Intelligent Fixed Income Investment System

Jiayi Liu, Xuan Liu, Chi-Yuan Peng, Zhenyu Ma, Muya Liu

1. Introduction

The project is developed to support decision-making in fixed income investment by integrating quantitative financial modeling and machine learning techniques to enhance fixed income investment decision-making. The objective is to build a data-driven framework that can systematically model yield curves, forecast bond returns, and construct optimized bond portfolios under risk constraints.

1.1 Components

The system is structured into three core modules:

- **Yield Curve Construction** focuses on analyzing and forecasting how interest rates vary across different bond maturities.
- **Bond Yield Prediction** combines economic indicators and detailed bond data to estimate future bond returns. And then we select bond with higher predicted return for the next step.
- **Portfolio Optimization** uses the predicted bond returns and risk profiles to construct investment portfolios that aim to balance return and stability.

1.2 Data

To support the analytical process, a comprehensive dataset has been assembled. This includes daily Treasury yield and bill rate data from the U.S. Department of the Treasury, macroeconomic indicators (such as GDP, CPI, and unemployment rate) from the Federal Reserve Economic Database (FRED), and bond-level data from Bloomberg, including maturity, bid/ask prices, and credit spreads. The table below shows what kind of data we need to grab and their sources.

Table 1: Raw data and sources

<table>
<tr><th>Section</th><th colspan="2">Data Concept</th><th>Time Period</th><th>Frequency</th><th>Source</th></tr>
<tr><td rowspan="3">Yield Curve Construction</td><td rowspan="2">Treasury Par Yield Curve Rates</td><td>1, 2, 3, 6 Mo</td><td rowspan="2">2015-2024</td><td rowspan="2">daily</td><td rowspan="3">U.S. Department of the Treasury</td></tr>
<tr><td>1, 2, 3 Mo, …, 20, 30,Yr</td></tr>
<tr><td>Treasury Bill Rates</td><td>4, 6, …, 26, 52weeks</td><td>2019-2024</td><td>monthly</td></tr>
</table>

Bond Return & Portfolio Optimization	Macroeconomic data	GDP	2019-2024	monthly	Federal Reserve Economic Data (FRED)
		CPI			
		Unemployment Rate			
		PCE			
	Bond data	maturity	2019-2024	monthly	Bloomberg
		bid price			
		ask price			
		bond yield			
		credit spread			

2. Yield Curve Modeling and Forecasting

This section describes how we modeled the U.S. Treasury yield curve using the Nelson-Siegel framework and forecasted its evolution through time. The objective is to capture both the cross-sectional shape and the temporal dynamics of the yield curve in a way that is interpretable and suitable for simulation and scenario analysis. We first fit the Nelson-Siegel model to historical yield data to extract daily parameter estimates, and then apply a VAR(1) model to those parameters to generate forward-looking yield curves.

2-1. Nelson-Siegel Model Fitting

To capture the cross-sectional shape of the yield curve at each point in time, we applied the Nelson-Siegel model. The model expresses the yield $y(\tau)$ at maturity τ as:

$$y(\tau) = \beta_0 + \beta_1 \times \left(\frac{1 - e^{-\lambda\tau}}{\lambda\tau}\right) + \beta_2 \times \left(\frac{1 - e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau}\right)$$

Where:

