THE UNIVERSITY OF CHICAGO

Forecasting Chinese Government Bond Yield Curves: An Empirical Comparison of DNS (Dynamic-Nelson-Siegel) Model and Machine Learning Approaches

By

Jingnan Zhang

November 5, 2024

A paper submitted in partial fulfillment of the requirements for the
Master of Arts degree in Social Sciences with a Concentration in Economics

Preceptor: Rui Zhao

Abstract

This paper investigates the efficacy of the Dynamic Nelson-Siegel (DNS) model in forecasting Chinese government bond yields. Our sample comprises daily zero-coupon yields from March 2006 to April 2024. To tailor the DNS model to the Chinese market condition, we optimize the DNS decay parameter λ to 0.42625. The DNS model achieves a parsimonious representation through exponential-form factor loadings, and its AR(1) specification effectively captures factor evolution, demonstrating robust predictive capability ($R^2 = 0.9780$). Given the recent advances in computational methods, we explore whether machine learning approaches can enhance yield curve forecasting. Recognizing that yield curves exhibit complex cross-sectional and non-linear patterns, we introduce eXtreme Gradient Boosting (XGBoost) to model these aspects. XGBoost's ensemble of decision trees allows it to effectively capture non-linear relationships and potential regime shifts in the term structure by recursively partitioning the feature space. We also employ Long Short-Term Memory (LSTM) networks to capture temporal dependencies inherent in the data, which can leverage their memory cells and gating mechanisms for adaptive information processing. We optimize the network architecture by tuning hyperparameters such as the number of layers, neurons, and the look-back window length. The empirical results show that, while both machine learning approaches demonstrate marginally superior predictive power (R^2: 0.9861 and 0.9801 respectively) compared to DNS (R^2: 0.9780), their modest improvements suggest that the traditional framework's parsimony and economic interpretability remain valuable in yield curve modeling.

Keywords: Dynamic Nelson-Siegel model, Yield curve forecasting, Term structure, Chinese Government Bond Yields, LSTM, XGBoost

1 Introduction

The modeling and forecasting of term structure of interest rates play a pivotal role in financial markets, monetary policy implementation, and economic analysis. Among various methodologies, the Dynamic Nelson-Siegel (DNS) model stands out for its parsimonious yet flexible specification in capturing yield curve dynamics. We examine the efficacy of the DNS model in forecasting Chinese government bond yields while exploring potential enhancements through machine learning techniques.

First, we conduct a detailed examination of Chinese government bond yield through the DNS model. Second, we enhance the traditional DNS methodology by optimizing the decay parameter λ specifically for Chinese market conditions, improving both the model's fit and forecasting accuracy. Third, we perform out-of-sample testing to evaluate the model's predictive performance, comparing it against both simpler benchmark models (Random Walk and VAR model) and more sophisticated machine learning approaches (XGBoost and LSTM). This comparative analysis helps assess whether the relative parsimonious DNS framework remains competitive with newer methodologies in capturing yield curve dynamics.

By analyzing sample data from March 2006 to April 2024, we find that the DNS-AR(1) framework demonstrates robust predictive capability ($R^2 = 0.9780$). The level factor exhibits non-stationary characteristics and the strongest persistence (autocorrelation of 0.9277 at lag 30), while the slope and curvature factors show stationarity. Correlation analysis among factors reveals a significant positive correlation (0.4365) between the level and slope factors, suggesting that yield spreads between short and long-term rates tend to widen during periods of rising interest rate levels. Through our calibration of the optimal decay parameter λ (0.42625), the model achieves strong predictive performance (average MSE of 0.0028) while maintaining structural parsimony, outperforming traditional VAR(1) and random walk models.

Furthermore, we compare the DNS framework with machine learning methods. Empirical results show that while LSTM (MSE: 0.0019, R^2: 0.9861) and XGBoost (MSE: 0.0027, R^2: 0.9801) slightly outperform the DNS-AR(1) model in predictive accuracy, the improvements are relatively modest. Notably, model performance differences are insignificant in the medium-term segment (1-10 years). Machine learning methods demonstrate their advantages primarily in forecasting short-term (0.25-0.5 years) and long-term (20-30 years) yields, likely due to their ability to capture more complex non-linear relationships at these maturities. This finding suggests that while machine learning methods offer certain predictive improvements, the traditional DNS framework remains a highly practical choice when considering model complexity, interpretability, and implementation costs. Under normal market conditions, the DNS approach may sufficiently meet most application requirements, while machine learning methods may serve as valuable complements during special market environments (such as high volatility periods).

2 Data Processing

2.1 Data Sources and Processing

Our study utilizes zero-coupon government bond yield data provided by China Central Depository & Clearing Co., Ltd. (CCDC), of which the data is highly authoritative as the custodian for Chinese government bonds. The dataset encompasses daily zero-coupon yields for 21 different maturities (ranging from overnight to 30 years) over the period from March 2006 to April 2024, covering a total of 4,528 trading days.

2.2 Hermite Interpolation Model

CCDC employs Hermite interpolation model to construct the yield curve. Hermite interpolation is a method of interpolating between given data points that takes into account not only the values at these points but also their derivative values. This approach ensures the continuity and smoothness of the yield curve while maintaining a high degree of accuracy in fitting.

Given a set of maturities $x_1 < \cdots < x_n$ and their corresponding yields $y_1 < \cdots < y_n$, the slope of the yield curve at the corresponding maturities is denoted as $d_1 \cdots d_n$.

