

Article

# Forecasting Term Structure of Interest Rates in Japan

Hokuto Ishii

Graduate School of Economics, Nagoya University, Furo-cho, Chikusa-ku, Nagoya 464-8601, Japan;  
ishii.hokuto@h.mbox.nagoya-u.ac.jp

Received: 15 May 2019; Accepted: 1 July 2019; Published: 8 July 2019



**Abstract:** In this paper, we examined and compared the forecast performances of the dynamic Nelson–Siegel (DNS), dynamic Nelson–Siegel–Svensson (DNSS), and arbitrage-free Nelson–Siegel (AFNS) models after the financial crisis period. The best model for the forecast performance is the DNSS model in the middle and long periods. The AFNS is inferior to the DNS model for long-period forecasting. In U.S. bond markets, AFNS is shown to be superior to DNS in the U.S. However, for Japanese data, there is no evidence that the AFNS is superior to the DNS model in the long forecast horizon.

**Keywords:** dynamic Nelson–Siegel; arbitrage-free Nelson–Siegel; affine term structure; forecasting

**JEL Codes:** G12; G14; G17

## 1. Introduction

After the financial crisis, the Japanese economy faced an economic environment unlike it had ever experienced before. In January 2013, under the new governor of the central bank, the Bank of Japan began carrying out a new monetary policy framework. In general, this is called the “unconventional monetary policy”.

In this paper, we examine and compare forecast performances for Japanese Government Bonds (JGB) from the financial crisis onwards. To analyze the dynamics of JGB yield, we adopt the Dynamic Nelson–Siegel (DNS), Dynamic Nelson–Siegel–Svensson (DNSS), and Arbitrage-free Nelson–Siegel (AFNS) models. The former two models are curve-fitting models, which do not impose theoretical fundamentals such as arbitrage pricing theory. On the other hand, the last model imposes the arbitrage pricing theory and, thus, it is theoretically more rigorous than the former two models.

To forecast the future dynamics of the yield curves is a key concern of economists, policymakers, and those in the private sector, as the dynamics of the yields are considered to include the expectations of the market. If we can forecast the dynamics of the yield curve or term structure of interest rates, we may capture future fluctuations of the economic fundamentals. Therefore, forecasting the term structure of interest rates is a major step towards discerning the economic activity of a country.

This study provides one main result: The DNSS model is the best, among the three models, to forecast future yields; it outperforms the random walk model (a benchmark model) for 6- and 12-month forecasts. Moreover, the AFNS model is inferior to the DNS model for long forecast horizons, with evidence suggesting that the arbitrage-free assumption is not appropriate for long-period forecasting; although, [Christensen et al. \(2011\)](#) showed that the AFNS model is superior to the DNS in terms of forecasting performance.

The next sub-section reviews the literature. The organization of the paper is at end of the section.

### Literature Review

The literature of the curve-fitting model started with the pioneering work of [Nelson and Siegel \(1987\)](#). [Diebold and Li \(2006\)](#) extended the cross-sectional Nelson–Siegel model to the time series. We

refer to the Diebold–Li extension of Nelson–Siegel model as the dynamic Nelson–Siegel (DNS) model. The DNS model is a three-factor model. On the other hand, Diebold and [Diebold and Rudebusch \(2013\)](#) introduced a four-factor<sup>1</sup> model as an extension of the DNS. We refer to the four-factor model as the dynamic Nelson–Siegel–Svensson (DNSS) model.

A fundamental characteristic of curve-fitting models, such as DNS and DNSS, is that theoretically rigorous restrictions (like the arbitrage-free theory) are not imposed. Although this point has attracted criticism, [Gasha et al. \(2010\)](#) showed that these models have empirically better performances<sup>2</sup> than arbitrage-free models; in particular, the affine term structure model, which was pioneered by [Vasicek \(1977\)](#). [Cox et al. \(1985\)](#) and [Hull and White \(1990\)](#) presented representative models that extend upon the Vasicek model. These arbitrage-free models impose the rigorous arbitrage-free theory. After [Duffie and Kan \(1996\)](#) derived the analytical solution of the affine term structure model, arbitrage-free models have rapidly developed ([Ang and Piazzesi 2003](#)). Since then, a lot of studies ([Rudebusch and Wu 2008](#); [Hördahl et al. 2005](#); [Wu 2006](#); and [Moench, Emanuel 2008. Forecasting the yield curve in a data-rich environment: a no-arbitrage factor augmented VAR approach 2008](#)) have been written on arbitrage-free macro-finance equilibrium models.

[Fama \(2006\)](#) discussed the forecasting of future spot rates by using the forward interest rate. [Al-Al-Zoubi \(2009\)](#) examined a wide range of the interest rate models and reported the model parameter by using the generalized method of moment (GMM). [De Rezende and Ferreira \(2013\)](#) reported the improvement of the out-of-sample forecasts of the Nelson–Siegel class of models by using the quantile autoregression approach. There are a lot of the applied studies for the Nelson–Siegel class of models. [Chen and Tsang \(2013\)](#) applied the DNS model to domestic and foreign interest rate differential. [Ishii \(2018\)](#) extended the original [Chen and Tsang \(2013\)](#) model to the four- and five-factor models.

[Christensen et al. \(2009, 2011\)](#) set the Nelson–Siegel model into an affine term structure framework. Their model is called the arbitrage-free Nelson–Siegel (AFNS) model, which is considered to be both an empirically and theoretically good model. In fact, [Christensen et al. \(2011\)](#) and [Diebold and Rudebusch \(2013\)](#) showed that the AFNS was superior to the DNS for forecasting the performance of U.S. bond markets. Hence, they concluded that the imposition of the arbitrage-free theory improves the DNS.

There are many varieties of AFNS models. As [Christensen et al. \(2009, 2011\)](#) used, some are continuous time models<sup>3</sup>. Others (e.g., [Alfaro 2011](#) and [Niu 2012](#)) proposed discrete-time AFNS models. The advantage of the discrete model is that matching the data with the model is easier, such that we can estimate the model easily and simply.

A Kalman filter, through maximum likelihood estimation, is often used when estimating the parameters of the DNS and AFNS models. However, [Hamilton and Wu \(2012\)](#) pointed out that the maximum likelihood estimator might be a local estimator and, so, they suggested another method, such as the ordinary least squares (OLS) method. [Diebold and Li \(2006\)](#) and [Niu \(2012\)](#) used the multiple regression method. Following these studies, our study employs the multiple regression method. In the first step, using the non-linear least-squares (NLS) method, the scale parameters of DNS, DNSS, and AFNS are estimated. In the second step, using OLS, the state variables are estimated.

[Gurkaynak and Wright \(2012\)](#) surveyed affine term structure models, including the Nelson–Siegel frameworks. [Diebold and Rudebusch \(2013\)](#) explained the framework of the DNS and the AFNS in detail.

Few studies have investigated JGB yields. One of the contributions of this study is to examine JGB yields and to compare the forecast performance of AFNS with those of DNS and DNSS. The second contribution is to provide evidence that the arbitrage-free assumption is not appropriate for the forecasting of JGB yields, unlike the forecasting of U.S. government bond yields.

---

<sup>1</sup> [Svensson \(1994\)](#) extended the three-factor model of [Nelson and Siegel \(1987\)](#) to the four-factor model.

<sup>2</sup> [Gurkaynak et al. \(2007\)](#) showed that Nelson–Siegel–Svensson model captured U.S. yields well. [Chan et al. \(1992\)](#) compared a variety of affine term structure models.

<sup>3</sup> [Christensen et al. \(2011\)](#) and [Diebold and Rudebusch \(2013\)](#) use the continuous-time model.

The rest of this paper is organized as follows. In Section 2, we show the main theoretical models: DNS, DNSS, and AFNS. Section 3 characterizes the empirical models derived from the theoretical models. In Section 4, we show the data used in the analyses. Section 5 describes the empirical results. Our conclusions are drawn in Section 6.

## 2. Theoretical Model

This section explains three yield curve models: DNS, DNSS, and the discrete- and continuous-time versions of the AFNS model.

### 2.1. The Framework of the DNS and the DNSS Model

In the continuous-time model, the bond price  $P(t, T)$  of the maturity  $(T - t)$  and the yield  $y(t, T)$  at time  $t$  are represented as:

$$P(t, T) = e^{-(T-t)y(t, T)}. \quad (1)$$

The instantaneous forward rate is represented as:

$$f(t, T) = -\frac{\partial P(t, T) / \partial (T - t)}{P(t, T)}. \quad (2)$$

Then, the yield is written as:

$$y(t, T) = \frac{1}{T-t} \int_0^{T-t} f(s, T) ds. \quad (3)$$

Nelson and Siegel (1987) proposed the following forward rate:

$$f(t, T) = \beta_1 + \beta_2 e^{-\lambda(T-t)} + \beta_3 (T-t) \lambda e^{-\lambda(T-t)}, \quad (4)$$

where  $\beta$  and  $\lambda$  are parameters. This model is a three-factor model.

Substituting Equation (4) into Equation (3), we obtain the Nelson–Siegel yield function:

$$y(t, T) = \beta_1 + \left( \frac{1 - e^{-\lambda(T-t)}}{\lambda(T-t)} \right) \beta_2 + \left( \frac{1 - e^{-\lambda(T-t)}}{\lambda(T-t)} - e^{-\lambda(T-t)} \right) \beta_3. \quad (5)$$

Svensson (1994) extended the three-factor Nelson–Siegel model to a four-factor model. In general, the extended version of the Nelson–Siegel model, which was proposed by Svensson (1994), is called the generalized Nelson–Siegel model. Its forward rate is written as:

$$f(t, T) = \beta_1 + \beta_2 e^{-\lambda_1(T-t)} + \beta_3 (T-t) \lambda_1 e^{-\lambda_1(T-t)} + \beta_4 (T-t) \lambda_2 e^{-\lambda_2(T-t)}. \quad (6)$$

Substituting Equation (6) into Equation (3), we obtain the Nelson–Siegel–Svensson yield:

$$y(t, T) = \beta_1 + \left( \frac{1 - e^{-\lambda_1(T-t)}}{\lambda_1(T-t)} \right) \beta_2 + \left( \frac{1 - e^{-\lambda_1(T-t)}}{\lambda_1(T-t)} - e^{-\lambda_1(T-t)} \right) \beta_3 + \left( \frac{1 - e^{-\lambda_2(T-t)}}{\lambda_2(T-t)} - e^{-\lambda_2(T-t)} \right) \beta_4. \quad (7)$$

Diebold and Li (2006) proposed the dynamic Nelson–Siegel model by allowing time-varying parameters. For convenience, we denote  $y(t, T)$  by  $y_t(\tau)$  with  $\tau = T - t$ . The DNS model is written as:

$$y_t(\tau) = \beta_{1t} + \beta_{2t} \left( \frac{1 - e^{-\lambda_t \tau}}{\lambda_t \tau} \right) + \beta_{3t} \left( \frac{1 - e^{-\lambda_t \tau}}{\lambda_t \tau} - e^{-\lambda_t \tau} \right), \quad (8)$$

where  $t$  denotes time.

Similarly, the DNSS is described as:

$$y_t(\tau) = \beta_{1t} + \beta_{2t} \left( \frac{1 - e^{-\lambda_{1t}\tau}}{\lambda_{1t}\tau} \right) + \beta_{3t} \left( \frac{1 - e^{-\lambda_{1t}\tau}}{\lambda_{1t}\tau} - e^{-\lambda_{1t}\tau} \right) + \beta_{4t} \left( \frac{1 - e^{-\lambda_{2t}\tau}}{\lambda_{2t}\tau} - e^{-\lambda_{2t}\tau} \right). \tag{9}$$

In the next section, we show the empirical models of Equations (8) and (9).

### 2.2. Affine Term Structure Model

In this sub-section, following [Duffie and Kan \(1996\)](#), we show the properties of the affine term structure model. Suppose that  $N$  state variables follow a vector autoregression(VAR)(1) process under the physical measure denoted by  $\mathbb{P}$ . The continuous representation of the VAR (1) process is:

$$dX_t = \kappa^{\mathbb{P}}(\theta^{\mathbb{P}} - X_t)dt + \Sigma dW_t^{\mathbb{P}}, \tag{10}$$

where  $\kappa^{\mathbb{P}}$  is an  $N \times N$  matrix,  $\theta^{\mathbb{P}}$  is an  $N \times 1$  matrix,  $\Sigma$  is an  $N \times N$  matrix, and  $W_t^{\mathbb{P}}$  is a Brownian motion under the physical measure  $\mathbb{P}$ . This form is called the Ornstein–Uhlenbeck process, as used in [Vasicek \(1977\)](#).

Suppose that the risk premium  $\Gamma_t$  is defined in the essentially affine form, as proposed by [Duffee \(2002\)](#),

$$\Gamma_t = \Gamma_0 + \Gamma_1' X_t, \tag{11}$$

where the scalar  $\Gamma_0$  is a constant term, and  $\Gamma_1$  is an  $N \times 1$  matrix.

Under the risk-neutral measure  $\mathbb{Q}$ , the process  $dW_t^{\mathbb{Q}}$  is defined<sup>4</sup> as:

$$dW_t^{\mathbb{Q}} = dW_t^{\mathbb{P}} + \Gamma_t dt. \tag{12}$$

Substituting Equation (12) into Equation (10), the risk-neutral measure process of  $dX_t$  is derived as:

$$dX_t = \kappa^{\mathbb{Q}}(\theta^{\mathbb{Q}} - X_t)dt + \Sigma dW_t^{\mathbb{Q}}, \tag{13}$$

where  $\kappa^{\mathbb{Q}}$  is an  $N \times N$  matrix,  $\theta^{\mathbb{Q}}$  is an  $N \times 1$  matrix, and  $W_t^{\mathbb{Q}}$  is a Brownian motion under the risk-neutral measure  $\mathbb{Q}$ .

Suppose that the instantaneous risk-free rate ( $r$ ) is also an affine function:

$$r = \rho_0 + \rho_1 X_t, \tag{14}$$

where  $\rho_0$  is a scalar and  $\rho_1$  is a row vector.

The zero-coupon bond prices are an exponential affine function:

$$P(t, T) = \exp[B(t, T)' X_t + C(t, T)], \tag{15}$$

where  $B(t, T)$  and  $C(t, T)$  are the solutions of the ordinary differential equations:

$$\frac{dB(t, T)}{dt} = \rho_1 + (\kappa^{\mathbb{Q}})' B(t, T), \text{ and} \tag{16}$$

$$\frac{dC(t, T)}{dt} = \rho_0 - B(t, T)' \kappa^{\mathbb{Q}} \theta^{\mathbb{Q}} - \frac{1}{2} \sum_{j=1}^n (\Sigma' B(t, T) B(t, T) \Sigma), \tag{17}$$

---

<sup>4</sup> In real world, investors and market participations behave with risk-aversion, so that we need to consider the risk premium to ensure the model is risk-neutral.

where  $n$  denote the number of the factor.

When  $B(T, T) = 0$  and  $C(T, T) = 0$ , the zero-coupon bond price  $P(t, T)$  is equal to one. This implies that the bond price corresponds with the par value at maturity, so that the boundary conditions are  $B(T, T) = 0$  and  $C(T, T) = 0$ . From the price function of the zero-coupon bond prices, the yield function is written as:

$$y(t, T) = -\frac{\log P(t, T)}{T-t} = -\frac{B(t, T)'}{T-t} X_t - \frac{C(t, T)}{T-t}. \tag{18}$$

This function is also the affine function. The details of the derivation are in Appendix A.

### 2.3. Continuous-Time AFNS Framework

In the previous two sub-sections, we have shown the derivations of DNS models and the affine term structure model. In this sub-section, following Christensen et al. (2011), we derived the AFNS model by combining the DNS model with the affine term structure model. In the three-factor model, the  $j$ -th elements of  $B(t, T)$  in Equation (18), denoted by  $B^j$ , are as follows:

$$B^1(t, T) = -(T-t), \tag{19}$$

$$B^2(t, T) = -\frac{1 - e^{-\lambda(T-t)}}{\lambda}, \tag{20}$$

$$B^3(t, T) = (T-t)e^{-\lambda(T-t)} - \frac{1 - e^{-\lambda(T-t)}}{\lambda}. \tag{21}$$

Then, Equation (18) is written as

$$y(t, T) = X_t^1 + \left[ \frac{1 - e^{-\lambda(T-t)}}{\lambda(T-t)} \right] X_t^2 + \left[ \frac{1 - e^{-\lambda(T-t)}}{\lambda(T-t)} - e^{-\lambda(T-t)} \right] X_t^3 - \frac{C(t, T)}{T-t}. \tag{22}$$

This equation matches DNS model, except for the last term—which is often called the yield adjustment term or Jensen’s inequality term. This term arises from the convexity of the function and plays a role describing the non-linearity of the yield curve.