β_0 is the long-term level of interest rates,

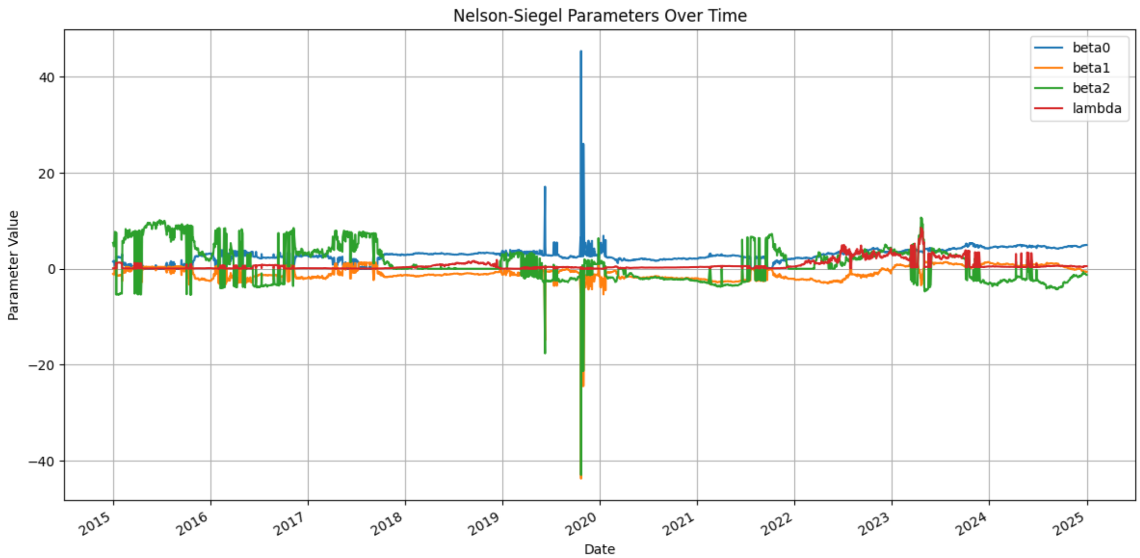
β_1 controls the short-term slope,

β_2 captures the medium-term curvature,

λ is a decay parameter determining how fast the slope and curvature terms decay.

These four parameters were estimated for each business day using the L-BFGS-B optimization algorithm, which minimizes the sum of squared differences between actual yields and model-implied yields. The fitting used standard maturities ranging from 0.25 to 30 years. The resulting parameter series (β_0, β_1, β_2, λ) provide a time-varying summary of the yield curve dynamics.

Figure 1: Parameters for Nelson-Siegel Model



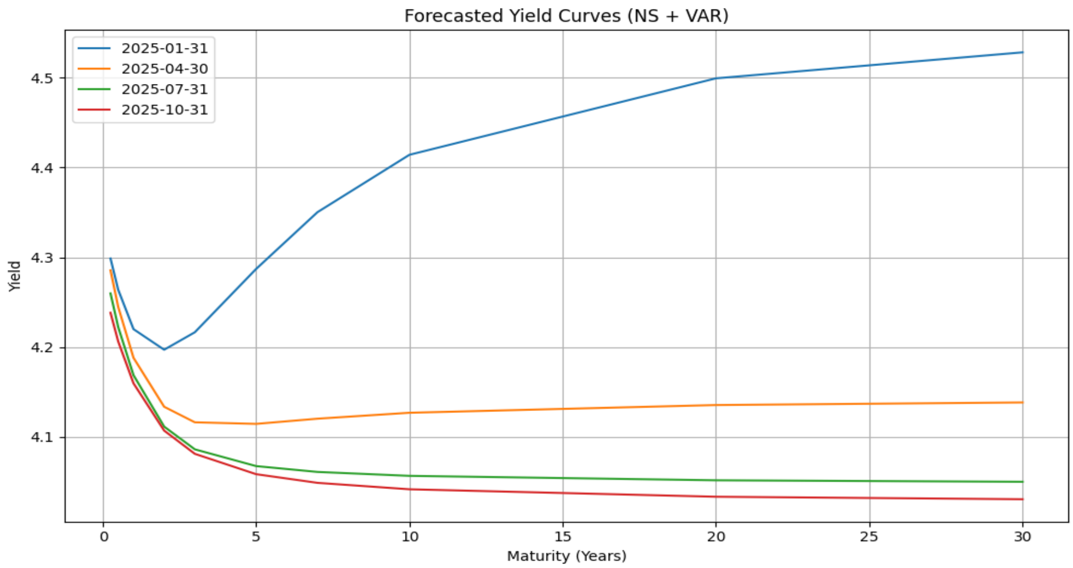
2-2. VAR(1) Modeling and Forecasted Yield Curves

After obtaining the time series of daily Nelson-Siegel parameters, we modeled their joint dynamics using a first-order Vector Auto Regression model (VAR(1)). This multivariate time series model captures how each parameter depends on its own lag and the lags of the other parameters.

