Spring 2024

MF 728 Fixed Income

Project Report

Yield Curve Decomposition and Dynamic Trading Strategy

Group Members:

Aaryan Sanghavi

Mansi Shilamkar

Furkan Karayel

Rijul Gupta

Hamza Zbiri

Introduction

In today's financial landscape, the ability to analyze and predict the behavior of the yield curve is crucial. This curve, which represents the relationship between bond interest rates and their respective maturities while holding credit quality constant, is more than just a financial tool; it is a barometer of economic health and investor sentiment. Typically, a normal upward-sloping yield curve suggests economic expansion, while an inverted curve may signal an impending recession.

The aim of this project was to delve into the dynamics of the yield curve, inspired by Johan Hagenbjörk's seminal work, "Optimization-Based Models for Measuring and Hedging Risks in Fixed Income Markets." Hagenbjörk's study is comprehensive, dissecting the yield curve into six distinct factors and applying sophisticated stochastic hedging processes to manage risks associated with fixed income securities. Our project, however, streamlines this approach by concentrating on three primary factors—Level, Slope, and Curvature—making the analysis more accessible and focused.

Level: This factor represents a uniform shift in the yield curve where all maturities move simultaneously in the same magnitude, indicative of overarching economic shifts like monetary policy changes.
Slope: Reflecting the differential between yields of short-term and long-term bonds, this factor is crucial for understanding the term structure of interest rates and forecasting economic activity.
Curvature: Often overlooked, this factor captures the nuanced inflections in the yield curve, particularly how medium-term rates react differently from short and long-term rates, offering deeper insights into market expectations and liquidity preferences.
By employing these simplified yet powerful components, our project proposes a dynamic trading strategy that leverages predictive analytics to capitalize on yield curve movements. This strategy integrates advanced statistical tools and machine learning models, including Principal Component Analysis (PCA) for dimensionality reduction, Gaussian Copulas for modeling dependencies, and predictive algorithms like Long Short-Term Memory (LSTM) networks and Gradient Boosting Machines (GBM) to forecast future movements.

This endeavor not only enhances our understanding of the structural nuances of the yield curve but also serves as a practical guide for traders and financial analysts. It bridges the gap between theoretical financial models and real-world applications, offering insights that could significantly influence trading decisions and risk management strategies in the volatile realm of fixed income markets.

Literature Review

Based on Johan Hagenbjörk's work outlined in his dissertation, "Optimization-Based Models for Measuring and Hedging Risk in Fixed Income Markets," we can enrich our literature review section significantly. Hagenbjörk's comprehensive approach to modeling fixed-income markets uses optimization to gauge and manage risks effectively. His dissertation emphasizes the importance of accurately measuring financial quantities through inverse optimization problems, highlighting how market prices are often observed with varying levels of noise. The focus on systematic risk factors is particularly pertinent; he identifies six risk factors for the interbank market and additional factors for credit markets, which collectively explain a significant portion of the variance in interest rates and default intensities. By applying stochastic optimization, including considerations for transaction costs, his methodologies improve hedging strategies over traditional approaches. This rigorous analytical framework provides a robust basis for our project's simplified model, where we adapt the concept of decomposing the yield curve into three main components—Level, Slope, and Curvature—enhancing our trading strategy's responsiveness to market dynamics.

Methodology

Data Collection

The foundation of this project's analysis is a comprehensive dataset of daily par rates sourced from the US Treasury's official website. This dataset spans the entire year of 2023 and extends into 2024 up to April 19th. By utilizing such a complete and up-to-date dataset, our analysis is grounded on robust and timely data, providing an accurate reflection of recent market dynamics. This extensive dataset is crucial for the reliability of our yield curve analysis and the effectiveness of our proposed trading strategy.

