Forecast Yield Curve of China's Government Bond with Machine Learning

Yadong Cao

Abstract

This paper takes the interest rates of China's government bond as the research object, and demonstrates the feasibility of the yield curve factor model by analyzing the Bootstrap, DNS, AFNS model of yield curve. Then based on the factor construction, construct the yield curve factor. In order to make the interest rate forecast, constructed 182 features and mined the features affecting the yield curves from multiple dimensions. Then used the traditional statistical models and machine learning models to predict yield curve. Through empirical analysis, it is found that machine learning models and DNS models can construct an ideal framework for predicting yield curve. Based on the analysis and research in this paper, we propose a new framework for yield curve forecast.

Contents

1 Introduction

The bond market is an important part of the capital market, and the development of the bond market can help corporate and government financing. It also helps improve the financing structure, increase the stability of the financial system. In China, the history of the bond market development is not long. In 1984, China started a bond issue. From then on, the bond market has continuously extended. The bond market promote the internationalization of the RMB and the reform of the financing mode of Chinese enterprises. So it’s very important to study the bond market.

From 2008 to 2017, Chinese bond market growth has been very rapid and the US bond market has developed very steadily. In 2008, the Chinese bond market capitalization over GDP rose from 35%, and in 2017, the ratio to more than 90%. For comparison, the US bond market capitalization over GDP keeps above 200%. With the development of China’s economy, we can know the size and scope of the bond market will increase(Marlene Amstad and Zhiguo He, 2019)[22].

Compared to the stock market capitalization with the bond market, the Chinese bond market and stock market all increase. From 2017, the China Bond/Stock has 90%, but in 2017, the ratio increased to 130%. In the US, the ratio is 260%, and in 2017, the bond/stock is 125%. The US bond/stock ratio has decreased, it says the US capital markets have shown contraction. China’s capital markets have shown a booming development trend in terms of volume and growth.

Even though China’s bond market achieved great success, it still has some risks. For example, the interest rates spread is very large fluctuations. From 2017 to 2022, the interest rate spread(2yrs vs. 10yrs) minimum value is 12bp, maximum value is 147 bp, average value is 53.75 bp and the standard deviation value is 23.36 bp.

And the yield volatility(such as 10yr bonds) is also very volatile. It has brought a very negative impact on the country, enterprises and individuals. It is difficult to form a reasonable expectation for the future, do a good job in financial planning.

With the rapid development of China’s bond market, it is very important to study the interest rate forecast for China’s financial market. Firstly, yield curve forecast can promote interest rate liberalization and help the central bank and government financial institutions to formulate appropriate monetary and fiscal policies from the macro-control level. Secondly, yield curve forecasting is of great significance for exploring China’s fixed income market. It can explore China’s interest rate transmission mechanism and interest rate market from an objective and empiric perspective. Thirdly, yield curve forecast can help financial institutions to make good financial plans and investment strategies to effectively avoid risks. To prevent a liquidity crisis caused by maturity mismatch.

Acknowledging the yield curve importance, and artificial intelligence brings the research paradigm revolution, the thesis work on a yield curve forecasting framework with artificial intelligence method. Because different fixed income products, such as Treasury bonds, interbank certificates of deposit, policy financial bonds, short and medium term bills, can be expressed by the interest rate term structure(that is, the yield curve). Structurally, these fixed income products have the same structural form. Therefore, the methodology for predicting the yield curve of some fixed income product can also be applied to other fixed income products. We want to build a general framework for forecasting yield curve, and here we use China’s government bond as an example, because these bonds are the most common of all fixed income products. It fits the requirements of our analysis based on a general yield curve forecast framework. The yield curve forecast framework can be applied to the interest rate analysis and forecasting in the fixed income market, and further serve the investment strategy and risk management in the fixed income market.

In our research, there are four parts. First, we start from descriptive statistical analysis, based on the interest rates of maturities and yield spread of China’s government bond, we find some facts from the characteristics of volatility and autocorrelation. For example, there is a certain degree of autocorrelation in interest rate, and the volatility decreases with the period. Second, we use principal component analysis to demonstrate the feasibility and interpretability of the yield curve factor model, so as to prove the DNS(Dynamic Nelson-Siegel model, Nelson and Siegel, 1987)[8] and AFNS(Arbitrage-Free Nelson-Siegel, Diebold et al. 2005; Christensen et al. 2011a)[16] are feasible in the construction of yield curve of China's government bond. Thirdly, we use DNS and AFNS models to build yield curves, and use yield curve constructed by Bootstrap method as the benchmark to analyze the error between the model value and the real value, and prove that both DNS and AFNS perform better in yield curve fitting. Moreover, due to the acceptability of factor mode and arbitrage free equilibrium analysis in financial practice, DNS and AFNS have great advantages in financial analysis. Finally, we use the traditional statistical method and machine learning method to predict the factors of the yield curve model, and realize the prediction of the yield curve through the prediction of the factors. The best combination of yield curve model and forecasting method is obtained by comparing forecasting error. Through the research, we found that the combination of DNS yield curve model and machine learning mix model can form a generally yield curve forecasting framework.

Through this research, we want to empirically prove the feasibility of a generalized financial analysis model based on artificial intelligence. The model and forecasting method of yield curve in this thesis can be applied to any fixed income asset of the same type. Next, there are few studies on China's fixed income yield curve at present. These yield curve models and forecasting methods used in this thesis provide a new methodology for the interest rate analysis of China's financial market, which can widely serve the financial transactions and risk management of the inter-bank market, and promote the RMB liquidity and interest rate marketization.

The structure of this thesis is as follows: Chapter 2 is the literature review, which introduces the interest rate involved in this paper (different types of interest rates and their relationships), interest rate model (Bootstrap model, DNS, AFNS, etc.), and prediction methods (traditional statistical methods, machine learning methods, deep learning methods). Chapter 3 systematically introduces our research objectives.

Chapter 4 builds the specific methods and models we use in this paper. Chapter 5 introduces our data sources, the data we use, and the methods of data processing. In Chapter 6, we analyze our empirical results and get a general yield curve forecast framework. Chapter 7 is the application part. The methodology and framework constructed in this paper are applied in the fields of financial engineering, investment and risk management. Chapter 8 is the final conclusion.

2 Literature Review

2.1 Interest Rate and Yield Curve

Usually, the interest rate has par rate, spot rate, and forward rate. The par rate is the coupon rate at the time the bond was issued. A coupon bond can be viewed as a bundle of zero-coupon bonds(zeros)(Solomon Brothers, 1995)[1].The spot rate is the discount rate of a single future cash flow such as a zero-coupon bond. Equation(1) shows the relation between an n-year zero's price P_n and the annualized n-year spot rate S_n.

$$P_n = \frac{100}{(1+S_n)^n} \tag{1}$$

Forward rate is an interest rate applicable to a financial transaction that will take place in the future. A given term structure of spot rates implies a specific term structure of forward rates. If the m-year and n-year spot rates are known, the annualized forward rate between maturities m and n, $f_{m,n}$is easily computed from Equation(2).

$$(1+f_{m,n})^{n-m} = \frac{(1+S_n)^n}{(1+S_\mathrm{m})^m} \tag{2}$$

A plot of the yields on bonds with differing terms to maturity but the same risk, liquidity, and tax considerations is called a yield curve, and it described the term structure of interest rates for particular types of bonds, such as government bonds(Lin Chen, 1996)[3]. Figure 1 shows the different types of yield curves, which show the shape of different yield curves.

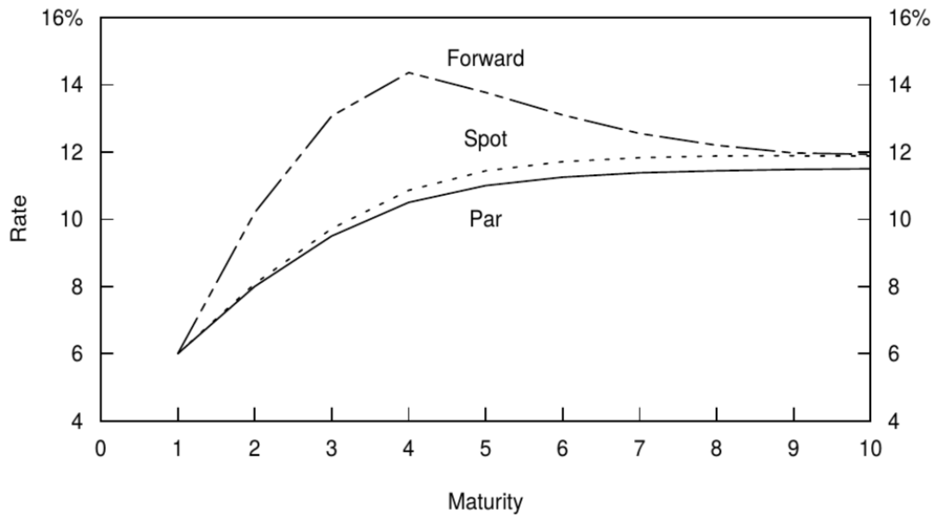


Figure 1. Par, Spot and One-year Forward Rate Curves

Every day, par, spot and forward rate curves would be published by some organization for fixed income market analysis. It present different types of yield curves on the same day have different shapes. Data source: Understanding the Yield curve[1].

The interest rate curve has the discount curve, the forward curve, and the yield curve. Let $P_{(\tau)}$ denote the price of a τ period discount bond, that is , the present value of $1 receivable τ periods ahead(Francis X. Diebold and Glenn D. Rudebusch, 2013)[8]. If $y(\tau)$ is its continuously compounded yield to maturity, then by definition

$$P(\tau) = e^{-\tau y(\tau)} \tag{3}$$

The forward rate curve can get from discount curve,

$$f(\tau) = \frac{-P'(\tau)}{-P(\tau)} \tag{4}$$

The yield curve can get from forward rate curve,

$$y(t) = \frac{1}{\tau}\int_0^\tau f(\mu)d\mu \tag{5}$$

For the discount curve, the forward curve and the yield curve, if we get one of them, we get the other two. Usually, we often use the yield curve to discuss. The yield curve will change over time, showing the trend of the time series. When we study yield curve forecast, we mainly study the change of interest rate curve. Then we can get interest rate from yield curve by tenor. Figure 2 shows the changed of yield curve from a three-dimensional dimension.

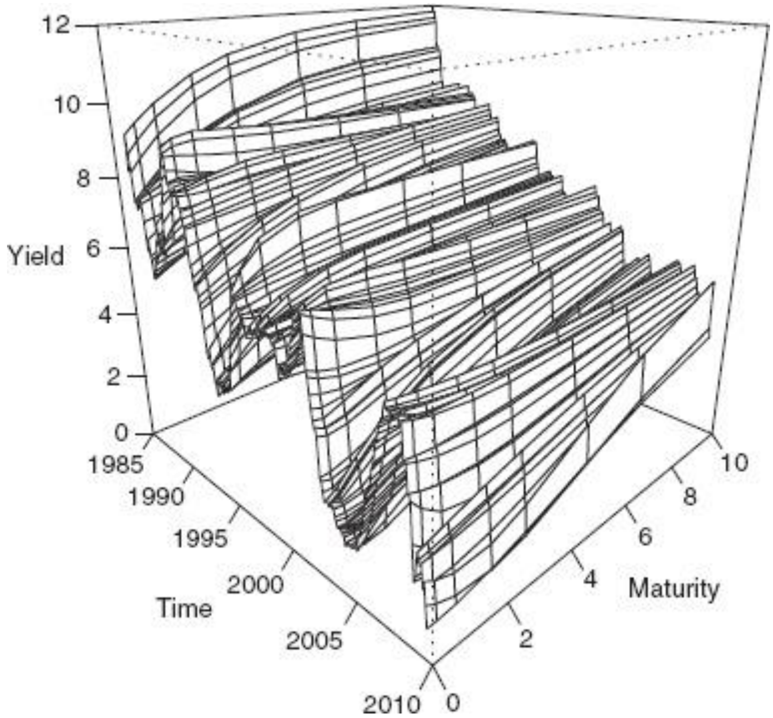


Figure 2. Bond Yield Curves in Three Dimensions.

It constructs three dimensions surface by date, maturity and yield. Source: Yield Curve Modeling And Forecasting[8].

2.2 Yield Curve Modeling

Yield curve models have three types, such as regression-type models, empirical models, and equilibrium models. Regression modeling is the method of fitting a yield curve, where a function is simply fitted to the yields to maturity of regular coupon-paying bonds. One of the main problems with the regression-type models is that the “coupon effect” is not taken into account. The coupon effect refers to the fact that different bonds with the same term to maturity may have very different yields to maturity because of differences in their coupon rates.