When we suppose the three-factor model, Equations (16) and (17) are rewritten as:

$$\frac{dB(t, T)}{dt} = \rho_1 + (\kappa^{\mathbb{Q}})' B(t, T), \tag{23}$$

$$\frac{dC(t, T)}{dt} = \rho_0 - B(t, T)' \kappa^{\mathbb{Q}} \theta^{\mathbb{Q}} - \frac{1}{2} \sum_{j=1}^3 (\Sigma' B(t, T) B(t, T) \Sigma). \tag{24}$$

From Equation (23) and the relationship

$$\frac{d}{dt} \left[ e^{(\kappa^{\mathbb{Q}})'(T-t)} B(t, T) \right] = e^{(\kappa^{\mathbb{Q}})'(T-t)} \frac{dB(t, T)}{dt} - e^{(\kappa^{\mathbb{Q}})'(T-t)} (\kappa^{\mathbb{Q}})' B(t, T), \tag{25}$$

we obtain

$$\frac{d}{dt} \left[ e^{(\kappa^{\mathbb{Q}})'(T-t)} B(t, T) \right] = e^{(\kappa^{\mathbb{Q}})'(T-t)} \rho_1. \tag{26}$$

Integrating Equation (26), we get:

$$\int_t^T \frac{d}{ds} \left[ e^{(\kappa^{\mathbb{Q}})'(T-s)} B(s, T) \right] ds = \int_t^T e^{(\kappa^{\mathbb{Q}})'(T-s)} \rho_1 ds. \tag{27}$$

From the boundary condition  $B(T, T) = 0$ , the solution is:

$$B(t, T) = -e^{-(\kappa^{\mathbb{Q}})'(T-s)} \int_t^T e^{(\kappa^{\mathbb{Q}})'(T-s)} \rho_1 ds. \tag{28}$$

When we impose the following structure<sup>5</sup>,

$$\kappa^{\mathbb{Q}} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & -\lambda & \lambda \end{bmatrix}, \tag{29}$$

and

$$\rho_0 = 0, \rho_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}. \tag{30}$$

From the structures (30) and (31),

$$e^{-(\kappa^{\mathbb{Q}})'(T-t)} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & e^{\lambda(T-t)} & 0 \\ 0 & \lambda(T-t)e^{-\lambda(T-t)} & e^{-\lambda(T-t)} \end{bmatrix}. \tag{31}$$

Substituting above results into the ordinary differential equation system (29), we obtain:

$$B(t, T) = - \begin{bmatrix} 1 & 0 & 0 \\ 0 & e^{-\lambda(T-t)} & 0 \\ 0 & \lambda(T-t)e^{-\lambda(T-t)} & e^{-\lambda(T-t)} \end{bmatrix} \times \int_t^T \begin{bmatrix} 1 \\ e^{-\lambda(T-t)} \\ -\lambda(T-t)e^{-\lambda(T-t)} \end{bmatrix} ds, \tag{32}$$

and the solution is:

$$B(t, T) = \begin{bmatrix} -(T-t) \\ -\frac{1-e^{-\lambda(T-t)}}{\lambda} \\ (T-t)e^{-\lambda(T-t)} - \frac{1-e^{-\lambda(T-t)}}{\lambda} \end{bmatrix}. \tag{33}$$

Each element of Equation (33) corresponds to Equations (19)–(21), respectively, such that we can set the DNS model into the affine term structure model, except for the term  $C(t, T)$ .

Next, we consider the yield adjustment term  $-C(t, T)/(T-t)$ . From Equation (24), we can analytically derive the yield adjustment term. In the risk-neutral measure  $\mathbb{Q}$ ,  $\theta^{\mathbb{Q}}$  is zero. Therefore, the yield adjustment term is:

$$\frac{C(t, T)}{T-t} = \frac{1}{2(T-t)} \sum_{j=1}^3 \int_t^T (\Sigma' B(t, T) B(t, T) \Sigma) ds. \tag{34}$$

As for the dynamics of the state transition equations, there are two patterns in the yield adjustment term, depending on whether the factor dynamics are independent or correlated. If the state factor has independent factor dynamics, the volatility matrix  $\Sigma$  is represented as:

$$\Sigma = \begin{bmatrix} \sigma_{11} & 0 & 0 \\ 0 & \sigma_{22} & 0 \\ 0 & 0 & \sigma_{33} \end{bmatrix}. \tag{35}$$

---

<sup>5</sup> The derivation of the restriction is in the Appendix A of Christensen et al. (2011).

Then, the yield adjustment term is written as:

$$\begin{aligned} \frac{C(t,T)}{T-t} = & \bar{A}\left(\frac{(T-t)^2}{6}\right) + \bar{B}\left(\frac{1}{2\lambda^2} - \frac{1}{\lambda^3} \frac{1-e^{-\lambda(T-t)}}{T-t} + \frac{1}{4\lambda^3} \frac{1-e^{-2\lambda(T-t)}}{T-t}\right) + \bar{C}\left(\frac{1}{2\lambda^2} + \frac{1}{\lambda^2} e^{-\lambda(T-t)} - \right. \\ & \left. \frac{1}{4\lambda}(T-t)e^{-2\lambda(T-t)} - \frac{3}{4\lambda^2} e^{-2\lambda(T-t)} - \frac{2}{\lambda^3} \frac{1-e^{-\lambda(T-t)}}{T-t} - \frac{2}{8\lambda^3} \frac{1-e^{-2\lambda(T-t)}}{T-t}\right), \end{aligned} \tag{36}$$

where:

$$\begin{aligned} \bar{A} &= \sigma_{11}^2 + \sigma_{12}^2 + \sigma_{13}^2, \\ \bar{B} &= \sigma_{21}^2 + \sigma_{22}^2 + \sigma_{23}^2, \\ \bar{C} &= \sigma_{31}^2 + \sigma_{32}^2 + \sigma_{33}^2. \end{aligned}$$

On the other hand, if the factor dynamics are correlated, the volatility matrix  $\Sigma$  is represented as<sup>6</sup>:

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{bmatrix}. \tag{37}$$

Then, the yield adjustment term is:

$$\begin{aligned} \frac{C(t,T)}{T-t} = & \bar{A}\left(\frac{(T-t)^2}{6}\right) + \bar{B}\left(\frac{1}{2\lambda^2} - \frac{1}{\lambda^3} \frac{1-e^{-\lambda(T-t)}}{T-t} + \frac{1}{4\lambda^3} \frac{1-e^{-2\lambda(T-t)}}{T-t}\right) + \bar{C}\left(\frac{1}{2\lambda^2} + \frac{1}{\lambda^2} e^{-\lambda(T-t)} - \right. \\ & \left. \frac{1}{4\lambda}(T-t)e^{-2\lambda(T-t)} - \frac{3}{4\lambda^2} e^{-2\lambda(T-t)} - \frac{2}{\lambda^3} \frac{1-e^{-\lambda(T-t)}}{T-t} + \frac{5}{8\lambda^3} \frac{1-e^{-2\lambda(T-t)}}{T-t}\right) + \bar{D}\left(\frac{1}{2\lambda}(T-t) + \right. \\ & \left. \frac{1}{\lambda^2} e^{-\lambda(T-t)} - \frac{1}{\lambda^3} \frac{1-e^{-\lambda(T-t)}}{T-t}\right) + \bar{E}\left(\frac{3}{\lambda^2} e^{-\lambda(T-t)} + \frac{1}{2\lambda}(T-t) + \frac{1}{\lambda}(T-t)e^{-\lambda(T-t)} - \right. \\ & \left. \frac{3}{\lambda^3} \frac{1-e^{-\lambda(T-t)}}{T-t}\right) + \bar{F}\left(\frac{1}{\lambda^2} + \frac{1}{\lambda^2} e^{-\lambda(T-t)} - \frac{1}{2\lambda^2} e^{-2\lambda(T-t)} - \frac{3}{\lambda^3} \frac{1-e^{-\lambda(T-t)}}{T-t} + \frac{3}{4\lambda^3} \frac{1-e^{-\lambda(T-t)}}{T-t}\right), \end{aligned} \tag{38}$$

where:

$$\begin{aligned} \bar{D} &= \sigma_{11}\sigma_{21} + \sigma_{12}\sigma_{22} + \sigma_{13}\sigma_{23}, \\ \bar{E} &= \sigma_{11}\sigma_{31} + \sigma_{12}\sigma_{32} + \sigma_{13}\sigma_{33}, \\ \bar{F} &= \sigma_{21}\sigma_{31} + \sigma_{22}\sigma_{32} + \sigma_{23}\sigma_{33}. \end{aligned}$$

The details of the derivation of Equations (36) and (38) are in Appendix A.

The yield adjustment term depends only on the time of maturity  $\tau (= T - t)$ . Therefore, it is a constant value with time  $t$ . Moreover, as the maturity becomes larger, the value of the yield adjustment term is larger.

#### 2.4. Discrete-Time AFNS

As it is easier to match data with a discrete AFNS model, rather than with the continuous version, we derive the discrete-time AFNS model, following Niu (2012). Under the physical measure  $\mathbb{P}$ , the dynamics of the state variables are written as:

$$X_t = \mu^{\mathbb{P}} + A^{\mathbb{P}} X_{t-1} + \eta_t^{\mathbb{P}}, \tag{39}$$

<sup>6</sup> There are the nine volatility parameters, but they are not separately identified. Only the six terms  $\bar{A}$ ,  $\bar{B}$ ,  $\bar{C}$ ,  $\bar{D}$ ,  $\bar{E}$ , and  $\bar{F}$  can be identified. Following Christensen et al. (2011), we specify the matrix as the triangular volatility matrix,

$\Sigma = \begin{bmatrix} \sigma_{11} & 0 & 0 \\ \sigma_{21} & \sigma_{22} & 0 \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{bmatrix}$ . Krippner (2015) suggested using the Cholesky decomposition representation.

where  $\eta_t^{\mathbb{P}} \stackrel{\mathbb{P}}{\sim} N(0, Q)$ . Under the risk-neutral measure  $\mathbb{Q}$ , the dynamics of the state variables are written as:

$$X_t = \mu^{\mathbb{Q}} + A^{\mathbb{Q}}X_{t-1} + \eta_t^{\mathbb{Q}}, \tag{40}$$

where  $\eta_t^{\mathbb{Q}} \sim N(0, Q)$ . The discrete versions<sup>7</sup> of Equations (23) and (24) are represented as:

$$B(t, T)' = -\rho_1' \sum_{k=0}^{T-t-1} (\Phi^{\mathbb{Q}})^k \text{ and} \tag{41}$$

$$C(t, T) = C(t, T-1) + C(t, T-1)' \theta^{\mathbb{Q}} + \frac{1}{2} C(t, T-1) Q C(t, T-1), \tag{42}$$

where:

$$\rho_1' = \begin{bmatrix} 1 \\ \frac{1-e^{-\lambda}}{\lambda} \\ \frac{1-e^{-\lambda}}{\lambda} - e^{-\lambda} \end{bmatrix}, \tag{43}$$

and:

$$(\Phi^{\mathbb{Q}})^k = \begin{bmatrix} 1 & 0 & 0 \\ 0 & e^{-\lambda} & \lambda e^{-\lambda} \\ 0 & 0 & e^{-\lambda} \end{bmatrix}^k = \begin{bmatrix} 1 & 0 & 0 \\ 0 & e^{-k\lambda} & k\lambda e^{-k\lambda} \\ 0 & 0 & e^{-k\lambda} \end{bmatrix}. \tag{44}$$

From Equations (41) and (44), we obtain:

$$\begin{aligned} B(t, T)' &= -\rho_1' \sum_{k=0}^{T-t-1} (\Phi^{\mathbb{Q}})^k \\ &= \begin{bmatrix} -n \\ -\frac{1-e^{-\lambda n}}{\lambda} \\ ne^{-\lambda n} - \frac{1-e^{-\lambda n}}{\lambda} \end{bmatrix}. \end{aligned} \tag{45}$$

We have that Equation (45) is the discrete version of Equation (33).

Moreover, in the risk-neutral measure  $\mathbb{Q}$ ,  $\theta^{\mathbb{Q}} = 0$ . Therefore, the yield adjustment term of the discrete version is written as:

$$C(t, T) = C(t, T-1) + \frac{1}{2} C(t, T-1) Q C(t, T-1). \tag{46}$$

Substituting Equations (43) and (44) into Equation (19), we obtain the discrete-time AFNS model as:

$$y(t, T) = X_t^1 + \left[ \frac{1 - e^{-\lambda(T-t)}}{\lambda(T-t)} \right] X_t^2 + \left[ \frac{1 - e^{-\lambda(T-t)}}{\lambda(T-t)} - e^{-\lambda(T-t)} \right] X_t^3 - \frac{C(t, T)}{T-t}, \tag{47}$$

where  $C(t, T)$  is given by Equation (46).

Until now, we have shown theoretical models of DNS, DNSS, and AFNS (continuous- and discrete-time versions). In the next section, we show empirical models of Equations (8), (9), and (45). Moreover, we compare the forecast performances of the yields.

### 3. Empirical Models

For the purpose of evaluating forecasting performance of the three yield curve models, it is convenient to represent the empirical counterparts of the theoretical models explicitly.

---

<sup>7</sup> The elements of the column vector  $\rho_1$  are a difference between the continuous version and the discrete version. However, it is  $[1 \quad (1 - e^{-\lambda})/\lambda \quad (1 - e^{-\lambda})/\lambda - e^{-\lambda}]' \approx [1 \quad 1 \quad 0]'$ .

As we cannot observe the state variables, we must estimate them first. As Piazzesi (2010) used, the most frequently used way of estimating state variables is to use the principal component method. As Litterman and Scheinkman (1991) first showed, it is usually considered that the yield curve is explained almost by the first principal components. Moreover, the first, second, and third components are usually identified as the level, slope, and curvature of the yield curve, respectively.

In this paper, we supposed that the transition equation of the state variables followed a VAR (1) process, and that the state variables were uncorrelated or correlated with each other. The uncorrelated transition equation of the state variables is represented as:

$$\begin{bmatrix} X_t^{1,DNS} \\ X_t^{2,DNS} \\ X_t^{3,DNS} \end{bmatrix} = (I - A)\mu + \begin{bmatrix} A_{11} & 0 & 0 \\ 0 & A_{22} & 0 \\ 0 & 0 & A_{33} \end{bmatrix} \begin{bmatrix} X_{t-1}^{1,DNS} \\ X_{t-1}^{2,DNS} \\ X_{t-1}^{3,DNS} \end{bmatrix} + \begin{bmatrix} \eta_t^1 \\ \eta_t^2 \\ \eta_t^3 \end{bmatrix}, \tag{48}$$

where  $\mu$  is a vector of means. The variance–covariance matrix of the stochastic shocks is represented as:

$$Q^{DNS} = \begin{bmatrix} Q_{11} & 0 & 0 \\ 0 & Q_{22} & 0 \\ 0 & 0 & Q_{33} \end{bmatrix}. \tag{49}$$

From Equation (8), the empirical model of DNS is represented as:

$$\begin{bmatrix} y_t(\tau_1) \\ y_t(\tau_2) \\ \vdots \\ y_t(\tau_N) \end{bmatrix} = \begin{bmatrix} 1 & \frac{1-e^{-\lambda\tau_1}}{\lambda\tau_1} & \frac{1-e^{-\lambda\tau_1}}{\lambda\tau_1} - e^{-\lambda\tau_1} \\ 1 & \frac{1-e^{-\lambda\tau_2}}{\lambda\tau_2} & \frac{1-e^{-\lambda\tau_2}}{\lambda\tau_2} - e^{-\lambda\tau_2} \\ \vdots & \vdots & \vdots \\ 1 & \frac{1-e^{-\lambda\tau_N}}{\lambda\tau_N} & \frac{1-e^{-\lambda\tau_N}}{\lambda\tau_N} - e^{-\lambda\tau_N} \end{bmatrix} \begin{bmatrix} X_t^{1,DNS} \\ X_t^{2,DNS} \\ X_t^{3,DNS} \end{bmatrix} + \begin{bmatrix} \varepsilon_t(\tau_1) \\ \varepsilon_t(\tau_2) \\ \vdots \\ \varepsilon_t(\tau_N) \end{bmatrix}, \tag{50}$$

where  $\varepsilon_t(\tau_i)$  is the measurement error and i.i.d. white noise and  $\tau_N$  denotes a maturity  $N$ .

Similarly, we consider the uncorrelated state dynamics for the DNSS,

$$\begin{bmatrix} X_t^{1,DNSS} \\ X_t^{2,DNSS} \\ X_t^{3,DNSS} \\ X_t^{4,DNSS} \end{bmatrix} = (I - A)\mu + \begin{bmatrix} A_{11} & 0 & 0 & 0 \\ 0 & A_{22} & 0 & 0 \\ 0 & 0 & A_{33} & 0 \\ 0 & 0 & 0 & A_{44} \end{bmatrix} \begin{bmatrix} X_{t-1}^{1,DNSS} \\ X_{t-1}^{2,DNSS} \\ X_{t-1}^{3,DNSS} \\ X_{t-1}^{4,DNSS} \end{bmatrix} + \begin{bmatrix} \eta_t^1 \\ \eta_t^2 \\ \eta_t^3 \\ \eta_t^4 \end{bmatrix}, \tag{51}$$

where the variance–covariance matrix of the stochastic shocks is represented as:

$$Q^{DNSS} = \begin{bmatrix} Q_{11} & 0 & 0 & 0 \\ 0 & Q_{22} & 0 & 0 \\ 0 & 0 & Q_{33} & 0 \\ 0 & 0 & 0 & Q_{44} \end{bmatrix}. \tag{52}$$

From Equation (9), the empirical model of DNSS is represented as:

$$\begin{bmatrix} y_t(\tau_1) \\ y_t(\tau_2) \\ \vdots \\ y_t(\tau_N) \end{bmatrix} = \begin{bmatrix} 1 & \frac{1-e^{-\lambda_1\tau_1}}{\lambda_1\tau_1} & \frac{1-e^{-\lambda_1\tau_1}}{\lambda_1\tau_1} - e^{-\lambda_1\tau_1} & \frac{1-e^{-\lambda_2\tau_1}}{\lambda_2\tau_1} - e^{-\lambda_2\tau_1} \\ 1 & \frac{1-e^{-\lambda_1\tau_2}}{\lambda_1\tau_2} & \frac{1-e^{-\lambda_1\tau_2}}{\lambda_1\tau_2} - e^{-\lambda_1\tau_2} & \frac{1-e^{-\lambda_2\tau_2}}{\lambda_2\tau_2} - e^{-\lambda_2\tau_2} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & \frac{1-e^{-\lambda_1\tau_N}}{\lambda_1\tau_N} & \frac{1-e^{-\lambda_1\tau_N}}{\lambda_1\tau_N} - e^{-\lambda_1\tau_N} & \frac{1-e^{-\lambda_2\tau_N}}{\lambda_2\tau_N} - e^{-\lambda_2\tau_N} \end{bmatrix} \begin{bmatrix} X_t^{1,DNSS} \\ X_t^{2,DNSS} \\ X_t^{3,DNSS} \\ X_t^{4,DNSS} \end{bmatrix} + \begin{bmatrix} \varepsilon_t(\tau_1) \\ \varepsilon_t(\tau_2) \\ \vdots \\ \varepsilon_t(\tau_N) \end{bmatrix}. \tag{53}$$

From Equation (39), suppose that the state transition equation of AFNS also follows a VAR (1) process, as in DNS:

$$\begin{bmatrix} X_t^{1,AFNS} \\ X_t^{2,AFNS} \\ X_t^{3,AFNS} \end{bmatrix} = (I - A)\mu + \begin{bmatrix} A_{11} & 0 & 0 \\ 0 & A_{22} & 0 \\ 0 & 0 & A_{33} \end{bmatrix} \begin{bmatrix} X_{t-1}^{1,AFNS} \\ X_{t-1}^{2,AFNS} \\ X_{t-1}^{3,AFNS} \end{bmatrix} + \begin{bmatrix} \eta_t^1 \\ \eta_t^2 \\ \eta_t^3 \end{bmatrix}, \tag{54}$$

where the stochastic shock variance–covariance matrix is represented as:

$$Q^{AFNS} = \begin{bmatrix} Q_{11} & 0 & 0 \\ 0 & Q_{22} & 0 \\ 0 & 0 & Q_{33} \end{bmatrix}. \tag{55}$$

From Equation (45), the empirical model of AFNS is represented as:

$$\begin{bmatrix} y_t(\tau_1) \\ y_t(\tau_2) \\ \vdots \\ y_t(\tau_N) \end{bmatrix} = \begin{bmatrix} 1 & \frac{1-e^{-\lambda\tau_1}}{\lambda\tau_1} & \frac{1-e^{-\lambda\tau_1}}{\lambda\tau_1} - e^{-\lambda\tau_1} \\ 1 & \frac{1-e^{-\lambda\tau_2}}{\lambda\tau_2} & \frac{1-e^{-\lambda\tau_2}}{\lambda\tau_2} - e^{-\lambda\tau_2} \\ \vdots & \vdots & \vdots \\ 1 & \frac{1-e^{-\lambda\tau_N}}{\lambda\tau_N} & \frac{1-e^{-\lambda\tau_N}}{\lambda\tau_N} - e^{-\lambda\tau_N} \end{bmatrix} \begin{bmatrix} X_t^{1,AFNS} \\ X_t^{2,AFNS} \\ X_t^{3,AFNS} \end{bmatrix} - \frac{C(t, T)}{T - t} + \begin{bmatrix} \varepsilon_t(\tau_1) \\ \varepsilon_t(\tau_2) \\ \vdots \\ \varepsilon_t(\tau_N) \end{bmatrix}. \tag{56}$$

There is a yield adjustment term  $C(t, T)/(T - t)$  in this equation, which generates a different forecast than the DNS model. Simply stated, the arbitrage-free condition imposes this term, which is a feature of the AFNS and affine term structure models that distinguishes them from curve-fitting models. The derivation of the AFNS model, in detail, is in Appendix A.