For $i = 1, 2, \cdots, n-1$, assume that the yield curve of $[x_i, x_{i+1}]$ can be represented as a cubic polynomial function of x, and $f(x)$ satisfies the following conditions:

$$f(x_i) = y_i, f(x_{i+1}) = y_{i+1} \tag{2-2-1}$$

$$f'(x_i) = d_i, f'(x_{i+1}) = d_{i+1} \tag{2-2-2}$$

The specific form of Hermite interpolation model can be derived as follows:

$$f(x) = y_i H_1(x) + y_{i+1} H_2(x) + d_i H_3(x) + d_{i+1} H_4(x) \tag{2-2-3}$$

In which:

$$\begin{aligned}
H_1(x) &= 3\left(\frac{x_{i+1} - x}{x_{i+1} - x_i}\right)^2 - 2\left(\frac{x_{i+1} - x}{x_{i+1} - x_i}\right)^3 \\
H_2(x) &= 3\left(\frac{x - x_i}{x_{i+1} - x_i}\right)^2 - 2\left(\frac{x - x_i}{x_{i+1} - x_i}\right)^3 \\
H_3(x) &= \frac{(x_{i+1} - x)^2}{x_{i+1} - x_i} - \frac{(x_{i+1} - x)^3}{(x_{i+1} - x_i)^2} \\
H_4(x) &= \frac{(x - x_i)^3}{(x_{i+1} - x_i)^2} - \frac{(x - x_i)^2}{x_{i+1} - x_i}
\end{aligned} \tag{2-2-4}$$

Hermite interpolation model ensures the continuity and smoothness of the yield curve, avoiding unnatural jumps or inflection points at the interpolation nodes.

2.3 Descriptive Statistical Analysis

In order to comprehensively understand the statistical characteristics of Treasury bond yields, we conducted a descriptive statistical analysis of the zero-coupon yield data from March 2006 to April 2024. Table 2-3-1 presents the statistical indicators for government bond yields across some key maturities, including the number of observations, mean, standard deviation, minimum value, 25th percentile, median, 75th percentile, maximum value, variance, and skewness.

Table 2-3-1: Descriptive Statistics of Treasury Bond Yields for Key Maturities

Maturity	2M	3M	1Y	2Y	5Y	10Y
Count	4528	4528	4527	4528	4527	4527
Mean	2.3116	2.3608	2.5608	2.7717	3.1362	3.4496
Std	0.7469	0.7313	0.6932	0.6580	0.5536	0.5341
Min	0.7623	0.7989	0.8911	1.0744	1.7591	2.2848
25%	1.8056	1.8698	2.0917	2.2897	2.6845	3.0295
50%	2.1990	2.2673	2.4857	2.7120	3.0600	3.4070
75%	2.8665	2.9058	3.0803	3.2245	3.5368	3.7567
Max	5.6411	5.1132	4.2503	4.4744	4.5901	4.8048
Variance	0.5579	0.5348	0.4806	0.4330	0.3065	0.2853
Skewness	0.4106	0.2639	0.0306	0.1000	0.4366	0.4282

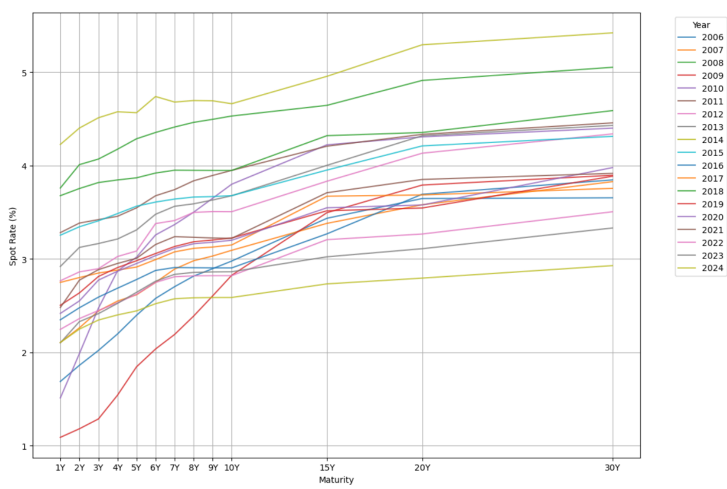

Figure 2-3-1: Time Series of Treasury Spot Rates

The yields show a clear upward trend as maturity extends, increasing from 2.3116% for 2-month bonds to 3.4496% for 10-year bonds. This pattern reflects a typical positive term structure, indicating liquidity and risk premiums for longer-dated bonds. We can also find that interest rate volatility tends to decrease with longer maturities. Short-term rates exhibit significantly higher standard deviations compared to long-term rates. For instance, the standard deviations for 2-month and 3-month rates are 0.7469 and 0.7313 respectively, while the 10-year rate's standard deviation is only 0.5341. This phenomenon suggests that short-term rates are more sensitive to changes in market liquidity and monetary policy adjustments, whereas long-term rates are more influenced by economic fundamentals and long-term expectations.

2.4 Term Structure Factors

In the study of interest rate term structures, traditionally we use three key statistics to summarize the main features of the yield curve: Level (L), Slope (S), and Curvature (C). These factors effectively capture the overall shape and dynamic changes of the interest rate term structure, providing important clues for understanding market expectations and economic conditions.

- Level Factor (L): L is defined as the 30-year interest rate. This factor reflects the long-term interest rate level, providing an indication of the overall yield environment. Changes in the level factor are typically closely related to long-term inflation expectations and economic growth prospects.
- Slope Factor (S): S is defined as the difference between the 30-year rate and the 3-month rate. The slope factor measures the difference between long-term and short-term interest rates, offering insights into market expectations of interest rates and economic outlook. A positive slope typically indicates that long-term rates are higher than short-term rates, which may suggest market expectations of economic growth and potential inflationary pressures.
- Curvature Factor (C): C is defined as twice the 2-year rate minus the sum of the 3-month rate and the 30-year rate. The curvature factor reflects the degree of bending in the yield curve and is particularly helpful in capturing changes in medium-term rates that might not be easily discerned when considering only level and slope.

Based on these definitions, we can represent these three factors as follows:

$$\begin{aligned} L(t) &= y(t, 30) \\ S(t) &= y(t, 30) - y(t, 0.25) \\ C(t) &= 2 \cdot y(t, 2) - y(t, 0.25) - y(t, 30) \end{aligned} \tag{2-4-1}$$

The descriptive statistics for term structure factors are as follows:

Table 2-4-1: Descriptive Statistics of Term Structure Factors

Parameter	L	S	C
Count	4528	4528	4528
Mean	4.1269	1.7661	-0.9443
Std	0.5439	0.5933	0.4984
Min	2.4594	-0.7644	-2.7610
25%	3.7780	1.3012	-1.1812
50%	4.1759	1.6964	-0.8942
75%	4.4481	1.9939	-0.6017
Max	5.4815	3.5263	0.1606
Autocorr(1)	0.9988	0.9959	0.9923
Autocorr(12)	0.9792	0.9295	0.9192
Autocorr(30)	0.9351	0.8178	0.8288

All three traditional factors showed high autocorrelation.