Empirical yield curve models usually specify a functional form for the discount function. The models often have two types, for example, Bootstrap method and

Nelson-Siegel method. Equilibrium models that are discussed have closed-form solutions and can be fitted to bond prices in a similar way to empirical yield curve models. These models are usually not flexible enough in practice to provide an adequate fit the underlying instrument used to derive the yield curve. In our situation, we mainly use empirical yield curve models. The models are cubic spline model, dynamic Nelson-Siegel model(DNS), and arbitrage-free Nelson-Siegel model(AFNS).

2.2.1 Cubic Spline Model

Say we have N bonds available to derive a yield curve and we let m_N denote the term to maturity of the longest-term bond(in years). We divided the maturity range $[0，\mathrm{m}_N]$ into subintervals by specifying k knot points $\kappa_1, \kappa_2, ..., \kappa_k$ such that $\kappa_1 = 0$ and $\kappa_k = \mathrm{m}_N$. A separate function for df_m is then fitted to each subinterval.

The cubic spline model is defined in terms of basis functions and is given by (McCulloch, 1975; Anderson et al., 1997)[2]:

$$f_j(m) = \begin{cases} 0 & \text{for } m < \kappa_{j-1} \\ \dfrac{(m-\kappa_{j-1})^3}{6(\kappa_j - \kappa_{j-1})} & \text{for } \kappa_{j-1} \le m < \kappa_j \\ \dfrac{c^2}{6} + \dfrac{ce+e^2}{2} + \dfrac{e^3}{6(\kappa_{j+1} - \kappa_j)} & \text{for } \kappa_j \le m < \kappa_{j+1} \\ (\kappa_{j+1} - \kappa_{j-1})\left[\dfrac{2\kappa_{j+1} - \kappa_j - \kappa_{j-1}}{6} + \dfrac{m - \kappa_{j+1}}{2}\right] & \text{for } \kappa_{j+1} \le m \end{cases} \tag{6}$$

When $j < k$ and with c = $\kappa_j - \kappa_{j-1}$ and $e = m - \kappa_j$. When $j = k$ we have that $f_j(m) = m$ for all m. k denotes the chosen number of knot points.

2.2.2 Nelson-Siegel Model

Nelson and Siegel(1987) begin with a forward rate curve and fit the function from a large set of yields(Francis X. Diebold and Glenn D. Rudebusch, 2013)[8]. The function is,

$$f(\tau) = \beta_1 + \beta_2 e^{-\lambda\tau} + \beta_3 \lambda\tau e^{-\lambda\tau} \tag{7}$$

The corresponding static Nelson-Siegel yield curve is

$$y(\tau) = \beta_1 + \beta_2 \left(\frac{1 - e^{-\lambda\tau}}{\lambda\tau} \right) + \beta_3 \left(\frac{1 - e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau} \right) \tag{8}$$

Even though Nelson-Siegel functional form is not better than bootstrapping method(such as cubic spline model) for approximating an arbitrary yield curve. But Nelson-Siegel turns out to have some very appealing features. First, it desirably enforces some basic constraints from financial economic theory. Second, the Nelson-Siegel form provides a parsimonious approximation. Third, the Nelson-Siegel form also provides a flexible approximation.

2.2.3 Dynamic Nelson-Siegel model

Diebold and Li (2006) found that the Nelson-Siegel parameters must be time-varying if the yield curve is to be time-varying[8]. This leads to dynamic Nelson-Siegel model.

$$y_t(\tau) = \beta_{1t} + \beta_{2t} \left(\frac{1 - e^{-\lambda\tau}}{\lambda\tau} \right) + \beta_{3t} \left(\frac{1 - e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau} \right) \tag{9}$$

This is a time-series linear projection of y_t on variables $\beta_{1t}, \beta_{2t}, \beta_{3t}$ with parameters $(1,\ ((1 - e^{-\lambda\tau}) / \lambda\tau),\ ((1 - e^{-\lambda\tau}) / \lambda\tau - e^{-\lambda\tau}))$.

2.2.4 Arbitrage-Free Nelson-Siegel Model

Suppose that the instantaneous risk-free rate is(Francis X. Diebold and Glenn D. Rudebusch, 2013)[16]

$$r_t = X_t^1 + X_t^2 \tag{10}$$

Where the state variables $X_t = (X_t^1, X_t^2, X_t^3)$ have risk-neutral(Q) dynamics:

$$\begin{pmatrix} dX_t^1 \\ dX_t^2 \\ dX_t^3 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & \lambda & -\lambda \\ 0 & 0 & 0 \end{pmatrix} \left[\begin{pmatrix} \theta_1^Q \\ \theta_2^Q \\ \theta_3^Q \end{pmatrix} - \begin{pmatrix} X_t^1 \\ X_t^2 \\ X_t^3 \end{pmatrix} \right] dt + \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{pmatrix} \begin{pmatrix} dW_t^{1,Q} \\ dW_t^{2,Q} \\ dW_t^{3,Q} \end{pmatrix} \tag{11}$$

Then zero-coupon bond prices are

$$\begin{aligned} \mathrm{P}(t,T) &= E_t^Q(\exp(-\int_t^T r_u du)) \\ &= \exp(B^1(t,T)X_t^1 + B^2(t,T)X_t^2 + B^3(t,T)X_t^3 + \mathrm{C}(t,T)) \end{aligned} \tag{12}$$

Where $B^1(t,T)$, $B^2(t,T)$, $B^3(t,T)$, and $\mathrm{C}(t,T)$ are governed by the ODEs:

$$\begin{pmatrix} \frac{dB^1(t,T)}{dt} \\ \frac{dB^2(t,T)}{dt} \\ \frac{dB^3(t,T)}{dt} \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & -\lambda & \lambda \end{pmatrix} \begin{pmatrix} B^1(t,T) \\ B^2(t,T) \\ B^3(t,T) \end{pmatrix} \tag{13}$$

$$\begin{aligned} \frac{dC(t,T)}{dt} &= -B(t,T)'^Q \theta^Q \\ &\quad - \frac{1}{2} \sum_{j=1}^{3} (\Sigma' B(t,T) B(t,T)' \Sigma)_{j,j} \end{aligned} \tag{14}$$

(with boundary conditions $B^1(T,T) = B^2(T,T) = B^3(T,T) = C(T,T) = 0$), with solution

$$
\begin{aligned}
B^1(t,T) &= -(T-t) \\
B^2(t,T) &= -\frac{1-e^{-\lambda(T-t)}}{\lambda}, \\
B^3(t,T) &= (T-t)e^{-\lambda(T-t)} - \frac{1-e^{-\lambda(T-t)}}{\lambda}, \\
C(t,T) &= (K^Q\theta^Q)_2 \int_t^T B^2(s,T)ds \\
&\quad + (K^Q\theta^Q)_3 \int_t^T B^3(s,T)ds \\
&\quad + \frac{1}{2}\sum_{j=1}^{3} \int_t^T \left(\Sigma' B(s,T)B(s,T)'\Sigma\right)_{j,j} ds.
\end{aligned} \tag{15}
$$

Hence zero-coupon bond yields are

$$
\begin{aligned}
y(t,T) &= X_t^1 + \frac{1-e^{-\lambda(T-t)}}{\lambda(T-t)} X_t^2 \\
&\quad + \left[\frac{1-e^{-\lambda(T-t)}}{\lambda(T-t)} - e^{-\lambda(T-t)}\right] X_t^3 - \frac{C(t,T)}{T-t}.
\end{aligned} \tag{16}
$$

2.2.5 Reasons of Factor Models for Yield Curve.

Factor Models for yields prove appealing for three key reasons.

First, factor structure generally provides a high accurate empirical description of yield curve data. Because only a small number of systematic risks appear to underlie the pricing of the myriad of tradeable financial assets, nearly all bond price information can be summarized with just a few constructed variables or factors. Therefore, yield curve models almost invariably employ a structure that consists of a small set of factors and the associated factor loadings that relate yields of different maturities to those factors.

Second, factor models prove tremendously appealing for statistical reasons. They provide a valuable compression of information, effectively collapsing an intractable high-dimensional modeling situation into a tractable low-dimensional situation. This would be small consolation if the yield data were not well approximated with factor structure, but again, they are. Hence we’re in a most fortunate situation. We need low-dimensional factor structure for statistical tractability, and, mercifully, the data actually have factor structure.

Last, financial economic theory suggests factor structure. Many financial assets in the markets, but for a variety of reasons we view the risk premiums that separate their expected returns as driven by a much smaller number of components, or risk factors. Yield curve factor models are a natural bond market parallel.

2.3 Forecast Models

Forecast models usually have statistical models, statistical learning models, and deep learning models. We used to use statistical models to predict time series. However, with the development of big data technology and artificial intelligence(especially machine learning and deep learning) in recent years, statistical learning and deep learning have more and more advantage in the prediction of time series.

2.3.1 Forecast with Statistical Model

Statistical models have these series types, such as MA(q) model(moving average process), AR(p) model(autoregressive process), ARMA(p, q) model(autoregressive moving average), ARIMA(p, d, q) model(autoregressive integrated moving average model), and SARIMA(p, d, q)(P, D, Q)m for seasonal time series(the seasonal autoregressive integrated moving average model)(Marco Peixeiro, 2022)[23]. In our situation, the yield curve does not have a seasonal trend. So we use ARIMA(p, d, q) with external variables model.

An autoregressive integrate moving average(ARIMA) process is the combination of the AR(p) and MA(q) process, but in terms of the difference series. It is denoted as ARIMA(p, d, q), where p is the order of the AR(p) process, d is the order of integration, and q is the order of the MA(q) process. Integration is the reverse of differencing, and the order of integration d is equal to the number of times the series has been differenced to be rendered stationary. The general equation of the ARIMA(p, d, q) process is

$$y'_t = C + \varphi_1 y'_{t-1} + \ldots + \varphi_p y'_{t-p} + \theta_1 \varepsilon'_{t-1} + \ldots + \theta_q \varepsilon'_{t-q} + \varepsilon_t \tag{17}$$

Add any number of exogenous variables X_t ,

$$y'_t = C + \varphi_1 y'_{t-1} + \ldots + \varphi_p y'_{t-p} + \theta_1 \varepsilon'_{t-1} + \ldots + \theta_q \varepsilon'_{t-q} + \varepsilon_t + \sum_{i=1}^{n} \beta_i X_t^{i} \tag{18}$$

2.3.2 Forecast with Statistical Learning

For statistical learning models, some models excels at extrapolating trends, but cannot learn interactions. Some models excels at learning interactions, but cannot extrapolate trends. So we need create hybrid forecasters that combine complementary learning algorithms and let the strengths of one make up for the weakness of the other. The forecast series has four components,

$$series = trend + seasons + cycles + error \ .$$

In order to forecast with hybrid models, we should have these steps,

First, create time series as features. Decompose time series to trend, seasons, and cycles to get new features. Let trend feature is T, seasons feature is S， and cycles feature is C.

Second, create external variable as features, let them is X.

Third, use the model that is excels at extrapolating trends, let the model is M_1 , use the model that is excels at learning interactions, let the mode is M_2 .

For M_1,

$$M_1.fit([X,T,S,C],Y)$$

$$Y_{M1} = M_1.predit([X,T,S,C])$$

For M2,

$$M_{2.}.fit([X,T,S,C],Y - Y_{M1})$$

$$Y_{M2} = M_2.predit([X,T,S,C],Y - Y_{M1})$$

The predictions is,

$$Y_{pred} = Y_{M1} + Y_{M2}$$

Linear regression models excels at extrapolating trends. Let input vector is $\mathrm{X}^T = (X_1, X_2, ..., X_n)$, and the output is Y.The linear regression model has the form

$$f(X) = \beta_0 + \sum_{j=1}^{p} X_j \beta_j$$

The linear model either assumes that the regression function $E(Y|X)$ is linear, or that the linear model is reasonable approximation. The β_j's are unknown parameters or coefficients, ant the variables X_j can come from different sources:

1. quantitative inputs;
2. transformations of quantitative inputs, such as log, square-root or square;
3. basic expansions, such as $X_2 = X_1^2$, $X_3 = X_1^3$, leading to a polynomial representations;
4. numeric or “dummy” coding of the levels of qualitative inputs. For example, if G is five-level factor input, we might create X_j, j =1,…,5, such that $X_j = I(G = j)$.Together this group of X_j represents the effect G by a set of level-dependent constants, since in $\sum_{j=1}^{5} X_j \beta_j$, one of the X_j s is one, the others are zero.

