UPPSALA
UNIVERSITET

U.U.D.M. Project Report 2021:45

Fitting Yield Curve with Dynamic Nelson-Siegel Models: Evidence from Sweden

Zhe Huang

Examensarbete i matematik, 30 hp
Handledare: Yukai Yang
Ämnesgranskare: Rolf Larsson
Examinator: Magnus Jacobsson
Juni 2021

Department of Mathematics
Uppsala University

Abstract

In this thesis, the yield curve of the Swedish treasury bill and government bond is modelled by five dynamic Nelson-Siegel models separately. Estimations of the latent factors are given by the Kalman filter optimization. According to the estimation results, the dynamic Nelson-Siegel model is the most stable and convenient model to fit the yields with all maturities. The arbitrage-free Nelson-Siegel model makes good performance on the short-term yields. The dynamic generalized Nelson-Siegel model has the advantage for estimating medium-term and long-term yields.

Acknowledgement

I would like to extend my sincerest and deepest thanks to my supervisor Yukai Yang for offering me this exciting topic, for his patient guidance, constant encouragement and support. I learnt a lot from his supervision. I believe that will be the fortune in my life. I would like to thank my subject reviewer Rolf Larsson for giving me comments to improve my thesis. I would like to acknowledge Jens Christensen for explaining his idea on AFNS models. I would also like to give my thanks to Yuqiong Wang for helping me understand the arbitrage theory.

In addition, I would like to thank my parents and my friends for their company during this tough period.

Contents

1 Introduction

In many financial areas, it is important to get a good estimation of the yield curve. From the last century, economists attempt to develop plenty of models to fit the yield curve with different forms. One class of the most widely used models are the Nelson-Siegel models.

Nelson and Siegel [1987] suggest a simple static model with three latent factors to fit the yield curve of the bond market. The most significant contribution of the Nelson-Siegel model is using only three variables to describe a complex curve with great performance. The three factors can be interpreted as the level, the slope and the curvature of the yield curve. Since then many different versions of Nelson-Siegel model have been offered for building zero-coupon yield curves by researchers.

To study the dynamic evolution of the yield curve, Diebold and Li [2006] offer a dynamic version of the three-factor Nelson-Siegel model, which is named as dynamic Nelson-Siegel model. In this model, Diebold and Li develop the three Nelson-Siegel factors to latent time-varying parameters. Diebold et al. [2006] use the Kalman filter maximum log-likelihood optimization method to estimate the Nelson-Siegel parameters, which has become the common method to deal with this kind of problems now.

Empirically, the dynamic Nelson-Siegel model has good achievement on in-sample fit, but it does not set the restrictions necessary for the absence of arbitrage. To solve the theoretical restrictions, Christensen et al. [2011] develop the affine arbitrage-free class of dynamic Nelson-Siegel term structure models, referred to as the Arbitrage-free Nelson-Siegel model.

Besides the traditional three-factor model, Diebold and Rudebusch [2013] collect some extended Nelson-Siegel models with additional factors. For example, the four-factor Svensson model developed by Svensson [1995], the five-factor dynamic generalized Nelson-Siegel model created by Christensen et al. [2009] and so on. Following the study of Christensen et al. [2009], the extra factors can improve the fitting performance on long-term bond yields.

Although there is a vast of papers about the Nelson-Siegel model and its empirical study, they seldom focus on Swedish data. While in my thesis, I investigate Nelson-Siegel models and figure out appropriate models for the Swedish treasury bill and government bond yields. I mainly apply three versions of the Nelson-Siegel model in this thesis, the dynamic Nelson-Siegel model, the arbitrage-free Nelson-Siegel model and the dynamic generalized Nelson-Siegel model. In the following section, I review some financial concepts that are related to this thesis. In Section 3, I introduce the state space models and the Kalman filter. In Section 4, I explain the Nelson-Siegel term structure models, because they are the key models for estimations in next section. Section 5 is the most important in my thesis, all the estimation results are contained in this section and I provide in-sample fit and out-of-sample forecast tests on different models. The R Codes of my thesis can be founded on my GitHub page `https://github.com/ZheHuang96/Master-Thesis-Code`.