3 Introduction of Dynamic Nelson-Siegel Model

3.1 Theoretical Background

While traditional factors can help us describe the term structure, they have certain limitations. Firstly, these factors are directly calculated from interest rates at specific maturities, they may potentially fail to fully capture information across the entire yield curve. Secondly, the static nature of traditional factors makes it challenging to describe the dynamic evolution process of the term structure. To solve these limitations, Nelson and Siegel (1987) proposed the famous Nelson-Siegel model which uses a small number of parameters to flexibly fit various shapes of yield curves, which provides a powerful yet concise framework for modeling the interest rate term structure.

3.2 Nelson-Siegel Model Construction

The core of the NS model lies in the construction of the forward interest rate function:

$$f(t,\tau) = \beta_{0t} + \beta_{1t}e^{-\lambda\tau} + \beta_{2t}\lambda\tau e^{-\lambda\tau} \tag{3-2-1}$$

By integrating Eq. (3-2-1) to obtain the immediate yield function under continuous compounding at any point in time:

$$R(t,\tau) = \frac{1}{\tau}\int_0^\tau f(t,s)ds = \beta_{0t} + \beta_{1t}\left[\frac{1-e^{-\lambda\tau}}{\lambda\tau}\right] + \beta_{2t}\left[\frac{1-e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau}\right] \tag{3-2-2}$$

Where $R(t,\tau)$ denotes the spot rate with a maturity of τ, β_0, β_1 and β_2 are the parameters to be evaluated. As the four parameters change, the interest rate term structure takes on different shapes. λ determines how fast the index function decays and when the loadings of β_1 and β_2 are maximized. β_0, β_1, and β_2 have different effects on the interest rates of different maturities, and the specific effects can be reflected by the changes in the loadings of the factors.

The limit of the loading of the slope factor β_1 is:

$$\lim_{\tau \to 0} \frac{1 - e^{-\lambda\tau}}{\lambda\tau} = 1 \tag{3-2-3}$$

$$\lim_{\tau \to \infty} \frac{1 - e^{-\lambda\tau}}{\lambda\tau} = 0 \tag{3-2-4}$$

The limit of the loading of the slope factor β_2 is:

$$\lim_{\tau \to 0} \left(\frac{1 - e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau} \right) = 0 \tag{3-2-5}$$

$$\lim_{\tau \to \infty} \left(\frac{1 - e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau} \right) = 0 \tag{3-2-6}$$

β_0 affects the rate of return to the same extent at any tenor, which is referred to as the level factor. The load on β_1 decreases rapidly with increasing maturity and decreases less and less, so it has a larger effect on the short-term interest rate and we call it the slope factor. The loading of β_2 increases and then decreases with maturity, so it has a greater impact on the medium-term interest rate and we call it the curvature factor. Changes in β_0, β_1, and β_2 affect changes in the overall term structure of interest rates, so changes in β_0, β_1, and β_2 can be used to measure changes in the term structure of interest rates.

The traditional approach to estimating the Nelson-Siegel model is to minimize the weighted sum of squares of the difference between the actual and theoretical prices of all bonds, i.e.:

$$\min \sum_{i=1}^{n} w_i (P_i - PV_i)^2 \tag{3-2-7}$$

where P_i denotes the actual price of bond i; PV_i denotes the theoretical price of bond i estimated using the Nelson-Siegel model, and w_i denotes the weighted weights of bond i. By solving the nonlinear optimization problem, we can estimate the parameters of the Nelson-Siegel model. However, accurately solving this complex nonlinear optimization problem not only requires the selection of a reasonable algorithm, but also relies on the choice of initial values of the parameters.

3.3 Dynamic Nelson-Siegel Model Construction

3.3.1 Introduction to DNS Model

Diebold and Li (2006) pointed out that this type of estimation leads to distortion of the parameter estimation results. They enhance the Nelson-Siegel model by introducing a dynamic factor analysis approach, making it more effective for forecasting yield curves. In the Diebold-Li method, the solution of the nonlinear optimization problem formulation is transformed into a linear least squares regression problem by determining the value of the parameter λ. They reset the original model to the following form:

$$y_t(\tau) = L_t + S_t\left(\frac{1-e^{-\lambda\tau}}{\lambda\tau}\right) + C_t\left(\frac{1-e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau}\right) \tag{3-3-1}$$

Where, $y_t(\tau)$ denotes the spot interest rate with maturity period τ at time t, L_t, S_t, C_t are the level factor, slope factor and curvature factor respectively, which have the same meaning as β_0, β_1, β_2 in the above equation, and their coefficients are called factor loadings.

The DNS model assumes that the potential factors obey a first-order vector autoregressive form, and the model is written in the form of a state-space model as follows:

The equation of state of the model is:

$$\begin{pmatrix} L_t \\ S_t \\ C_t \end{pmatrix} = \begin{pmatrix} \mu_L \\ \mu_S \\ \mu_C \end{pmatrix} + \begin{pmatrix} a_{11} & 0 & 0 \\ 0 & a_{22} & 0 \\ 0 & 0 & a_{33} \end{pmatrix} \begin{pmatrix} L_{t-1} \\ S_{t-1} \\ C_{t-1} \end{pmatrix} + \begin{pmatrix} \eta_t(L) \\ \eta_t(S) \\ \eta_t(C) \end{pmatrix} \tag{3-3-2}$$

The corresponding measurement equation is:

$$\begin{pmatrix} y_t(\tau_1) \\ y_t(\tau_2) \\ \vdots \\ y_t(\tau_N) \end{pmatrix} = \begin{pmatrix} 1 & \frac{1-e^{-\lambda\tau_1}}{\lambda\tau_1} & \frac{1-e^{-\lambda\tau_1}}{\lambda\tau_1} - e^{-\lambda\tau_1} \\ 1 & \frac{1-e^{-\lambda\tau_2}}{\lambda\tau_2} & \frac{1-e^{-\lambda\tau_2}}{\lambda\tau_2} - e^{-\lambda\tau_2} \\ \vdots & \vdots & \vdots \\ 1 & \frac{1-e^{-\lambda\tau_N}}{\lambda\tau_N} & \frac{1-e^{-\lambda\tau_N}}{\lambda\tau_N} - e^{-\lambda\tau_N} \end{pmatrix} \begin{pmatrix} L_t \\ S_t \\ C_t \end{pmatrix} + \begin{pmatrix} \varepsilon_t(\tau_1) \\ \varepsilon_t(\tau_2) \\ \vdots \\ \varepsilon_t(\tau_N) \end{pmatrix} \tag{3-3-3}$$