Principal Component Analysis (PCA)

PCA is a powerful statistical technique used to simplify the complexity in high-dimensional data by reducing its number of dimensions without significant loss of information. In the context of our project, PCA was applied to decompose the yield curve into three principal components: Level, Slope, and Curvature. This decomposition is vital for understanding the underlying movements and traits of the yield curve:

Level: This principal component accounts for 82.4% of the variance and represents a parallel shift across all maturities of the yield curve. It reflects broad, market-wide movements often driven by changes in policy or macroeconomic conditions.
Slope: Capturing 10.66% of the variance, this component measures the steepness between short-term and long-term interest rates. It is crucial for assessing the economic outlook, as a steeper slope generally indicates expectations of rising interest rates and economic expansion.
Curvature: This component explains 4.06% of the variance and is essential for capturing shifts where medium-term rates diverge from the movements of short-term and long-term rates. It often indicates nuanced investor expectations about future rate movements and economic conditions.

Gaussian Copulas

We employed Gaussian Copulas to model the dependencies between the different maturities of the bonds. Gaussian Copulas allow us to understand and simulate the joint distribution of bond yields, which is pivotal for accurately modeling the relationships and dependencies within the yield curve data. This statistical tool is crucial for our simulation processes and enhances the robustness of our trading strategy by providing a deeper insight into the yield dynamics across different maturities.

Trading Strategy

The trading strategy developed leverages the insights derived from PCA and Gaussian Copulas:

Data Windowing: A 30-day rolling window of yield curve data is used for analysis to ensure that our strategy is responsive to recent market conditions.
Forecasting: Within this window, we apply PCA to identify the dominant movement patterns and use Gaussian Copulas to model the interdependencies among different maturities.
Prediction and Execution: A Long-Short Term Memory (LSTM) model predicts the future rates for each maturity on the 31st day. Based on these predictions, trades are executed to buy bonds if their rates are expected to decrease the next day (profiting from rising prices), or sell if rates are expected to increase (profiting from falling prices).

Predictive Modeling

Two advanced predictive models were employed to enhance the accuracy and reliability of our trading signals:

Long-Short Term Memory (LSTM) Model: This model is specifically designed to handle time-series data like yield curves. By utilizing sequences of yield data processed through PCA and Gaussian Copulas, the LSTM model can capture complex temporal patterns and dependencies, essential for predicting future movements in the yield curve.
Gradient Boosting Machine (GBM) Model: This model uses a series of decision trees to predict changes in bond prices for the next day across various maturities. The GBM model is particularly effective due to its ability to handle multiple outputs simultaneously, crucial for detailed forecasting across different segments of the yield curve.
Each model is finely tuned with the Mean Squared Error (MSE) loss function, optimizing them for precision in predicting yield curve dynamics.

Libraries Used

GradientBoostingRegressor (from sklearn.ensemble):
This component of scikit-learn provides a powerful machine learning technique that builds an additive model in a forward stage-wise fashion. It allows for the optimization of arbitrary differentiable loss functions, making it suitable for both regression tasks.

MultiOutputRegressor (from sklearn.multioutput):
This wrapper enables any single-target regression estimators from scikit-learn to be extended to also support multi-target regression. It creates one estimator per target, fitting them separately, which allows using any regression model for multi-output tasks.

mean_squared_error (from sklearn.metrics):
A metric from scikit-learn used to measure the average of the squares of the errors—that is, the average squared difference between the estimated values and the actual value. It's widely used as a measure of the performance of regression models.

StandardScaler (from sklearn.preprocessing):
Part of scikit-learn, this tool standardizes features by removing the mean and scaling to unit variance, which is crucial for many machine learning algorithms to perform well.

PCA (from sklearn.decomposition):
Principal Component Analysis (PCA) is a technique used to emphasize variation and bring out strong patterns in a dataset. It's often used to reduce the dimensions of data while retaining as much information as possible.

Sequential (from tensorflow.keras.models):
This Keras API provides a linear stack of layers to create models. It simplifies the model building process in deep learning by allowing developers to add layers sequentially to build various neural network architectures.