5. Interactions between variables, for example, $X_3 = X_1 \cdot X_2$.

Boosting methods excels at learning interactions, Freund and Schapire (1997) put forward “AdaBoost.M1” model. Consider a two-class problem, with the output variable coded as $Y \in \{-1,1\}$. Given a vector of predictor variables X, a classifier $G(X)$ produces a prediction taking one of the two values $\{-1,1\}$. The error rate on the training sample is

$$\overline{err} = \frac{1}{N}\sum_{i=1}^{N} I(y_i \neq G(x_i))$$

And the expected error rate on the future predictions is $E_{XY} I(Y \neq G(X))$.

A week classifier is one whose error rate is only slightly better than random guessing. The purpose of boosting is to sequentially apply the weak classification algorithm to repeatedly modified versions of the data, thereby producing a sequence of weak classifiers $G_m(x), m = 1, 2, ..., M$ (Tianqi Chen and Carolos Guestrin, 2016)[37].

Figure 3 shows schematic of AdaBoost. Classifiers are trained on weight versions of the dataset, and then combine to produce a final prediction.

The predictions from all of the them are then combined through a weighted majority vote to product the final prediction:

$$G(x) = sign(\sum_{m=1}^{M} a_m G_m(x))$$

Here $\alpha_1, \alpha_2, ..., \alpha_M$ are computed by the boosting algorithm, and weight the contribution of each respective $G_m(x)$. Their effect is to give higher influence to the more accurate classifiers in the sequence.

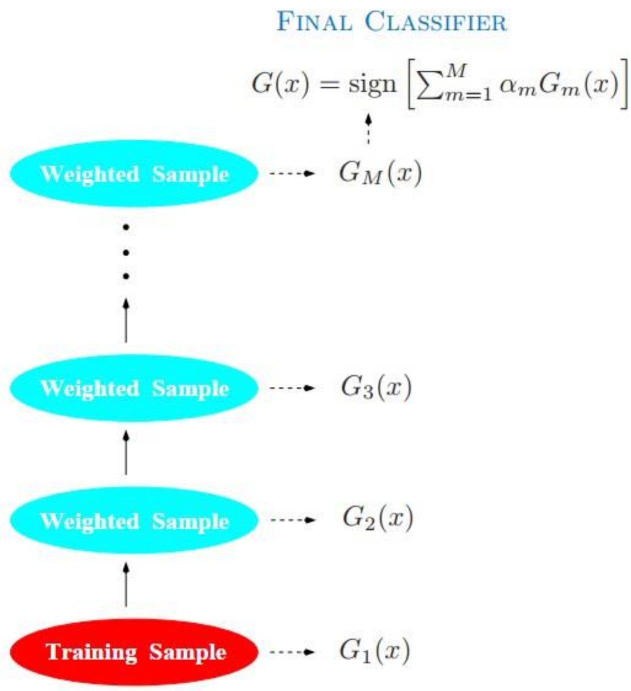


Figure 3. Schematic of AdaBoost.

Classifiers are trained on weighted versions of the dataset, and then combined to produce a final prediction. Source: The Elements of Statistical Learning.

2.3.3 Forecast with Deep Learning

Time series forecasting models predict future values of a target $y_{i,t}$ for a given entity i at time t. Each entity represents a logical grouping of temporal information. In the simplest case, one-step-ahead forecasting models take the form:

$$\hat{y}_{i,t+1} = f(y_{i,t-k:t}, x_{i,t-k:t}, s_i) \tag{19}$$

Where $\hat{y}_{i,t+1}$ is the model forecast, $y_{i,t-k:t} = \{y_{i,t-k}, ..., y_{i,t}\}$, $x_{i,t-k:t} = \{x_{i,t-k}, ..., x_{i,t}\}$ are observations of the target and exogenous inputs respectively over a look-back window k, s_i is static metadata associated with entity, and $f(.)$ is the prediction function learnt by the model. Such as long short-term memory model is very general time series forecast model. LSTM is a deep learning architecture that is a subtype of RNN. LSTM addressed the problem of short-term memory by adding the cell state. This allows for past information to flow through the network for a long period of time, meaning that the network still carries information from early values in the sequence. LSTM is made up of three gates, forget gate, input gate, and output gate.

Input gate: $i_t = \sigma\left(W_{i1}z_{t-1} + W_{i2}y_t + W_{i3}x_t + W_{i4}s + b_i\right)$,

Output gate: $o_t = \sigma\left(W_{o1}z_{t-1} + W_{o2}y_t + W_{o3}x_t + W_{04}s + b_o\right)$

Forget gate: $f_t = \sigma\left(W_{f1}z_{t-1} + W_{f2}y_t + W_{f3}x_t + W_{f4}s + b_f\right)$

Where z_{t-1} is the hidden state of the LSTM, and $\sigma(.)$ is the sigmoid activation function.

2.4 Yield Curve Forecasting Models

Yield curve forecasting models have three types: factor forecast models and functional principal component model. Factor forecast models base on Nelson-Siegel factor models, split the yield curve into factors, then predict the factors to realize yield curve forecast. Functional principal component models use the principal component analysis of the yield curve get yield curve FPCA factors. Then use FPCA model analysis the feature data which effect yield curve get feature data FPCA factors. Then combine yield curve FPCA factors and feature data FPCA factors, by predict feature FPCA feature data to forecast yield curve FPCA factors.

2.4.1 Factor Forecast Models

Diebold and Li (2005) based on the Nelson-Siegel model and statistical model make forecast yield curve. They model and forecast the Nelson-Siegel factors as univariate AR(1) processes. The AR(1) models can be viewed as natural benchmarks determined a priori: the simplest great workhouse autoregressive models. The yield forecasts based on underlying univariate AR(1) factor specifications are:

$$y_{t+h/t}(\tau) = \beta_{1,t+h/t} + \beta_{2,t+h/t}\left(\frac{1-e^{-\lambda\tau}}{\lambda\tau}\right) + \beta_{3,t+h/t}\left(\frac{1-e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau}\right) \tag{20}$$

Where

$$\beta_{1,t+h/t} = \hat{c}_i + \lambda_i \beta_{it}, i = 1, 2, 3,$$

$\hat{c}$ and λ_i are obtained by regressing β_{it} on an intercept and $\beta_{i,t-h}$.

For comparison, they also produce yield forecasts based on an underlying multivariate VAR(1) specification, as

$$y_{t+h/t}(\tau) = \beta_{1,t+h/t} + \beta_{2,t+h/t}\left(\frac{1-e^{-\lambda\tau}}{\lambda\tau}\right) + \beta_{3,t+h/t}\left(\frac{1-e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau}\right) \qquad (21)$$

Where

$$\beta_{t,t+h/t} = \hat{c} + \Gamma\beta_t$$

2.4.2 Functional Principal Component Analysis Model

Feng and Qian (2017)[38] use functional principal component analysis Chinese term structure of interest rates. First, they use the B-spline basis transform the discrete data into functional data by smoothing. Second, they estimate the eigenfunctions(FPCs) and the corresponding principal component scores for function data. Given the functional time series (f_t) estimate the mean curve μ by $\mu = 1/T\sum_{t=1}^{T} f_t$. The first FPC ξ_1 can be obtained by solving the following optimization problem:

$$\max_{\xi} 1/T\sum_{t=1}^{T} <\xi, f_t - \mu > 2, \|\xi\|^2 = 1$$

The k th FPC ξ_k can be obtained by solving:

$$\max_{\xi} 1/T\sum_{t=1}^{T} <\xi, f_t - \mu > 2$$

s.t.:

$$\|\xi\|^2 = 1, < \xi, \xi_j \geq 0, j = 1, \ldots, k-1$$

Approximate f_t using the first K FPCs:

$$f_t \approx \mu + \sum_{k=1}^{K} \beta_{t,k} \hat{\xi}_k$$

Compare the statistical models, such as AR, ARIMA, the machine learning models have absolute advantages.

First, the statistical models can realize the forecast results, and machine learning can also do. The statistical models are conventional linear regression models. Machine learning has many other methods that can achieve the effects than linear regression models, such as Ridge regression, Boosting methods, Neural networks, and so on.

Second, machine learning can provide more flexibility to forecast. In the machine learning methods, you can use single model and also use hybrid model. These models can provide some new viewpoints on the data.

Third, machine learning method can support more large data and more large calculations. When you use the statistical model, if you input big data and run the models, it would cost many times. But the machine learning method can run fast.

3 Research Objectives

In this theirs our aim is to build a yield curve forecast framework, take China's government bond as an example. In particular, the empirical section of this thesis is formed of three parts on quite unrelated problems.

The first analysis uses descriptive statistical analysis, we found some facts of the interest rates of China's government bond. Such as the movement trend, volatility, autocorrelation, and interpretability. From the analysis, we found we can use factor model view to construct yield curve.

In the second part, we use principal component analysis to analyze China's Government bond. We want to research the factor model is fitted yield curve. We mainly explore the first three factors how can explain the interest rates.
The third part is to use DNS and AFNS to build yield curve, and get yield curve factors from these models. On the point of view, factors of one day can present the yield curve of one day. In a date interval, we can get factors of time series.

Finally, we use statistical models and machine learning methods to forecast the factors, then from factors of predict to get yield curve of forecast. From errors of the real value and forecast model value, we can explore which combination method(yield curve model and predict method) is the best.

From the above works, the framework which can predict the yield curve will be presented.

4 Methodology

We make yield curve forecast mainly based on the yield curve factor models(the dynamic Nelson-Siegel model and the arbitrage-free Nelson-Siegel model). In the yield curve forecast based on factor model, many people used statistical models, such as AR(p), VAR(p). Our research uses the machine learning model represented by statistical learning method, and based on the time series prediction scenario, combined with the characteristics of different statistical methods of learning models to present hybrid model. Here we do not consider the prediction of deep learning, because we forecast based on the daily frequency yield curve of China's government bond, which does not meet the requirements of deep learning in terms of data magnitude. Our research combines the most advanced interest rate curve model research and artificial intelligence method forecasting practice on time series.

Our objects:

1. We will discuss the advantages and disadvantages of the dynamic Nelson-Siegel model in forecasting the yield curve of China's government bond.
2. We will discuss the advantages and disadvantages of statistical model and machine learning model in forecasting the yield curve of China's government bond.
3. Present a better framework of yield curve forecasting for China's government bond compared with the current method of yield curve forecast.
4. Based on the yield curve forecast, make risk management methods and investment strategies, and apply to other financial practices.

The specific methods are:

1. Calculate the optimal parameters of the models. For the dynamic Nelson-Siegel model, we will calculate factor loadings, get the λ_t value. In time series, we will get the $(\lambda_1, \ldots, \lambda_N)$. Let λ is constant value, and $\lambda = average(\lambda_1, \ldots, \lambda_N)$. For the arbitrage-free Nelson-Siegel model, we use initial parameters,

$$P_{AFNS}(Initial) = (\kappa_1, \kappa_2, \kappa_3, \sigma_{11}, \sigma_{21}, \sigma_{22}, \sigma_{31}, \sigma_{32}, \sigma_{33}, \theta_1, \theta_2, \theta_3)$$

Input the AFNS model and get optimal parameters.

$$P_{AFNS} = (\kappa_1, \kappa_2, \kappa_3, \sigma_{11}, \sigma_{21}, \sigma_{22}, \sigma_{31}, \sigma_{32}, \sigma_{33}, \theta_1, \theta_2, \theta_3)$$

2. Use the optimal parameters to get yield curve model. The dynamic Nelson-Siegel model, the factors are:

$$F_t^{\mathrm{DNS}} = \left(\beta_{t,1}, \beta_{t,2}, \beta_{t,3}\right)$$

The arbitrage-free Nelson-Siegel model factor, the factors are:

$$F_t^{\mathrm{AFNS}} = \left(\mathrm{Level}_t, Slope_t, Curvature_t\right)$$

3. Construct features from the interest rate transmission, monetary policy, fiscal policy, economic and financial, factor tend and other dimensions to construct characteristics. These features can be divided into two main categories, internal features(X_{in}) and external features(X_{ex}).

$$(X_t^{\mathrm{in}}, X_t^{ex}) \to (F_{1,t}, F_{2,t}, F_{3,t})$$

4. Use single forecast model, for statistical model use ARIMA(p, d, q), and for machine learning method use hybrid model. For machine learning method, it needs let factors lag N day so that it can forecast the date after N days.

$$(X_t^{\mathrm{in}}, X_t^{ex}) \to (F_{1,t+N}, F_{2,t+\mathrm{N}}, F_{3,t+N})$$

5. Use predict factors data input model and get key tenor interest rate, then validate the actual value. The paper uses Mean Square Error(MAE), Mean Absolute Percentage Error(MAPE), Root Mean Square Error(RMSE) to compare the actual value and forecast value.

$$\left(F_{1,t}, F_{2,t}, F_{3,t}\right) \xrightarrow{Yield\ \mathrm{Curve\ Model}} Yield\ \mathrm{Curve} \to Y$$

Let the actual value is:

$$y = \{y_1, y_2, ..., y_{\mathrm{n}}\}$$

The predicted value is:

$$y = \left\{y_1, y_2, ..., y_{\mathrm{n}}\right\}$$

The Mean Square Error is:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - y_i\right|$$

MSE in $[0,+\infty)$, the MAE close to 0 for the perfect model ; The bigger the MAE value, the bigger the error.