In their empirical analysis, Diebold and Li used end-of-month price quotes (bid-ask averages) for U.S. Treasury bonds from January 1985 to December 2000. The dataset was sourced from the CRSP government bond files. In our paper, we use the zero-coupon government bond yield data provided by CCDC and the curve is constructed by Hermite Interpolation since we do not have full access to the bond prices. To ensure data quality and accuracy, CRSP applied rigorous filtering criteria. First, they excluded bonds with option features, such as callable bonds, which could distort the yield calculations due to embedded options. Additionally, bonds with special liquidity issues, such as those nearing maturity (less than one year) or with irregular liquidity patterns, were also removed to prevent anomalies in the yield curve fitting process. After the filtering, the dataset consists of U.S. Treasury bonds with maturities ranging from 3 months to 120 months.

Then they used the unsmoothed Fama-Bliss forward rate method to convert bond prices into zero-coupon yields. They calculated forward rates for different maturities and constructs zero-coupon yields by averaging the forward rates.

3.3.2 Empirical Attempt

First, we loaded the historical spot yield data and sorted the data by date. We defined the Nelson-Siegel yield curve function and established a maturity mapping dictionary to convert all tenors into annual units. We utilized Eq. (3-3-3) to estimate the factors (L, S, and C) and computed the corresponding factor loadings. Based on the characteristics of Chinese government bonds, we experimented with multiple decay factors to minimize prediction error. Specifically, for each maturity, we computed the factor loadings using various lambda values and estimated the factors for each date by regressing yields on these factor loadings at each time point.

We used data prior to January 1, 2021, as the training set and subsequent data as the test set. For each λ value (ranging from 0.05 to 0.5 with an increment of 0.00001), we first conducted cross-sectional estimation at each time point t using measurement equation (3-3-3) through OLS regression to obtain the time series of factors (L_t, S_t, C_t). Then we estimated AR(1) coefficients using these factor series according to state equation (3-3-2). With these estimations, we conducted out-of-sample predictions on the test set by forecasting the factors using Eq.(3-3-2) and reconstructing the yields using Eq.(3-3-3). We measure the prediction error by MSE calculated from the differences between actual and predicted yields across all maturities in the test set. The optimization process yielded an optimal λ value of 0.42625.

Then we conducted statistical analysis of descriptive statistics for both yields and estimated factors, along with autocorrelations at lags 1, 12, and 30. Furthermore, we employed ADF tests to examine the stationarity properties of the factors. We examine the model result by calculating MSE and MAE between estimated factors and traditional factors.

Moreover, we conducted model's residual analysis including descriptive statistics and autocorrelation patterns, facilitating the assessment of model fit and potential systematic patterns.

Finally, we fitted AR(1) models to the estimated factors to examine their dynamic properties. Specifically, we evaluate 40-day-ahead factor predictions.

Table 3-3-1: Descriptive Statistics of DNS Parameters

Parameter	L(t)	S(t)	C(t)
Count	4528	4528	4528
Mean	4.4502	2.1349	-0.5805
Std	0.7846	0.8330	1.1097
Min	2.4926	-4.6027	-10.9421
25%	4.0419	1.7060	-0.8176
50%	4.5047	1.9793	0.0000
75%	4.8755	2.4231	0.0000
Max	18.8300	15.1435	0.0263
Autocorr(1)	0.9974	0.9966	0.9746
Autocorr(12)	0.9711	0.9459	0.7862
Autocorr(30)	0.9277	0.8599	0.6112

Our analysis aligns with Diebold-Li's findings. While all three factors exhibit significant persistence, their degrees vary notably. The level factor demonstrates the highest persistence, maintaining an autocorrelation of 0.9277 even at the 30th lag, reflecting the stability of the overall interest rate level. The slope factor shows the second-highest persistence, while the curvature factor exhibits relatively weaker autocorrelation, particularly at longer lags, consistent with Diebold-Li's observations. Overall level changes occur gradually, while curve shape exhibits greater flexibility.

Table 3-3-2: ADF Test of DNS Parameters

	Level	**Slope**	**Curvature**
ADF Statistic	-1.6302	-3.4494	-4.8119
p-value	0.4674	0.0094	0.0001
Critical Values	-2.8622	-2.8622	-2.8622
Stationary	FALSE	TRUE	TRUE

The ADF test results reveal that the Level factor is non-stationary, consistent with Diebold-Li's characterization that it is near-unit-root processes. Our analysis indicates that the Slope factor is stationary, which is slightly different from Diebold-Li's findings. However, Curvature factor demonstrates significant stationarity, aligning with Diebold-Li's observation of faster mean reversion. While these subtle differences might reflect variations in sample periods or market environments, the overall pattern supports the fundamental assumptions of the Nelson-Siegel model. The non-stationarity of the Level factor suggests that long-term interest rates are susceptible to persistent macroeconomic shocks, whereas the stationarity of Slope and Curvature factors indicates that yield curve shape changes are predominantly driven by short-term market dynamics.

Diebold-Li recognized that the Level and Slope factors may be near-unit-root processes, but they chose to model and predict directly on the level data using AR(1) model rather than conducting differencing process. They argue that distinguishing between a true unit root process and a highly persistent near-unit root process in a finite sample is both difficult and possibly unnecessary.

Table 3-3-3: Factor Correlation Analysis

	Level	**Slope**	**Curvature**
Level	1.0000	0.4365	-0.2427
Slope	0.4365	1.0000	-0.2864
Curvature	-0.2427	-0.2864	1.0000

The factor correlation analysis reveals that the positive correlation (0.4365) between Level and Slope factors indicates a tendency for the spread between short-term and long-term rates to widen during periods of rising interest rate levels. This pattern may reflect market adjustments to expectations of future economic growth and inflation. Simultaneously, the negative correlations of the Curvature factor with Level and Slope factors (-0.2427 and -0.2864, respectively) suggest that the prominence of medium-term rates relative to short and long-term rates tends to diminish when interest rate levels rise or the yield curve steepens. This dynamic likely stems from variations in market participants' risk preferences across different maturities.