LSTM (from tensorflow.keras.layers):
Long Short-Term Memory (LSTM) layers are a type of recurrent neural network (RNN) suitable for sequence prediction problems. LSTMs are widely used in deep learning because they effectively capture time dependencies in sequence data.

Dense (from tensorflow.keras.layers):
Dense layers are the regular deeply connected neural network layers, which are used extensively in neural networks. Each neuron receives input from all the neurons in the previous layer, thus densely connected.

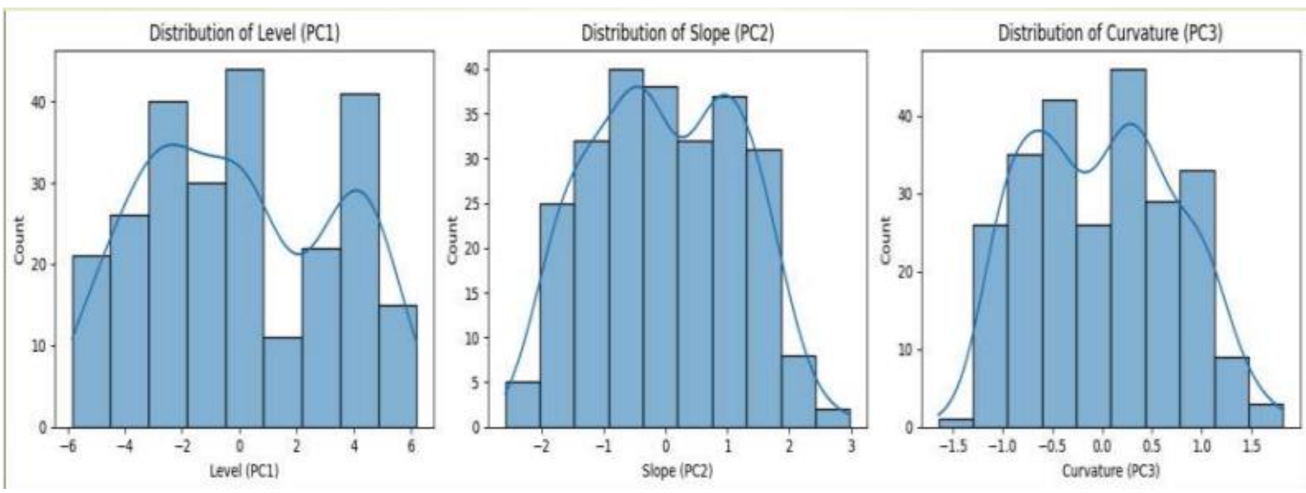
GaussianMultivariate (from copulas.multivariate):
This model is part of the copulas library, which is used for modeling and sampling from multivariate distributions. The Gaussian Multivariate method assumes a Gaussian copula, which can capture the linear correlation between multiple variables efficiently.

Results and Conclusion

Principal Component Analysis of the Yield Curve

In our project, we analyzed the yield curve through Principal Component Analysis (PCA) to understand its dynamics effectively. We focused on three primary factors: Level, Slope, and Curvature, which are instrumental in assessing changes in the yield curve.

Level (PC1): This component accounts for 82.4% of the variance in the yield curve data, as demonstrated in the first histogram. It represents a parallel shift in the yield curve across all maturities, indicating broad economic changes such as policy shifts or macroeconomic trends. The histogram shows a bimodal distribution, suggesting two distinct states in market conditions.
Slope (PC2): Capturing 10.66% of the variance, this component reflects the differential between short-term and long-term bond yields. The histogram exhibits a roughly normal distribution with a slight skew, indicating variability in investor expectations about future interest rates. This factor is crucial for understanding the economic outlook, as changes in the slope can signal shifts in economic growth expectations.
Curvature (PC3): Accounting for 4.06% of the variance, Curvature captures nuances in the yield curve, especially how medium-term rates behave relative to the short and long ends. The histogram shows a slightly skewed bimodal distribution, reflecting specific market conditions where medium-term rates diverge from the expected norm.