2 Financial Preliminaries

In this section, I follow Diebold and Rudebusch [2013] and introduce some financial facts as a basis of my thesis. The facts are interest rates, the zero-coupon bond yields, and the factor structure of a yield curve.

2.1 Interest Rates

There are three interest rates in the bond market, the discount rate, the forward rate and the yield. The discount rate of a bond is the interest for calculating bond prices with the present value. The yield to maturity is the interest rate that represents the total return when the bond reaches maturity. The forward rate is the future yield of the bond. Following Diebold and Rudebusch [2013], there are three definitions.

Definition 2.1. *The relationship between the discount rate and the yield is defined as:*

$$P(\tau) = e^{-\tau y(\tau)} \tag{2.1}$$

where $P(\tau)$ is the price of a τ-period discount bond and $y(\tau)$ is the continuously compound yield to maturity of this bond.

Definition 2.2. *The relationship between the discount rate and the forward rate is defined as:*

$$f(\tau) = -\frac{P'(\tau)}{P(\tau)} \tag{2.2}$$

where $f(\tau)$ is the forward rate of a bond.

It is easy to get the connection between the yield and the forward rate by inserting Equation (2.1) into Equation (2.2).

Definition 2.3. *The relationship between the yield and the forward rate is defined as:*

$$y(\tau) = \frac{1}{\tau} \int_{0}^{\tau} f(u)du \tag{2.3}$$

Hence, researchers construct the unobserved yield with two different approaches, the forward rate and the discount rate.

2.2 Zero-Coupon bond Yields

A zero-coupon bond, as named, is a kind of bond without any coupons or interest payments during the bond holding period. When the bond attains maturity, the bondholder can obtain the face value of the bond. Hence the zero-coupon bond is the simplest bond in the bond market. In general, the government bonds of countries are zero-coupon bonds. The most common example is the US Treasury Bills, which is widely used in vast papers.

Definition 2.4. *The relationship between the price and yield of a zero-coupon bond is defined by*

$$y = [\frac{F}{p}]^{\frac{1}{n}} - 1 \tag{2.4}$$

where y represents the yield of a zero-coupon bond, F represents the face value, p stands for the current price, and n is the compounding period to maturity.

Practically, the yields are not observed and need to be estimated by the prices of zero-coupon bonds. The most popular way to construct yields is using the estimated forward rates at related maturities. Following Diebold and Rudebusch [2013], the zero-coupon yield is an equally-weighted average of forward rates.

2.3 Yield Curve Factors

A yield curve, also called term structure, is a curve that graphs many different yields to different maturities at any time. The evolution of the yield curve is supposed to be dynamic. To describe the yield curve better, many studies suggest the imposition of "factors".

Normally, the bond market yields are described by multivariate models. According to many pieces of evidence shown, the financial asset returns can be modelled by a typical restricted vector autoregressive process, illustrated as a factor structure. The basic method behind the factor structure is using a set of limited latent objects ("factors") to drive a large set of objects, e.g., the bond yields in this thesis. Hence, the complicated observations can be explained by easy dynamic variables. Researchers think the term structure data can have various factors in the real world. They also prove that the dynamic factor models can fit the observed yield data with high accuracy as a whole.

The difficulty of fitting the yield curve is how to figure out the dynamic evolution of latent models. To build and estimate the unknown dynamic factor models from observations, state space models and the Kalman Filter method are necessary. Hence they will be introduced in details in the next section.

3 State Space Models and Kalman Filter

This section considers how to fit the yield curve with latent factor models. Dynamic models are widely used to describe the evolution of the financial market. In this section, I introduce some methods to build and estimate dynamic models.

3.1 State Space Models

In many economic applications, the evolution of input and output variables cannot be measured directly. To study the evolution of the internal variables, the common way is to apply state space models, suggested by Kunst [2007].

A state space model is a time series model that describes an observed time series, Y_t, by the unobserved state vector, X_t. Usually, the dynamic system of X_t is a Markov process, which means X_t only depends on the history of X_{t-1}. Hence, a linear state space model has two equations: the measurement equation

$$Y_t = BX_t + \epsilon_t \tag{3.1}$$

and the state transition equation

$$X_t = AX_{t-1} + C\mu_{t-1} + \eta_t \tag{3.2}$$

where μ is the mean vector and the measurement error ϵ_t and the state error η_t are i.i.d white noise. Normally, ϵ_t is assumed to be a zero-mean Guassian with the covariance H, i.e. $\epsilon_t \sim N(0, H)$ and η_t is assumed to be a zero-mean Guassian with the covariance Q, i.e. $\eta_t \sim N(0, Q)$.