Next, we considered the correlated transition equation of the state variables. The representations of the state transition equation of the correlated state variables are as follows:

$$A^{DNS} = \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{bmatrix}, \tag{57}$$

$$A^{DNSS} = \begin{bmatrix} A_{11} & A_{12} & A_{13} & A_{14} \\ A_{21} & A_{22} & A_{23} & A_{23} \\ A_{31} & A_{32} & A_{33} & A_{34} \\ A_{41} & A_{42} & A_{43} & A_{44} \end{bmatrix}, \tag{58}$$

$$A^{AFNS} = \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{bmatrix}. \tag{59}$$

The variance-covariance matrixes of  $\eta_t$  are written as:

$$Q^{DNS} = \begin{bmatrix} Q_{11} & Q_{12} & Q_{13} \\ Q_{21} & Q_{22} & Q_{23} \\ Q_{31} & Q_{32} & Q_{33} \end{bmatrix}, \tag{60}$$

$$Q^{DNSS} = \begin{bmatrix} Q_{11} & Q_{12} & Q_{13} & Q_{14} \\ Q_{21} & Q_{22} & Q_{23} & Q_{23} \\ Q_{31} & Q_{32} & Q_{33} & Q_{34} \\ Q_{41} & Q_{42} & Q_{43} & Q_{44} \end{bmatrix}, \tag{61}$$

$$Q^{AFNS} = \begin{bmatrix} Q_{11} & Q_{12} & Q_{13} \\ Q_{21} & Q_{22} & Q_{23} \\ Q_{31} & Q_{32} & Q_{33} \end{bmatrix}. \quad (62)$$

Even though each state variable is correlated, the empirical DNS, DNSS, and AFNS models have the same representation.

In summary, we can write the above equation simply, in matrix representation. Each state transition equation of the DNS, DNSS, and AFNS models, in matrix form, is represented as:

$$X_t = (I - A)\mu + AX_{t-1} + H_t, \quad (63)$$

where  $X_t$  is a  $3 \times 1$  matrix (DNS and AFNS) or  $4 \times 1$  matrix (DNSS),  $A$  is a state transition matrix of  $3 \times 3$  (DNS and AFNS) or  $4 \times 4$  (DNSS),  $H_t$  is a  $N \times 1$  matrix such that  $H_t \sim N(0, Q)$ , and  $Q$  is a variance–covariance matrix of  $3 \times 3$  (DNS and AFNS) or  $4 \times 4$  (DNSS).

In matrix representation, the empirical DNS and DNSS models are represented as:

$$Y_t = BX_t + t, \quad (64)$$

where  $Y_t$  is a  $N \times 1$  matrix,  $B_t$  is a matrix of  $3 \times N$  (DNS) or  $4 \times N$  (DNSS), and  $t$  is a  $N \times 1$  matrix such that  $t \sim N(0, I)$ .

On the other hand, the empirical model of AFNS is represented as:

$$Y_t = BX_t + C + t, \quad (65)$$

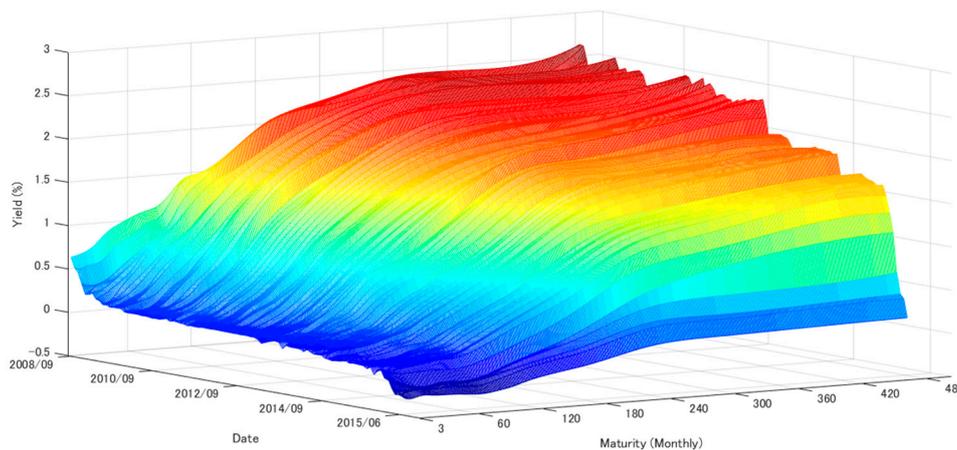
where  $C$  is the yield adjustment term, which is a  $N \times 1$  matrix corresponding to the term  $-C(t, T)/(T - t)$  in Equation (23) or (45).

#### 4. Estimation

In this section, we explained the details of the data used in the empirical analysis and report the estimated parameters and the state variables of the DNS, DNSS, and AFNS models. We employed monthly JGB zero-coupon yield from September 2009 to June 2015. The zero-yields were calculated<sup>8</sup> using the treasury discount bill and the prevailing interest rates, published by the Ministry of Finance. The maturities are 3, 6, 12, 24, 36, 48, 60, 72, 84, 96, 108, 120, 180, 240, 300, 360, and 480 months. Figure 1 is the JGB zero-yield curve (September 2009 to June 2015).

In the first step, we estimated the scale parameters  $\lambda$ ,  $\lambda_1$ , and  $\lambda_2$  for the DNS and DNSS models using non-linear least-squares. In the second step, we estimated the state variables for the DNS and DNSS models using the estimated parameters  $\lambda$ ,  $\lambda_1$ , and  $\lambda_2$ , and OLS. Moreover, we estimated the transition matrix  $A$  and the variance–covariance matrix  $Q$  for each independent and correlated state variable. The estimated state variables for DNS are invalid for using for the AFNS, since the constant term includes a bias. Using the estimated variance–covariance matrix, we could obtain the values of  $-C(t, T)/(T - t)$  from Equation (42). Subtracting the yield adjustment term  $-C(t, T)/(T - t)$  from the yields, we excluded the constant term bias, so that we could estimate the state variable for the AFNS model.

<sup>8</sup> We used the MATLAB financial and financial instruments toolboxes to translate the par-yield (the treasury discount bill and the prevailing interest rates) into the zero-coupon yield.



**Figure 1.** Japanese government bond zero-coupon yield curve (2008/09–2015/06). Note: Figure 1 plotted the Japanese government bond zero-coupon yield curve. We used the yields of maturities of 3, 6, 12, 24, 36, 48, 60, 72, 84, 96, 108, 120, 180, 240, 300, 360, and 480 months, and the term was from September 2009 to June 2015 monthly data. To calculate the zero-yield of the Japanese government bond, the treasury bill and the prevailing interest rates, which are published by the Japanese Ministry of Finance were used. To interpolate of the missing maturities, the cubic spline was used.

#### 4.1. Estimation of the Values of $\lambda$

In this sub-section, we estimated the scale parameters  $\lambda$ ,  $\lambda_1$ , and  $\lambda_2$  of the DNS and DNSS models (Equations (5) and (7)). The estimation method was non-linear least-squares. As this model is a numerical approximation method, we needed to determine the initial values of the parameters. We set the initial values as  $X_1 = X_2 = X_3 = X_4 = 0$ ,  $\lambda = \lambda_1 = 0.0327^9$ , and  $\lambda_2 = 0.01635^{10}$ . The Levenberg–Marquardt algorithm was used to approximate the values numerically.

As the DNS and DNSS are time-series models, we approximated each parameter, using non-linear least-squares, at each time point. The estimated DNS and DNSS models are written as:

$$\hat{y}_t(\tau) = \hat{\beta}_{1,t} + \hat{\beta}_{2,t} \left( \frac{1 - e^{-\hat{\lambda}_t \tau}}{\hat{\lambda}_t \tau} \right) + \hat{\beta}_{3,t} \left( \frac{1 - e^{-\hat{\lambda}_t \tau}}{\hat{\lambda}_t \tau} - e^{-\hat{\lambda}_t \tau} \right), \tag{66}$$

and:

$$\hat{y}_t(\tau) = \hat{\beta}_{1,t} + \hat{\beta}_{2,t} \left( \frac{1 - e^{-\hat{\lambda}_{1,t} \tau}}{\hat{\lambda}_{1,t} \tau} \right) + \hat{\beta}_{3,t} \left( \frac{1 - e^{-\hat{\lambda}_{1,t} \tau}}{\hat{\lambda}_{1,t} \tau} - e^{-\hat{\lambda}_{1,t} \tau} \right) + \hat{\beta}_{4,t} \left( \frac{1 - e^{-\hat{\lambda}_{2,t} \tau}}{\hat{\lambda}_{2,t} \tau} - e^{-\hat{\lambda}_{2,t} \tau} \right), \tag{67}$$

respectively.

In the DNS and DNSS models, the parameters  $\hat{\lambda}$ ,  $\hat{\lambda}_1$ , and  $\hat{\lambda}_2$  changed over time. However, in previous studies (Diebold and Li (2006); Christensen et al. (2011), and Diebold and Rudebusch (2013)), the parameters were defined to be constant throughout time, as the fluctuation of the values is small in each time point and to ensure the stability of the estimation results. Therefore, we adopted the averages of  $\hat{\lambda}_t$ ,  $\hat{\lambda}_{1,t}$ , and  $\hat{\lambda}_{2,t}$  as the parameters  $\lambda$ ,  $\lambda_1$ , and  $\lambda_2$ , respectively. From the estimation, we obtained the estimated parameters<sup>11</sup>  $\lambda = 0.004$ ,  $\lambda_1 = 0.016$ , and  $\lambda_2 = 0.012$ . Moreover, suppose that the value  $\lambda$  of the DNS model was identical to  $\lambda$  in the AFNS model<sup>12</sup>. From the estimated  $\lambda$ ,  $\lambda_1$ , and  $\lambda_2$ , the factor loadings on  $\beta_1$ ,  $\beta_2$ ,  $\beta_3$ , and  $\beta_4$  of the DNS and DNSS models were calculated. Then,  $X_1$ ,  $X_2$ ,  $X_3$ , and  $X_4$  was simply a rewriting of each  $\beta$ .

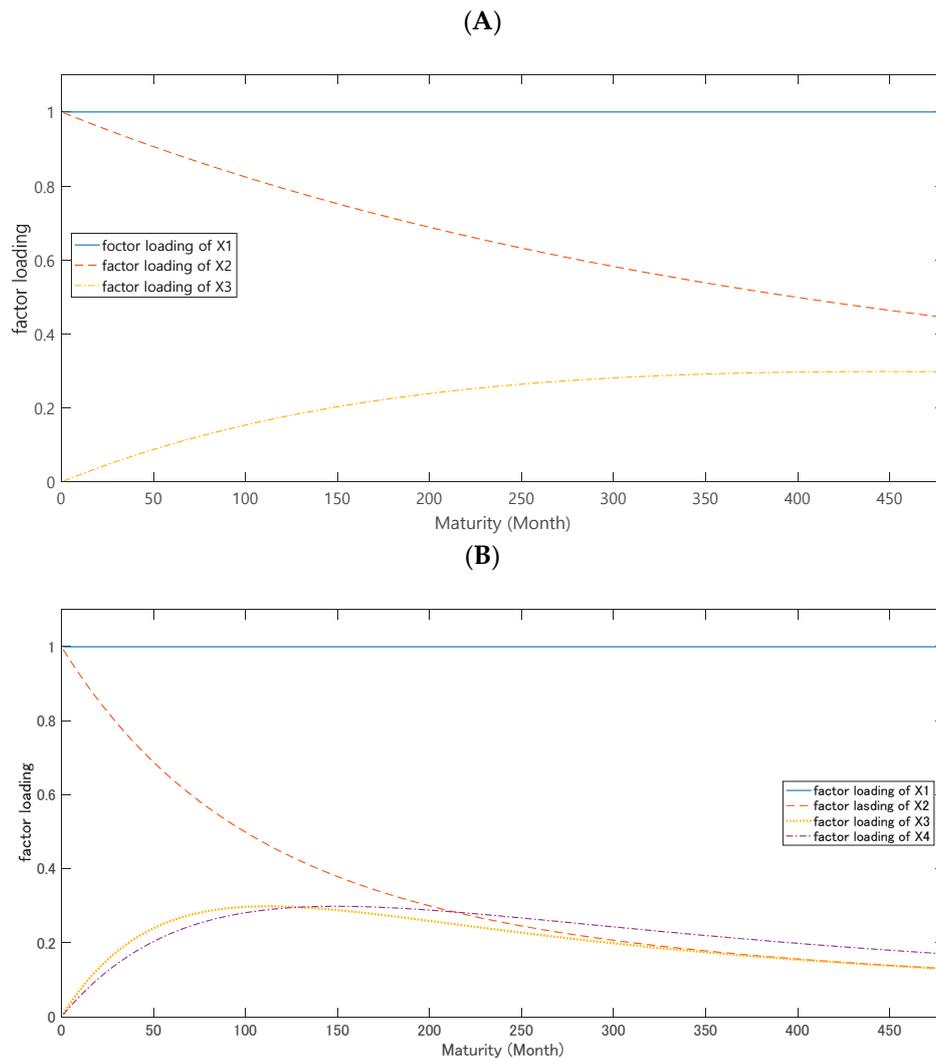
<sup>9</sup> Fujii and Takaoka (2007) reported the estimated parameter  $\lambda = 0.0327$ .

<sup>10</sup> Bjork and Christensen (1999) imposed the restriction that  $\lambda_1 = 2\lambda_2$ .

<sup>11</sup> For U.S. data, Diebold and Rudebusch (2013) adopted  $\lambda_1 = 0.0609$  and  $\lambda_2 = 0.0295$ .

<sup>12</sup> Christensen et al. (2011) estimated each  $\lambda$  of the DNS and the AFNS model separately; however, the values were almost the same.

The dynamics of the factor loadings through the maturity  $\tau$  are described in Figure 2. Figure 2A shows the dynamics of the factor loadings of the DNS model and Figure 2B shows the dynamics of those of the DNSS model. The first factor loading is a constant through the maturity. The second factor loading starts at 1 and goes to 0 gradually, so that it affects short-term yields. The third factor loading starts at 0 and rises in the medium-term, after which it goes to 0 again. The fourth factor loading also starts 0. Its dynamics are initially the same as the third factor. However, the fourth factor loading affects longer-term yields than the third factor loading. As we can see from the above results, the second factor loading was a short-term factor, the third was a medium-term factor, and the fourth was a medium- to long-term factor.



**Figure 2.** Factor loadings of the dynamic Nelson–Siegel (DNS) and the dynamic Nelson–Siegel–Svensson (DNSS) models. Note: The first plot represents the factor loadings of the DNS, and the second represents the factor loadings of the DNSS when  $\lambda = 0.004$ ,  $\lambda_1 = 0.016$ , and  $\lambda_2 = 0.012$ . **A:** the dynamics of the factor loadings of the DNS model; **B:** the dynamics of those of the DNSS model.

#### 4.2. Estimation of Latent State Variables

In the previous sub-section, we specified the values of  $\lambda$ ,  $\lambda_1$ , and  $\lambda_2$ . In this sub-section, we estimated the unobservable state variables of the DNS, DNSS, and AFNS models. However, the AFNS

model includes a constant term. Hence, we cannot use the state variables of the DNS model<sup>13</sup> for the AFNS model, due to the constant term bias. The estimation method for the AFNS model was described later. Here, we carried on the estimation of the second step. If we estimated the parameters without separating between the first step and the second step (i.e., using one-step estimation), we needed to use a state space estimation, such as a Kalman filter. This method imposes a large burden numerically and it is possible that the estimation result is only locally convergent. On the other hand, if we used the multiple regression method, once the value of  $\lambda$  was determined, we could estimate the state variables easily using OLS and, so, we separated the estimation: The first step<sup>14</sup> estimated the scale parameter  $\lambda$ , and the second step estimated the state variables using the value obtained in the first step.