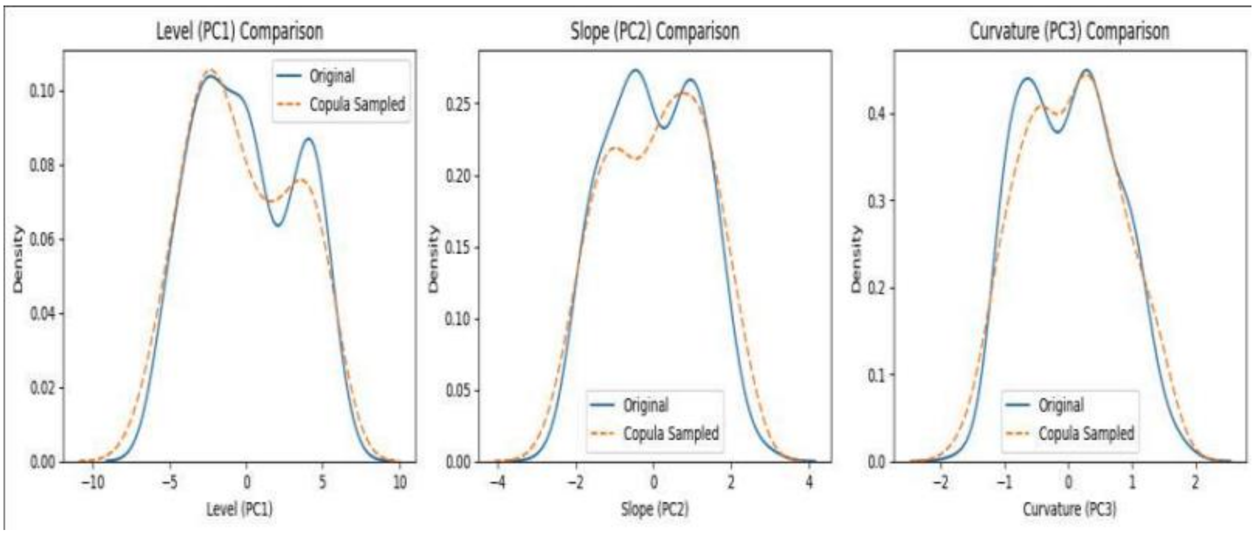


Gaussian Copula Analysis of the Yield Curve

We employed Gaussian Copulas to model dependencies between different maturities of the yield curve, helping to understand and simulate bond yield distributions effectively.

Level (PC1) Comparison: Both the original and Copula-sampled data show a similar bimodal distribution, with the Copula model smoothing the transitions slightly, indicating its effectiveness in capturing essential market conditions.
Slope (PC2) Comparison: The original and Copula-sampled distributions are closely aligned,