The measurement equation illustrates the relationship between the observed time series Y_t and the unobserved state X_t. The state transition equation illustrates the evolution of the state vector from time $t-1$ to time t. As the equations show, several latent variables need to be estimated. To solve this problem, researchers have created an effective method, called the Kalman filter.

3.2 Kalman Filter

In practical utilization, the Kalman filter is a useful way to estimate the latent variables of the linear state space model. The Kalman filter can help to construct the log-likelihood function related to the state space model. The optimal predictions can be obtained by using the maximum log-likelihood method.

The procedure of the Kalman filter is to estimate X_t, given the initial estimate of X_0, the observed series of measurement Y_t, and the information of the system described by A, B, C, H and Q.

According to Kim and Bang [2018], the Kalman filter algorithm includes two steps: prediction and update. Consider the period $t-1$ and suppose the state update X_{t-1} and its covariance matrix Σ_{t-1}, and the update step is in time t. The Kalman filter algorithm is briefed as follows:

Prediction:

$$X_{t|t-1} = AX_{t-1} + C\mu_{t-1} \quad \text{predicted state estimate} \tag{3.3}$$
$$\Sigma_{t|t-1} = A\Sigma_{t-1}A' + Q \quad \text{predicted error covariance} \tag{3.4}$$

Update:

$$X_{t|t} = X_{t|t-1} + K_t v_t \quad \text{updated state estimate} \tag{3.5}$$
$$\Sigma_{t|t} = (I - K_t B)\Sigma_{t|t-1} \quad \text{updated error covariance} \tag{3.6}$$

where $v_t = y_t - BX_{t|t-1}$ is the measurement residual and $K_t = \Sigma_{t|t-1}B'(H + B\Sigma_{t|t-1}B')^{-1}$ is called the Kalman gain.

Because the Kalman filter is a recursive procedure, initialization is needed. The initial guess of the state estimate and the error covariance matrix can be the initial values, with Q and H. Finally, a Kalman filter can be completed by implementing the prediction and update steps for each time step after the initialization.

Kalman filter offers all the parameters to construct the log-likelihood function for the observations. But the maximization of the log-likelihood function is complicated. Prediction error decomposition is a common format of log-likelihood function to simply the problem. The prediction error decomposition of the log-likelihood is

$$\log l(Y_T|\varphi) = \sum_{t=1}^{T} \log l(y_t|Y_{t-1}, \varphi) \tag{3.7}$$

$$= -\sum_{t=1}^{T} (\frac{N}{2} \log 2\pi + \frac{1}{2} \log |\Omega_{t|t-1}| + \frac{1}{2} v_t^{'} \Omega_{t|t-1}^{-1} v_t) \tag{3.8}$$

where N is the number of observations, φ represents the unknown parameters, $Y_T = (y_1, \ldots, y_T)$, $v_t = y_t - y_{t-1} = y_t - Bx_{t|t-1}$, and $\Omega_{t|t-1} = B\Sigma_{t|t-1}B' + H$. It is easy to maximize the log-likelihood function and get the optimization of unknown parameters by this decomposition.

Up to now, I have proposed the general methods to build and estimate latent factor models. I will introduce the specific dynamic models for yields in the next section.

4 The Nelson-Siegel Term Structure Models

This section introduces the important underlying models used in the thesis. To make everything clear, I start this part with the original Nelson-Siegel model created by Nelson and Siegel [1987], then move to three dynamic models, the dynamic Nelson-Siegel model defined by Diebold and Li [2006], the arbitrage-free Nelson-Siegel model illustrated by Christensen et al. [2011] and the dynamic generalized Nelson-Siegel model designed by Christensen et al. [2009].