The Mean Absolute Percentage Error is:

$$MAPE = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{y_i - y_i}{y_i}\right|$$

MAPE in $[0,+\infty)$, the MAPE close to 0% for the perfect model, the MAPE close to 100% for the poor model.

The Root Mean Square Error is:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - y_i\right)^2}$$

RMSE in $[0,+\infty)$, the RMSE close to 0 for the perfect model ; The bigger the MAE value, the bigger the error.

5 Data

In our research, our data comes from the Wind database. The data come from Wind Inc., which is the largest financial data service provider in China. We mainly explore the yield curve of China's government bond. The interest rates(yield to maturity) come from China Foreign Exchange Trade System(CFETS), the institution would public the interest rates of maturities every day, then Wind Inc. would gather these data. So we can get recent ten years data from the Wind database, the date from January 1, 2012 to December 31, 2022. The figures, being official, are remarkable neat, it can be used directly for analysis. And we do the research on daily frequency data.

In order to forecast the yield curve, we need to build features. We used data from eight dimensions to construct the features, such as economic data, inflation data, monetary policy, fiscal policy, interest rate data(interest rate transmission), China's stock market data, international finance related data, and people's livelihood data. There are 182 feature data. These data from the Wind database. Some of the data are daily frequency, for the data we use directly. Some of the data is annual, seasonal, monthly, or weekly frequency, and we use forward fill method to make them as daily time series. The forward fill method is propagate last valid observation forward. And the missing value is replaced by forward fill method. All data are dated from January 1, 2012 to December 31, 2022. Table 1 shows the specific data details.

Table 1. Dataset of Feature Engineering

We mainly based on monetary policy, fiscal policy, economic development to build model features. Data from Wind.

Feature Type	Example
Economic data	GDP, Fixed asset investment, PMI, Business Inventory, consumption data, China customs data……
Inflation data	CPI, PPI, Industrial output, DTD, LME……
Monetary policy	M0, M1, M2, Social financing, Benchmark interest rate for RMD deposits, Interest rate on deposit reserves, Loan Rate……
Fiscal policy	Chinese fixed interest rate issue rate……
Interest rate data	Treasury futures……
China's Stock market data	Shanghai Stock Index, Shenzhen Stock Index, Growth Enterprise Index……
International finance related data	USA treasury yield to maturity, Exchange rate data……
People's livelihood data	Employment data……

6 Results and Analysis

In our empirical analysis, we mainly do these works: First, from the different interest rates, we will explore the differences between types of interest rates. From the interest rates to the yield curves, analyze different models to get the difference after the yield curve. We will use the principal component analysis method to mine the principal components of yield curves. From the analysis, we will demonstrate the feasibility of the factor model for China's government bond yield curve from an empirical point of view. Second, we will use the economic and financial indicators, monetary policy, fiscal policy, interest rate transmission and people's livelihood to mine the relationship between these indicator data and interest rate, so as to construct effective influence features. Third, based on the DNS model and AFNS model, we will construct the yield curve of time series and analyze the trend feature of these factors. Finally, combined with the above analysis, we demonstrate the feasibility of forecasting yield curve from statistical model and machine learning model.

6.1 Interest Rates Analysis

China's government bond usually have three types of interest rate, such as spot rate, forward rate, and yield to maturity(YTM). From yield to maturity, it can get spot rate and forward rate.

An example based on interest rates on March 24, 2023 is shown in Figure 4, in which X axis represents the maturity, and Y axis represents the yields. From yield to maturity, we can get spot Rate and forward Rate(which from the Formula(1)(2)). For example, the 5 years yield to maturity is 2.71%, then we can the spot rate is 2.72%, and 5 years forward rate is 3.26%. In our situation, we mainly research yield to maturity. If we can get the yield curve of maturity, we can get spot yield curve, and forward yield curve. From the Figure 4, we can get different types of yield curve.

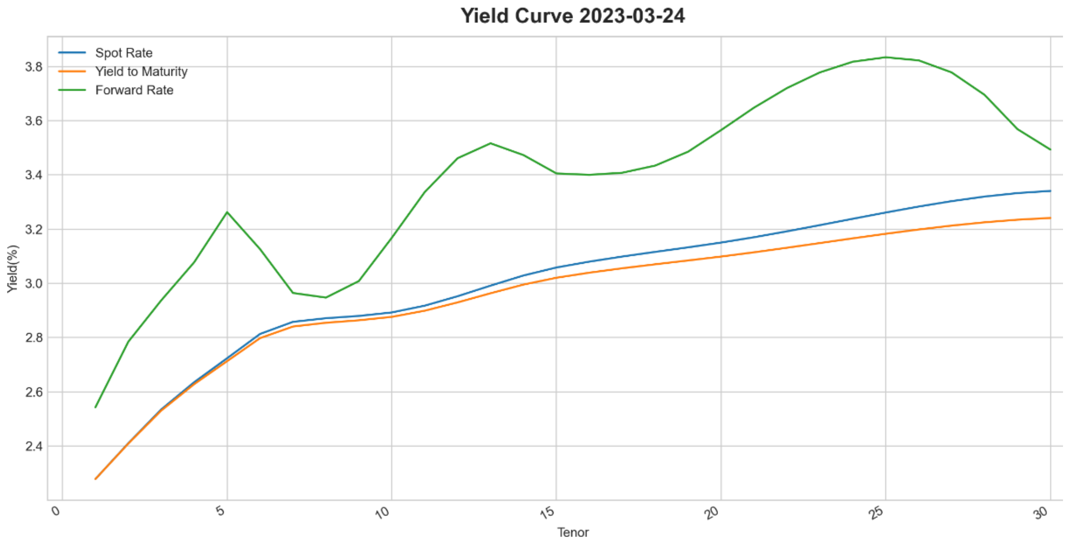


Figure 4. Yield Curve Example

The yield curves sample, date is 2023-03-24. On the same day, different types of yield curves have very different shapes. Data from Wind.

From time series, we can get different yields of different bond maturities. And the yield curves would evolve dynamically in cross section and temporal dimension. And we want research to yield trends to behave across different maturities and over time. Figure 5 is the resulting three-dimensional surface, with yields shown as a function of maturity, over time. The figure reveals a key yield curve facto: yield curves move a lot, shifting among different shapes: increasing at increasing or decreasing rates, decreasing at increasing or decreasing rates, flat, U-shaped, and so on.

Figure 5 is the three-dimensional surface which as a function show change of yields of maturity by time. The figure reveals a key yield curve factor: yield curves move different shapes: 1. Normal Yield Curve: A normal or upward-sloping yield curve is the most common shape where longer-term bonds have higher yields than shorter-term bonds. This shape indicates a growing economy with higher inflation expectations, and investors are compensated for the higher risks associated with longer-term investments.2. Flat Yield Curve: A flat yield curve occurs when the yields on short-term and long-term bonds are almost similar. This shape indicates uncertainty in the market about future economic growth, inflation, and other factors.3. Inverted Yield Curve: An inverted or downward-sloping yield curve is when the yields on long-term bonds are lower than the yields on short-term bonds. This shape is considered a predictor of an economic recession, where investors expect lower inflation or even deflation, and prefer long-term bonds.4. Humped Yield Curve: A humped yield curve is when the yields for medium-term bonds are higher than short-term and long-term bonds. This shape shows uncertain market sentiments, where investors expect higher inflation in the near future, and prefer shorter-term investments.

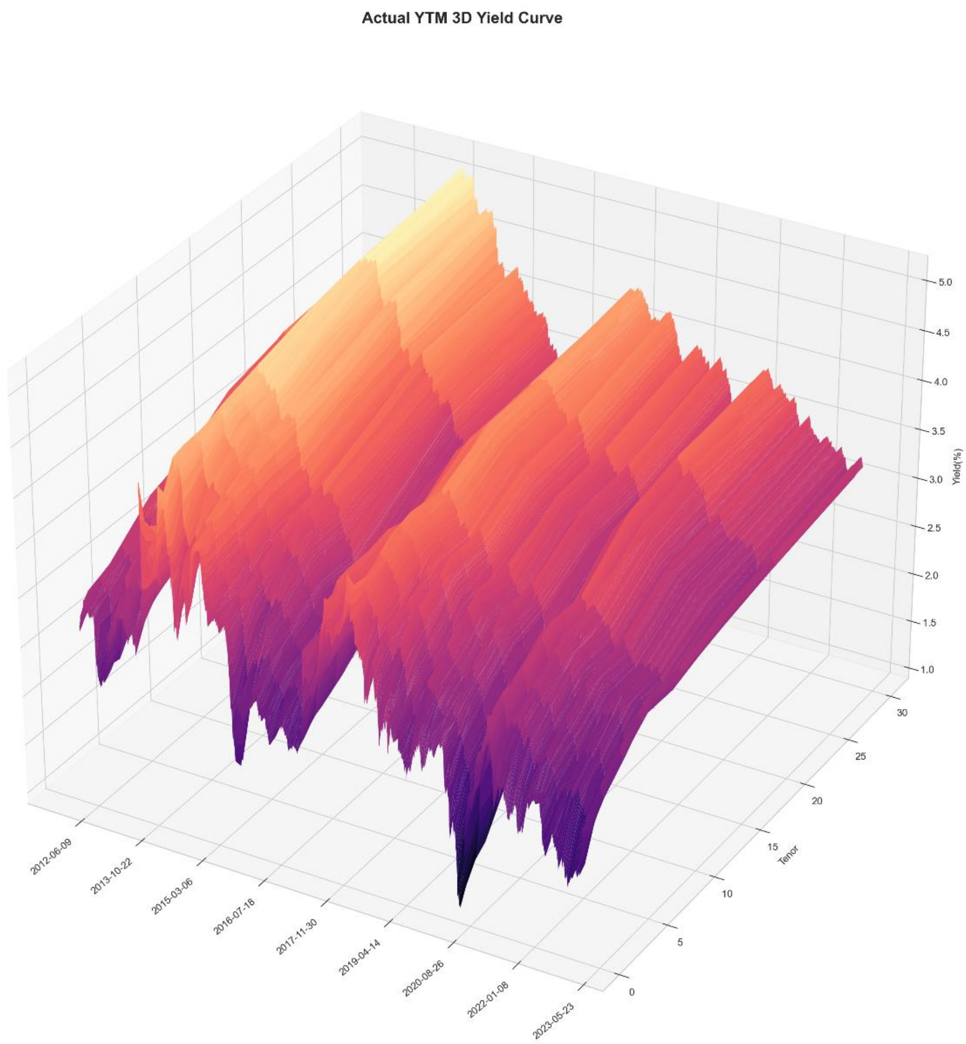


Figure 5. Bond Yields in Three Dimensions.

A three-dimensional plot of the interest rate curve from January 1, 2012 to December 31, 2022. X-axis is the Date, Y-axis is Tenor, and Z-axis is yield. From the 3-D graph we can see the yield curve is movement over time. Data from Wind.

Table 2 present descriptive statistics for yields at various maturities. Several well-known and important yield curve factors emerge. First, time-average yields increase with maturity(Such as, 6M yield is 2.61, 1Y yield is 2.71%, 5Y yield is 3.14%, …, the yield is increase with maturity.); that is, term premia appear to exist, perhaps dur to risk aversion, liquidity preferences, or preferred habitats. Second, yield volatilities decrease with maturity, presumably because long rates involve averages of expected short rates. Third, yields are highly persistent, as evidenced not only by the very large 1-day(1_Lags) autocorrelations but also by the sizable 1-month(20_Lags) autocorrelations.

Table 2. Bond Yield Statistics

We present descriptive statistics for yields at various maturities. We show sample mean, sample standard deviation, and 1-order and 20-order sample autocorrelations. Std is sample standard deviation. 1_Lags is first-order sample autocorrelations(1 day autocorrelations). 20-Lags is twentieth-order sample autocorrelations(1 month autocorrelations). Data are from Wind.