First, we considered the estimation of the state variables in the DNS and DNSS models. Taking the values of  $\lambda$ ,  $\lambda_1$ , and  $\lambda_2$  as above and substituting them into Equations (66) and (67), the parameters  $\hat{\beta}_{1,t}$ ,  $\hat{\beta}_{2,t}$ ,  $\hat{\beta}_{3,t}$ , and  $\hat{\beta}_{4,t}$  were estimated using ordinary least-squares at any one point. Moreover, since the estimated  $\hat{\beta}_{i,t}$  corresponds to the unobservable variables at each time-point, the time-series of each estimated parameter was regarded as the time series of the state variables. Therefore, we redefined the parameters,  $(\hat{\beta}_{1,t}^{DNS}, \hat{\beta}_{2,t}^{DNS}, \hat{\beta}_{3,t}^{DNS})$  and  $(\hat{\beta}_{1,t}^{DNSS}, \hat{\beta}_{2,t}^{DNSS}, \hat{\beta}_{3,t}^{DNSS}, \hat{\beta}_{4,t}^{DNSS})$  as  $(\hat{X}_{1,t}^{DNS}, \hat{X}_{2,t}^{DNS}, \hat{X}_{3,t}^{DNS})$  and  $(\hat{X}_{1,t}^{DNSS}, \hat{X}_{2,t}^{DNSS}, \hat{X}_{3,t}^{DNSS}, \hat{X}_{4,t}^{DNSS})$ .

Next, we explained how we estimated the state variables in AFNS. Although the number of the state variables was the same as the DNS model, the AFNS includes the constant term  $C$ , as shown in Equation (63). Therefore, we could not directly apply the estimated state variables in the DNS to the AFNS, as such a method causes the constant term bias. To correct the bias, we needed to estimate the state variables after subtracting the constant term from the yield.

Defining  $\hat{Z}_t(\tau) \equiv \hat{y}_t(\tau) - \hat{C}$ , we have:

$$\hat{Z}_t(\tau) = \hat{X}_{1,t} + \hat{X}_{2,t} \left( \frac{1 - e^{-\lambda\tau}}{\lambda\tau} \right) + \hat{X}_{3,t} \left( \frac{1 - e^{-\lambda_1\tau}}{\lambda_1\tau} - e^{-\lambda\tau} \right), \quad (68)$$

from Equation (47).

We obtained the estimate  $\hat{C}$  from Equation (46), assuming that the variance–covariance matrix  $Q$  of the AFNS model was the same as that of the DNS model. As we assumed that the parameter  $\lambda$  of the AFNS model should be the same as that of the DNS model, we could apply the OLS method to Equation (68).

Table 1 reports the mean, standard deviation, and the 1- and 12-month lag autocorrelations of the estimated state variables for the DNS, DNSS, and AFNS models. In the state variables for the DNS and DNSS models, there was no difference between the independent or correlated cases, since the state variables were identical.

We saw two remarkable features in the AFNS model. First, regardless of whether the yield adjustment term was independent or correlated, the state variable had a strong auto-correlation with the 1-month lag. Second, the auto-correlation decayed with the 12-month lag. The sign of the mean was the same between the first and second state variables of the DNS and AFNS models.

#### 4.3. Estimation of the Transition Matrix

Now, we proceeded to the estimation of the transition matrix  $A$  in Equations (48), (51), and (54). In this estimation, the data set was the estimated state variables of the previous sub-section. When estimating Equation (48), the covariance matrix  $Q$  was estimated as the sample covariance matrix of

<sup>13</sup> Coroneo et al. (2011) showed that the constant term bias was negligibly small for U.S. data; so, they used the state variables of DNS for AFNS.

<sup>14</sup> In the first step,  $\beta_i$  was estimated; however, it is possible that the value of  $\beta_i$  is a locally convergent value as we used non-linear least-squares in the first step. For this reason, we estimated the value of  $\beta_i$  again, using OLS in the second step. The OLS estimator had more consistency.

the error terms. Table 2 reports the estimation results for the transition matrix and Table 3 reports the variance–covariance matrix.

**Table 1.** Mean, standard deviation, and auto-correlation of the estimated state variables.

	Mean	Std. Dev.	$\rho(1)$	$\rho(12)$
<b>DNS</b>				
$X_1^{DNS}$	0.028	0.016	0.936	0.564
$X_2^{DNS}$	−0.029	0.016	0.939	0.582
$X_3^{DNS}$	0.025	0.033	0.961	0.746
<b>DNSS</b>				
$X_1^{DNSS}$	0.022	0.003	0.783	−0.043
$X_2^{DNSS}$	−0.021	0.003	0.815	0.008
$X_3^{DNSS}$	−0.090	0.027	0.895	0.219
$X_4^{DNSS}$	0.080	0.038	0.932	0.531
<b>AFNS (Ind)</b>				
$X_1^{AFNS}$	0.345	0.016	0.936	0.564
$X_2^{AFNS}$	−0.342	0.016	0.939	0.582
$X_3^{AFNS}$	−0.364	0.033	0.961	0.746
<b>AFNS (Corr)</b>				
$X_1^{AFNS}$	0.210	0.016	0.936	0.564
$X_2^{AFNS}$	−0.210	0.016	0.939	0.582
$X_3^{AFNS}$	−0.195	0.033	0.961	0.746

Note: The values of  $\rho(L)$  represent an autocorrelation of lag  $L$ . The state variables were estimated using OLS at any one point.

**Table 2.** The transition matrix  $A$  of the DNS, DNSS, and arbitrage-free Nelson–Siegel (AFNS) models.

	DNS (Independent)				DNS (Correlated)			
	$A_{.,1}$	$A_{.,2}$	$A_{.,3}$	$A_{.,4}$	$A_{.,1}$	$A_{.,2}$	$A_{.,3}$	$A_{.,4}$
$A_{1.}$	0.936 (0.041)	0	0	-	−0.389 (0.570)	−1.151 (0.529)	−0.071 (0.084)	-
$A_{2.}$	0	0.939 (0.040)	0	-	1.203 (0.571)	1.977 (0.530)	0.070 (0.084)	-
$A_{3.}$	0	0	0.962 (0.034)	-	2.036 (0.976)	1.952 (0.906)	0.984 (0.144)	-
	DNSS (Independent)				DNSS (Correlated)			
$A_{1.}$	0.782 (0.063)	0	0	0	0.419 (0.194)	−0.291 (0.202)	−0.024 (0.034)	−0.025 (0.027)
$A_{2.}$	0	0.815 (0.064)	0	0	0.379 (0.204)	1.060 (0.212)	0.034 (0.035)	0.034 (0.028)
$A_{3.}$	0	0	0.895 (0.049)	0	−1.075 (1.134)	−0.497 (1.184)	0.174 (0.201)	−0.603 (0.155)
$A_{4.}$	0	0	0	0.932 (0.041)	2.316 (1.372)	0.412 (1.431)	−0.840 (0.243)	1.608 (0.188)
	AFNS (Independent)				AFNS (Correlated)			
$A_{1.}$	0.936 (0.041)	0	0	-	−0.389 (0.570)	−1.151 (0.529)	−0.071 (0.084)	-
$A_{2.}$	0	0.939 (0.040)	0	-	1.203 (0.571)	1.977 (0.530)	0.070 (0.084)	-
$A_{3.}$	0	0	0.962 (0.034)	-	2.036 (0.976)	1.952 (0.906)	0.984 (0.144)	-

Note: Standard error is represented in parenthesis.

**Table 3.** The variance–covariance matrix  $Q$  of the DNS and DNSS models.

	DNS (Independent)				DNS (Correlated)			
	$Q_{.,1}$	$Q_{.,2}$	$Q_{.,3}$	$Q_{.,4}$	$Q_{.,1}$	$Q_{.,2}$	$Q_{.,3}$	$Q_{.,4}$
$Q_{1.}$	$3.429 \times 10^{-5}$	0	0	-	$3.213 \times 10^{-5}$	$-3.212 \times 10^{-5}$	$-5.147 \times 10^{-5}$	-
$Q_{2.}$	0	$3.394 \times 10^{-5}$	0	-	$-3.212 \times 10^{-5}$	$3.216 \times 10^{-5}$	$5.141 \times 10^{-5}$	-
$Q_{3.}$	0	0	$0.941 \times 10^{-5}$	-	$-5.147 \times 10^{-5}$	$5.141 \times 10^{-5}$	$9.406 \times 10^{-5}$	-
	DNSS (Independent)				DNSS (Correlated)			
	$Q_{1.}$	$Q_{2.}$	$Q_{3.}$	$Q_{4.}$	$Q_{1.}$	$Q_{2.}$	$Q_{3.}$	$Q_{4.}$
$Q_{1.}$	$3.409 \times 10^{-6}$	0	0	0	$3.183 \times 10^{-6}$	$-3.276 \times 10^{-6}$	$8.170 \times 10^{-6}$	$-1.363 \times 10^{-5}$
$Q_{2.}$	0	$3.777 \times 10^{-6}$	0	0	$-3.276 \times 10^{-6}$	$3.511 \times 10^{-6}$	$-1.091 \times 10^{-5}$	$1.625 \times 10^{-5}$
$Q_{3.}$	0	0	$1.332 \times 10^{-5}$	0	$8.170 \times 10^{-6}$	$-1.091 \times 10^{-5}$	$1.093 \times 10^{-4}$	$-1.260 \times 10^{-4}$
$Q_{4.}$	0	0	0	$1.963 \times 10^{-4}$	$-1.363 \times 10^{-5}$	$1.625 \times 10^{-5}$	$-1.260 \times 10^{-4}$	$1.597 \times 10^{-4}$

Note: The variance–covariance matrix  $Q$  of AFNS is identical to that of DNS.

As Table 2 indicates, almost all of the diagonal elements of the transition matrix were statistically significant in all models. Note that, by assumption, the transition matrix of each DNS model was identical to that of the AFNS model, and that the covariance matrix  $Q$  of the AFNS model was identical to that of the DNS model.

As already explained in Section 4.2, we could calculate the yield adjustment term  $C$  of the AFNS model from the estimated variance–covariance matrix  $Q$ . However, due to large volatility, the yield adjustment term of the AFNS model with independent error term became too huge. Each element of the variance–covariance matrix was from two to ten times larger than what Christensen et al. (2011)<sup>15</sup> obtained using U.S. data. For example, our estimated  $Q$  results in a yield adjustment term of about 20,000 bps for a 480-month maturity. As this is extremely unrealistic, in what follows, we omitted the results of the AFNS model with independent error terms.

### 5. Results: Forecasting Performances of the Three Models

In this section, we estimated the root mean square forecast error (RMSFE) of the 1-month, 3-month, 6-month, and 12-month forecasts for each of the three models.

#### 5.1. Out-of-Sample Forecasting: Comparison of Forecasting Performances

We defined the  $h$  period forecast of the yield with maturity  $\tau$  at time  $t$  as  $y_{t+h,t}(\tau)$ . As in Christensen et al. (2011), the forecasting measurement equations of the three models are written, respectively, as:

$$y_{t+h,t}^{DNS}(\tau) = \mathbb{E}^{\mathbb{P}}[y_{t+h}(\tau)] = \mathbb{E}^{\mathbb{P}}[\hat{X}_{t+h}^1] + \mathbb{E}^{\mathbb{P}}[\hat{X}_{t+h}^2] \left( \frac{1-e^{-\lambda\tau}}{\lambda\tau} \right) + \mathbb{E}^{\mathbb{P}}[\hat{X}_{t+h}^3] \left( \frac{1-e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau} \right). \tag{69}$$

$$y_{t+h,t}^{DNSS}(\tau) = \mathbb{E}^{\mathbb{P}}[\hat{X}_{t+h}^1] + \mathbb{E}^{\mathbb{P}}[\hat{X}_{t+h}^2] \left( \frac{1-e^{-\lambda_1\tau}}{\lambda_1\tau} \right) + \mathbb{E}^{\mathbb{P}}[\hat{X}_{t+h}^3] \left( \frac{1-e^{-\lambda_1\tau}}{\lambda_1\tau} - e^{-\lambda_1\tau} \right) + \mathbb{E}^{\mathbb{P}}[\hat{X}_{t+h}^4] \left( \frac{1-e^{-\lambda_1\tau}}{\lambda_1\tau} - e^{-\lambda_1\tau} \right), \tag{70}$$

<sup>15</sup> Christensen et al. (2011) reported that the variance-covariance matrix was  $Q^{DNS} = \begin{bmatrix} 6.17 \times 10^{-6} & 0 & 0 \\ 0 & 1.11 \times 10^{-5} & 0 \\ 0 & 0 & 5.58 \times 10^{-5} \end{bmatrix}$  using U.S. bond data.

$$y_{t+h,t}^{AFNS}(\tau) = \mathbb{E}^{\mathbb{P}}[y_{t+h}(\tau)] = \mathbb{E}^{\mathbb{P}}[\hat{X}_{t+h}^1] + \mathbb{E}^{\mathbb{P}}[\hat{X}_{t+h}^2] \left(\frac{1-e^{-\lambda\tau}}{\lambda\tau}\right) + \mathbb{E}^{\mathbb{P}}[\hat{X}_{t+h}^3] \left(\frac{1-e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau}\right) - \frac{\hat{C}(\tau)}{\tau}, \tag{71}$$

where  $-\frac{\hat{C}(\tau)}{\tau}$  is the yield adjustment term. As the transition dynamics are first-order, we can write the expectation of the state variables at  $t + h$  as:

$$\mathbb{E}^{\mathbb{P}}[X_{t+h}] = \left(\sum_{i=0}^{h-1} A^i\right)(I - A)\mu + A^h X_t, \tag{72}$$

where  $X_t$  denotes the column vector of the state variables. In this paper, we employed the random walk model (without a drift term) as a benchmark model. This paper investigated the 1-, 3-, 6-, and 12-month ahead forecast ( $h = 1,3,6,12$ ). In out-of-sample forecast, these models were estimated only one time over the full sample periods (September 2009 to June 2015<sup>16</sup>) and then used for forecasting  $h$ -month ahead horizon ( $h = 1,3,6,12$ ).

Tables 4–7 report the root mean square forecast error (RMSFE)<sup>17</sup> of the 1-, 3-, 6-, and 12-month forecasts, respectively. In the tables, the smallest RMSFE is boxed for each maturity and horizon. As these tables show, DNSS was clearly the best forecasting model among the three models for 6- and 12-month forecast horizons. Between two types of the DNSS models, the DNSS model based on the transition equation of the independent (uncorrelated) state variable was superior to the correlated model. It is clear to see that all of the off-diagonal elements of the transition matrix were statistically highly significant in the DNSS model. This superiority is plausible, because the DNSS model is theoretically the generalized model of the DNS model. We can say that the generality of the DNSS model was confirmed by our result. On the other hand, the random walk model was superior to the Nelson–Siegel class of models in the short-period forecast horizons. Table 8 shows the Diebold and Mariano test (Diebold and Mariano (1995)) for testing forecast accuracy. We compared the forecast accuracy of the DNSS model with the random walk model for the 6- and 12-month forecast horizons. The positive bold value shows that the DNSS model outperformed the random walk model at a 5% statistically significant level.

<sup>16</sup> This paper employed the monthly frequency. The number of the observations is 82.

<sup>17</sup> The RMSFE is the forecast error using the estimated value from Equation (72). We calculate the RMSFE as  $\frac{1}{h} \sum_{k=1}^h \sqrt{(Y_k - \hat{Y}_{k|k=1})^2}$ .

**Table 4.** Root mean squared forecast error of a 1-month forecast.

Maturity	RW	DNS		DNSS		AFNS
		Ind	Corr	Ind	Corr	Corr
3-month	0.20	13.56	12.10	4.71	4.77	2.84
6-month	0.05	12.19	10.73	2.03	2.40	1.97
1-year	0.50	8.98	7.51	1.73	0.80	1.09
2-year	0.80	1.91	0.40	4.74	2.94	1.33
3-year	0.20	3.73	5.28	5.14	2.73	1.80
4-year	0.20	6.73	8.33	5.19	2.37	1.12
5-year	1.51	9.40	11.05	3.42	0.33	1.40
6-year	3.02	13.44	15.14	1.24	4.49	4.16
7-year	3.73	14.53	16.29	3.97	7.30	4.92
8-year	4.46	11.72	13.55	3.43	6.80	2.55
9-year	4.16	10.17	12.07	4.50	7.86	2.08
10-year	3.76	8.65	10.62	5.74	9.08	2.14
15-year	3.62	1.90	0.47	6.87	10.09	3.41
20-year	2.79	19.25	16.42	5.84	2.58	1.50
25-year	3.07	13.21	9.92	1.64	1.84	13.07
3-year	1.39	4.35	0.57	1.21	5.00	22.73
40-year	2.31	17.987	22.69	5.96	10.39	7.28

Note: The unit of the values is the basis points (bps). The boxed value is the smallest RMSFE for each maturity. RW denotes the random walk model (the benchmark model).

**Table 5.** Root mean squared forecast error of a 3-month forecast.

Maturity	RW	DNS		DNSS		AFNS
		Ind	Corr	Ind	Corr	Corr
3-month	1.96	13.18	10.63	6.90	6.97	4.79
6-month	0.83	12.34	9.72	3.22	3.87	3.21
1-year	0.71	9.85	7.12	1.94	0.47	1.50
2-year	1.28	3.05	0.25	5.99	2.85	1.59
3-year	0.33	3.46	6.28	6.23	2.12	2.85
4-year	1.60	7.42	10.42	5.55	1.39	3.43
5-year	3.63	10.83	13.94	3.22	2.66	4.64
6-year	5.01	14.81	18.01	1.40	7.05	7.26
7-year	7.23	17.28	20.63	5.56	11.35	9.61
8-year	8.52	15.05	18.55	5.79	11.57	8.21
9-year	8.20	13.53	17.16	6.90	12.68	7.82
10-year	7.29	11.56	15.32	7.75	13.46	7.39
15-year	6.44	2.22	5.67	9.01	14.34	8.14
20-year	4.83	18.14	13.39	4.63	3.46	3.77
25-year	4.20	12.95	7.53	1.62	5.80	16.21
3-year	3.29	3.85	4.44	3.50	9.59	26.54
40-year	3.84	17.97	26.54	7.51	14.84	4.85

Note: The unit of the values is the basis points (bps). The boxed value is the smallest RMSFE for each maturity. RW denotes the random walk model (the benchmark model).