4.1 Nelson-Siegel Model

The Nelson-Siegel model is a three-factor exponential parametric model founded by Nelson and Siegel [1987]. They discover that the classic yield curve shape functions follow the solutions to differential and difference equations and the forward rates are the solutions to the differential equations generated by spot rates. Hence, they give the function to fit the forward rate curve as follows:

$$f(\tau) = \beta_1 + \beta_2 e^{-\lambda\tau} + \beta_3 \lambda e^{-\lambda\tau} \tag{4.1}$$

The zero-coupon yield curve is equal to the average of the forward rates, it can be implied as:

$$y(\tau) = \beta_1 + \beta_2(\frac{1-e^{-\lambda\tau}}{\lambda\tau}) + \beta_3(\frac{1-e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau}) \tag{4.2}$$

where $y(\tau)$ is the zero-coupon yield with τ months to maturity, λ is a constant the controls the decay rate and β_1, β_2, β_3 are latent parameters. For the yield curve, β_1 is the long-term factor and imposes the level of it, β_2 is the short-term factor and interprets the slope, β_3 is the medium-term factor and is related to the curvature.

The Nelson-Siegel model is the basic model for the DNS model, the AFNS model and the DGNS model.

4.2 Dynamic Nelson-Siegel Model(DNS)

In order to find the progression of the bond market time by time, Diebold and Li [2006] improve the Nelson-Siegel model by adding dynamic behaviour to the factors of the model. The new model is:

$$y_t(\tau) = L_t + S_t(\frac{1-e^{-\lambda\tau}}{\lambda\tau}) + C_t(\frac{1-e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau}) \tag{4.3}$$

where L_t, S_t, C_t are latent time-varying parameters which have the same interpretation as β_1, β_2, β_3 in the original Nelson-Siegel model. L_t, S_t, C_t represent level, slope and curvature of the yield curve. The factor loading is shown as Figure 4.1. The dynamic movements of the latent parameters follow the time series process. For example, independent AR(1) processes and a VAR(1) process.

The dynamic Nelson-Siegel model is a state-space model and the parameters are state variables. In this thesis, two versions of the DNS model are considered, the independent-factor DNS model and the correlated-factor DNS model.

For the independent-factor DNS model, the three latent factors follow independent first-order auto-regressive processes (AR(1)). The state transition equation is

$$\begin{pmatrix} L_t - \mu_L \\ S_t - \mu_S \\ C_t - \mu_C \end{pmatrix} = \begin{pmatrix} a_{11} & 0 & 0 \\ 0 & a_{22} & 0 \\ 0 & 0 & a_{33} \end{pmatrix} \begin{pmatrix} L_{t-1} - \mu_L \\ S_{t-1} - \mu_S \\ C_{t-1} - \mu_C \end{pmatrix} + \begin{pmatrix} \eta_t(L) \\ \eta_t(S) \\ \eta_t(C) \end{pmatrix} \tag{4.4}$$

where the stochastic shocks $\eta_t(L)$, $\eta_t(S)$ and $\eta_t(C)$ have diagonal covariance matrix

$$Q = \begin{pmatrix} q_{11}^2 & 0 & 0 \\ 0 & q_{22}^2 & 0 \\ 0 & 0 & q_{33}^2 \end{pmatrix} \tag{4.5}$$