Tenor	Mean(%)	Std(%)	1_Lags	20_Lags
6M	2.613144	0.610506	0.997510	0.919155
1Y	2.709559	0.582337	0.998238	0.925805
2Y	2.880922	0.535774	0.998401	0.939065
3Y	2.981476	0.524701	0.998538	0.944185
4Y	3.062613	0.508023	0.998779	0.950003
5Y	3.143736	0.494070	0.998302	0.952103
6Y	3.228773	0.474643	0.998766	0.957371
7Y	3.313797	0.457345	0.998616	0.959091
8Y	3.325143	0.459516	0.998824	0.961836
9Y	3.336487	0.462534	0.998873	0.963651
10Y	3.347833	0.466380	0.998769	0.964543
15Y	3.640543	0.434715	0.998493	0.960222
20Y	3.757715	0.461164	0.998640	0.964811
30Y	3.906467	0.453997	0.998661	0.964091

From the yield spread can reflect the volatility of the yield curve. Table 3 shows the sample descriptive statistics for yield spreads relative to the rates of same maturity on the previous day. Yield spread dynamics contrast rather sharply with those of yield levels; in particular, spreads are noticeable less volatile and less persistent. As with yields, the 1-day spread autocorrelations are modest values, and they decay very quickly, so that 1-month spread autocorrelations are noticeably smaller than those for yields.

Table 3. Yield Spread Statistics

We present descriptive statistics for yield spreads at various maturities. We show sample mean, sample standard deviation, and 1-order and 20-order sample autocorrelations. Data are from Wind.

Tenor	Mean(BP)	Std(BP)	1_Lags	20_Lags
6M	-2.05	449.80	0.036352	0.020208
1Y	-1.37	364.66	0.191969	0.032645
2Y	-1.73	319.11	0.142251	0.024719
3Y	-1.68	297.64	0.144978	0.015655
4Y	-1.53	266.27	0.232445	0.015648
5Y	-1.37	301.19	0.134865	0.008393
6Y	-1.63	250.16	0.221349	0.018819
7Y	-1.89	252.56	0.161451	0.019787
8Y	-1.99	235.60	0.218196	0.024410
9Y	-2.10	232.45	0.229011	0.029916
10Y	-2.20	243.64	0.186833	0.033999
15Y	-3.05	230.79	0.167991	0.052679
20Y	-3.30	230.73	0.152733	0.033087
30Y	-3.19	225.59	0.157730	0.055650

From the descriptive statistics analysis, we can know the multivariate models are required for sets of interest rates. An obvious model is a vector autoregression or some close relative. Fortunately, it turns out that financial asset returns typically conform to a certain type of restricted vector autoregression, displaying factor structure. Factor structure is said to be operative in situations where one sees a high-dimensional object(e.g. a large set of interest rates), but where that high-dimensional object is driven by an underlying lower—dimensional set of object, or “factors”. Thus, we can further use principal component analysis method to research the interest rates.

6.2 PCA of Interest Rates

In Figure 6, we show a time-series plot of a standard set of bond yields. Clearly they do tend to move noticeably together, but at the same time, it's clear that more common level factor is operative.

Figure 6. Bond Yields in Two Dimensions

It shows interest rates of maturities(from 6M to 30Y) from January 1, 2012 to December 31, 2022. Data are from the Wind.

Figure 7 is show the first, second, and third principal components of bond yields from principle component analysis. The first factor is borderline nonstationary. The first factor is the most variable but also the most predictable, due to its very high persistence. The second factor is also high persistent and displays a clear business cycle rhythm. The second factor is less variable, less persistent, and less predictable than level factor. The third factor is the least variable, least persistent, and least predictable.

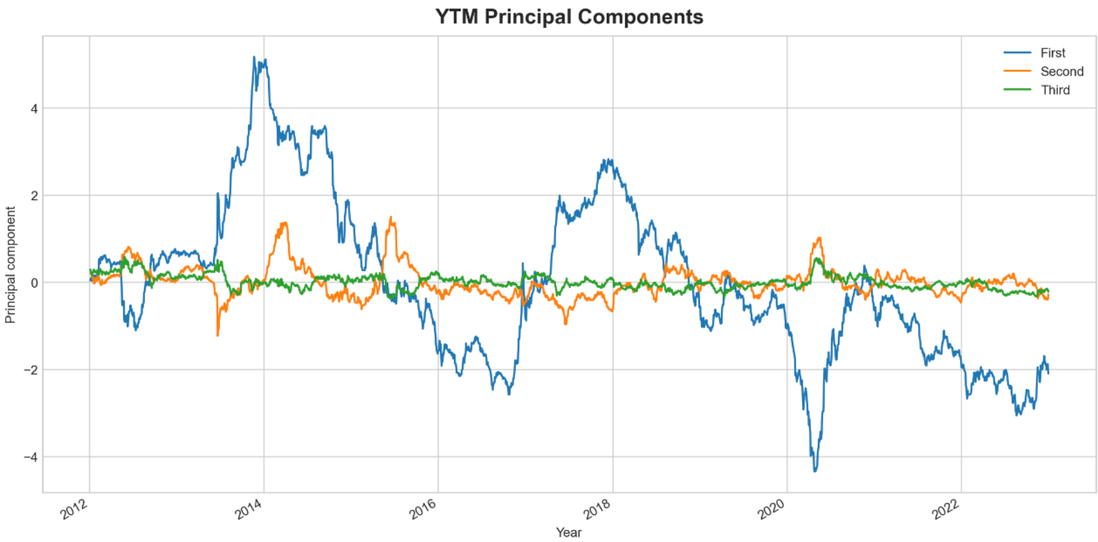


Figure 7. Bond Yield Principal Components

It shows time series of the first, second, and third principal components of bond yields from January 1, 2012 to December 31, 2022.

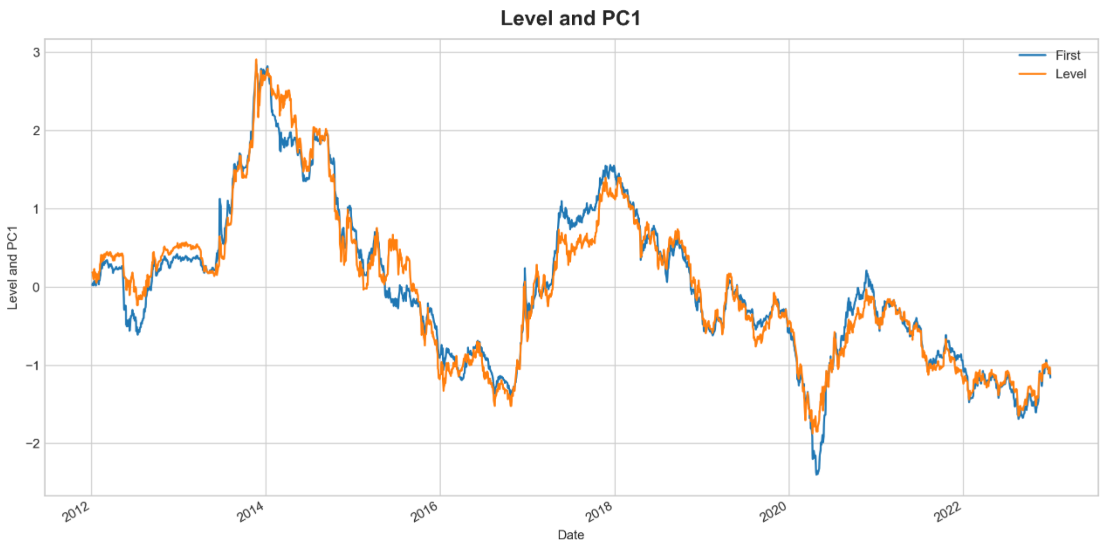
Typically three factors, or principal components, are all that one needs to explain most yield variation. From Table 4, we can know in our data set the first three principal components explain almost 100 percent of the variation in bond yields. Calculate the factors sample deviation, first- and twentieth- order principal component sample autocorrelations, and the predictive R^2.

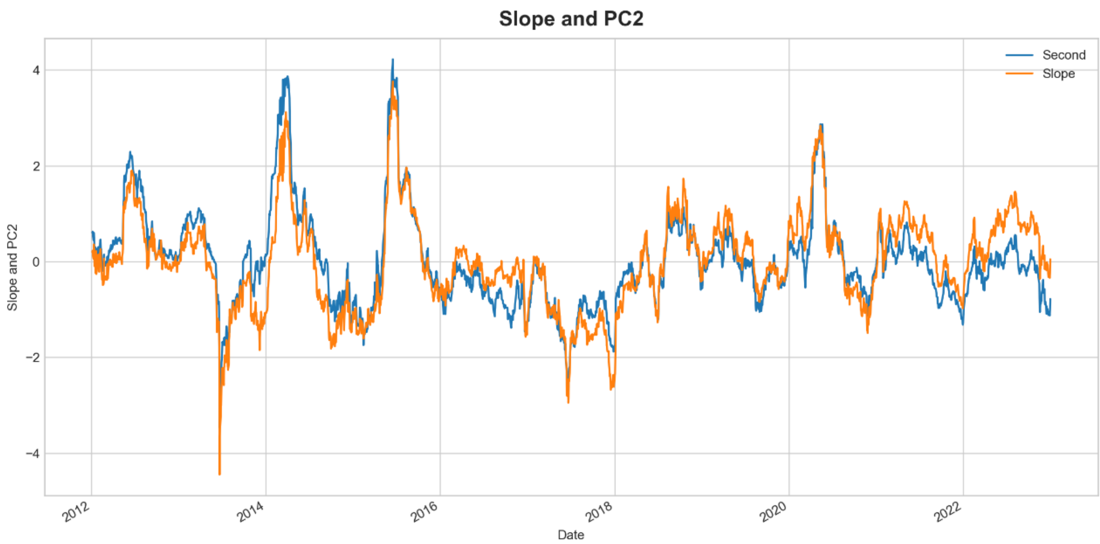
Table 4. Yield Principal Components Statistics

We present descriptive statistics for the first three principal components of interest rate of China's government bond. The R^2 (Diebold and Kilian, 2001) from an AR(p) approximating models with p selected using the Schwartz criterion. Data are from Wind.

PC	Std	1_Lags	20_Lags	R^2
First	1.813091	0.999187	0.960629	0.994374
Second	0.355658	0.992832	0.751748	0.97979
Third	0.162599	0.982732	0.705034	0.977811

In Figure 8 we plot three principal components(factors) against standard empirical yield curve level, slope, and curvature measures(the 10-year yield, the 10-6M spread, and a 6M + 10Y-2 *5Y butterfly spread, respectively). The figure reveals that the three bond yield factors effectively are level, slope, and curvature. This is important, because it implies that the different factors likely have different and specific macroeconomic determinants. Inflation, for example, is clearly related to the yield curve level, and the stage of the business cycle is relevant for the slope. It is also noteworthy that the yield factors are effectively orthogonal due to their exceptionally close links to the principal components, which are orthogonal by construction.





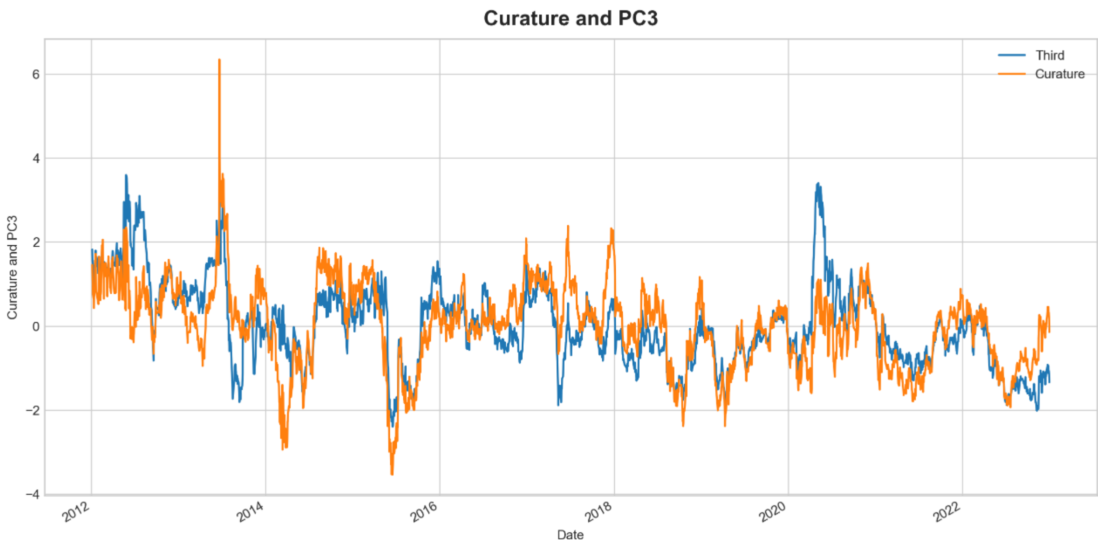


Figure 8. Level, Slope, Curvature, and First Three Principal Components

These data from January 1, 2012 to December 31, 2022. It shows the standardized empirical level, slope, and curvature, and the first three standardized principal components of bond yields. Data from Wind.