**Table 6.** Root mean squared forecast error of a 6-month forecast.

Maturity	RW	DNS		DNSS		AFNS
		Ind	Corr	Ind	Corr	Corr
3-month	3.83	13.18	9.42	9.35	9.37	6.93
6-month	2.50	12.12	8.28	5.41	6.22	5.32
1-year	1.87	9.50	5.77	1.98	2.86	4.07
2-year	1.55	2.78	2.80	5.63	2.34	4.08
3-year	1.40	3.62	8.16	6.54	1.76	4.86
4-year	3.23	8.30	13.01	5.46	2.39	6.14
5-year	5.84	12.41	17.26	2.67	5.33	8.05
6-year	7.88	17.02	22.03	2.91	10.53	11.44
7-year	10.73	20.28	25.46	7.81	15.58	14.53
8-year	13.05	19.12	24.53	9.29	16.96	14.22
9-year	13.10	17.96	23.58	10.99	18.54	14.31
10-year	12.32	16.12	21.95	12.17	19.54	14.13
15-year	13.27	8.15	14.62	16.28	22.69	17.08
20-year	11.99	14.29	10.31	7.98	12.28	13.09
25-year	11.05	10.33	8.79	9.52	15.13	24.89
3-year	8.91	4.60	12.52	10.57	17.54	33.92
40-year	8.87	20.42	33.47	13.26	22.14	8.17

Note: The unit of the values is the basis points (bps). The boxed value is the smallest RMSFE for each maturity. RW denotes the random walk model (the benchmark model).

**Table 7.** Root mean squared forecast error of a 12-month forecast.

Maturity	RW	DNS		DNSS		AFNS
		Ind	Corr	Ind	Corr	Corr
3-month	13.07	11.15	10.05	19.61	19.29	16.64
6-month	14.32	10.48	11.09	17.96	18.55	17.51
1-year	15.64	10.36	13.54	15.51	17.44	18.73
2-year	16.05	13.04	18.75	12.80	16.21	20.04
3-year	17.91	18.68	25.12	13.25	17.64	22.33
4-year	20.00	23.34	30.16	14.24	19.67	24.14
5-year	23.56	28.53	35.63	17.44	23.90	27.36
6-year	26.92	34.47	41.82	23.09	30.24	32.19
7-year	30.90	39.10	46.69	29.14	36.42	36.61
8-year	35.61	40.82	48.63	33.57	40.50	39.08
9-year	37.73	41.92	49.95	37.67	44.19	41.38
10-year	39.02	42.29	50.53	41.18	47.23	43.34
15-year	51.59	46.73	55.91	58.44	62.41	58.24
20-year	62.88	47.83	56.93	62.85	65.42	67.12
25-year	69.03	56.74	67.71	72.27	75.30	84.70
3-year	73.22	66.78	79.86	79.36	83.34	98.00
40-year	78.08	84.73	102.04	85.58	91.80	79.54

Note: The unit of the values is the basis points (bps). The boxed value is the smallest RMSFE for each maturity. RW denotes the random walk model (the benchmark model).

**Table 8.** Diebold–Mariano test for 6- and 12-month forecasting of each maturity of the zero-yield government bond (random walk model and DNSS (Ind) model).

<b>6-Month</b>	<b>3-Month</b>	<b>6-Month</b>	<b>1-Year</b>	<b>2-Year</b>	<b>3-Year</b>	<b>3-Year</b>	<b>5-Year</b>	<b>6-Year</b>	<b>7-Year</b>	<b>8-Year</b>
DM	−3.61	−3.13	−0.26	−4.81	−7.43	−4.32	<b>2.59</b>	<b>3.87</b>	<b>3.61</b>	<b>4.04</b>
<i>p</i> -value	0.000	0.001	0.791	0.000	0.000	0.000	0.010	0.000	0.000	0.000
	9-year	10-year	15-year	20-year	25-year	30-year	40-year			
DM	<b>3.57</b>	0.69	−3.44	<b>2.26</b>	<b>3.98</b>	−1.73	−3.03			
<i>p</i> -value	0.000	0.488	0.000	0.023	0.000	0.082	0.002			
<b>12-Month</b>	3-month	6-month	1-year	2-year	3-year	3-year	5-year	6-year	7-year	8-year
DM	−3.91	−3.35	1.494	<b>2.42</b>	<b>2.47</b>	<b>2.92</b>	<b>3.68</b>	<b>4.59</b>	<b>4.49</b>	<b>4.74</b>
<i>p</i> -value	0.000	0.001	0.134	0.015	0.013	0.004	0.000	0.000	0.000	0.000
	9y	10y	15y	20y	25y	30y	40y			
DM	3.45	−2.30	−3.17	0.04	−2.19	−2.73	−2.94			
<i>p</i> -value	0.001	0.021	0.002	0.965	0.030	0.001	0.003			

Note: The bold value shows that the DNSS model outperformed the random walk model at a 5% statistically significant level. The table shows the Diebold–Mariano statistic value (DM) and *p*-value.

Next, we compared the forecasting performance of the DNS model with the AFNS model<sup>18</sup>. Theoretically, an arbitrage opportunity should be exploited by the market force. If there remains an arbitrage opportunity in the real world, the market is considered inefficient.

From the results for the 1-, 3-, and 6-month forecast horizons, the correlated AFNS was superior to each DNS model in almost all of the maturities. However, as Table 7 shows, the 12-month forecast performance of the AFNS model was inferior to that of the DNS model. The results suggest that the imposition of an arbitrage-free condition does not improve long forecast performance.

Our results provide evidence against those in Christensen et al. (2011) and Diebold and Rudebusch (2013). They argued that the AFNS model improved the out-of-sample forecast for almost all maturities and horizons for the U.S. bond market. Our results indicated that the imposition of an arbitrage-free condition restricts the long-forecasting performance of the AFNS model for JGB yields.

The first plot in Figure 3 depicts the 3-month forecast yields of each model.

Each model captured the actual yield well, as shown by the first and second plots for the 3- and 6-month forecast yields of each model. The forecast performance became poorer as the forecasting horizon becomes longer<sup>19</sup>.

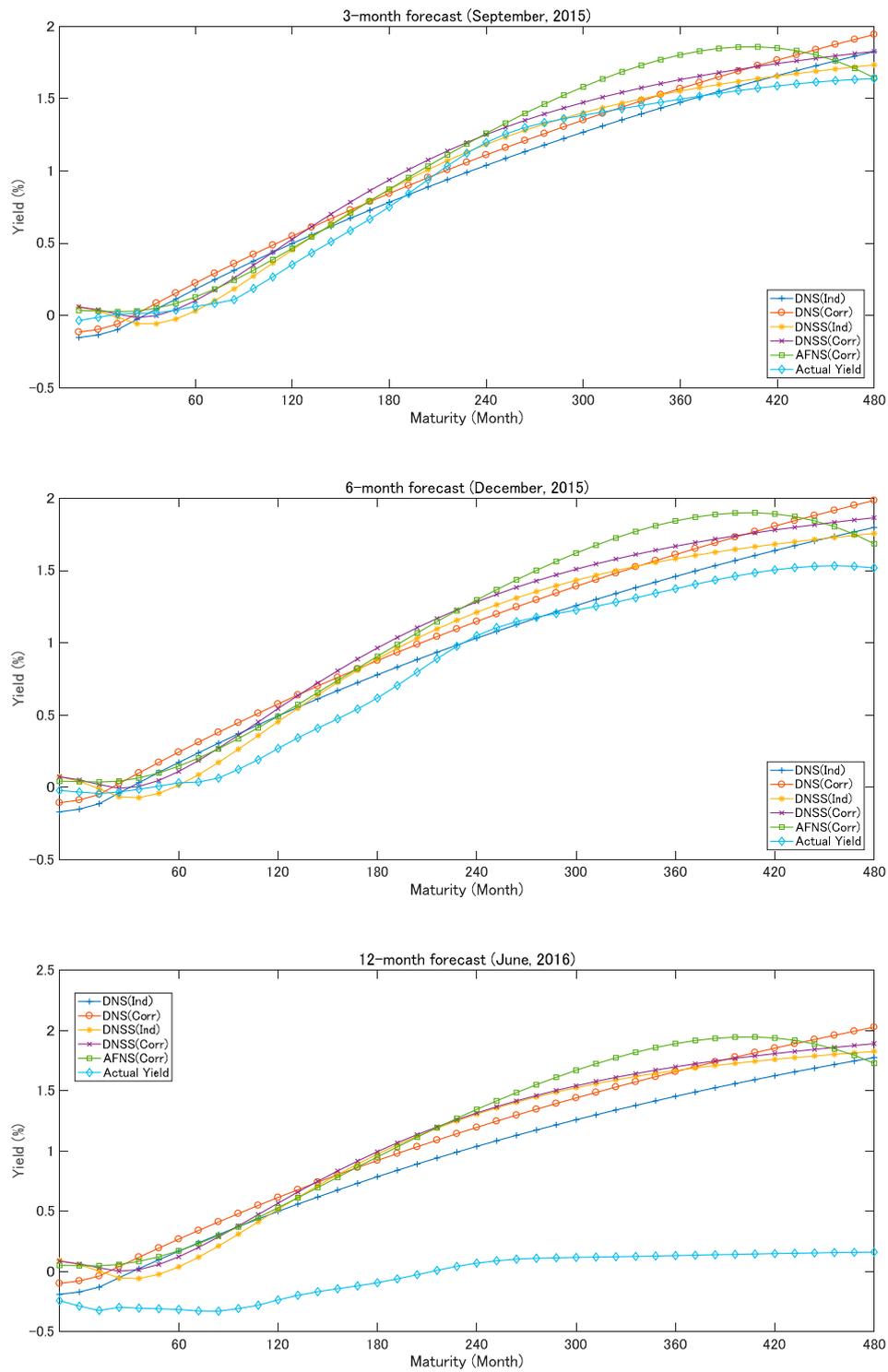
The correlated AFNS captured the convexity due to the yield adjustment term, while the DNS model and the DNSS model did not capture that effect. We could see a large difference between the two curve-fitting models and the affine term structure model. Moreover, the plot of the DNS model was U-shaped and that of the DNSS model was S-shaped; such a difference in shape between models arose from the fourth term of the DNSS model. The figures suggest that the DNSS model captured the dynamics of the yield more flexibly than the DNS model.

In addition to the above analysis of the forecast accuracy, this paper investigated the cumulative of the squared forecast errors to compare the forecast accuracy of the models with the benchmark model. Figure 4 shows the cumulative of the squared forecast errors of the JGB yields of the 3-month, 1-, 5-, and 10-year maturity. The cumulative of the squared forecast errors imply that the upward (downward) trend in the plot suggest that the models are inferior (superior) to the benchmark model (the random walk model). In the forecasting the JGB yields of 3-month maturity and 1-year maturity, the cumulative of the squared errors of the DNS model shows the downward trend for long-forecast horizon. On the other hand, the cumulative of the squared errors of the DNSS model shows the downward trend for the long forecast horizon. These results provide the evidence that the DNS and the DNSS model is more useful for forecasting the long-term horizon than the random walk model from the perspective of the cumulative of the squared forecast errors.

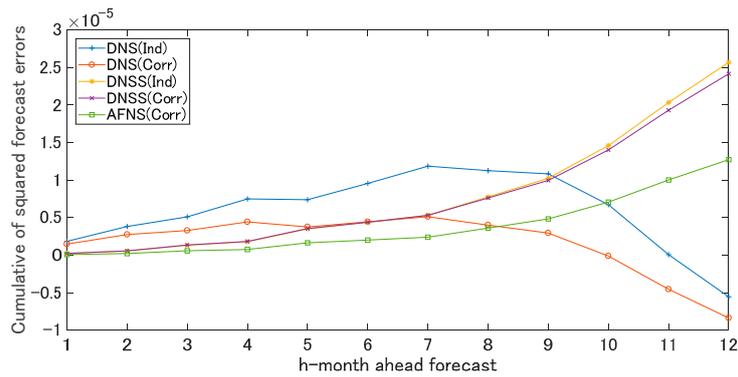
---

<sup>18</sup> DNSS is a four-factor model, so it might be unfair to compare DNNS with AFNS directly. Therefore, it may be more appropriate to compare the forecast performance of DNS with that of AFNS.

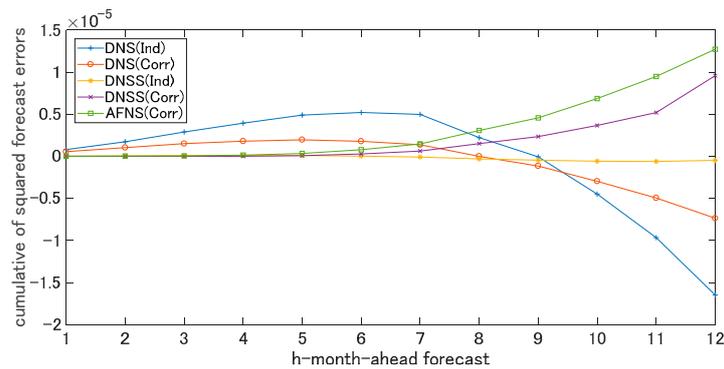
<sup>19</sup> Bank of Japan employed the unconventional monetary policy and the negative interest rate policy after 2000s, so the term structure of interest rates is very sensitive to the changes of the policy stance. Especially, the long-term forecast is affected by the changes of the policy stance.



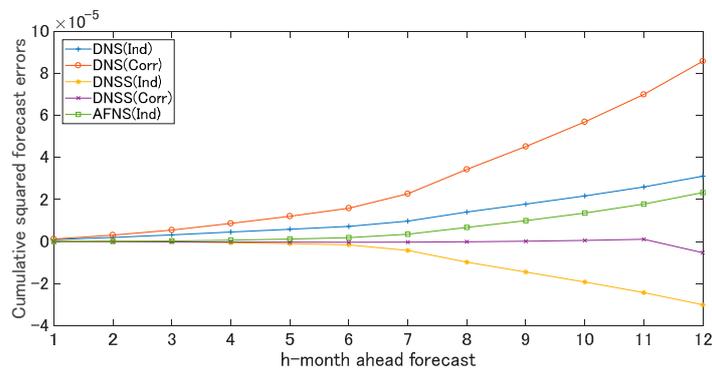
**Figure 3.** Forecast yields for 3-, 6-, and 12-month forecasts. Note: The notation “Ind” denotes the independent factor model and the notation “Corr” denotes the correlated factor model. The plots show the forecast yields of each model and the actual yield in September 2015; December 2015; and June 2016.



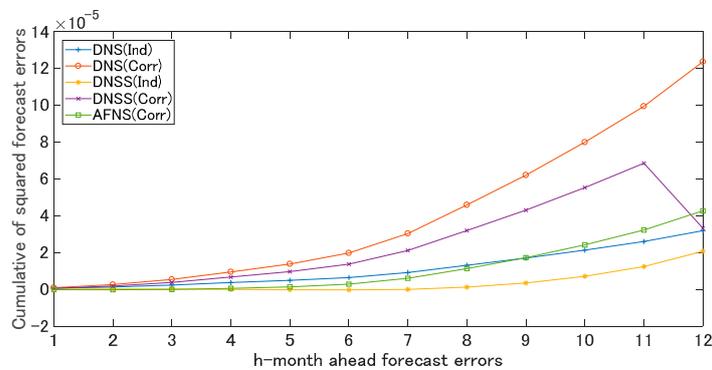
(A) 3-month maturity



(B) 1-year maturity

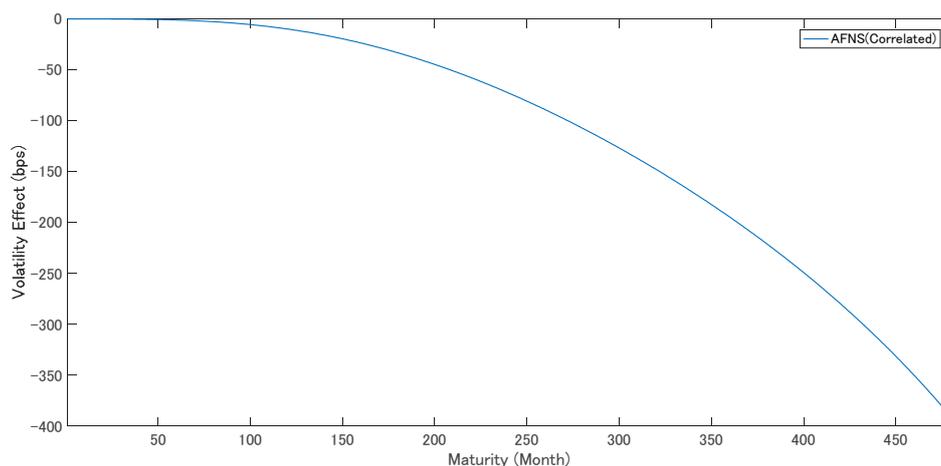


(C) 5-year maturity



(D) 10-year maturity

**Figure 4.** Cumulative of the squared forecast errors. Note: Figure 5 shows the cumulative of the squared errors of the forecasting of the JGB yields of the 3-month, 1-, 5-, and 10-year maturity. The cumulative of the squared forecast errors imply that the upward (downward) trend in the plot suggest that the models are inferior (superior) to the benchmark model (the random walk model).



**Figure 5.** The volatility effect of the correlated AFNS. Note: The unit of the volatility effect is basis points (bps).

## 5.2. Volatility Effect

As the yield adjustment term depends on the volatility matrix, it is often called the volatility effect. In this sub-section, we refer to the yield adjustment term as the volatility effect and discuss it for JGB markets.

Diebold and Rudebusch (2013) argued that the volatility effect can be negligible in most periods, but may become larger in some periods. They also argued that the effect becomes larger with longer maturities.

The volatility effect generated a large difference between the DNS, DNSS, and AFNS models, since it affects the value of the state variables. As discussed in Section 4.1, it caused the constant term bias. The volatility effect arose from the non-linearity of the yield curve. It can be paraphrased mathematically as the Jensen's inequality term. In the economic literature, it has often been ignored, as a negligible quantity. However, with the exception of in finance research, this concept is typically not discussed at length.