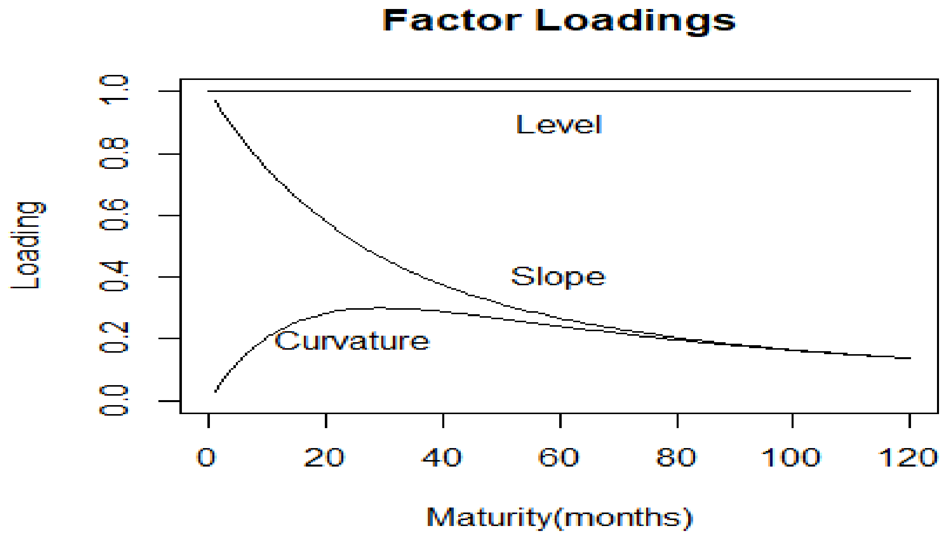

Figure 4.1: The DNS factor loadings for $\lambda = 0.0609$.

For the correlated-factor DNS model, the factors follow a first-order vector autoregressive process (VAR(1)). Both the mean and the covariance matrix are correlated, hence the transition matrix and the state error matrix are non-diagonal. The transition equation is

$$\begin{pmatrix} L_t - \mu_L \\ S_t - \mu_S \\ C_t - \mu_C \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \begin{pmatrix} L_{t-1} - \mu_L \\ S_{t-1} - \mu_S \\ C_{t-1} - \mu_C \end{pmatrix} + \begin{pmatrix} \eta_t(L) \\ \eta_t(S) \\ \eta_t(C) \end{pmatrix} \tag{4.6}$$

where the stochastic shocks $\eta_t(L)$, $\eta_t(S)$ and $\eta_t(C)$ have non-diagonal state errors matrix $Q = qq^{'}$, where

$$q = \begin{pmatrix} q_{11} & 0 & 0 \\ q_{21} & q_{22} & 0 \\ q_{31} & q_{32} & q_{33} \end{pmatrix} \tag{4.7}$$

The measurement equation for two DNS models is

$$\begin{pmatrix} y_t(\tau_1) \\ y_t(\tau_2) \\ \vdots \\ y_t(\tau_N) \end{pmatrix} = \begin{pmatrix} 1 & \frac{1-e^{-\lambda\tau_1}}{\lambda\tau_1} & \frac{1-e^{-\lambda\tau_1}}{\lambda\tau_1} - e^{-\lambda\tau_1} \\ 1 & \frac{1-e^{-\lambda\tau_2}}{\lambda\tau_2} & \frac{1-e^{-\lambda\tau_2}}{\lambda\tau_2} - e^{-\lambda\tau_2} \\ \vdots & \vdots & \vdots \\ 1 & \frac{1-e^{-\lambda\tau_N}}{\lambda\tau_N} & \frac{1-e^{-\lambda\tau_N}}{\lambda\tau_N} - e^{-\lambda\tau_N} \end{pmatrix} \begin{pmatrix} L_t \\ S_t \\ C_t \end{pmatrix} + \begin{pmatrix} \epsilon_t(\tau_1) \\ \epsilon_t(\tau_2) \\ \vdots \\ \epsilon_t(\tau_N) \end{pmatrix} \tag{4.8}$$

where the measurement errors $\epsilon_t(\tau_i)$ are i.i.d. white noise.

4.3 Arbitrage-free Nelson-Siegel Model(AFNS)

Christensen et al. [2011] develop the Arbitrage-free Nelson-Siegel (AFNS) model by imposing the affine arbitrage-free term structure into the DNS model.

Here I briefly introduce the affine arbitrage-free term structure by following Björk [2009] and Duffie and Kan [1996].

Definition 4.1. *A stochastic process W is named a Wiener process when it satisfies the conditions as follows:*

- $W(0) = 0$
- *The process has independent increments.*
- *For $s < t$, the stochastic variable $W(t) - W(s)$ has the Gaussian distribution $N[0, \sqrt{t-s}]$.*
- *W has continuous trajectories.*

Proposition 4.1. *For a multidimensional stochastic process X_t, the short rate r_t is a deterministic function of X_t.*

$$r_t = \rho_0(t) + \rho_1(t)' X_t \tag{4.9}$$

where $\rho_0 : [0, T] \to \mathbf{R}$ and $\rho_0 : [0, T] \to \mathbf{R^n}$ are continuous bounded functions.