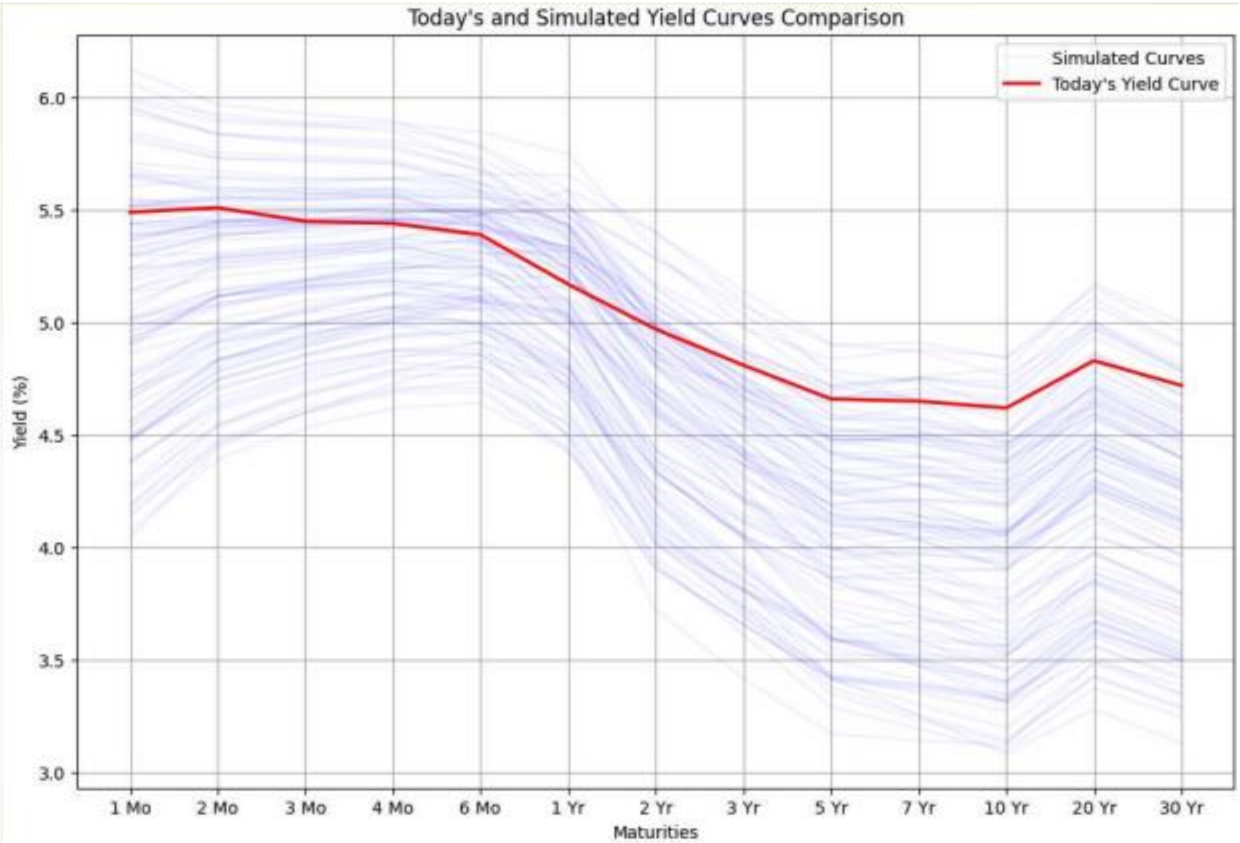
with the Copula model maintaining the shape and asymmetry of the original, confirming its accuracy in reflecting real-world yield curve behaviors.
Curvature (PC3) Comparison: The Copula model reproduces the original data's sharp peak and slight skewness well, demonstrating its capability to capture nuanced investor sentiments regarding medium-term rates.
These comparisons confirm that Gaussian Copulas are effective in accurately modeling yield curve dependencies, ensuring that our predictive models and trading strategies are based on realistic and robust simulations.



Yield Curve Comparison

This graph displays a comparison between today's actual yield curve and multiple yield curves simulated using Gaussian Copulas across various maturities from 1 month to 30 years.

Today's Yield Curve (Red Line): Shows a typical upward-sloping curve which flattens towards longer maturities, indicating normal economic conditions with stable long-term expectations.
Simulated Curves (Blue Lines): Illustrate a range of possible scenarios, providing a visual dispersion around today's curve. These simulations capture the variability in yields across different economic scenarios, reflecting the effectiveness of Gaussian Copulas in generating diverse yield curve outcomes.
This comparison highlights how well the Gaussian Copula method models the full spectrum of yield behaviors, useful for stress testing and scenario analysis in financial planning and risk management.



Sharpe Ratio Analysis for LSTM Model Performance

In evaluating the performance of our LSTM model for trading based on yield curve predictions, we calculated the Sharpe Ratio, which is a measure of risk-adjusted return. A Sharpe Ratio of 0.1175 was obtained for our strategy. This metric is particularly useful as it helps us understand how well the return of an asset compensates for the risk taken.