From the analysis, we have known the high persistent of yields, the lesser persistence of yield spread, the good empirical approximation afforded by a low-dimensional three-factor structure with highly persistent level and slope factors. Thus, we can further use Bootstrap model and factor model to build yield curve.

6.3 Yield Curve Models

The yield curve models mainly refers to bootstrap model and factors. The bootstrap model has better fitting effect, while the factor model has stronger financial interpretability. Based on different financial views, different factor models will be generated. We can use the bootstrap model as the benchmark to measure the error between the model value and the actual value generated by different factor models. The fit of factor model is evaluated from the fitting level.

6.3.1 Bootstrap Models

Figure 9 is the yield curve, which data is 2023-01-08. For Bootstrap model, it use cubic spline model. This yield curve is mainly obtained by using numerical fitting method, which has certain advantages in fitting effect. So it can be used as the benchmark of yield curve fitting performance.

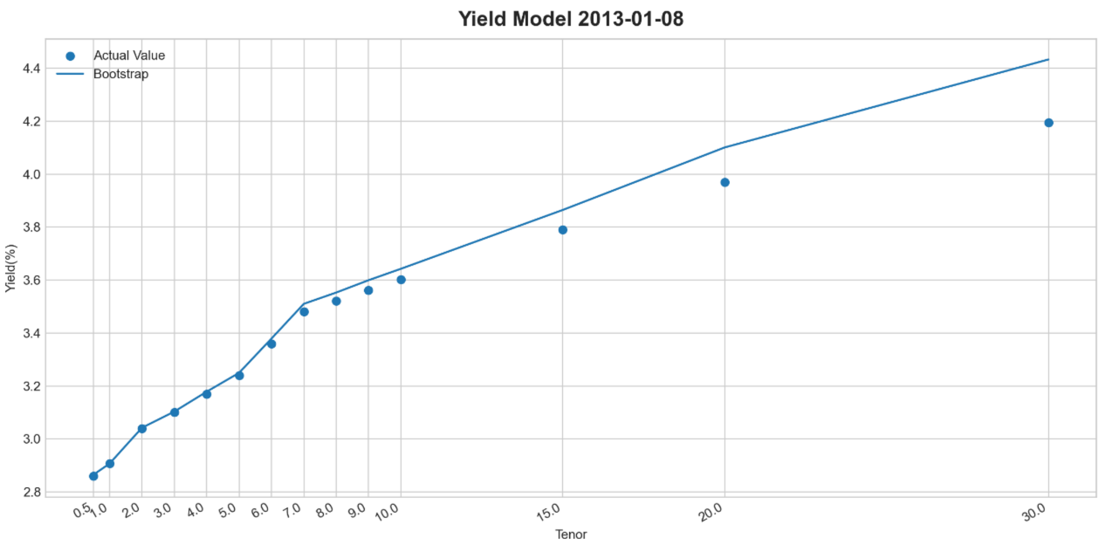


Figure 9. Yield Curve With Bootstrap Model

Sample yield curve, date is 2023-01-08, build by Bootstrap model(Cubic Spline Model). Data from Wind.

Figure 10 shows that there is still an error between the interest rate generated by numerical fitting and the real interest rate. Therefore, for any interest rate model, the error between model interest rate and real interest is inevitable. Interest rate model(especially interest rate factor model) is to stand in a relatively high dimension to research the change of yield curve.

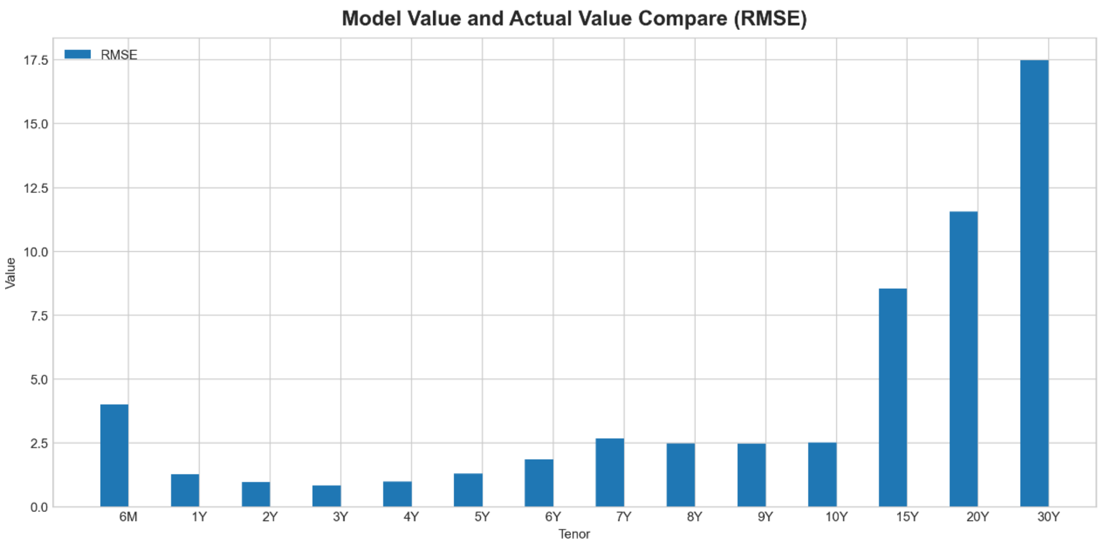


Figure 10. RMSE of Actual Values and Bootstrap Model Values

6.3.2 Dynamic Nelson-Siegel Model

For dynamic Nelson-Siegel model, we should set the λ value. For every day, we can get the best paraments, such as β_1, β_2, β_3, λ.So the λ values are time series values. Figure 11 shows the best λ value in time series. We can know the best λ values are relatively concentrated. So we can set mean value as λ constant value, then λ=0.385711.

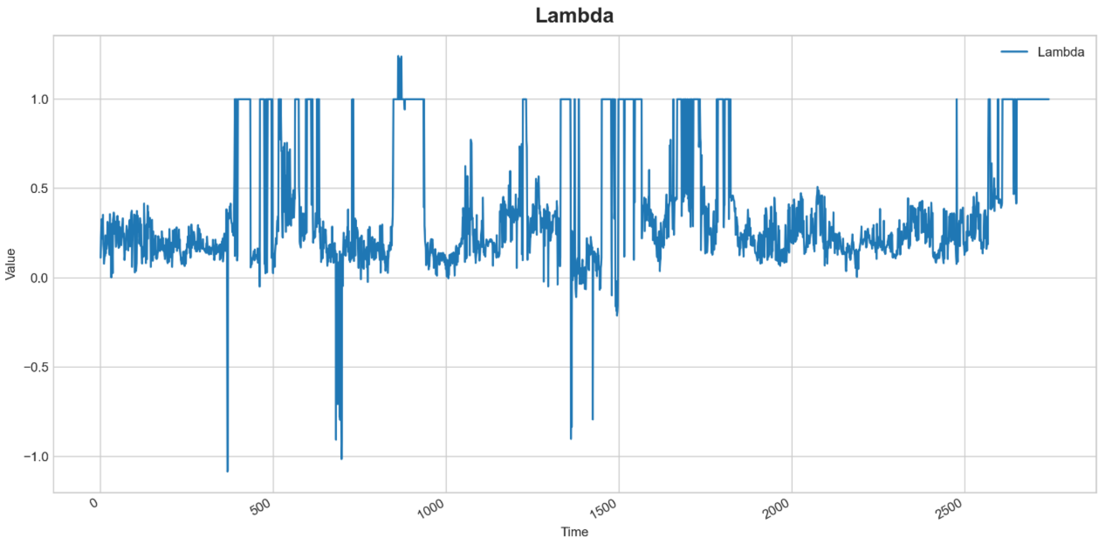


Figure 11. Optimal λ Value in Time Series

Figure 12 present the factor loadings, which plot as a function of maturity. And the factor loadings of β_1, β_2, β_3 are,

$$\left(1,\left(\left(1-e^{\lambda\tau}\right)/\lambda\tau\right),(1-e^{-\lambda\tau})/\lambda\tau-e^{-\lambda\tau}\right)$$

The loading on β_1,which is constant at 1, and so not decaying to zero in the limit. Hence, unlike the other two factor, it affects long yields and therefore was often a long term factor. The loading on β_2, $\left(1-e^{\lambda\tau}\right)/\lambda\tau$, a function that start at 1 but decays monotonically and quickly to 0. Hence is was often called a medium-term factor, mostly affecting medium-term yields. The loading on β_3, $(1-e^{-\lambda\tau})/\lambda\tau - e^{-\lambda\tau}$, which starts at 0 (and is thus not short term). Increases, and then decays to zero(and thus is not long term). Hence it was often called a short-term factor, mostly affecting short-term yields. Figure 13 is show the β_1, β_2, β_3 values of time series when set the constant λ value.

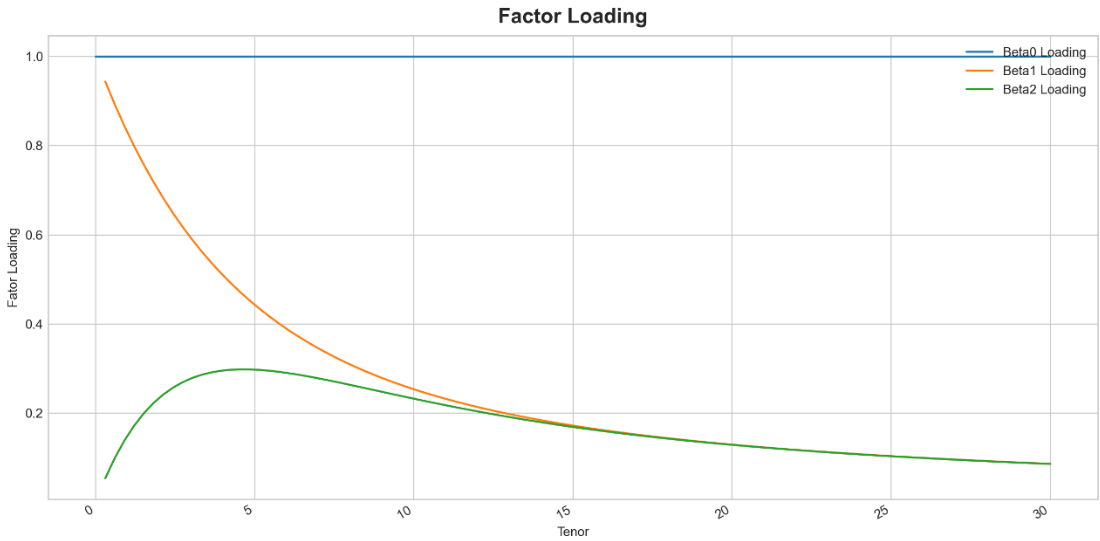


Figure 12. DNS Factor Loadings

It plots DNS factor loading as a function of maturity, for λ =0.385711.

From the yield curve, we can get β_1, β_2, β_3, and λ. Similarly, if we can get the factors value, we also can get the yield curve. And the factors value have financial significance, for example, β_1 is a long term factor, β_2 is a medium-term factor, and β_3 is short-term factor. If we want to forecast the yield curve, we can the method by forecasting the factors then get the yield curve of the future.