If the bond market is arbitrage-free, there is a risk-neutral measure Q, and the zero-coupon bond prices with an affine term structure can be represented as follows:

$$P(t, T) = E^Q[\exp(-\int_t^T r_u du)] \tag{4.10}$$

$$= \exp(A(t, T) - B(t, T)' X_t) \tag{4.11}$$

where the Q-dynamics of the process X_t are given by

$$dX_t = K^Q(t)[\theta^Q(t) - X_t]dt + \Sigma(t) D(X_t, t) dW_t^Q \tag{4.12}$$

In Equation (4.12), W_t^Q is a Q-Wiener process, K^Q, θ^Q, Σ are continuous bounded functions. D is a diagonal matrix and the ith diagonal element is $\sqrt{\gamma^i(t) + \delta_1^i(t) X_t^1 + \cdots + \delta_n^i(t) X_t^n}$, where γ and δ are continuous bounded functions.

The main idea to build the AFNS model is to construct a new term structure model that not only follows the affine arbitrage-free structure but also satisfies the three-factor dynamic Nelson-Siegel term structure. Hence, Christensen et al. [2011] prove the proposition of the AFNS model as follows:

Proposition 4.2. *Define that the instantaneous risk-free rate (short rate) is*

$$r_t = X_t^1 + X_t^2 \tag{4.13}$$

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

$$\begin{pmatrix} dX_t^1 \\ dX_t^2 \\ dX_t^3 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & \lambda & -\lambda \\ 0 & 0 & \lambda \end{pmatrix} \left[\begin{pmatrix} \theta_1^Q \\ \theta_2^Q \\ \theta_3^Q \end{pmatrix} - \begin{pmatrix} X_t^1 \\ X_t^2 \\ X_t^3 \end{pmatrix} \right] dt + \Sigma \begin{pmatrix} dW_t^{1,Q} \\ dW_t^{2,Q} \\ dW_t^{3,Q} \end{pmatrix}, \quad \lambda > 0 \tag{4.14}$$

where X_t is a Markov process, W_t^Q is a Wiener process under the Q-measure, Σ is the volatility matrix. Then following Duffie and Kan [1996], the zero-coupon bond prices are:

$$P(t,T) = E_t^Q[\exp(-\int_t^T r_u du)] \tag{4.15}$$

$$= \exp(B^1(t,T)X_t^1 + B^2(t,T)X_t^2 + B^3(t,T)X_t^3 + A(t,T)) \tag{4.16}$$

where $B^1(t,T)$, $B^2(t,T)$, $B^3(t,T)$, and $A(t,T)$ are solutions to the system of ODEs:

$$\begin{pmatrix} \frac{dB^1(t,T)}{dt} \\ \frac{dB^2(t,T)}{dt} \\ \frac{dB^3(t,T)}{dt} \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & -\lambda & \lambda \end{pmatrix} \begin{pmatrix} B^1(t,T) \\ B^2(t,T) \\ B^3(t,T) \end{pmatrix} \tag{4.17}$$

and

$$\frac{dA(t,T)}{dt} = -B(t,T)' K^Q \theta^Q - \frac{1}{2} \sum_{j=1}^{3} (\Sigma' B(t,T) B(t,T)' \Sigma)_{j,j} \tag{4.18}$$

with the boundary conditions $B^1(t,T) = B^2(t,T) = B^2(t,T) = A(t,T) = 0$. In this system of ODEs, K^Q is the mean reversion matrix, θ^Q is the means of factors. The solution to this system of ODEs is:

$$B^1(t,T) = -(T-t) \tag{4.19}$$

$$B^2(t,T) = \frac{1 - e^{-\lambda(T-t)}}{\lambda} \tag{4.20}$$

$$B^3(t,T) = (T-t)e^{-\lambda(T-t)} - \frac{1-e^{-\lambda(T-t)}}{\lambda} \tag{4.21}$$

$$A(t,T) = (K^Q\theta^Q)_2 \int_t^T B^2(s,T)ds + (K^Q\theta^Q)_3 \int_t^T B^3(s,T)ds$$
$$+ \frac{1}{2}\sum_{j=1}^{3}\int_t^T (\Sigma' B(t,T)B(t,T)'\Sigma)_{j,j}ds. \tag{4.22}$$

Finally, zero-coupon bound yields are:

$$y(t,T) = X_t^1 + \frac{1-e^{-\lambda(T-t)}}{\lambda(T-t)}X_t^2 + \left[\frac{1-e^{-\lambda(T-t)}}{\lambda(T-t)} - e^{-\lambda(T-t)}\right]X_t^3 - \frac{A(t,T)}{T-t} \tag{4.23}$$

where X_t^1 represents the level factor, X_t^2 represents the slope factor, X_t^3 represents the curvature factor, and λ is the decay parameter.