The VAR(1) model was fit to the historical β-series and used to forecast the next 12 months of parameters on a monthly basis. Each set of forecasted β-values was then substituted back into the Nelson-Siegel formula to reconstruct the forecasted yield curve for that month.

We evaluated yields at standard maturities: [0.25, 0.5, 1, 2, 3, 5, 7, 10, 20, 30] years. Each curve provides a snapshot of the expected term structure at a future month-end. For example, the January forecasted curve represents what the market might expect the full yield curve to look like on January 31st. These forecasted yield curves serve as the foundation for scenario analysis, portfolio construction, and risk simulations in the next phase of the project.

Figure 2: Forecasted Yield Curve



3. Factor-Based Machine Learning Prediction of Bond Returns

3.1 Objective and Motivation

This section presents a predictive modeling framework that combines financial domain knowledge with machine learning techniques to forecast the 12-month forward return of bonds. In the fixed income market, predicting future returns is challenging due to complex interactions between interest rates, credit risk, and liquidity conditions. Our objective is to leverage interpretable factors and data-driven techniques to identify attractive investment opportunities ahead of time

3.2 Data Sources and Preprocessing

Our dataset includes:

- Bond-level data: mid-price, bid, ask, yield, credit spread.
- Macroeconomic indicators: GDP growth, CPI, UNRATE, and PCE.

Key steps:

- Chronologically sort data by bond ID and date.
- Compute credit spread as bond yield minus Treasury yield of similar maturity.
- Compute bid-ask spread as a proxy for liquidity.

- Align macroeconomic indicators to monthly frequency and merge with bond data.

For each bond-date pair, we ensure all features are available before calculating the return.

3.3 Factor Construction

We construct a parsimonious yet comprehensive set of predictive features:

(1) Term Structure (Yield Curve Slope)

$$Slope = Y_{10Yr} - Y_{2Yr}$$

Reflects the market's economic outlook and rate expectations.

(2) Credit Spread

$$Credit\ Spread = Y_{corporate} - Y_{treasury}$$

Captures risk premium due to creditworthiness concerns.

(3) Liquidity (Bid-Ask Spread)

$$Bid - ask\ Spread = Bid\ price - ask\ price$$

Indicates transaction cost and market efficiency.

(4) Macroeconomic Indicators

- GDP Growth: Proxy for overall economic momentum.
- CPI and PCE: Inflation measures.
- Unemployment Rate (UNRATE): Labor market health.

3.4 Target Variable: Forward Return

We define the target variable as the annualized forward return based on mid-prices:

$$r_{t+12} = \left(\frac{P_{t+12}}{P_t}\right) - 1$$

Where P_t and P_{t+12} are mid-prices 12 months apart. This return is meaningful for long-term bond investors and can be compared across bonds.

3.5 Model Choice: XGBoost Regressor

We adopt **XGBoost** regressor model, with hyper-parameter : n_estimators = 100, max_depth = 3, learning_rate = 0.1 to fit the bond data we have. We split the data using train_test_split while preserving time order, ensuring no data leakage.

3.6 Prediction and Strategy

After model training:

- We applied the model to the latest data point for each bond.
- Predicted the forward 12-month return.
- Ranked the bonds by predicted returns.
- Selected the top 7 bonds as the most promising investments.

3.7 Bond Selection

Based on our bond prediction model above, we select the following bonds with better performance for portfolio optimization:

Table 2: Bond Selection

BOND	ONE-YEAR RETURN	MATURITY TIME
XS2339398971 ISIN	2.37%	5
US48128G3D03 ISIN	2.36%	5
US06654DAF42 ISIN	2.04%	7
US06406RAH03 ISIN	6.67%	10
XS1451090192 ISIN	4.65%	15
XS1451093709 ISIN	3.97%	15
XS1451122573 ISIN	3.58%	15

4. Bond Portfolio Optimization and Back-test

4.1 Duration and Convexity Matching

This task aims to construct a bond portfolio that is robust to interest rate fluctuations. The portfolio is optimized to:

- Match a specified target Macaulay Duration.
- Maximize portfolio convexity under the duration constraint.