Key Points from the Analysis:

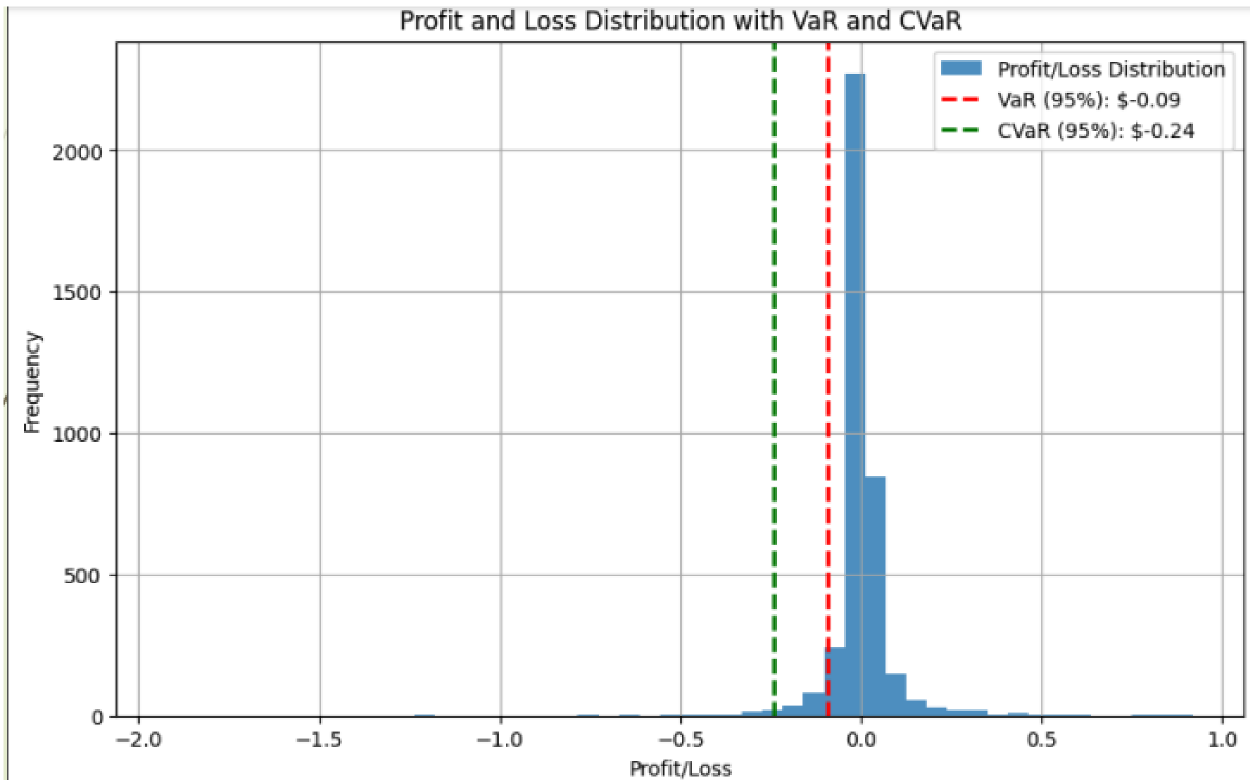
Sharpe Ratio Value: The computed Sharpe Ratio of 0.1175 indicates that the excess return per unit of risk is moderate. While not exceedingly high, this value suggests that the strategy offers a reasonable return for the level of risk involved, especially considering the complexity and variability of bond markets.

Implication: A Sharpe Ratio greater than zero means that the strategy has generated a return above the risk-free rate, which is an encouraging sign. However, it also points to room for improvement, either by enhancing the return or reducing the volatility of the strategy.

Strategy Performance: This Sharpe Ratio reflects the outcomes from our LSTM model's predictions across various bond maturities, showing effectiveness in some areas (notably with medium to long-term bonds) but also potential volatility or underperformance in others (e.g., the negative results in shorter maturities and the one-year maturity).

