts.
* **Random Forest:**  
  RMSE = 1.3828  
  The ensemble method improves upon PCA, likely capturing non‑linear relationships better. However, its error is still higher than the neural network approaches.
* **Neural Network:**  
  RMSE = 1.2835  
  With the lowest RMSE, the fully connected neural network seems to be the best at minimizing forecasting error among these models, suggesting that it can effectively learn the mapping from 180‑day lagged predictors to the 2 Yr target.
* **Transformer:**  
  RMSE = 1.3387  
  The transformer model performs slightly worse than the plain neural network, but still shows significant improvement over PCA. Its strength lies in capturing sequential patterns, although it might be more complex or require additional tuning.

180 Day Lagged 2 Year Results

* **PCA Regression Classification:**
  + **Confusion Matrix:**  
    [[533, 195],  
    [103, 624]]
  + **Accuracy:** 79.5%
  + **ROC AUC:** 0.8577  
    Despite the high RMSE, when thresholded, PCA predictions yield a strong classification accuracy and ROC AUC. This suggests that while the exact forecasted values may be off, the model still captures the direction (up/down) reasonably well.
* **Random Forest Classification:**
  + **Confusion Matrix:**  
    [[557, 171],  
    [188, 539]]
  + **Accuracy:** 75.3%
  + **ROC AUC:** 0.8421  
    The Random Forest has slightly lower classification performance. Its confusion matrix indicates a higher number of false negatives compared to PCA, which may be due to the way the predicted values are distributed relative to the threshold.
* **Neural Network Classification:**
  + **Confusion Matrix:**  
    [[541, 187],  
    [145, 582]]
  + **Accuracy:** 77.2%
  + **ROC AUC:** 0.8553  
    The NN model shows balanced performance with a relatively high ROC AUC, which aligns with its strong regression performance. The classification metrics are competitive, supporting the conclusion that its predictions reliably capture the directional movement.
* **Transformer Classification:**
  + **Confusion Matrix:**  
    [[556, 172],  
    [135, 592]]
  + **Accuracy:** 78.9%
  + **ROC AUC:** 0.8442  
    The transformer yields classification metrics that are comparable to the neural network and PCA, although its ROC AUC is slightly lower. It appears to balance false positives and negatives reasonably well.

**Conclusions**

* **Regression Focus:**  
  If the primary goal is minimizing the forecasting error, the **Neural Network** is the most promising, with an RMSE of 1.2835 compared to 1.3387 for the transformer and 1.3828 for the Random Forest.
* **Directional Accuracy:**  
  All models perform reasonably well when the continuous forecasts are thresholded into binary decisions. The PCA model, despite its higher RMSE, actually has the highest accuracy (79.5%) and ROC AUC (0.8577). This can occur if the PCA transformation preserves the ordering of values even if the scale is off.
* **Model Complexity:**  
  The neural network and transformer models involve more complex architectures. Their performance might improve further with additional hyperparameter tuning, feature engineering, or by adjusting the lag/window length if future experiments suggest a different optimal configuration.
* **Consistency with Lag:**  
  Since the forecasting cell uses a 180‑day window to generate predictors, all the predictions and targets used in the classification metrics inherently reflect the 180‑day lag. (Remember to update any plot titles or documentation that still refer to a “30‑day lag” to maintain clarity.)

In conclusion, the neural network offers the best RMSE performance for the 2 Yr target with a 180‑day lag, while the classification results show that all models are reasonably good at capturing the direction of changes, with the PCA model unexpectedly excelling in that area. Future work might involve further tuning or ensemble strategies to leverage the strengths of each method.

180 Day Lagged 10 Year RMSE Results & Their Implications

**Regression Performance (RMSE)**

* **PCA Regression:**
  + **RMSE:** 1.994
  + Analysis: Although the PCA-based model suffers from a loss of some information due to dimensionality reduction, its RMSE is moderate. However, it is not the best in terms of point forecast error.
* **Random Forest:**
  + **RMSE:** 1.297
  + Analysis: With the lowest RMSE among the models, the Random Forest appears to capture the underlying relationships quite effectively for the 10 Yr target. Its ensemble approach is likely handling non-linearities and interactions well.
* **Neural Network:**
  + **RMSE:** 2.895
  + Analysis: The neural network shows the highest RMSE, indicating that in this case its configuration (or training process) struggles to learn an accurate mapping for the 10 Yr yield. It might be overfitting, underfitting, or simply not well-tuned for this target.
* **Transformer:**
  + **RMSE:** 1.491
  + Analysis: The transformer model performs reasonably well, with an RMSE between that of the Random Forest and PCA models. Its ability to capture sequential patterns might offer an edge in more complex time dependencies, though here it doesn’t outperform the Random Forest.