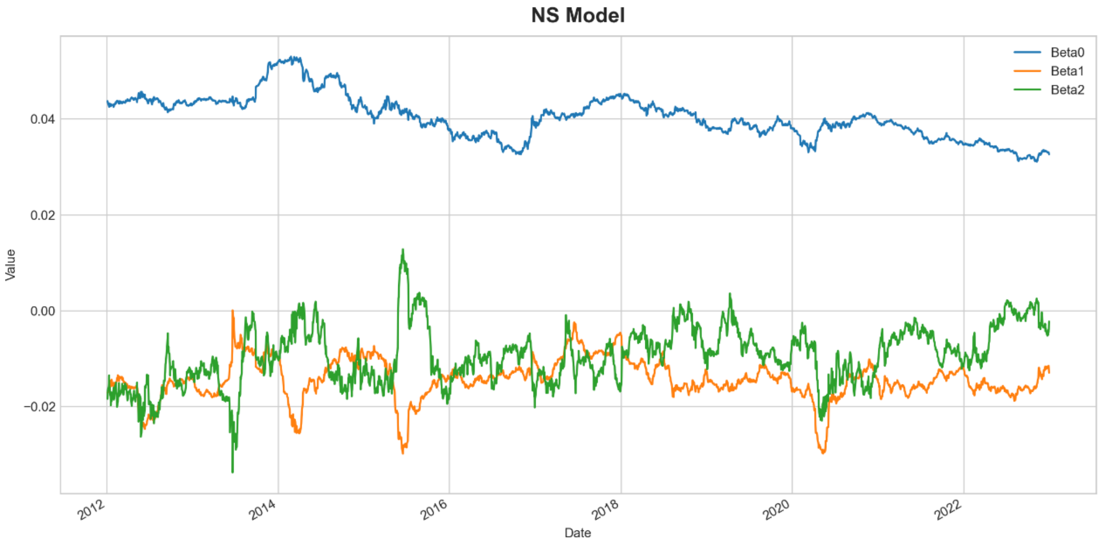


Figure 13. β_1, β_2, β_3 Values of Time Series

Fixed λ value, each β_1, β_2, β_3 of day can represent the yield curve of the same day. Data are from DNS model.

6.3.3 Arbitrage-Free Nelson-Siegel Model

For riskless arbitrage across maturities and over time, it can get arbitrage-free Nelson-Siegel model from Nelson-Siegel model. For AFNS(Arbitrage-Free Nelson-Sigel Model), first we need set optimal parameters,

$$P_{AFNS} = (\kappa_1, \kappa_2, \kappa_3, \sigma_{11}, \sigma_{21}, \sigma_{22}, \sigma_{31}, \sigma_{32}, \sigma_{33}, \theta_1, \theta_2, \theta_3)$$

We can use numerical method get model parameters. Table 5 is the model parameters value. From the model parameters, we can get the AFNS factors in time series. Figure 14 shows AFNS factors of time series. The dataset, from 2012-01-01 to 2022-12-31 bond yield(10 years)v.

Table 5. Parameters of AFNS

These parameters are obtained by numerical calculation (optimization method). They came from the R code.

Parameter Name	Parameter Value
κ_1	2.946924
κ_2	0.845110
κ_3	0.634485
σ_{11}	-0.087337
σ_{21}	0.165153
σ_{22}	0.406481
σ_{31}	0.119850
σ_{32}	-0.169886
σ_{33}	0.285420
θ_1	3.209325
θ_2	0.007235
θ_3	0.030360
λ	0.380626

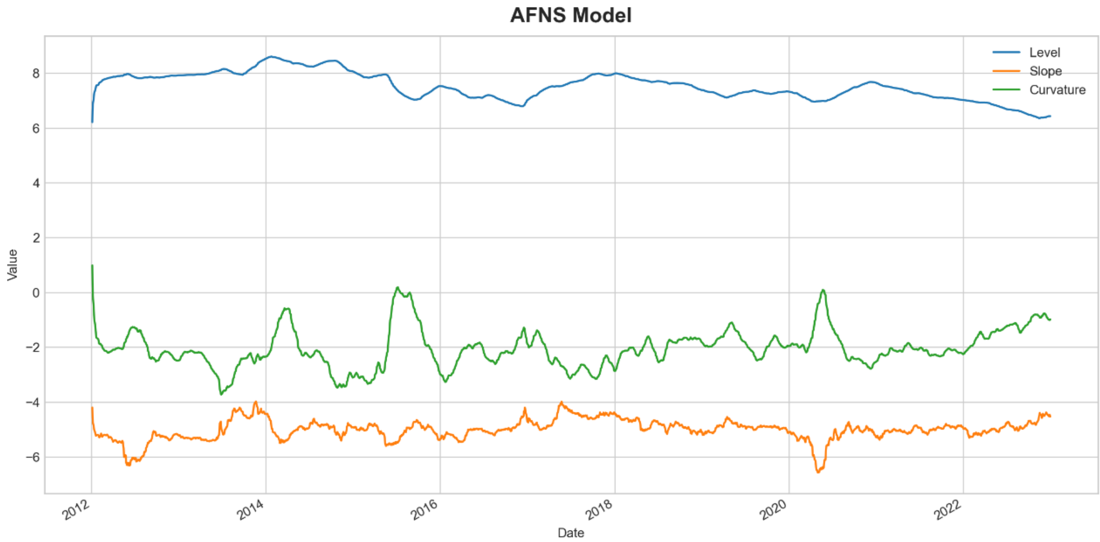


Figure 14. AFNS Factors of Time Series.

Fixed parameters of AFNS, factors(Level, Slope, Curvature) of day can represent the yield curve of the same day. Data are from AFNS model.

By comparing model values with actual values in RAME(Figure 15). Discovered these facts. First, in terms of fitting effect, the Bootstrap method has no advantage over DNS and AFNS. It indicates that both DNS and AFNS have good results in the fitting of interest rate curve. Second, by comparing the fitting effects of DNS and AFNS in different maturities, it is found that DNS has a better fitting effect on interest rates with shorter maturities, while AFNS has a better fitting effect on interest rates with longer maturities.

Since AFNS based on no arbitrage analysis is more explanatory in finance, and both AFNS and DNS have their own advantages and disadvantages in terms of model fitting effect, we need to conduct prediction analysis on both factor models. Compare them against each other by predicting the results.

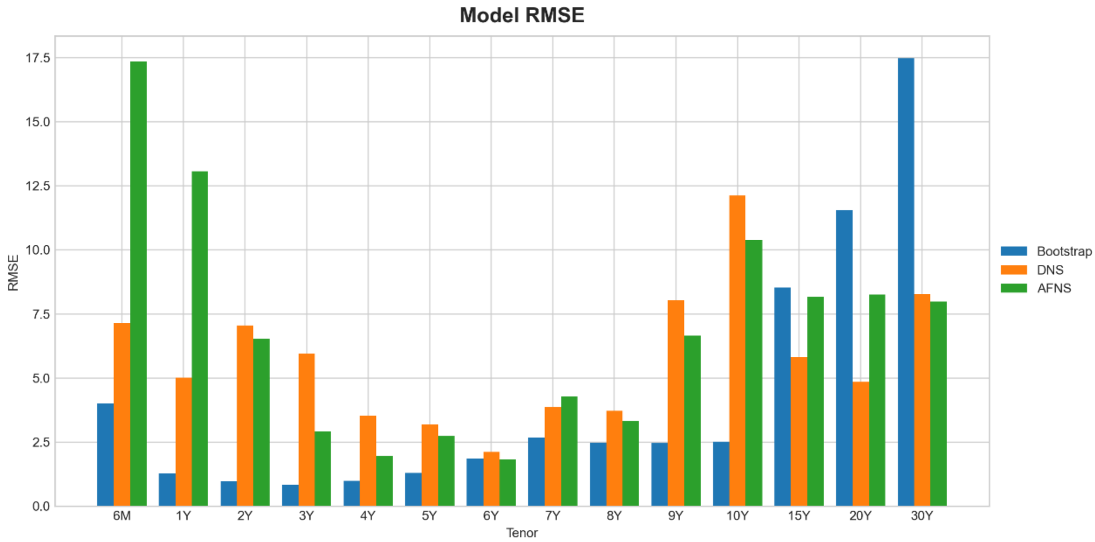


Figure 15. RMSE of Actual Values and Models Values

6.4 Forecasting Results

In order to forecast yield curve, we can predict the factors. Firstly, we use DNS and AFNS models to get the factors value of time series. Secondly, we use statistical models(ARIMA) and machine learning to predict the factors values in the future. Thirdly, we use the yield curve model to restore the yield curve by the factors. Lastly, we compare the predict value and actual value errors. In our thesis, we mainly forecast 6- and 12-month yield curve, and use singe-step forecast method.

Table 6. Forecasting Factors Performance for 6 Months

It shows out-of-sample root mean squared forecast errors for DNS vs AFNS, ARIMA vs ML. Bold numbers are the winners.

Factor	DNS		AFNS	
	ARIMA	ML	ARIMA	ML
Level	0.000231	**0.000231**	0.004006	0.002719
Slope	0.001184	**0.000490**	0.101593	0.035605
Curvature	0.002677	**0.000952**	0.142307	0.025444

Table 6 shows predict factors performance for 6 months, compare RMSE values and the number of winning items, ML vs. ARIMA is 3:0, DNS vs. AFNS is 3:0. Thus, ML method is better than ARIMA, and DNS is better than AFNS in predicting next 6 months factors. So the DNS + ML method is the best choice in forecast 6 months factors.

Table 7. Forecasting Factors Performance for 12 Months

It shows out-of-sample root mean squared forecast errors for DNS vs AFNS, ARIMA vs ML. Bold numbers are the winners.

Factor	DNS		AFNS	
	ARIMA	ML	ARIMA	ML
Level	0.000232	**0.000229**	0.004060	0.002750
Slope	0.001198	**0.000478**	0.095836	0.036328
Curvature	0.002126	**0.000957**	0.188173	0.026016

Table 7 shows predict factors performance for 12 months, compare RMSE values and the number of winning items, ML vs. ARIMA is 3:0, DNS vs. AFNS is 3:0. Thus, ML method is better than ARIMA, and DNS is better than AFNS in predicting next 12 months factors. So the DNS + ML method is the best choice in forecast 12 months factors.

To sum up, the DNS+ML method is the best combination of methods for the prediction of factors, whether it is a short-term prediction of 6 months or a long-term prediction of 12 months.

The accuracy of the predict factors of yield curve gives us a better idea of how yield curves change, but our goal is to predict the yield curve. From the above empirical analysis, we know that there is a one-to-one mapping between factors and yield curve. So, we can further analyze the predictive performance of yield curve of different yield curve models and predict methods.

Table 8. Forecasting Yield Curve Performance for 6 Months

It shows out-of-sample root mean squared forecast errors for DNS vs AFNS, ARIMA vs ML. Bold numbers are the winners.

Tenor	DNS		AFNS	
	ARIMA	ML	ARIMA	ML
6M	0.128182	**0.092998**	0.200704	0.195934
1Y	0.115573	**0.062392**	0.141401	0.122221
2Y	0.143474	0.096542	0.086402	**0.048102**
3Y	0.113425	**0.061019**	0.088898	0.067556
4Y	0.095466	**0.037618**	0.100040	0.087935
5Y	0.088673	**0.034494**	0.104425	0.095447
6Y	0.085863	**0.033256**	0.090956	0.082596
7Y	0.092738	**0.048666**	0.075383	0.065286
8Y	0.076992	**0.046276**	0.124956	0.125361
9Y	0.097731	**0.087579**	0.180671	0.184587
10Y	0.130630	**0.128937**	0.235791	0.241227
15Y	0.073925	**0.053056**	0.194760	0.198523
20Y	0.065129	0.048930	**0.320163**	0.324949
30Y	0.069830	**0.066167**	0.423382	0.427128

Table 8 shows forecast 6 months yield curve, compare RMSE values and numbers of winning items, ML(machine learning) vs. ARIMA(Autoregressive Integrated Moving Average) is 13:1, and DNS vs. ANFS is 12:2. Thus, in the yield curve forecast for the next six months, the machine learning method is better than statistical method(such as ARIMA(p, q)), and the DNS model is better than AFNS model. So the DNS + ML method is the best choice in forecast 6 months yield curve.

Table 9. Forecasting Yield Curve Performance for 12 Months

It shows out-of-sample root mean squared forecast errors for DNS vs AFNS, ARIMA vs ML. Bold numbers are the winners.

Tenor	DNS		AFNS	
	ARIMA	ML	ARIMA	ML
6M	0.145481	**0.096507**	0.221277	0.200284
1Y	0.105843	**0.059375**	0.148512	0.121816
2Y	0.110747	0.096325	0.082496	**0.047639**
3Y	0.084632	**0.060115**	0.092278	0.066742
4Y	0.075038	**0.037123**	0.105624	0.086889
5Y	0.074906	**0.034644**	0.109973	0.093979
6Y	0.068115	**0.033770**	0.097947	0.081767
7Y	0.067802	**0.049094**	0.083019	0.065018
8Y	0.078410	**0.047529**	0.135018	0.124873
9Y	0.113390	**0.087894**	0.190676	0.183915
10Y	0.150010	**0.128406**	0.245182	0.240286
15Y	0.057740	**0.052353**	0.204673	0.199846
20Y	0.051645	**0.045820**	0.328849	0.326833
30Y	0.062076	0.067905	0.424005	**0.423467**

Table 9 shows forecast 12 months yield curve, compare RMSE values and numbers of winning items, ML vs. ARIMA is 14:0, and DNS vs. ANFS is 12:2. Therefore, in the yield curve forecast for the next 12 months, ML method is better than ARIMA method, and the DNS model is better than AFNS model. So the DNS + ML method is the best choice in forecast 12 months yield curve.