Table 3-3-4: Error Analysis between NS and Traditional Factors

	Level	Slope	Curvature
MSE	0.0447	0.0517	0.6047
MAE	0.1492	0.0127	0.5897

As shown in Figure 3-3-1 to 3-3-3, we plot the time series of the traditional values and NS values of the three factors to visually demonstrate the dynamic evolution of the term structure.

Figure 3-3-1: L(t) Over time

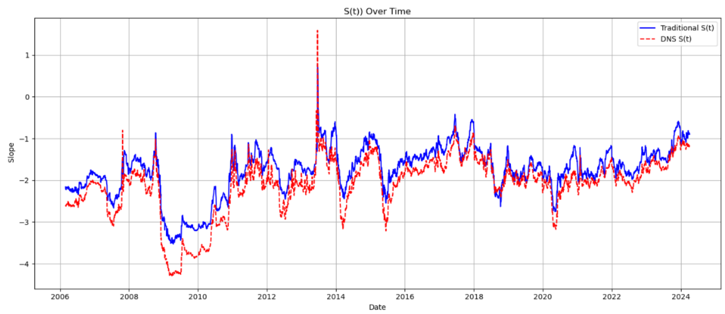

Figure 3-3-2: S(t) Over time

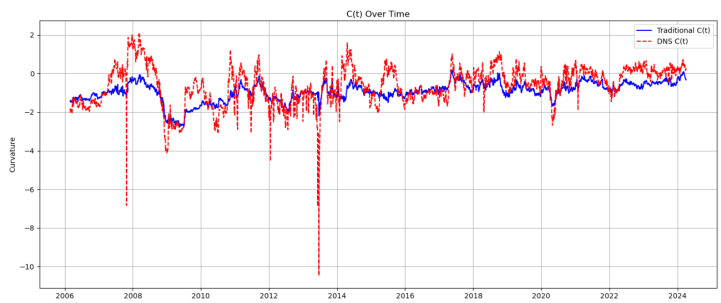

Figure 3-3-3: C(t) Over time

Next, we fit the AR(1) model. The results show that, after controlling for the other variables, the current values of Slope and Curvature are significantly influenced mainly by their first- and second-order lagged values, which have obvious dynamic characteristics, while the effect of the Level variable is not significant.

Table 3-3-5: AR(1) Model Statistical Significance

	Const	Level.L1	Const	Slope.L1	Const	Curvature.L1
coef	0.0067	0.9983	-0.0062	0.9967	-0.0157	0.9745
std err	0.005	0.001	0.003	0.001	0.004	0.003
z	1.481	935.563	-2.379	818.253	-3.954	292.812
P> |z|	0.139	0.000	0.017	0.000	0.000	0.000
[0.025	-0.002	0.996	-0.011	0.994	-0.023	0.968
0.975]	0.016	1.000	-0.001	0.999	-0.008	0.981

Then we conduct the residual analysis.

Table 3-3-6: AR(1) Model Residual Analysis

Residual Autocorr	Level AR(1)	Slope AR(1)	Curvature
1	0.0865	0.1614	0.0369
12	-0.0143	-0.0363	-0.0139
30	-0.0148	-0.0224	-0.0186

The analysis reveals relatively low autocorrelations suggest that the AR(1) model effectively captures the dynamic characteristics of the factors. At longer lags (2-60 periods), the residual autocorrelations for all three factors fluctuate primarily within a ±0.05 band without exhibiting significant systematic patterns, further supporting the adequacy of the AR(1) specification. Notably, the Slope factor displays weak but persistent positive autocorrelation (approximately 0.05) at lags 20-24, potentially indicating subtle monthly cyclical patterns in yield curve steepness.