This analysis underscores the practical utility of our LSTM-based strategy in navigating the fixed income market, providing insights into economic conditions and investor sentiment through yield curve analysis. However, the moderate Sharpe Ratio also highlights the need for continuous refinement of the model to better handle market dynamics and potentially increase the risk-adjusted returns.

```
Total P&L by Maturity:
1 Mo     -0.303551
2 Mo      -0.16941
3 Mo      6.241824
4 Mo      0.789583
6 Mo     -1.422649
1 Yr     -2.099851
2 Yr     -0.114557
3 Yr      0.029741
5 Yr     -0.070687
7 Yr      0.179026
10 Yr     0.256188
20 Yr     0.178165
30 Yr     0.017201
```



Profit & Loss Results from the GBM Model

The Gradient Boosting Machine (GBM) Model has displayed diverse performance across different bond maturities:

Short-Term Maturities (1 to 6 Months): No gains were recorded for the 1 to 4 month maturities, indicating potentially flat or non-responsive market conditions for these terms. A modest gain was seen in the 6-month maturity.
Medium to Long-Term Maturities (1 to 30 Years): More substantial profits were observed in

longer-term maturities, with the highest gains noted at 1 year. This suggests that the GBM model may be more effective in capturing and leveraging trends or patterns at these longer durations. The calculated Sharpe Ratio of 0.6331 for this model indicates a favorable risk-adjusted return, suggesting that while there are profits to be gained, they come with a reasonable level of risk. This performance profile points to the effectiveness of the GBM model in navigating the complexities of fixed-income markets, especially for medium to long-term investments.

Maturity	Total P&L
1 Mo	0.000000
2 Mo	0.000000
3 Mo	0.000000
4 Mo	0.000000
6 Mo	0.500348
1 Yr	1.859743
2 Yr	0.999891
3 Yr	0.313441
5 Yr	0.154188
7 Yr	0.161772
10 Yr	0.092596
20 Yr	0.073056
30 Yr	0.108659

Conclusion

Throughout our project, we employed advanced statistical and machine learning techniques to analyze and predict movements in the yield curve. Utilizing Principal Component Analysis (PCA), Gaussian Copulas, Long Short-Term Memory (LSTM) networks, and Gradient Boosting Machines (GBM), we developed a robust framework for trading strategies in the fixed-income market.

Key Findings:

PCA and Gaussian Copulas proved instrumental in decomposing the yield curve and understanding the dependencies across different maturities. This foundational analysis facilitated more accurate modeling and simulation of yield curve dynamics.

LSTM Model: This model displayed a moderate risk-adjusted performance, as indicated by a Sharpe Ratio of 0.1175. It was particularly effective in predicting yield curve movements but highlighted areas for potential improvement in risk management and return optimization.

GBM Model: Showed a more favorable Sharpe Ratio of 0.6331, particularly excelling in medium to long-term maturities. This suggests that GBM may be better suited for capturing longer-term trends in the bond market.

Implications for Future Research:

Our results underscore the potential for integrating machine learning with traditional financial analysis to enhance trading strategies. Future work could explore deeper into machine learning algorithms or incorporate additional factors like macroeconomic indicators to refine predictions and improve profitability.

Expanding the model to include more comprehensive risk assessment tools could further optimize the trading strategies, providing a more balanced approach between risk and return.

Practical Applications:

The strategies and models developed in this project can serve as a valuable guide for traders and financial analysts, offering insights into both the theoretical and practical aspects of yield curve trading.

The methodologies outlined could be adapted to other financial markets, enhancing the versatility of the tools and strategies employed.

In conclusion, this project not only advanced our understanding of the structural nuances of the yield curve but also demonstrated the practical applications of sophisticated statistical tools and machine learning in real-world financial scenarios. The integration of these technologies presents a promising avenue for future financial innovations.