Figure 4 shows the volatility effect, which was calculated<sup>20</sup> from the yield adjustment terms of the AFNS model. The volatility effect depended only on maturity, independent of the time. We can confirm that the magnitude of the volatility effect grew larger with longer maturities, in agreement with Diebold and Rudebusch (2013). The magnitude of the volatility effect was 9.8 bp for a 120-month maturity and 126 bp for a 300-month maturity. Hence, we might be able to ignore the effect for maturities from 3- to 120-months, which were practically well used. However, when we are interested in longer-term maturities, such as 300- and 480-month maturities, the volatility effect should be taken into account in the analysis.

## 6. Conclusions

This study investigated the forecasting performances for yields by the DNS, DNSS, and AFNS models. The DNS and DNSS models are curve-fitting models, which do not impose an arbitrage-free condition. On the other hand, the AFNS model is an affine term structure model based on the arbitrage-free theory.

Following Christensen et al. (2011) and Diebold and Rudebusch (2013), we first introduced the theoretical and empirical framework of the three models. The AFNS was different from the other two models, in that includes a yield adjustment term in the yield curve equation.

<sup>20</sup> The volatility effect is identical to the yield adjustment term.

Second, our study provided evidence that the DNSS model was the best model, in terms of forecast performance for yields in JGB markets. The forecast performance of the AFNS model was inferior to the DNS model for the 12-month forecast horizon. Our results were in contrast with those of Christensen et al. (2011) and Diebold and Rudebusch (2013), which showed that the AFNS model was superior to the DNS model for U.S. bond data. For JGB markets, the imposition of the arbitrage-free condition did not improve the forecast performance for relatively long forecasting horizons.

The volatility effect was negligibly small for the short maturities. On the other hand, the longer the maturity was, the bigger the volatility effect became. Therefore, when we consider yields at long maturities, we need to pay attention to the existence of the volatility effect.

Finally, we shortly discussed the remaining issues and the directions of future researches. This study adopted the period from August 2009 to June 2015. However, in February 2016, the Bank of Japan carried out the “minus” interest rate policy, which should have created a large, unpredictable shock in the JGB markets. As a large shock generates arbitrage opportunities, the analyses may be biased without taking this into account. Therefore, we need to replicate the analysis over another period.

In this paper, although we used only the three-factor AFNS model, it is also worth using the generalized model of the AFNS model. If so, we may compare the DNSS model with the arbitrage-free Nelson–Siegel model appropriately.

**Funding:** This research received no external funding.

**Conflicts of Interest:** The authors declare no conflicts of interest associated with this article.

## Appendix A. Derivation of the Theoretical Models

### Appendix A.1. Yield of the Nelson–Siegel Model

The forward yield function of the Nelson–Siegel model,

$$f(\tau) = \beta_1 + \beta_2 e^{-\lambda\tau} + \beta_3 \lambda \tau e^{-\lambda\tau}, \quad (\text{A1})$$

can be described as:

$$y(\tau) = \beta_1 + \beta_2 \left( \frac{1 - e^{-\lambda\tau}}{\lambda\tau} \right) + \beta_3 \left( \frac{1 - e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau} \right). \quad (\text{A2})$$

**Proof.**

$$\begin{aligned} y(\tau) &= \frac{1}{\tau} \int_0^\tau (\beta_1 + \beta_2 e^{-\lambda u} + \beta_3 \lambda u e^{-\lambda u}) du \\ &= \frac{1}{\tau} \left\{ \int_0^\tau \beta_1 du + \int_0^\tau \beta_2 e^{-\lambda u} du + \int_0^\tau \beta_3 \lambda u e^{-\lambda u} du \right\} \\ &= \frac{1}{\tau} \left\{ \beta_1 \int_0^\tau 1 du + \beta_2 \int_0^\tau e^{-\lambda u} du + \beta_3 \int_0^\tau \lambda u e^{-\lambda u} du \right\} \\ &= \frac{1}{\tau} \left\{ \beta_1 [u]_0^\tau + \beta_2 \left[ -\frac{e^{-\lambda u}}{\lambda} \right]_0^\tau + \beta_3 \left[ -u e^{-\lambda u} \right]_0^\tau + \beta_3 \int_0^\tau e^{-\lambda u} du \right\} \\ &= \frac{1}{\tau} \left\{ \beta_3 \tau + \beta_2 \frac{1 - e^{-\lambda\tau}}{\lambda} - \beta_3 \tau e^{-\lambda\tau} + \beta_3 \left[ -\frac{e^{-\lambda u}}{\lambda} \right]_0^\tau \right\} \\ &= \frac{1}{\tau} \left\{ \beta_1 \tau + \beta_2 \frac{1 - e^{-\lambda\tau}}{\lambda} - \beta_3 \tau e^{-\lambda\tau} + \beta_3 \frac{1 - e^{-\lambda\tau}}{\lambda} \right\} \\ &= \beta_1 + \beta_2 \left( \frac{1 - e^{-\lambda\tau}}{\lambda\tau} \right) + \beta_3 \left( \frac{1 - e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau} \right). \end{aligned}$$

□

### Appendix A.2. Yield of the Nelson–Siegel–Svensson Model

The forward yield function of the Nelson–Siegel–Svensson model,

$$f(\tau) = \beta_1 + \beta_2 e^{-\lambda_1\tau} + \beta_3 \lambda_1 \tau e^{-\lambda_1\tau} + \beta_4 \lambda_2 \tau e^{-\lambda_2\tau}, \quad (\text{A3})$$

can be described as:

$$y(\tau) = \beta_1 + \beta_2 \left( \frac{1 - e^{-\lambda_1 \tau}}{\lambda_1 \tau} \right) + \beta_3 \left( \frac{1 - e^{-\lambda_1 \tau}}{\lambda_1 \tau} - e^{-\lambda_1 \tau} \right) + \beta_4 \left( \frac{1 - e^{-\lambda_2 \tau}}{\lambda_2 \tau} - e^{-\lambda_2 \tau} \right). \tag{A4}$$

**Proof.**

$$\begin{aligned} y(\tau) &= \frac{1}{\tau} \int_0^\tau \{ \beta_1 + \beta_2 e^{-\lambda_1 s} + \beta_3 \lambda_1 s e^{-\lambda_1 s} + \beta_4 \lambda_2 s e^{-\lambda_2 s} \} ds \\ &= \frac{1}{\tau} \left\{ \int_0^\tau \beta_1 ds + \int_0^\tau \beta_2 e^{-\lambda_1 s} ds + \int_0^\tau \beta_3 \lambda_1 s e^{-\lambda_1 s} ds + \int_0^\tau \beta_4 \lambda_2 s e^{-\lambda_2 s} ds \right\} \\ &= \frac{1}{\tau} \left\{ [\beta_1 s]_0^\tau + \beta_2 \left[ -\frac{e^{-\lambda_1 s}}{\lambda_1} \right]_0^\tau + \beta_3 \left( -\tau e^{-\lambda_1 \tau} + \int_0^\tau e^{-\lambda_1 s} ds \right) \right. \\ &\quad \left. + \beta_4 \left( -\tau e^{-\lambda_2 \tau} + \int_0^\tau e^{-\lambda_2 s} ds \right) \right\} \\ &= \frac{1}{\tau} \left\{ \beta_1 \tau + \beta_2 \left( \frac{1 - e^{-\lambda_1 \tau}}{\lambda_1} \right) + \beta_3 \left( -\tau e^{-\lambda_1 \tau} + \left[ -\frac{e^{-\lambda_1 s}}{\lambda_1} \right]_0^\tau \right) \right. \\ &\quad \left. + \beta_4 \left( -\tau e^{-\lambda_2 \tau} + \left[ -\frac{e^{-\lambda_2 s}}{\lambda_2} \right]_0^\tau \right) \right\} \\ &= \frac{1}{\tau} \left\{ \beta_1 \tau + \beta_2 \left( \frac{1 - e^{-\lambda_1 \tau}}{\lambda_1} \right) + \beta_3 \left( \frac{1 - e^{-\lambda_1 \tau}}{\lambda_1} - \tau e^{-\lambda_1 \tau} \right) + \beta_4 \left( \frac{1 - e^{-\lambda_2 \tau}}{\lambda_2} - \tau e^{-\lambda_2 \tau} \right) \right\} \\ &= \beta_1 + \beta_2 \left( \frac{1 - e^{-\lambda_1 \tau}}{\lambda_1 \tau} \right) + \beta_3 \left( \frac{1 - e^{-\lambda_1 \tau}}{\lambda_1 \tau} - e^{-\lambda_1 \tau} \right) + \beta_4 \left( \frac{1 - e^{-\lambda_2 \tau}}{\lambda_2 \tau} - e^{-\lambda_2 \tau} \right). \end{aligned}$$

□

*Appendix A.3. Affine Term Structure Model of Duffie and Kan*

We considered the Duffie–Kan framework (Duffie and Kan (1996)) for defining the arbitrage-free Nelson–Siegel model. By substituting the Nelson–Siegel factors into the state variable of the Duffie–Kan framework, we can make the Nelson–Siegel model arbitrage-free.

In Duffie and Kan, the filtered probability space is  $(\Omega, \mathcal{F}_t, Q)$ , where the filtration is:

$$\mathcal{F}_t = \{ \mathcal{F}_t, t \geq 0 \}, \tag{A5}$$

and the state variables are  $X_t$  on a set  $M \subset \mathbb{R}^n$ , which is a Markov process and the solution of the stochastic differential equation,

$$dX(t) = \kappa^Q(t) (\theta^Q(t) - X(t)) dt + \Sigma(t) D(X_t, t) dW^Q(t), \tag{A6}$$

where  $W^Q$  is the standard Brownian motion in  $\mathbb{R}^n$ .

In Duffie and Kan (1996), the definition of  $D(X_t, t)$  is:

$$D(X_t, t) = \begin{bmatrix} \sqrt{\gamma^1(t) + \delta^1(t) X_t} & \cdots & 0 \\ \vdots & \cdots & \vdots \\ 0 & \cdots & \sqrt{\gamma^n(t) + \delta^n(t) X_t} \end{bmatrix}, \tag{A7}$$

which is a diagonal matrix where  $\delta^i(t)$  denotes the  $i$ th row of the matrix:

$$\delta^i(t) = \begin{bmatrix} \delta_1^1(t) & \cdots & \delta_n^1(t) \\ \vdots & \cdots & \vdots \\ \delta_1^n(t) & \cdots & \delta_n^n(t) \end{bmatrix}. \tag{A8}$$

The instantaneous risk-free rate is represented by an affine function; that is,

$$r_t = \rho_0(t) + \rho_1(t)' X_t, \tag{A9}$$

where each parameter is bounded and continuous, respectively; that is,  $\theta^{\mathbb{Q}} : [0, T] \rightarrow \mathbb{R}^n, \kappa^{\mathbb{Q}} : [0, T] \rightarrow \mathbb{R}^{n \times n}, \Sigma : [0, T] \rightarrow \mathbb{R}^{n \times n}, D : M \times [0, T] \rightarrow \mathbb{R}^{n \times n}, \gamma : [0, T] \rightarrow \mathbb{R}^{n \times n}, \delta : [0, T] \rightarrow \mathbb{R}^{n \times n}, \rho_0 : [0, T] \rightarrow \mathbb{R}, \rho_1 : [0, T] \rightarrow \mathbb{R}^n$ .

**Proposition A1.** *The zero-coupon price is represented as*

$$P(t, T) = \mathbb{E}_t^{\mathbb{Q}} \left[ e^{-\int_t^T r_u du} \right] = e^{B(t, T)' X_t + C(t, T)}, \tag{A10}$$

under the above assumptions, where  $B(t, T)$  and  $C(t, T)$  are the solutions of following ordinary differential equation systems under the boundary conditions  $B(T, T) = 0$  and  $C(T, T) = 0$ ,

$$\frac{dB(t, T)}{dt} = \rho_1 + (\kappa^{\mathbb{Q}})' B(t, T) - \frac{1}{2} \sum_{j=1}^n (\Sigma' B(t, T) B(t, T) \Sigma)_{j,j} (\delta^j)', \tag{A11}$$

$$\frac{dC(t, T)}{dt} = \rho_0 - B(t, T)' \kappa^{\mathbb{Q}} \theta^{\mathbb{Q}} - \frac{1}{2} \sum_{j=1}^n (\Sigma' B(t, T) B(t, T) \Sigma)_{j,j} \gamma^j. \tag{A12}$$

Then, the yield is given by:

$$\begin{aligned} Y(t, T) &= -\frac{1}{T-t} \log P(t, T) \\ &= -\frac{B(t, T)'}{T-t} X_t - \frac{C(t, T)}{T-t}. \end{aligned} \tag{A13}$$

**Proof.**

Here, we consider three factors (or three state variables). The bond price is described as:

$$P_t^{(\tau)} = \mathbb{E}_t^{\mathbb{Q}} \left[ e^{-\int_t^{\tau} r_u du} \right]. \tag{A14}$$

From the Feynman–Kac formula, we can describe the bond price as:

$$P_t^{(\tau)} = F(X_t, \tau), \tag{A15}$$

where  $F(x_t, \tau)$  solves the partial differential equation:

$$\begin{aligned} F(X(t), \tau) r(t) &= \frac{\partial F(X(t), \tau)}{\partial \tau} + \sum_{i=1}^3 \left[ \frac{\partial F(X(t), \tau)}{\partial x_i} \kappa^{\mathbb{Q}}(t) (\theta^{\mathbb{Q}}(t) - X(t)) + \right. \\ &\left. \frac{1}{2} \frac{\partial^2 F(X(t), \tau)}{\partial x_i^2} \Sigma D(X(t), t) D(X(t), t)' \Sigma \right], \end{aligned} \tag{A16}$$

and we can write it as:

$$F(X(t), \tau) = e^{B(t, T)' X_t + C(t, T)}. \tag{A17}$$

From this equation, the first differential of  $X_i$  and  $\tau$ , and the second differential of  $X_i$  are given by:

$$\frac{\partial F(X(t), \tau)}{\partial x_i} = B(t, T) F(X(t), \tau), \tag{A18}$$

$$\frac{\partial^2 F(X(t), \tau)}{\partial x_i^2} = B(t, T) B(t, T)' F(X(t), \tau), \tag{A19}$$

$$\frac{\partial F(X(t), \tau)}{\partial \tau} = [B'(t, T) x_t + C'(t, T)] F(X(t), \tau). \tag{A20}$$

Substituting these equations into the Equation (A16), we obtain:

$$\begin{aligned}
 F(X(t), \tau)r(t) &= \frac{\partial F(X(t), \tau)}{\partial \tau} \\
 &\quad + \sum_{i=1}^3 \left[ \frac{\partial F(X(t), \tau)}{\partial x_i} \kappa^Q(t) (\theta^Q(t) - X(t)) \right. \\
 &\quad \left. + \frac{1}{2} \frac{\partial^2 F(X(t), \tau)}{\partial x_i^2} \Sigma D(X(t), t) D(X(t), t)' \Sigma \right] \\
 &= [B'(t, T)X_t + C'(t, T)]F(X(t), \tau) + \sum_{i=1}^3 [B(t, T)F(X(t), \tau) \kappa^Q(t) (\theta^Q(t) - X(t)) + \\
 &\quad \frac{1}{2} F(X(t), \tau) (\Sigma B(t, T) B(t, T)' \Sigma')_{j,j} D(X(t), t) D(X(t), t)'] \tag{A21}
 \end{aligned}$$

From the definition of the matrix  $D(X(t), t)$ ,

$$\begin{aligned}
 r(t) &= [B'(t, T)X_t + C'(t, T)] \\
 &\quad + \sum_{i=1}^3 [B(t, T) \kappa^Q(t) (\theta^Q(t) - X(t)) \\
 &\quad + \frac{1}{2} (\Sigma B(t, T) B(t, T)' \Sigma')_{j,j} (\gamma^j(t) + \delta^j(t) X_t)] \\
 &= [B'(t, T)X_t + C'(t, T)] \\
 &\quad + \sum_{i=1}^3 [B(t, T) \kappa^Q(t) (\theta^Q(t) - X(t)) \\
 &\quad + \frac{1}{2} (\Sigma B(t, T) B(t, T)' \Sigma')_{j,j} (\gamma^j(t)) \\
 &\quad + \frac{1}{2} (\Sigma B(t, T) B(t, T)' \Sigma')_{j,j} (\delta^j(t) X_t)] \\
 &= \left[ \frac{B(t, T)}{dt} - B(t, T) \kappa^Q(t) + \frac{1}{2} \sum_{i=1}^3 (\Sigma B(t, T) B(t, T)' \Sigma')_{j,j} \delta^j(t) \right] X_t + \left[ \frac{C(t, T)}{dt} + \right. \\
 &\quad \left. B(t, T) \kappa^Q(t) \theta^Q(t) + \frac{1}{2} \sum_{i=1}^3 (\Sigma B(t, T) B(t, T)' \Sigma')_{j,j} (\gamma^j(t)) \right]
 \end{aligned}$$

From the definition of  $r_t = \rho_0(t) + \rho_1(t)' X_t$ , we obtain:

$$\begin{aligned}
 \rho_1(t) &= \frac{B(t, T)}{dt} - B(t, T) \kappa^Q(t) + \frac{1}{2} \sum_{i=1}^3 (\Sigma B(t, T) B(t, T)' \Sigma')_{j,j} \delta^j(t), \\
 \rho_0(t) &= \frac{C(t, T)}{dt} + B(t, T) \kappa^Q(t) \theta^Q(t) + \frac{1}{2} \sum_{i=1}^3 (\Sigma B(t, T) B(t, T)' \Sigma')_{j,j} (\gamma^j(t)).
 \end{aligned}$$

Arranging the equations, we get:

$$\frac{B(t, T)}{dt} = \rho_1(t) + B(t, T) \kappa^Q(t) - \frac{1}{2} \sum_{i=1}^3 (\Sigma B(t, T) B(t, T)' \Sigma')_{j,j} \delta^j(t), \tag{A22}$$

$$\frac{C(t, T)}{dt} = \rho_0(t) - B(t, T) \kappa^Q(t) \theta^Q(t) - \frac{1}{2} \sum_{i=1}^3 (\Sigma B(t, T) B(t, T)' \Sigma')_{j,j} (\gamma^j(t)), \tag{A23}$$

where boundary condition is:

$$B(T, T) = C(T, T) = 0. \tag{A24}$$

Then, the solutions of the ordinary differential equations are  $B(t, T)$ ,  $C(t, T)$ , and the yield functions:

$$Y(t, T) = -\frac{1}{T-t} \log P(t, T) = -\frac{B(t, T)'}{T-t} X_t - \frac{C(t, T)}{T-t}. \tag{A25}$$

□

Appendix A.4. Arbitrage-Free Nelson–Siegel Model

In this appendix, we show the derivation of the arbitrage-free Nelson–Siegel model following Christensen et al. (2011).