From Proposition 4.2, the instantaneous interest rate is the sum of level and slope factors in the yield function. Comparing (4.23) with (4.3), it is clear to find that both the AFNS model and the DNS model have the same factor loadings in their yield functions. The key difference between the AFNS model and the DNS model is the $-\frac{A(t,T)}{T-t}$ term. Hence, it is defined as the yield-adjustment term.

Christensen et al. [2011] identify the yield-adjustment term in their AFNS models by giving zero value to the mean parameters of the state variables under the Q-measure, i.e. $\theta^Q = 0$. Hence, the yield-adjustment term can be represented as:

$$-\frac{A(t,T)}{T-t} = -\frac{1}{2}\frac{1}{T-t}\sum_{j=1}^{3}\int_t^T (\Sigma' B(s,T)B(s,T)'\Sigma)_{j,j}ds \tag{4.24}$$

Given a general volatility matrix

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{pmatrix} \tag{4.25}$$

the term can be derived in analytical form as

$$\frac{A(t,T)}{T-t} = \frac{1}{2}\frac{1}{T-t}\sum_{j=1}^{3}\int_t^T (\Sigma' B(s,T)B(s,T)'\Sigma)_{j,j}ds \tag{4.26}$$

$$
\begin{aligned}
&= \overline{A}\frac{(T-t)^2}{6} \\
&+ \overline{B}\left[\frac{1}{2\lambda^2} - \frac{1}{\lambda^3}\frac{1-e^{-\lambda(T-t)}}{T-t} + \frac{1}{4\lambda^3}\frac{1-e^{-2\lambda(T-t)}}{T-t}\right] \\
&+ \overline{C}\left[\frac{1}{2\lambda^2} + \frac{1}{\lambda^2}e^{-\lambda(T-t)} - \frac{1}{4\lambda}(T-t)e^{-2\lambda(T-t)}\right. \\
&\left. - \frac{3}{4\lambda^2}e^{-2\lambda(T-t)} - \frac{2}{\lambda^3}\frac{1-e^{-\lambda(T-t)}}{T-t} + \frac{5}{8\lambda^3}\frac{1-e^{-2\lambda(T-t)}}{T-t}\right] \\
&+ \overline{D}\left[\frac{1}{2\lambda}(T-t) + \frac{1}{\lambda^2}e^{-\lambda(T-t)} - \frac{1}{\lambda^3}\frac{1-e^{-\lambda(T-t)}}{T-t}\right] \\
&+ \overline{E}\left[\frac{1}{2\lambda}(T-t) + \frac{3}{\lambda^2}e^{-\lambda(T-t)} + \frac{1}{\lambda}(T-t)e^{-\lambda(T-t)}\right. \\
&\left. - \frac{3}{\lambda^3}\frac{1-e^{-\lambda(T-t)}}{T-t}\right] \\
&+ \overline{F}\left[\frac{1}{\lambda^2} + \frac{1}{\lambda^2}e^{-\lambda(T-t)} - \frac{1}{2\lambda^2}e^{-2\lambda(T-t)}\right. \\
&\left. - \frac{3}{\lambda^3}\frac{1-e^{-\lambda(T-t)}}{T-t} + \frac{3}{4\lambda^3}\frac{1-e^{-2\lambda(T-t)}}{T-t}\right]
\end{aligned} \tag{4.27}
$$

where

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

Because the volatility matrix is not separately identified and only the six terms above can be identified, Christensen et al. [2011] suggest that the volatility matrix of the maximally flexible AFNS specification is a triangular matrix (Both lower and upper triangular forms work well in the term, the choice is not important.)

$$
\Sigma = \begin{pmatrix} \sigma_{11} & 0 & 0 \\ \sigma_{21} & \sigma_{22} & 0 \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{pmatrix} \tag{4.28}
$$

In order to build the AFNS models in continuous time, it is necessary to change the measure under the real world dynamics (the P-measure) into the risk neutral measure (the Q-measure).