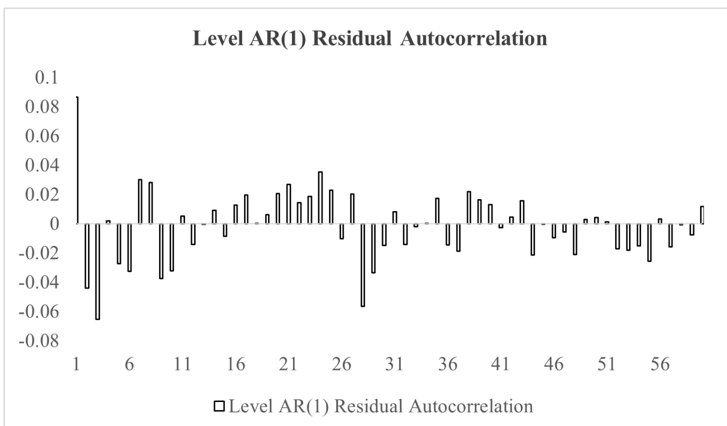

Figure 3-3-4: Level AR(1) Residual Autocorrelation

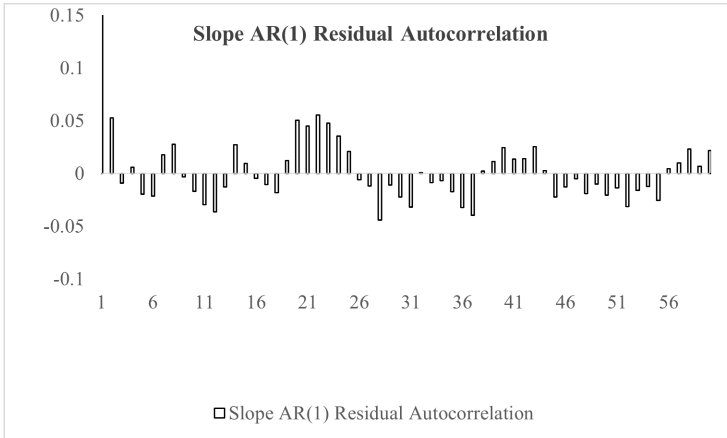

Figure 3-3-5: Slope AR(1) Residual Autocorrelation

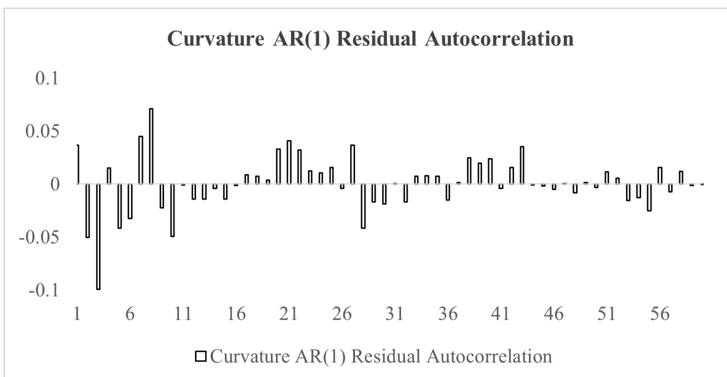

Figure 3-3-6: Curvature AR(1) Residual Autocorrelation

3.4 Model Fitting Effectiveness Test

In this paper, we choose MSE, RMSE, MAE and R^2 to be the benchmark of the fitting and prediction effects. Following are the formulas:

$$\text{MSE} = \frac{1}{m}\sum_{i=1}^{m}(y_i - \hat{y}_i)^2 \tag{3-4-1}$$

Where $y_i - \hat{y}_i$ is the difference between the predicted value and the actual value.

$$\text{RMSE} = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(y_i - \hat{y}_i)^2} \tag{3-4-2}$$

$$\text{MAE} = \frac{1}{m}\sum_{i=1}^{m} |y_i - \hat{y}_i| \tag{3-4-3}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{m}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{m}(y_i - \bar{y})^2} \tag{3-4-4}$$

Where $\bar{y}$ is the average of all actual values, for each date, m is the 11 key maturities, for the whole prediction, m is $11\times$ all the prediction period.

For the prediction, we employ AR(1) models to generate 40-day-ahead predictions for each factor using daily frequency data. The predictions start from 4/7/2024 to 5/31/2024.

We reconstruct forecast yield curves across eleven key maturities ranging from 3 months to 30 years with the optimal λ 0.42625. The forecasting accuracy is then evaluated by comparing the reconstructed yield curves with actual yields.

Table 3-4-1: DNS Predicting Performance

Date	MSE	RMSE	MAE
2024/4/7	0.0012	0.0343	0.0284
2024/4/8	0.0008	0.0286	0.0244
2024/4/9	0.0009	0.0292	0.0251
⋮	⋮	⋮	⋮
2024/5/28	0.0014	0.0373	0.0304
2024/5/29	0.0012	0.0343	0.0290
2024/5/30	0.0012	0.0343	0.0288
2024/5/31	0.0016	0.0401	0.0327

The DNS model exhibits stability and accuracy throughout the entire forecasting period, with Avg-MSE of 0.0028, Avg-RMSE of 0.0490, and Avg-MAE of 0.0422. Notably, the model's R^2 reached 0.9780, indicating a high degree of goodness-of-fit and suggesting that the DNS model can explain the vast majority of variations in the yield curve movements. In terms of specific values, the MSE fluctuated between 0.0008 and 0.0016, demonstrating that the model not only performs well overall but also maintains a high level of stability across individual forecasting dates.

In addition to using the DNS model, we employed the VAR(1) process and the Random Walk process to model and forecast the yield curve. The general form of VAR(1) is:

$$Y_t = c + \Phi Y_{t-1} + \epsilon_t \tag{3-4-5}$$

In our study, the forecasting performance of VAR(1) is relatively poor, with average MSE of 0.0682, RMSE of 0.2200, MAE of 0.2111, and R^2 of 0.4971, which suggests Its weak ability to explain yield movements. The general form of Random Walk is:

$$y_{t+1} = y_t + \epsilon_t \tag{3-4-6}$$

Random Walk outperforms VAR(1) with average MSE of 0.0392, RMSE of 0.1711, MAE of 0.1531, and R^2 of 0.7119, but it still underperforms our DNS model.

Table 3-4-2: Predicting Performance Comparison

Metric	DNS	VAR(1)	Random Walk
MSE	0.0028	0.0682	0.0392
RMSE	0.049	0.22	0.1711
MAE	0.0422	0.2111	0.1531
R^2	0.9780	0.4971	0.7119

Overall, the DNS model significantly outperforms, especially in explaining the shifts in the yield curve. This suggests that the Dynamic Nelson-Siegel model combined with the AR(1) process is able to better capture the characteristics of the yield curve and the time-series dynamics, but there is still room for improvement in terms of stability.

4 Compare DNS with Machine Learning Methods

In the prediction scenario of yield curve, we hope to explore the most superior method of prediction performance by comparing the prediction effect with other models under the same data source and the same scenario. Therefore, we choose two popular machine learning models for comparison.

4.1 XGBoost Method

Among various machine learning methods, XGBoost demonstrates unique advantages in predicting government bond yield curves. From the theoretical foundation of term structure of interest rates, whether based on Expectations Theory and Market Segmentation Theory, interest rate formation involves complex interactions between multiple market participants and economic factors. XGBoost can automatically capture these nonlinear relationships through its tree structure, with its objective function:

$$L = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k) \tag{4-1-1}$$

where L represents the objective function, $l(y_i, \hat{y}_i)$ is the training loss between the predicted value $\hat{y}_i$ and the true value y_i, $\Omega(f_k)$ is the regularization term for the k-th tree, and K is the total number of trees.

By integrating multiple decision trees, it can approximate arbitrarily complex nonlinear functions. There are numerous feature interactions in the interest rate formation process, such as the interactive effects between inflation and economic growth, and the joint impact of monetary policy and market liquidity. XGBoost's tree structure is naturally suited to capture these high-order feature interactions, with each split point effectively processing the conditional probability distribution of features. Its split gain calculation formula is:

$$\text{Gain} = \frac{1}{2}\left[\frac{(\sum_{i\in L_L} g_i)^2}{\sum_{i\in I_L} h_i+\lambda} + \frac{(\sum_{i\in I_R} g_i)^2}{\sum_{i\in I_R} h_i+\lambda} - \frac{(\sum_{i\in I} g_i)^2}{\sum_{i\in I} h_i+\lambda}\right] - \gamma \tag{4-1-2}$$

where g_i and h_i are the first and second order gradients of the loss function respectively, I_L and I_R represent the instance sets of left and right nodes after splitting, λ is the L2 regularization term, and γ is the minimum gain needed for a split.