We use Equation (24) from the main text:

$$\frac{dB(t, T)}{dt} = \rho_1 + (\kappa^{\mathbb{Q}})' B(t, T), \tag{A26}$$

and the relationship as follows,

$$\frac{d}{dt} \left[ e^{(\kappa^{\mathbb{Q}})'(T-t)} B(t, T) \right] = e^{(\kappa^{\mathbb{Q}})'(T-t)} \frac{dB(t, T)}{dt} - e^{(\kappa^{\mathbb{Q}})'(T-t)} (\kappa^{\mathbb{Q}})' B(t, T). \tag{A27}$$

Substituting the Equation (A26) into the Equation (A27), we obtain:

$$\begin{aligned} \frac{d}{dt} \left[ e^{(\kappa^{\mathbb{Q}})'(T-t)} B(t, T) \right] &= \rho_1 e^{(\kappa^{\mathbb{Q}})'(T-t)} + e^{(\kappa^{\mathbb{Q}})'(T-t)} (\kappa^{\mathbb{Q}})' B(t, T) - e^{(\kappa^{\mathbb{Q}})'(T-t)} (\kappa^{\mathbb{Q}})' B(t, T) \\ &= \rho_1 e^{(\kappa^{\mathbb{Q}})'(T-t)}. \end{aligned} \tag{A28}$$

The integration of the Equation (A28) is written as:

$$\begin{aligned} \int_t^T \frac{d}{ds} \left[ e^{(\kappa^{\mathbb{Q}})'(T-s)} B(s, T) \right] ds &= \int_t^T \rho_1 e^{(\kappa^{\mathbb{Q}})'(T-s)} ds. \\ \int_t^T d \left[ e^{(\kappa^{\mathbb{Q}})'(T-s)} B(s, T) \right] &= \int_t^T \rho_1 e^{(\kappa^{\mathbb{Q}})'(T-s)} ds. \\ \left[ e^{(\kappa^{\mathbb{Q}})'(T-s)} B(s, T) \right]_t^T &= \int_t^T \rho_1 e^{(\kappa^{\mathbb{Q}})'(T-s)} ds. \end{aligned} \tag{A29}$$

From the boundary condition  $B(T, T) = 0$ , the Equation (A29) can be rewritten as:

$$B(t, T) = -e^{-(\kappa^{\mathbb{Q}})'(T-t)} \int_t^T \rho_1 e^{(\kappa^{\mathbb{Q}})'(T-s)} ds. \tag{A30}$$

From the restrictions (30), (31), and (32) in the main text, we obtain:

$$B(t, T) = \begin{bmatrix} -(T-t) \\ -\frac{1-e^{-\lambda(T-t)}}{\lambda} \\ (T-t)e^{-\lambda(T-t)} - \frac{1-e^{-\lambda(T-t)}}{\lambda} \end{bmatrix}. \tag{A31}$$

Appendix A.5. Yield Adjustment Term of Arbitrage-Free Nelson–Siegel Model

In this appendix, we show the derivation of the yield adjustment the arbitrage-free Nelson–Siegel model following Christensen et al. (2011).

**Proposition A2.** Fixing the average levels of state variables, which, under  $\mathbb{Q}$  measure, we can identify the yield adjustment term of the three-factor arbitrage-free Nelson–Siegel model.

Under the  $\mathbb{Q}$  measure,  $\theta^{\mathbb{Q}} = 0$ , and so we can represent the yield adjustment term  $C(t, T)$  as:

$$\frac{C(t, T)}{T-t} = \frac{1}{2} \frac{1}{T-t} \sum_{j=1}^3 \int_t^T (\Sigma' B(s, T) B(s, T)' \Sigma)_{j,j} ds. \tag{A32}$$

Then, we can write above equation as:

$$\begin{aligned} \frac{C(t,T)}{T-t} = & \bar{A}\left(\frac{(T-t)^2}{6}\right) + \bar{B}\left(\frac{1}{2\lambda^2} - \frac{1}{\lambda^3} \frac{1-e^{-\lambda(T-t)}}{T-t} + \frac{1}{4\lambda^3} \frac{1-e^{-2\lambda(T-t)}}{T-t}\right) + \bar{C}\left(\frac{1}{2\lambda^2} + \frac{1}{\lambda^2} e^{-\lambda(T-t)} - \right. \\ & \left. \frac{1}{4\lambda} (T-t) e^{-2\lambda(T-t)} - \frac{3}{4\lambda^2} e^{-2\lambda(T-t)} - \frac{2}{\lambda^3} \frac{1-e^{-\lambda(T-t)}}{T-t} + \frac{5}{8\lambda^3} \frac{1-e^{-2\lambda(T-t)}}{T-t}\right) + \bar{D}\left(\frac{1}{2\lambda} (T-t) + \right. \\ & \left. \frac{1}{\lambda^2} e^{-\lambda(T-t)} - \frac{1}{\lambda^3} \frac{1-e^{-\lambda(T-t)}}{T-t}\right) + \bar{E}\left(\frac{3}{\lambda^2} e^{-\lambda(T-t)} + \frac{1}{2\lambda} (T-t) + \frac{1}{\lambda} (T-t) e^{-\lambda(T-t)} - \right. \\ & \left. \frac{3}{\lambda^3} \frac{1-e^{-\lambda(T-t)}}{T-t} + \bar{F}\left(\frac{1}{\lambda^2} + \frac{1}{\lambda^2} e^{-\lambda(T-t)} - \frac{1}{2\lambda^2} e^{-2\lambda(T-t)} - \frac{3}{\lambda^3} \frac{1-e^{-\lambda(T-t)}}{T-t} + \frac{3}{4\lambda^3} \frac{1-e^{-\lambda(T-t)}}{T-t}\right), \right. \end{aligned} \tag{A33}$$

where:

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{bmatrix},$$

and:

$$\begin{aligned} \bar{A} &= \sigma_{11}^2 + \sigma_{12}^2 + \sigma_{13}^2, \\ \bar{B} &= \sigma_{21}^2 + \sigma_{22}^2 + \sigma_{23}^2, \\ \bar{C} &= \sigma_{31}^2 + \sigma_{32}^2 + \sigma_{33}^2, \\ \bar{D} &= \sigma_{11}\sigma_{21} + \sigma_{12}\sigma_{22} + \sigma_{13}\sigma_{23}, \\ \bar{E} &= \sigma_{11}\sigma_{31} + \sigma_{12}\sigma_{32} + \sigma_{13}\sigma_{33}, \text{ and} \\ \bar{F} &= \sigma_{21}\sigma_{31} + \sigma_{22}\sigma_{32} + \sigma_{23}\sigma_{33}. \end{aligned}$$

**Proof.**

$$\begin{aligned} \frac{C(t,T)}{T-t} = & \frac{1}{2} \frac{1}{T-t} \sum_{j=1}^3 \int_t^T (\Sigma' B(S,T) B(S,T)' \Sigma)_{j,j} ds \\ = & \frac{\bar{A}}{2} \frac{1}{T-t} \int_t^T B^1(S,T)^2 ds + \frac{\bar{B}}{2} \frac{1}{T-t} \int_t^T B^2(S,T)^2 ds \\ & + \frac{\bar{C}}{2} \frac{1}{T-t} \int_t^T B^3(S,T)^2 ds + \frac{\bar{D}}{2} \frac{1}{T-t} \int_t^T B^1(S,T) B^2(S,T) ds \\ & + \frac{\bar{D}}{2} \frac{1}{T-t} \int_t^T B^2(S,T) B^1(S,T) ds + \frac{\bar{E}}{2} \frac{1}{T-t} \int_t^T B^1(S,T) B^3(S,T) ds \\ & + \frac{\bar{E}}{2} \frac{1}{T-t} \int_t^T B^3(S,T) B^1(S,T) ds + \frac{\bar{F}}{2} \frac{1}{T-t} \int_t^T B^2(S,T) B^3(S,T) ds \\ & + \frac{\bar{F}}{2} \frac{1}{T-t} \int_t^T B^3(S,T) B^2(S,T) ds \\ = & \frac{\bar{A}}{2} \frac{1}{T-t} \int_t^T B^1(S,T)^2 ds + \frac{\bar{B}}{2} \frac{1}{T-t} \int_t^T B^2(S,T)^2 ds + \frac{\bar{C}}{2} \frac{1}{T-t} \int_t^T B^3(S,T)^2 ds + \\ & \frac{\bar{D}}{T-t} \int_t^T B^1(S,T) B^2(S,T) ds + \frac{\bar{F}}{T-t} \int_t^T B^1(S,T) B^3(S,T) ds \quad a - 34 + \\ & \frac{\bar{F}}{T-t} \int_t^T B^2(S,T) B^3(S,T) ds, \end{aligned} \tag{A34}$$

where:

$$\begin{aligned} \bar{A} &= \sigma_{11}^2 + \sigma_{12}^2 + \sigma_{13}^2, \\ \bar{B} &= \sigma_{21}^2 + \sigma_{22}^2 + \sigma_{23}^2, \\ \bar{C} &= \sigma_{31}^2 + \sigma_{32}^2 + \sigma_{33}^2, \\ \bar{D} &= \sigma_{11}\sigma_{21} + \sigma_{12}\sigma_{22} + \sigma_{13}\sigma_{23}, \\ \bar{E} &= \sigma_{11}\sigma_{31} + \sigma_{12}\sigma_{32} + \sigma_{13}\sigma_{33}, \text{ and} \\ \bar{F} &= \sigma_{21}\sigma_{31} + \sigma_{22}\sigma_{32} + \sigma_{23}\sigma_{33}. \end{aligned}$$

We define each term of Equation (A34) as:

$$I_A \equiv \frac{\bar{A}}{2} \frac{1}{T-t} \int_t^T B^1(s,T)^2 ds, \tag{A35}$$

$$I_B \equiv \frac{\bar{B}}{2} \frac{1}{T-t} \int_t^T B^2(s, T)^2 ds, \tag{A36}$$

$$I_C \equiv \frac{\bar{C}}{2} \frac{1}{T-t} \int_t^T B^3(s, T)^2 ds, \tag{A37}$$

$$I_D \equiv \frac{\bar{D}}{T-t} \int_t^T B^1(s, T) B^2(s, T) ds, \tag{A38}$$

$$I_E \equiv \frac{\bar{E}}{T-t} \int_t^T B^1(s, T) B^3(s, T) ds, \text{ and} \tag{A39}$$

$$I_F \equiv \frac{\bar{F}}{T-t} \int_t^T B^2(s, T) B^3(s, T) ds, \tag{A40}$$

respectively.

From the previous section,

$$B^1(t, T) = -(T - t),$$

$$B^2(t, T) = -\frac{1 - e^{-\lambda(T-t)}}{\lambda}, \text{ and}$$

$$B^3(t, T) = (T - t)e^{-\lambda(T-t)} - \frac{1 - e^{-\lambda(T-t)}}{\lambda}.$$

Substituting these into Equations (A35)–(A40), we can derive the analytical forms:

$$I_A = \frac{\bar{A}}{2} \frac{1}{T-t} \int_t^T B^1(s, T)^2 ds = \frac{\bar{A}}{2} \frac{1}{T-t} \int_t^T (T-s)^2 ds = \frac{\bar{A}}{2} \frac{1}{T-t} \left[ -\frac{(T-s)^3}{3} \right]_t^T = \frac{\bar{A}}{2} \frac{1}{T-t} \frac{(T-t)^3}{3} = \frac{\bar{A}}{6} (T-t)^2,$$

$$\begin{aligned} I_B &= \frac{\bar{B}}{2} \frac{1}{T-t} \int_t^T B^2(s, T)^2 ds = \frac{\bar{B}}{2} \frac{1}{T-t} \int_t^T \left[ -\frac{1 - e^{-\lambda(T-s)}}{\lambda} \right]^2 ds \\ &= \frac{\bar{B}}{2} \frac{1}{T-t} \int_t^T \left[ \frac{1 - 2e^{-\lambda(T-s)} + e^{-2\lambda(T-s)}}{\lambda^2} \right] ds \\ &= \frac{\bar{B}}{2} \frac{1}{T-t} \left\{ \int_t^T \frac{1}{\lambda^2} ds - \int_t^T \frac{2e^{-\lambda(T-s)}}{\lambda^2} ds + \int_t^T \frac{e^{-2\lambda(T-s)}}{\lambda^2} ds \right\} \\ &= \frac{\bar{B}}{2} \frac{1}{T-t} \left\{ \left[ \frac{s}{\lambda^2} \right]_t^T - \left[ \frac{2e^{-\lambda(T-s)}}{\lambda^3} \right]_t^T + \left[ \frac{e^{-2\lambda(T-s)}}{2\lambda^3} \right]_t^T \right\} \\ &= \frac{\bar{B}}{2} \frac{1}{T-t} \left\{ \frac{T-t}{\lambda^2} - \frac{2 - 2e^{-\lambda(T-t)}}{\lambda^3} + \frac{1 - e^{-2\lambda(T-t)}}{2\lambda^3} \right\} = \bar{B} \left\{ \frac{1}{2\lambda^2} - \frac{1}{\lambda^3} \frac{1 - e^{-\lambda(T-t)}}{T-t} + \frac{1}{4\lambda^3} \frac{1 - e^{-2\lambda(T-t)}}{T-t} \right\}, \end{aligned}$$

$$\begin{aligned}
 I_C &= \frac{\bar{C}}{2} \frac{1}{T-t} \int_t^T B^3(s, T)^2 ds \\
 &= \frac{\bar{C}}{2} \frac{1}{T-t} \left[ \int_t^T (T-s)^2 e^{-2\lambda(T-s)} ds - \int_t^T \frac{2(T-s)e^{-\lambda(T-s)}}{\lambda} ds \right. \\
 &\quad \left. + \int_t^T \frac{2(T-s)e^{-2\lambda(T-s)}}{\lambda} ds + \int_t^T \frac{1}{\lambda^2} ds - \int_t^T \frac{2e^{-\lambda(T-s)}}{\lambda^2} ds \right. \\
 &\quad \left. + \int_t^T \frac{e^{-2\lambda(T-s)}}{\lambda^2} ds \right] \\
 &= \frac{\bar{C}}{2} \frac{1}{T-t} \left\{ \left[ (T-s)^2 \left( -\frac{e^{-2\lambda(T-s)}}{2\lambda} \right) \right]_t^T - \int_t^T 2(T-s) \frac{e^{-2\lambda(T-s)}}{2\lambda} ds \right. \\
 &\quad \left. - \frac{1}{\lambda} \left\{ \left[ 2(T-s) \left( \frac{e^{-\lambda(T-s)}}{\lambda} \right) \right]_t^T - \int_t^T \frac{2e^{-\lambda(T-s)}}{2\lambda} ds \right\} + \left[ \frac{s}{\lambda^2} \right]_t^T - \left[ \frac{2e^{-\lambda(T-s)}}{\lambda^3} \right]_t^T \right. \\
 &\quad \left. + \left[ \frac{e^{-2\lambda(T-s)}}{2\lambda^3} \right]_t^T \right\} \\
 &= \frac{\bar{C}}{2} \frac{1}{T-t} \left\{ -(T-t)^2 \frac{e^{-2\lambda(T-t)}}{2\lambda} - 2(T-t) \frac{e^{-2\lambda(T-t)}}{4\lambda^2} + \int_t^T \frac{e^{-2\lambda(T-s)}}{4\lambda^2} ds \right. \\
 &\quad \left. + \frac{1}{\lambda} \left\{ 2(T-t) \frac{e^{-\lambda(T-t)}}{\lambda} - \left[ \frac{2e^{-\lambda(T-t)}}{\lambda^2} \right]_t^T \right\} \right. \\
 &\quad \left. - \frac{1}{\lambda} \left\{ 2(T-t) \frac{e^{-2\lambda(T-s)}}{2\lambda} - \left[ \frac{e^{-2\lambda(T-t)}}{2\lambda^2} \right] \right\} + \frac{T-t}{\lambda^2} - \frac{2-2e^{-\lambda(T-t)}}{\lambda^3} \right. \\
 &\quad \left. + \frac{1-1e^{-2\lambda(T-t)}}{2\lambda^3} \right\} \\
 &= \frac{\bar{C}}{2} \frac{1}{T-t} \left\{ (T-t)^2 \frac{e^{-2\lambda(T-t)}}{2\lambda} - (T-t) \frac{e^{-2\lambda(T-t)}}{2\lambda^2} - \frac{1-e^{-2\lambda(T-t)}}{4\lambda^3} + \frac{2e^{-\lambda(T-t)}}{\lambda^2} (T-t) \right. \\
 &\quad \left. - \frac{2-2e^{-\lambda(T-t)}}{\lambda^3} - \frac{e^{-2\lambda(T-t)}}{\lambda^2} (T-t) + \frac{1-e^{-2\lambda(T-t)}}{2\lambda^3} + \frac{T-t}{\lambda^2} \right. \\
 &\quad \left. - \frac{2-2e^{-\lambda(T-t)}}{\lambda^3} + \frac{1-1e^{-2\lambda(T-t)}}{2\lambda^3} \right\} \\
 &= \frac{\bar{C}}{2} \frac{1}{T-t} \left\{ (T-t)^2 \frac{e^{-2\lambda(T-t)}}{2\lambda} - (T-t) \frac{e^{-2\lambda(T-t)}}{2\lambda^2} - \frac{1-e^{-2\lambda(T-t)}}{4\lambda^3} + \frac{2e^{-\lambda(T-t)}}{\lambda^2} (T-t) \right. \\
 &\quad \left. - \frac{2-2e^{-\lambda(T-t)}}{\lambda^3} - \frac{e^{-2\lambda(T-t)}}{\lambda^2} (T-t) + \frac{1-e^{-2\lambda(T-t)}}{2\lambda^3} + \frac{T-t}{\lambda^2} \right. \\
 &\quad \left. - \frac{2-2e^{-\lambda(T-t)}}{\lambda^3} + \frac{1-1e^{-2\lambda(T-t)}}{2\lambda^3} \right\} \\
 &= \bar{C} \left\{ -\frac{e^{-2\lambda(T-t)}}{4\lambda} (T-t) - \frac{e^{-2\lambda(T-t)}}{4\lambda^2} + \frac{1}{8\lambda^3} \frac{1-e^{-2\lambda(T-t)}}{T-t} + \frac{e^{-\lambda(T-t)}}{\lambda^2} - \frac{1}{\lambda^3} \frac{1-e^{-\lambda(T-t)}}{T-t} \right. \\
 &\quad \left. - \frac{e^{-2\lambda(T-t)}}{2\lambda^2} + \frac{4}{\lambda^3} \frac{1-e^{-2\lambda(T-t)}}{T-t} + \frac{1}{2\lambda^2} - \frac{1}{\lambda^3} \frac{1-e^{-\lambda(T-t)}}{T-t} \right. \\
 &\quad \left. + \frac{1}{4\lambda^3} \frac{1-e^{-2\lambda(T-t)}}{T-t} \right\} \\
 &= \bar{C} \left\{ \frac{1}{2\lambda^2} + \frac{e^{-\lambda(T-t)}}{\lambda^2} - \frac{e^{-2\lambda(T-t)}}{2\lambda^2} - \frac{e^{-2\lambda(T-t)}}{4\lambda} (T-t) - \frac{e^{-2\lambda(T-t)}}{4\lambda^2} - \frac{e^{-2\lambda(T-t)}}{2\lambda^2} \right. \\
 &\quad \left. - \frac{1}{\lambda^3} \frac{1-e^{-\lambda(T-t)}}{T-t} - \frac{1}{\lambda^3} \frac{1-e^{-\lambda(T-t)}}{T-t} + \frac{1}{8\lambda^3} \frac{1-e^{-2\lambda(T-t)}}{T-t} \right. \\
 &\quad \left. + \frac{2}{4\lambda^3} \frac{1-e^{-2\lambda(T-t)}}{T-t} \right\} \\
 &= \bar{C} \left\{ \frac{1}{2\lambda^2} + \frac{e^{-\lambda(T-t)}}{2\lambda^2} - \frac{e^{-2\lambda(T-t)}}{4\lambda} (T-t) - \frac{3}{4\lambda^2} e^{-2\lambda(T-t)} - \frac{2}{\lambda^3} \frac{1-e^{-\lambda(T-t)}}{T-t} + \frac{5}{8\lambda^3} \frac{1-e^{-2\lambda(T-t)}}{T-t} \right\},
 \end{aligned}$$