After the above analysis, we can clearly know that DNS + Machine Learning is an optimal solution. So we can use DNS and machine learning to build yield curve forecast framework. For forecast yield curve, we found AFNS is not better than DNS. It maybe AFNS often exhibit poor empirical time-series performance, especially when forecasting future yields(Francis X. Diebold and Glenn D. Rudebusch, 2013)[8].

Machine Learning method is better than statistical method, because machine learning method offer flexibility than statistical model in forecast. In machine learning method, we can do the mix of models, using the advantages of various models to achieve our better goals.

Based on the above empirical analysis, we can conclude that the DNS model and machine learning method can form a good framework for yield curve forecast.

7 Application

In our thesis, we devote ourselves to building a common method to forecast the yield curve. With the development of financial models, big data, algorithms, and computing power, the era of large financial models will come. It will bring about a paradigm change in the field of finance. Compared with traditional research methods, it ranges from the empirical study of phenomena to the empirical study of frameworks. Therefore, this research will have applications in there three fields:

1. Bond investment strategy and risk management. Bonds trade based on the interest rate curve because of the variety of maturity they have left. So if we can predict the interest rate curve, we can do a good job of investment and risk management. For example, if we know that interest rates are going to go up in the future, we can go short or sell in advance. If we know that interest rates are going down in the future, we can choose to be long or buy. We can also manage our duration in a bond investment account so that the account does not lose money due to interest rate fluctuations.
2. Other fixed income product yield curve forecast. Use the yield curve forecast framework is extended to other fixed income interest rate products. As the most common interest rate trading products, bonds have a certain universality. In our actual trading products, the fixed income market is mostly over-the-counter trading. So there are a variety of trading contracts, such as spot, forward, swap, repo, reverse repo and so on. If we can predict the interest rate curve of these interest rate products, we can choose the most favorable contract for our trading needs to hedge our risk due to the interest rate.
3. The innovation of financial research paradigm. At present, most financial empirical research is based on a certain phenomenon to put forward the hypothesis, and then based on the hypothesis to do empirical research. This study is a universal study based on big data and big models. From a more abstract dimension, a theoretical framework is proposed to form the discovery of

knowledge. For example, in asset pricing, we often study the impact of a certain factor on the rate of return and risk. But as larger models develop, we hope that computers will automatically discover the factors that influence asset pricing. In this study, we use traditional financial models, such as DNS, AFNS model and machine learning model, to form a new research paradigm. This method can provide some reference for the research of financial engineering.

8 Conclusion

Reviewing the entire work, it built some yield curve models based on interest rate data, such as the Bootstrap model, DNS model, and AFNS model. Then it made further analysis based on the factor model(DNS, AFNS), and found that realize the prediction of the yield curve through predict factors. Finally, it uses statistical model and machine learning model to forecast the yield curve through forecasting factors. Through these works, it can draw the following conclusion:

1) From the yield curve models, we can know the bootstrap model can fit the actual interest rate best. This is because the bootstrap models use numerical fitting method, so the bootstrap is the best model than factors models. But the bootstrap model doesn't have enough finance meanings. Compare the DNS model and AFNS model, it found the AFNS model can fit the actual interest rate better than DNS. It means China's government bond exists as a no-arbitrage phenomenon.
2) From the interest rates of China's government bonds principal component analysis, we found the factor model can fit the yield curve of China's government bonds. We can use DNS and AFNS to fit the yield curve of China's government bonds. At present, the bootstrap method is the mainstream method to study the yield curve of China. This work proves that the factor model will have great development space in the future.
3) In the work, it presents the framework of using the factors and machine learning to forecast the yield curve of China's government bond. It also compares the predict results of DNS vs. AFNS, statistical model and machine learning model. And it proved the DNS and machine learning is a better method than other methods.

In the future, we can build a fixed income investment research framework based on yield curve prediction on the basis of this model, analyze future interest rate trends, give trading strategies and risk management solutions.

9 References

[1] Salomon Brothers. (1995), "Understanding the Yield Curve", United States Fixed Income Research Portfolio Strategies.

[2] Lin Chen. (1996), "Interest Rate Dynamics Derivatives Pricing, and Risk Management", Springer.

[3] Damiano Brigo and Fabio Mercurio. (2006), "Interest Rate Models Theory and Practive: With Smile, Inflation and Credit", Springer Finance.

[4] Rene A. Carmona and Michael R. Tehranchi. (2006), “Interest Rate Models: an Infinite Dimensional Stochastic Analysis Persperctive“., Springer Finance.

[5] Yolanda S. Stander. (2005), “Yield Curve Modeling”, Palgrave Macmillian.

[6] Frederic S. Mishkin. (2019), “The Economics of Money, Banking, and Financial Markets(Twelfth Edition)”, Pearson.

[7] Riccardo Rebonato. (2018), “Bond Pricing and Yield-Curve Modelling: A Structural Approach”, Cambridge University Press.

[8] Francis X. Diebold and Glenn D. Rudebusch. (2013), “Yield Curve Modeling and Forecasting: The Dynamic Nelson-Siegel Approach”.

[9] Jessica James and Nick Webber. (2000), “Interest Rate Modelling”, Wiley Series In Financial Engineering.

[10] Campbell R. Harvey. (1986), “Recovering Expectations of Consumptions Growth From An Equilibrium Model of The Term Structure of Interest Rates”, The University of Chicago.

[11] Ferdinando M. Ametrano and Marco Bianchetti. (2013), “Everything You Always Wanted to Know About Multiple Interest Rate Curve Bootstrapping But Were Afraid To Ask”.

[12] Robert Ferstl and Josef Hayden. (2010), “Zero-Coupon Yield Curve Estimation with the Package termstrc”.

[13] Nelson, C.R., and A.F. Siegel (1987), “Parsimonious Modeling of Yield Curves,” Journal of Business, p 473–489.

[14] Diebold, F.X., Li, C., and Yue, V. (2008), "Global Yield Curve Dynamics and Interactions: A Generalized Nelson-Siegel Approach," Journal of Econometrics, 146, p 351–363.

[15] Christensen, J.H.E., Diebold, F.X., and Rudebusch, G.D. (2009), "An Arbitrage-Free Generalized Nelson-Siegel Term Structure Model," The Econometrics Journal, 12, p 33–64.

[16] Christensen, J.H.E., Diebold, F.X., and Rudebusch, G.D. (2011), "The Affine Arbitrage-Free Class of Nelson-Siegel Term Structure Models," Journal of Econometrics, 164, p 4–20.

[17] Diebold, F.X., and Li, C. (2006), "Forecasting the Term Structure of Government Bond Yields," Journal of Econometrics, 130, p 337–364.

[18] Diebold, F.X., Ji, L., and Li, C. (2006), "A Three-Factor Yield Curve Model: Non-Affine Structure, Systematic Risk Sources, and Generalized Duration," in L.R. Klein (ed.), LongRun Growth abd Short-Run Stabilization: Essays in Memory of Albert Ando, 240–274. Cheltenham, U.K.: Edward Elgar.

[19] Almeida, C., A. Duarte, and C. Fernandes (2003), "A Generalization of Principal Components Analysis for Non-Observable Term Structures in Emerging Markets," International Journal of Theoretical and Applied Finance, 6, p 895–893.

[20] Balduzzi, P., S.R. Das, S. Foresi, and R.K. Sundaram (1996), "A Simple Approach to ThreeFactor Affine Term Structure Models," Journal of Fixed Income, 6, 43–53.

[21] Alper, C.E., K. Kazimov, and A. Akdemir (2007), "Forecasting the Term Structure of Interest Rates for Turkey: A Factor Analysis Approach," Applied Financial Economics, 17, p 77–85.

[22] Marlene Amstad and Zhiguo He. (2019), "Chinese Bond Market And Interbank Market".

[23] Marco Peixeiro. "Time Series Forecasting in Python", Manning.

[24] Aileen Nielson. "Practical Time Series Analysis", O'Reilly.

[25] Yu Lin, Zixiao Lin, Ying Liao, Yizhuo Li, Jiali Xu, Yan Yan, “Forecasting the realized volatility of stock price index: A hybrid model integrating CEEMDAN and LSTM”, Expert Systems with Applications, Volume 206, 2022.

[26] Ha Young Kim, Chang Hyun Won, “Forecasting the volatility of stock price index: A hybrid model integrating LSTM with multiple GARCH-type models”, Expert Systems with Applications, Volume 103, 2018, Pages 25-37.

[27] Xuerong Li, Wei Shang, Shouyang Wang,”Text-based crude oil price forecasting: A deep learning approach”, International Journal of Forecasting,Volume 35, Issue 4, 2019,Pages 1548-1560,

[28] Sinan Ozdemir and Divya Susarla, “Feature Engineering Make Easy”, Packt Publishing.

[29] Hariom Tatsat, Sahil Puri and Brad Lookabaugh. “Machine Learning & Data Scinece Blueprints for Finance”, O’Reilly.

[30] Trevor Hastie, Robert Tibshirani and Jerome Friedman. “The Elements of Statistical Learning: Data Mining, Inference, and Prediciton(Second Edition)”, Springer Series in Statistics.

[31] Ian Googfello, Yoshua Bengio and Aaron Courville. “Deep Learing”.

[32] Periklis Gogas, Theophilos Papadimitriou, Maria Mathaiou, and Efthymia Chrysanthidou. (2014), “Yield Curve and Recession Forecasting in a Machine Learning Framework”. Comput Econ(2015) 45:635-645.

[33] Shijie Wu, Ozan Irsoy et al. (2023), “BloombergGPT: A Large Language Model for Finance”. arXiv:2303.17564v1 [cs.LG] 30 Mar 2023.

[34] Joseph G. Haubrich and Ann M. Dombrosky. (1996), “Predicting Real Growth Using the Yield Curve”, Economic Review 1996 Q1, p 26-35.

[35] Jeffrey L. Anderson and Stephen L. Anderson. (1999), “A Monte Carlo Implementation of the Nonlinear Filtering Problem of Produce Ensemble Assimilations and Forecasts”, American Meteorological Society, December 1999. p 2471-2758.

[36] Jens H.E. Christensen, Francis X. Diebold, Glenn D. Rudebusch. (2010), “The Affine Arbitrage-Free Class of Nelson-Siegel Term Structure Models”.
[37] Tianqi Chen and Carolos Guestrin. (2016), “XGBoost: A Scalable Tree Boosting System”, arXiv:1603.02754v3 [cs.LG] 10 Jun 2016.
[38] David Redfern and Douglas McLean, (2014), “Principal Component Analysis for Yield Curve Modelling: Reproduction of out-of-sample yield curves”. Moody’s Analytics, August 2014.
[39] Nicolo Marting, “Yield Curve Modelling and Forecasting Applications”.
[40] Selina Reinicke, “Modeling and forecasting yield curves: A comparison of published methods”.
[41] People's Bank of China(中国人民银行). (2020),” Participate in the reform of international benchmark interest rates and improve China's benchmark interest rate system(参与国际基准利率改革和健全中国基准利率体系)”.
[42] Monika Piazzesi, “Affile Term Strucutre Models”.
[43] Rita Pimentel, Morten Risstad, Sjur Westgaard. (2022), “Predicting interest rate distributions using PCA & quantile regression”, Digit Finance(2022)4, p291-311.
[44] Reuven Y. Rubinstein. “Simulation and The Monte Carlo Method (Second Edition)”.
[45] Zhuoshi Liu et al. “Construct China's super - long - term non - arbitrage bond yield curve”, IMI Working Paper No.1920.
[46] Bryan Lim and Stefan Zohren. “Time Series Forecasting With Deep Learning: A Survey”, arXiv:2004.13408v2 [stat.ML] 27 Sep 2020.
[47] Patrick S. Hagan and Graeme West. “Methods for Constructing a Yield Curve”, Wilmott magazine, Technical Article 2, p 70-81.
[48] Gianluca Bontempi, Souhaib Ben Taieb, and Yann-A¨el Le Borgne, “Machine Learning Strategies for Time Series Forecasting”. eBISS 2012, LNBIP 138, pp. 62–77, 2013.
[49] Mark Fisher, Douglas Nychka, and David Zervos. (1994), “Fitting the term structure of interest rates with smoothing splines”.

[50] Mohammad Irfan, Mohamed Elhoseny, Salina Kassim, and Noura Metawa, "Advanced Machine Learning Algorithms for Complex Financial Applications".