4.1.1 Methodology

Each bond's Macaulay Duration and Convexity are calculated based on its cash flow schedule and interpolated forward yield curve. The optimization objective minimizes the squared deviation between the portfolio duration and the target duration, while rewarding convexity:

$$min_{\omega}(D(\omega) - D^{*})^{2} - \lambda \times C(\omega)$$

Where:

$D(\omega)$ = weighted average duration

$C(\omega)$ = weighted average convexity

λ = convexity penalty coefficient

Optimization is subject to:

- Weight constraints: $\omega_i \epsilon [0,1]$
- Weight constraints:$\Sigma_i \omega_i = 1$

4.1.2 Results

The target duration for the bond portfolio was set at **5 years**, and the optimization process successfully achieved a **final portfolio duration of 5.06 years**, closely aligning with the target. The resulting portfolio also exhibits a **convexity of 33.43**, indicating its sensitivity to interest rate changes. The allocation across bonds was determined based on an optimization algorithm, resulting in a set of **optimal portfolio weights** that balance return expectations with duration and convexity constraints.

Table 3: Optimal Portfolio Weights (Duration &Convexity)

ISIN	Weight
US06406RAH03	0.5179
XS14510901920	0.1604
XS1451093709	0.1605
XS1451122573	0.1612

The back-test results and interest rate sensitivity analysis are shown in the table below:

Table 4: Back-test Results

Indicators	Value
Annualized Return	1.90%
Annualized Volatility	9.05%
Sharpe Ratio	0.2096

Table 5: Interest Rate Sensitivity Analysis

Interest Rate Change	Gain & Loss
-100bps	5.22%
-50bps	2.57%
50bps	-2.48%
100bps	-4.49%

4.1.3 Interpretation

The portfolio is duration-neutral and exhibits positive convexity. It performs symmetrically under moderate rate shocks, with slight asymmetry favoring falling rates. Suitable for rate-sensitive strategies such as liability-driven investing or stable value preservation.

4.2 Bond Portfolio Optimization with VaR, ES, and MVO

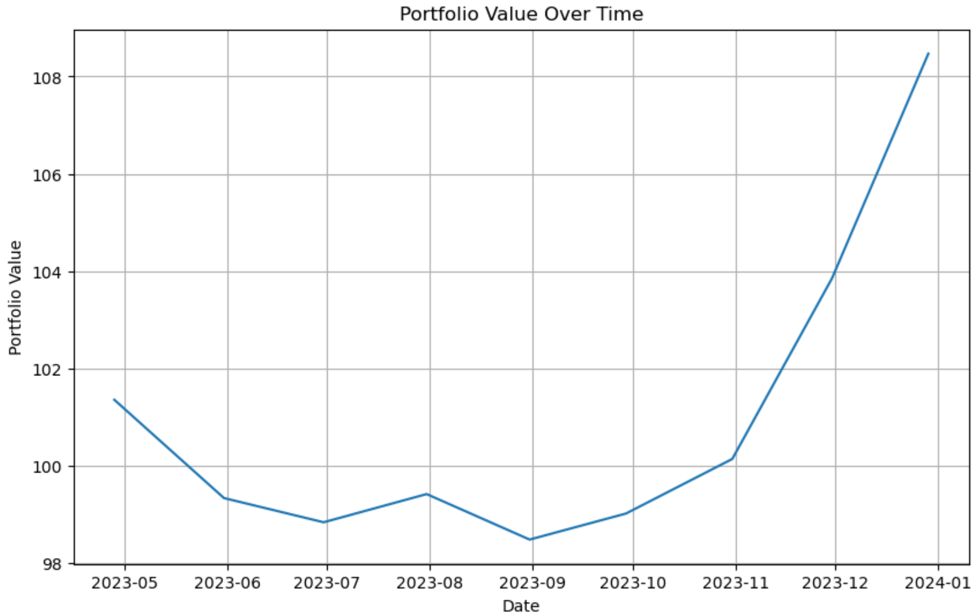
In this section, we construct a bond portfolio using a risk-aware optimization framework that incorporates Value-at-Risk (VaR), Expected Shortfall (ES), and Mean-Variance Optimization (MVO). After estimating historical monthly returns for a selected set of bonds, we calculated individual VaR and ES at the 95% confidence level to assess downside risk. VaR represents the maximum expected loss under normal conditions, while ES measures the average loss in the worst 5% of scenarios.