It can effectively identify and utilize these interaction effects. Moreover, financial market data generally contains noise and outliers, and XGBoost significantly enhances model robustness to noise through $\Omega(f_k)$ and column sampling mechanism. This characteristic is particularly valuable when predicting interest rates during periods of high market volatility.

4.2 LSTM Unit Structure

The advantages of LSTM in yield curve prediction stem from its grasp of time series characteristics, which highly aligns with the temporal properties of interest rates. Interest rates exhibit significant persistence characteristics, which is widely recognized in classical term structure theories. LSTM can effectively capture this persistence through its cell state update mechanism:

$$C_t = f_t \times C_{t-1} + i_t \times \tilde{C}_t \tag{4-2-1}$$

where C_t represents the cell state at time t, f_t is the forget gate output, C_{t-1} is the previous cell state, i_t is the input gate output, and $\tilde{C}_t$ is the candidate cell state.

The design of f_t allows the model to adaptively determine the importance of historical information, which highly corresponds with the autoregressive nature of interest rate formation. Moreover, the yield curve structure simultaneously contains short-term fluctuations and long-term trends. The gating mechanism can handle dependencies across different time scales:

$$h_t = o_t \times \tanh(C_t) \tag{4-2-2}$$

where the output gate o_t design allows the model to flexibly adjust weights between long-term and short-term information at different time points. From the perspective of Dynamic Term Structure Models, interest rate dynamics can be viewed as a state-space model. LSTM's hidden state update mechanism naturally aligns with this state-space representation but is not constrained by linear assumptions. Additionally, LSTM can simultaneously consider relationships between interest rates of different maturities, with its memory state storing information about the overall shape of the term structure, helping maintain consistency in predictions across different maturities.

4.3 Differences between XGBoost and LSTM

XGBoost and LSTM exhibit distinctive characteristics in predicting yield curve structures. XGBoost excels at extracting cross-sectional feature relationships, while LSTM specializes in capturing longitudinal time series patterns. Regarding modeling assumptions, XGBoost makes no explicit assumptions about the data generation process, instead discovering patterns through a data-driven approach; in contrast, LSTM inherently assumes sequential correlation in the data. This fundamental methodological difference results in distinct advantages under different market conditions.

In terms of prediction mechanisms, XGBoost constructs prediction rules through feature space partitioning, focusing on the hierarchical importance of various economic indicators. LSTM, however, generates predictions through state transmission, emphasizing the temporal evolution of interest rates. These contrasting prediction mechanisms reflect their different approaches to understanding yield curve dynamics. Compared to other machine learning methods such as random forests and conventional neural networks, each model offers unique strengths: XGBoost provides superior interpretability through its feature importance rankings and decision paths, while LSTM excels in capturing complex temporal dependencies that are crucial for interest rate dynamics.

While choosing between these models, It often depends on the specific aspects of yield curve prediction being prioritized - whether the focus is on capturing the impact of economic variables (where XGBoost may be more suitable) or modeling the temporal evolution of rates (where LSTM might be preferable).

4.4 Empirical Results

4.4.1 XGBoost Model Construction

For the XGBoost model, we adopted a similar data preparation approach, using the previous 10 days' factor values as features and the next day's factor value as the label. We set the initial number of training rounds to 100, but enable early stopping, halting training if the validation error fails to decrease for 10 consecutive rounds, to prevent overfitting.

In terms of hyperparameter tuning, we adjusted key parameters, including learning rate, max_depth, and subsampling ratio. We also use grid search and cross-validation on the training set to search for the optimal parameter combination that minimized validation error.

4.4.2 LSTM Model Construction

In implementing the LSTM model, we first applied Min-Max normalization to scale factor data values to the range (0, 1) to prevent saturation in the activation function. Then we set the hidden layer with 50 neurons, and set look_back to 10 based on AIC and BIC analysis and time series cross-validation, which means the model uses the factor values from the previous 10 days to predict the factor value for the next day.

We use MSE as the loss function and train the model with the Adam optimizer, setting the initial learning rate to 0.001. We apply early stopping to monitor the validation loss. The model will stop training if the validation loss does not decrease for 5 consecutive epochs in order to prevent overfitting. We set maximum number of training epochs to 100, with the batch size of 32 to balance computational efficiency and model convergence speed. To improve the model's generalization ability, we implemented grid search and cross-validation on the training set, we also tune key hyperparameters such as the number of units, learning_rate, and batch_size.

4.4.3 Reconstruction of the Yield Curve

During the model implementation, for each date in the test set, we first extracted the actual factor values and yield curve data, then used the trained LSTM and XGBoost models to predict the factor values for that date. With the predicted factor values, we reconstructed the yield curve over through the Nelson-Siegel function to generate the estimated yield curve. Also, we define a calculate_performance function to measure the MSE, RMSE and MAE between the predicted and actual yields.

Table 4-4-1: Predicting Performance Comparison

Maturity	Metric	LSTM	XGBoost	DNS
0.25Y	MAE	0.0232	0.0324	0.0026
	MSE	0.0009	0.0017	0.0394
	R^2	0.9415	0.8814	0.9485
0.5Y	MAE	0.0218	0.0394	0.0022
	MSE	0.0016	0.0023	0.0369
	R^2	0.9684	0.8963	0.9075
1.0Y	MAE	0.0108	0.0104	0.0012
	MSE	0.0015	0.0027	0.0269
	R^2	0.9776	0.9627	0.9165
2.0Y	MAE	0.0126	0.0039	0.0035
	MSE	0.0408	0.0423	0.0489
	R^2	0.9418	0.9235	0.9315
10.0Y	MAE	0.0116	0.0206	0.0012
	MSE	0.0028	0.0026	0.0300
	R^2	0.9787	0.9373	0.8772
20.0Y	MAE	0.0201	0.0286	0.0018
	MSE	0.0008	0.0014	0.0361
	R^2	0.9849	0.9012	0.9240
30.0Y	MAE	0.0338	0.0411	0.0065
	MSE	0.0332	0.0045	0.0681
	R^2	0.8945	0.8340	0.8793

While DNS demonstrates its theoretical advantages in short-term rate prediction, it is constrained by its linear assumptions in overall performance. LSTM demonstrates superior predictive performance across most maturities with better MAE and R^2 values, particularly in medium and long-term forecasts, showcasing its capability in complex market dynamics. Although XGBoost shows competitiveness at specific maturities, its predictive performance exhibits relatively higher volatility, indicating LSTM's more pronounced advantages in modeling the dynamic characteristics of the yield curve term structure.