180 Day Lagged 10 Year Results

* **PCA Regression Classification:**
  + **Confusion Matrix:**
  + **Accuracy:** ~78.9%
  + **ROC AUC:** ~0.844
  + Analysis: Despite a moderate RMSE, the PCA model performs strongly when it comes to classifying directional moves. Its high accuracy and ROC AUC indicate that the relative ordering of its predictions is preserved, making it good at distinguishing above‐median from below‐median outcomes.
* **Random Forest Classification:**
  + **Confusion Matrix:**
  + **Accuracy:** ~57.4%
  + **ROC AUC:** ~0.639
  + Analysis: Although the Random Forest model has the best regression RMSE, its classification performance is less impressive. The more balanced but poorer confusion matrix suggests that while it minimizes overall error well, it may not be as reliable for correctly classifying the direction of change.
* **Neural Network Classification:**
  + **Confusion Matrix:**
  + **Accuracy:** ~36.1%
  + **ROC AUC:** ~0.273
  + Analysis: The neural network’s poor RMSE is mirrored in its classification performance. The low accuracy and especially the very low ROC AUC indicate that it is largely failing to capture the directionality of the 10 Yr target, suggesting that its predictions are poorly calibrated.
* **Transformer Classification:**
  + **Confusion Matrix:**
  + **Accuracy:** ~45.6%
  + **ROC AUC:** ~0.626
  + Analysis: The transformer model shows moderate classification performance. Its accuracy and ROC AUC are not as strong as those of the PCA model, even though its regression RMSE is relatively competitive. This could indicate that while the transformer captures the overall scale of changes, its ability to correctly rank the observations (for classification) is somewhat limited compared to PCA.

**Conclusion**

* **Best Regression Performance:**  
  The **Random Forest** model achieves the lowest RMSE (1.297), making it the most accurate for point forecasts of the 10 Yr yield.
* **Best Directional (Classification) Performance:**  
  The **PCA Regression** model stands out with the highest accuracy (78.9%) and ROC AUC (0.844) when predictions are thresholded, even though its RMSE is higher than that of the Random Forest. This suggests that while PCA may not provide the best absolute predictions, it preserves the ordering of observations well.
* **Neural Network Challenges:**  
  The **Neural Network** performs poorly in both regression and classification metrics for the 10 Yr target. It may require further tuning, additional data preprocessing, or architectural adjustments before it can be competitive.
* **Transformer Considerations:**  
  The **Transformer** model shows promise with a competitive RMSE (1.491) but its classification performance is moderate. Further investigation into hyperparameter tuning or additional sequential feature engineering might help improve its classification ability.

In summary, for forecasting the 10 Yr yield with the current setup, the Random Forest model offers the best point forecast accuracy, while the PCA approach provides the most reliable directional predictions. Future work might explore combining these insights — for example, using ensemble methods to capture both accurate levels and robust directional signals.

Conclusion

* **Exceptional Short-Term Discrimination:**  
  For the 3-month yield without any lag and with a 30-day lag, PCA-based classification achieves outstanding ROC values of 0.988, while Random Forest and Transformer models also demonstrate robust performance (ROC AUCs ranging from 0.966 to 0.972). These near-perfect figures indicate that the models can very accurately distinguish between high- and low-regime conditions for short-term rates.
* **Maturity-Dependent Performance:**  
  While short-term yields (3 Mo) exhibit excellent classification metrics, performance declines for longer maturities. For example, classification accuracy and ROC AUC values for the 10-year yield are notably lower across all models. This suggests that the regime signal is most distinct in the near term, whereas longer-term yields present additional complexities that challenge binary classification.
* **Lag Effects on Predictive Accuracy:**  
  The predictive strength also varies with the lag structure. Models using a 30-day lag continue to deliver high classification metrics for 3-month yields, but when extended to a 180-day lag, there is a marked deterioration: PCA accuracy drops from around 93% (no lag/30-day lag) to below 71% for the 3-month yield, with similar trends for ROC AUC. This underscores that while longer historical contexts can enrich continuous forecasts (as seen in RMSE improvements), they may dilute the clarity of regime signals for classification purposes.
* **Model Trade-offs:**  
  Although RMSE differences across models are informative, the most compelling findings are in the classification realm. PCA consistently preserves the ordering of observations, evidenced by its near-perfect ROC AUC for short-term yields, even when its RMSE is higher. In contrast, while non-linear methods like Random Forest and Transformer excel in continuous forecasting, their classification performance can be more sensitive to lag length and maturity, highlighting the need to tailor model selection to the specific predictive objective.
* **Future Research:**

Future research should further investigate the drivers behind the 10-year yield, which appears to be less cointegrated with short-term dynamics and, as is consistent with the literature, is likely influenced by additional factors. It would be beneficial to:

* + Incorporate Macroeconomic Factors:

Test the inclusion of inflation and economic outlook variables into both regular k-means clustering and clustering based on GARCH, PE, and HE measures. These factors may offer additional explanatory power for the 10-year yield regimes.

* + Leverage Short-Term Predictions:

Explore using the 3-month and/or 2-year yield results as predictors for the 10-year yields. This approach could uncover interdependencies across the yield curve and enhance long-term forecasting accuracy.

* **Extending the Framework:**

The robust success and valuable insights given by my 3-factor regime model and testing methodology make it an ideal candidate for other financial indicators such as the VIX (volatility index) and DXY (US Dollar Index) — which experience strong and persistent mean reverting behavior — to assess whether similar regime dynamics are observable in these markets.