Table 6: Optimal Portfolio Weights (VaR & ES)

ISIN	Weight(1Y)	Weight(3Y)	Weight(4Y)	Coupon Rate	Last Price
US06406RAH03	0.5334	0.4715	0.4592	3.85	97.8513
XS2339398971	0	0.2024	0.1913	0.4	88.023
US48128G3D03	0	0.3261	0.3494	2	83.4854
US06654DAF42	0	0	0	2.48	83.2281
XS1451090192	0.2406	0	0	2.12	89.8042
XS1451093709	0.1416	0	0	2.08	89.5674
XS1451122573	0.0844	0	0	1.8	88.3829

The portfolio-level VaR was calculated to be **4.12%**, meaning there is a 5% probability of losing more than this in one period. The corresponding ES was **-5.57%,** indicating that average losses in the worst 5% of cases could reach 5.57%.

Figure 3: Portfolio Performance Plot



We then applied a Mean-Variance Optimization approach to construct a portfolio that maximizes expected return while minimizing risk, with short selling prohibited. To evaluate the risk profile further, we conducted an interest rate sensitivity analysis. The portfolio's weighted modified duration was **7.09**, and convexity was **59.38**, signaling high sensitivity to interest rate changes. Simulated parallel shifts in interest rates showed that:

- A 50 basis point increase would lead to a 3.47% loss, and a 100basis point increase would result in a 6.80% loss.
- In contrast, a 50 basis point decrease would yield a 3.62% gain, and a 100 basis point drop could produce a 7.39% gain.

To test the robustness of the strategy, we conducted a back-test over the period from January 2023 to January 2024, a relatively stable market environment following the peak of U.S. interest rate hikes in 2022. This time window was selected to isolate the effectiveness of the optimization model from macroeconomic shocks. The portfolio achieved a final cumulative return of 8.47%, with an annualized return of 11.45%, annualized volatility of 7.68%, and a Sharpe ratio of 1.49, indicating excellent risk-adjusted performance.

These results suggest that the VaR-ES-constrained MVO strategy can be highly effective under stable or declining interest rate environments. However, the high duration also implies exposure to rate volatility, making this strategy potentially vulnerable in tightening cycles. Overall, this approach offers a solid balance between return maximization and downside risk management, with its success depending largely on market regime and interest rate expectations.

5. Strategy Comparison: Duration & Convexity vs. VaR & ES

Compared to the duration and convexity matching strategy in Section 4.1, the VaR-ES-MVO approach used in Section 4.2 places more emphasis on historical return distributions rather than just interest rate sensitivity. The 4.1 strategy is designed to keep the portfolio stable under parallel rate shifts by tightly matching duration and maximizing convexity. As a result, it performs more symmetrically during interest rate shocks, but its back-test showed relatively modest returns and a lower Sharpe ratio (0.21).

In contrast, the 4.2 strategy achieved stronger performance in back-testing, with a significantly higher Sharpe ratio of 1.49 and an annualized return of 11.45%. However, it also comes with higher sensitivity to rate movements due to its longer duration. This makes the 4.2 strategy more suitable for environments with stable or falling interest rates, while the 4.1 strategy may be preferred when the priority is minimizing rate risk.

Overall, the choice between these two approaches depends on the investor’s objective: Section 4.1 is more defensive and rate-hedged, while Section 4.2 is return-focused with stronger risk-adjusted performance in favorable conditions.

Reference

1. Nelson, C. R., & Siegel, A. F. (1987). Parsimonious modeling of yield curves. Journal of Business, 60(4), 473–489. https://doi.org/10.1086/296409

2. Diebold, F. X., & Li, C. (2006). Forecasting the term structure of government bond yields. Journal of Econometrics, 130(2), 337–364. https://doi.org/10.1016/j.jeconom.2005.03.005

3. Diebold, F. X., Rudebusch, G. D., & Aruoba, S. B. (2006). The macroeconomy and the yield curve: A dynamic latent factor approach. Journal of Econometrics, 131(1–2), 309–338.

4. Litterman, R., & Scheinkman, J. (1991). Common factors affecting bond returns. *Journal of Fixed Income*, *1*(1), 54–61.