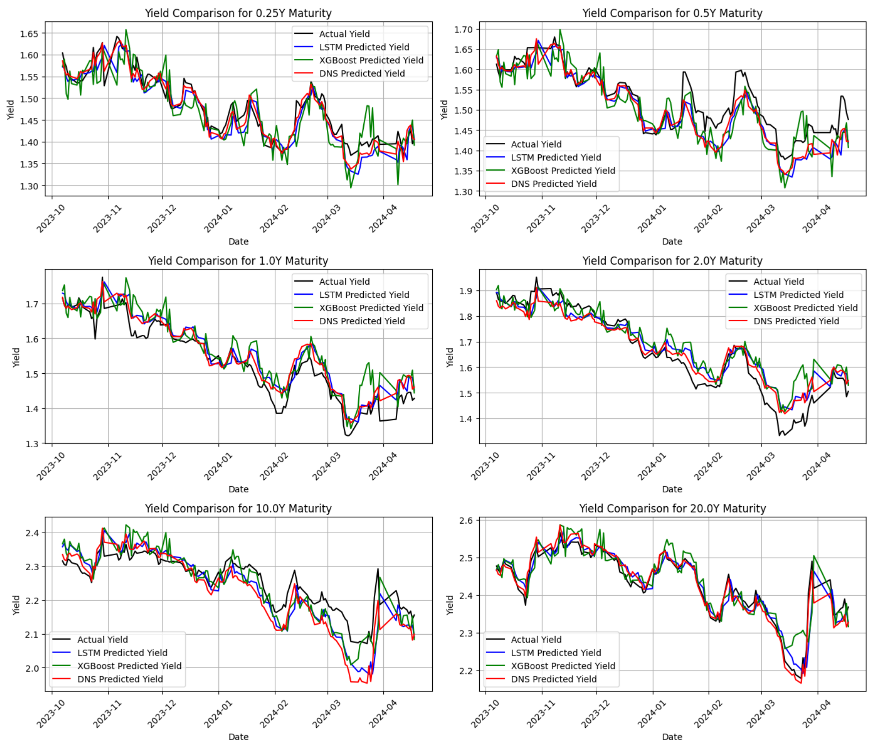

Figure 4-4-1: Predicting Performance Comparison of Specific Tenors

We can see that, from the comparison of the overall prediction performance, all three models show strong prediction ability, but there are still slight differences. the LSTM model shows the best performance in all the evaluation indexes, with the lowest Avg-MSE (0.0019), Avg-RMSE (0.0399), and Avg-MAE (0.0328) and the highest Avg-R^2 (0.9861).

However, the traditional DNS model is comparable to XGBoost in terms of its predictive ability, although its overall performance is slightly inferior to that of machine learning model, suggesting that the classical econometric approach is still of considerable practical value in the field of yield curve prediction.

Table 4-4-2: Predicting Performance Comparison with DNS

Metric	DNS	XGBoost	LSTM
MSE	0.0028	0.0027	0.0019
RMSE	0.0490	0.0472	0.0399
MAE	0.0422	0.0393	0.0328
R^2	0.9780	0.9801	0.9861

5 Conclusion

Our findings demonstrate that the DNS Model effectively captures the dynamic evolution of yield curves, where the state equation characterizes the autoregressive properties of factors, while the measurement equation establishes the mapping between yields and factors. This framework not only shows excellent statistical performance but, more importantly, provides clear economic interpretations - the level factor reflects long-term equilibrium, the slope factor embodies monetary policy expectations, and the curvature factor captures medium-term economic cycle fluctuations.

Theoretically, our study validates and extends traditional term structure theories through machine learning approaches. The LSTM model, with its gating mechanism and memory cell structure, can adaptively determine the importance of historical information, which is inherently similar to the autoregressive characteristics of factor dynamics in the DNS model but with enhanced nonlinear fitting capabilities. Meanwhile, XGBoost, through its tree ensemble structure and automatic discovery of feature interactions, complements our understanding of potential nonlinear and cross-maturity relationships in yield curves. Compared to the DNS model, while these machine learning methods show improvements in prediction accuracy, they sacrifice model interpretability and economic intuition. The findings that the level factor exhibits non-stationarity and high persistence (30th-order autocorrelation of 0.9277) confirm the structural characteristics of long-term interest rate formation, while the significant positive correlation (0.4365) between level and slope factors reveals the inherent mechanism of term premiums varying with interest rate levels.

From the methodological perspective, we optimize the decay parameter λ to 0.42625 through minimizing prediction MSE. The DNS model captures yield curve dynamics through AR(1) factor processes and exponential-form loadings $[1, \frac{1-e^{-\lambda\tau}}{\lambda\tau}, \frac{1-e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau}]$, providing theoretical foundations for cross-maturity risk premia formation. For the same factor prediction task, LSTM enhances the modeling of long-term time series dependencies through its gating mechanism and memory cell structure, while XGBoost captures non-linear interactions between yields through its tree ensemble structure. This theoretical comparison of different prediction methods not only deepens our understanding of yield curve dynamics but also suggests new directions for developing more comprehensive term structure theories. Given the advantages demonstrated by machine learning methods in predicting short-term and long-term yields, our future research could explore incorporating alternative factor dynamics normalizations (such as higher-order AR processes) within the DNS framework and dynamic adjustments of λ under different market conditions to enhance model performance in extreme market environments.

References

[1] Nelson, C.R. and Siegel, A.F. (1987), "Parsimonious Modeling of Yield Curves," *The Journal of Business*, Vol. 60, No. 4, pp. 473–489.

[2] Diebold, F.X. and Li, C. (2006), "Forecasting the term structure of government bond yields," *Journal of Econometrics*, Vol. 130, No. 2, pp. 337–364.