$$\begin{aligned}
 I_D &= \frac{\bar{D}}{T-t} \int_t^T B^1(s, T) B^2(s, T) ds = \frac{\bar{D}}{T-t} \int_t^T [-(T-s)] \left[ -\frac{1-e^{-\lambda(T-s)}}{\lambda} \right] ds \\
 &= \frac{\bar{D}}{T-t} \int_t^T \frac{(T-s)-(T-s)e^{-\lambda(T-s)}}{\lambda} ds \\
 &= \frac{\bar{D}}{T-t} \left\{ \int_t^T \frac{(T-s)}{\lambda} ds - \int_t^T \frac{(T-s)e^{-\lambda(T-s)}}{\lambda} ds \right\} \\
 &= \frac{\bar{D}}{T-t} \left\{ \left[ -\frac{(T-s)^2}{2\lambda} \right]_t^T - \left[ (T-s) \frac{e^{-\lambda(T-s)}}{\lambda^2} \right]_t^T - \int_t^T \frac{e^{-\lambda(T-s)}}{\lambda^2} ds \right\} \\
 &= \frac{\bar{D}}{T-t} \left\{ \frac{(T-t)^2}{2\lambda} + \frac{e^{-\lambda(T-s)}}{\lambda^2} (T-t) - \left[ \frac{e^{-\lambda(T-s)}}{\lambda^3} \right]_t^T \right\} \\
 &= \frac{\bar{D}}{T-t} \left\{ \frac{(T-t)^2}{2\lambda} + \frac{e^{-\lambda(T-s)}}{\lambda^2} (T-t) - \frac{1-e^{-\lambda(T-s)}}{\lambda^3} \right\} \\
 &= \bar{D} \left\{ \frac{T-t}{2\lambda} + \frac{1}{\lambda^2} e^{-\lambda(T-t)} - \frac{1}{\lambda^3} \frac{1-e^{-\lambda(T-t)}}{T-t} \right\},
 \end{aligned}$$

$$\begin{aligned}
 I_E &= \frac{\bar{E}}{T-t} \int_t^T B^1(s, T) B^3(s, T) ds \\
 &= \frac{\bar{E}}{T-t} \left\{ \int_t^T [-(T-s)] \left[ (T-s)e^{-\lambda(T-s)} - \frac{1-e^{-\lambda(T-s)}}{\lambda} \right] ds \right\} \\
 &= \frac{\bar{E}}{T-t} \left\{ \int_t^T -(T-s)^2 e^{-\lambda(T-s)} ds + \int_t^T \frac{(T-s)-(T-s)e^{-\lambda(T-s)}}{\lambda} ds \right\} \\
 &= \frac{\bar{E}}{T-t} \left\{ - \left[ (T-s)^2 \frac{e^{-\lambda(T-s)}}{\lambda} \right]_t^T \right. \\
 &\quad \left. - \int_t^T 2(T-s) \frac{e^{-\lambda(T-s)}}{\lambda} ds + \int_t^T \frac{T-s}{\lambda} ds - \int_t^T \frac{(T-s)e^{-\lambda(T-s)}}{\lambda} ds \right\} \\
 &= \frac{\bar{E}}{T-t} \left\{ (T-t)^2 \frac{e^{-\lambda(T-t)}}{\lambda} - \left[ \frac{2(T-s)e^{-\lambda(T-s)}}{\lambda^2} \right]_t^T \right. \\
 &\quad \left. - \int_t^T \frac{2e^{-\lambda(T-s)}}{\lambda^2} ds - \left[ \frac{(T-s)^2}{2\lambda} \right]_t^T - \left[ \frac{(T-s)e^{-\lambda(T-s)}}{\lambda^2} \right]_t^T - \int_t^T \frac{e^{-\lambda(T-s)}}{\lambda^2} ds \right\} \\
 &= \frac{\bar{E}}{T-t} \left\{ (T-t)^2 \frac{e^{-\lambda(T-t)}}{\lambda} + \frac{2(T-t)e^{-\lambda(T-t)}}{\lambda^2} - \left[ \frac{2e^{-\lambda(T-s)}}{\lambda^3} \right]_t^T + \frac{(T-t)^2}{2\lambda} \right. \\
 &\quad \left. - \left[ \frac{(T-t)e^{-\lambda(T-s)}}{\lambda^2} \right]_t^T - \left[ \frac{e^{-\lambda(T-s)}}{\lambda^3} \right]_t^T \right\} \\
 &= \frac{\bar{E}}{T-t} \left\{ (T-t)^2 \frac{e^{-\lambda(T-t)}}{\lambda} + \frac{2(T-t)e^{-\lambda(T-t)}}{\lambda^2} - \frac{2-2e^{-\lambda(T-s)}}{\lambda^3} + \frac{(T-t)^2}{2\lambda} \right. \\
 &\quad \left. + \frac{(T-t)e^{-\lambda(T-t)}}{\lambda^2} - \frac{1-e^{-\lambda(T-s)}}{\lambda^3} \right\} \\
 &= \bar{E} \left\{ \frac{e^{-\lambda(T-t)}}{\lambda} (T-t)e^{-\lambda(T-t)} + \frac{3}{\lambda^2} e^{-\lambda(T-t)} + \frac{T-t}{2\lambda} - \frac{3}{\lambda^3} \frac{1-e^{-\lambda(T-s)}}{T-t} \right\}, \text{ and}
 \end{aligned}$$

$$\begin{aligned}
 I_F &= \frac{\bar{F}}{T-t} \int_t^T B^2(s, T) B^3(s, T) ds \\
 &= \frac{\bar{F}}{T-t} \left\{ - \int_t^T \frac{(T-s)e^{-\lambda(T-s)}}{\lambda} ds + \int_t^T \frac{(T-s)e^{-2\lambda(T-s)}}{\lambda} ds + \int_t^T \frac{1}{\lambda^2} ds \right. \\
 &\quad \left. - \int_t^T \frac{2e^{-\lambda(T-s)}}{\lambda^2} ds + \int_t^T \frac{e^{-2\lambda(T-s)}}{\lambda^2} ds \right\} \\
 &= \frac{\bar{F}}{T-t} \left\{ - \left[ (T-s) \frac{e^{-\lambda(T-s)}}{\lambda^2} \right]_t^T - \int_t^T \frac{e^{-\lambda(T-s)}}{\lambda^2} ds + \left[ \frac{(T-s)e^{-2\lambda(T-s)}}{2\lambda^2} \right]_t^T + \int_t^T \frac{e^{-2\lambda(T-s)}}{2\lambda^2} ds \right. \\
 &\quad \left. + \left[ \frac{s}{\lambda^2} \right]_t^T - \left[ \frac{2e^{-\lambda(T-s)}}{\lambda^3} \right]_t^T + \left[ \frac{e^{-2\lambda(T-s)}}{\lambda^2} \right]_t^T \right\} \\
 &= \frac{\bar{F}}{T-t} \left\{ (T-t) \frac{e^{-\lambda(T-t)}}{\lambda^2} - \left[ \frac{e^{-\lambda(T-s)}}{\lambda^3} \right]_t^T - (T-t) \frac{e^{-2\lambda(T-t)}}{2\lambda^2} + \left[ \frac{e^{-2\lambda(T-s)}}{4\lambda^3} \right]_t^T + \frac{T-t}{\lambda^2} \right. \\
 &\quad \left. - \frac{2-2e^{-\lambda(T-t)}}{\lambda^3} + \frac{1-e^{-\lambda(T-t)}}{2\lambda^3} \right\} \\
 &= \frac{\bar{F}}{T-t} \left\{ (T-t) \frac{e^{-\lambda(T-t)}}{\lambda^2} - \frac{1-e^{-\lambda(T-t)}}{\lambda^3} - (T-t) \frac{e^{-2\lambda(T-t)}}{2\lambda^2} + \frac{1-e^{-2\lambda(T-t)}}{4\lambda^3} + \frac{T-t}{\lambda^2} \right. \\
 &\quad \left. - \frac{2-2e^{-\lambda(T-t)}}{\lambda^3} + \frac{1-e^{-\lambda(T-t)}}{2\lambda^3} \right\} \\
 &= \bar{F} \left\{ \frac{1}{\lambda^2} + \frac{1}{\lambda^2} e^{-\lambda(T-t)} - \frac{1}{2\lambda^2} e^{-2\lambda(T-t)} - \frac{3}{\lambda^3} \frac{1-e^{-\lambda(T-t)}}{T-t} + \frac{3}{4\lambda^3} \frac{1-e^{-2\lambda(T-t)}}{T-t} \right\}.
 \end{aligned}$$

□

## References

- Alfaro, Rodrigo A. 2011. Affine Nelson-Siegel model. *Economics Letters* 110: 1–3. [\[CrossRef\]](#)
- Al-Zoubi, Haitham A. 2009. Short-term spot rate models with nonparametric deterministic drift. *Quarterly Review of Economics and Finance* 49: 731–47. [\[CrossRef\]](#)
- Ang, Andrew, and Monika Piazzesi. 2003. A no-arbitrage vector autoregression of term structure dynamics with macroeconomy and latent variables. *Journal of Monetary Economics* 50: 745–87. [\[CrossRef\]](#)
- Bjork, Tomas, and Bent Jesper Christensen. 1999. Interest rate dynamics and consistent forward rate curves. *Mathematical Finance* 9: 323–48. [\[CrossRef\]](#)
- Chan, Ka-keung Ceajer, Andrew Karolyi, Francis A. Longstaff, and Anthony B. Sanders. 1992. An empirical comparison of alternative models of the short-term interest rate. *Journal of Finance* 47: 1209–27. [\[CrossRef\]](#)

- Chen, Yu-chin, and Kwok Ping Tsang. 2013. What does the yield curve tell us about exchange rate predictability? *Review of Economics and Statistics* 95: 185–205. [\[CrossRef\]](#)
- Christensen, Jens H. E., Francis X. Diebold, and Glenn D. Rudebusch. 2009. An arbitrage-free generalized Nelson-Siegel term structure model. *Econometrics Journal* 12: C33–C64. [\[CrossRef\]](#)
- Christensen, Jens H.E., Francis X. Diebold, and Glenn D. Rudebusch. 2011. The affine arbitrage-free class of Nelson-Siegel term structure models. *Journal of Econometrics* 164: 4–20. [\[CrossRef\]](#)
- Coroneo, Laura, Ken Nyholm, and Rositsa Vidova-Koleva. 2011. How arbitrage-free is the Nelson-Siegel model? *Journal of Empirical Finance* 18: 393–407. [\[CrossRef\]](#)
- Cox, John, Jonathan E. Ingersol, and Stephen A. Ross. 1985. A theory of the term structure of interest rates. *Econometrica* 53: 385–407. [\[CrossRef\]](#)
- De Rezende, Rafael B., and Mauro S. Ferreira. 2013. Modeling and forecasting the yield curve by an extended Nelson-Siegel class of models: A quantile autoregression approach. *Journal of Forecasting* 32: 111–23. [\[CrossRef\]](#)
- Diebold, Francis X., and Canlin Li. 2006. Forecasting the term structure of government bond yields. *Journal of Econometrics* 130: 337–64. [\[CrossRef\]](#)
- Diebold, Francis X., and Robert S. Mariano. 1995. Comparing predictive accuracy. *Journal of Business and Economic Statistics* 13: 134–44.
- Diebold, Francis X., and Glenn D. Rudebusch. 2013. *Yield Curve Modeling and Forecasting*. Princeton: Princeton University Press.
- Duffee, Gregory R. 2002. Term premia and interest rate forecasts in affine models. *Journal of Finance* 57: 405–43. [\[CrossRef\]](#)
- Duffie, Darrell, and Rui Kan. 1996. A yield-factor model of interest rates. *Mathematical Finance* 6: 379–406. [\[CrossRef\]](#)
- Fama, Eugene F. 2006. The behavior of interest rates. *Review of Financial Studies* 19: 359–79. [\[CrossRef\]](#)
- Fujii, Mariko, and Makoto Takaoka. 2007. *Kinri no Kikannkoso to Macroeizai: Nelson—Siegel Moderu wo Motiita Approach [Term Structure of Interest Rates and Macroeconomy: Nelson—Siegel Model Approach]*. FSA Research Review. Tokyo: Financial Services Agency, pp. 219–48.
- Gasha, Giancarlo, Ying He, Carlos Medeiros, Marco Rodriguez, Jean Salvati, and Jiangbo Yi. 2010. *On the Estimation of Term Structure Models and an Application to the United State*. IMF Working Paper WP/10/258. Washington: International Monetary Fund.
- Gurkaynak, Refet S., Brian Sack, and Jonathan H. Wright. 2007. The U.S. treasury yield curve: 1961 to the present. *Journal of Monetary Economics* 54: 2291–304. [\[CrossRef\]](#)
- Gurkaynak, Refet S., and Jonathan H. Wright. 2012. Macroeconomics and the term structure. *Journal of Economic Literature* 50: 331–67. [\[CrossRef\]](#)
- Hamilton, James D., and Jing Cynthia Wu. 2012. Identification and estimation of Gaussian affine term structure models. *Journal of Econometrics* 168: 315–31. [\[CrossRef\]](#)
- Hördahl, Peter, Oreste Tristani, and David Vestin. 2005. A joint econometric model of macroeconomic and term-structure dynamics. *Journal of Econometrics* 131: 405–44. [\[CrossRef\]](#)
- Hull, John, and Alan White. 1990. Pricing interest rate derivative securities. *Review of Financial Studies* 3: 573–92. [\[CrossRef\]](#)
- Ishii, Hokuto. 2018. Modeling and predictability of exchange rate changes by the extended relative Nelson-Siegel class of models. *International Journal of Financial Studies* 6: 68. [\[CrossRef\]](#)
- Krippner, Leo. 2015. *Zero Lower Bound Term Structure Modeling*. New York: Palgrave Macmillan.
- Litterman, Robert, and Jose Scheinkman. 1991. Common factors affecting bond returns. *Journal of Fixed Income* 1: 54–61. [\[CrossRef\]](#)
2008. Moench, Emanuel 2008. Forecasting the yield curve in a data-rich environment: a no-arbitrage factor augmented VAR approach. *Journal of Econometrics* 146: 26–43. [\[CrossRef\]](#)
- Niu, Linlin, and Gengming Zeng. 2012. The discrete-time framework of the arbitrage-free Nelson-Siegel class of term structure models. *SSRN 2015858*: 1–68. [\[CrossRef\]](#)
- Nelson, Charles, and Andrew Siegel. 1987. Parsimonious modeling of yield curves. *Journal of Business* 60: 473–89. [\[CrossRef\]](#)
- Piazzesi, Monika 2010. Affine term structure models. In *Handbook of Financial Econometrics*. Amsterdam, New York and Oxford: Elsevier, pp. 691–766.

- Rudebusch, Glenn D., and Tao Wu. 2008. A macro-finance model of the term-structure, monetary policy, and the economy. *Economic Journal* 118: 906–26. [[CrossRef](#)]
- Svensson, Lars E. O. 1994. *Estimating and Interpreting Forward Interest Rates: Sweden 1992–1994*. NBER Working Paper #4871. Cambridge: National Bureau of Economic Research.
- Vasicek, Oldrich. 1977. An equilibrium characterization of the term structure. *Journal of Financial Economics* 5: 177–88. [[CrossRef](#)]
- Wu, Tao. 2006. Macro Factors and the Affine Term Structure of Interest Rates. *Journal of Money, Credit and Banking* 38: 1847–75. [[CrossRef](#)]



© 2019 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).