The P-measure is the real-world measure that measured the probability with actual data in real market rather than the hypothesis of imposing absence of arbitrage into the market. The Q-measure is the risk-neutral measure with no arbitrage opportunities that has convenience on pricing assets. The arbitrage-free condition is important to the existence of a Q-measure. The relationship is:

$$dW_t^Q = dW_t^P + \Gamma_t dt \tag{4.29}$$

where Γ_t is the risk premium, W_t represents the standard Brownian motion. Following Duffee [2002], the essentially affine risk premium specification can keep the affine dynamics under the P-measure, and Γ_t is

$$\Gamma_t = \begin{pmatrix} \gamma_1^0 \\ \gamma_2^0 \\ \gamma_3^0 \end{pmatrix} + \begin{pmatrix} \gamma_{11}^1 & \gamma_{12}^1 & \gamma_{13}^1 \\ \gamma_{21}^1 & \gamma_{22}^1 & \gamma_{23}^1 \\ \gamma_{31}^1 & \gamma_{32}^1 & \gamma_{33}^1 \end{pmatrix} \begin{pmatrix} X_t^1 \\ X_t^2 \\ X_t^3 \end{pmatrix} \tag{4.30}$$

With this specification of Γ_t, the required affine structure under the Q-measure is retained and I can also utilize any mean vector θ^P and mean-reversion matrix K^P to construct the SDE equation under the P-measure.

$$dX_t = K^P[\theta^P - X_t]dt + \Sigma dW_t^P \tag{4.31}$$

For convenience, the independent-factor and correlated-factor AFNS models are introduced in this thesis.

In the independent-factor AFNS model, the state transition equation is

$$\begin{pmatrix} dX_t^1 \\ dX_t^2 \\ dX_t^3 \end{pmatrix} = \begin{pmatrix} \kappa_{11}^P & 0 & 0 \\ 0 & \kappa_{22}^P & 0 \\ 0 & 0 & \kappa_{33}^P \end{pmatrix} \left[\begin{pmatrix} \theta_1^P \\ \theta_2^P \\ \theta_3^P \end{pmatrix} - \begin{pmatrix} X_t^1 \\ X_t^2 \\ X_t^3 \end{pmatrix} \right] dt + \begin{pmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & \sigma_3 \end{pmatrix} \begin{pmatrix} dW_t^{1,p} \\ dW_t^{2,p} \\ dW_t^{3,p} \end{pmatrix} \tag{4.32}$$

In the correlated-factor AFNS model, the state transition equation is

$$\begin{pmatrix} dX_t^1 \\ dX_t^2 \\ dX_t^3 \end{pmatrix} = \begin{pmatrix} \kappa_{11}^P & \kappa_{12}^P & \kappa_{13}^P \\ \kappa_{21}^P & \kappa_{22}^P & \kappa_{23}^P \\ \kappa_{31}^P & \kappa_{32}^P & \kappa_{33}^P \end{pmatrix} \left[\begin{pmatrix} \theta_1^P \\ \theta_2^P \\ \theta_3^P \end{pmatrix} - \begin{pmatrix} X_t^1 \\ X_t^2 \\ X_t^3 \end{pmatrix} \right] dt + \begin{pmatrix} \sigma_{11} & 0 & 0 \\ \sigma_{21} & \sigma_{22} & 0 \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{pmatrix} \begin{pmatrix} dW_t^{1,p} \\ dW_t^{2,p} \\ dW_t^{3,p} \end{pmatrix} \tag{4.33}$$

Equation (4.33) is more flexible because all the parameters are identified.

The measurement equation for AFNS models is

$$\begin{pmatrix} y_t(\tau_1) \\ y_t(\tau_2) \\ \vdots \\ y_t(\tau_N) \end{pmatrix} = \begin{pmatrix} 1 & \frac{1-e^{-\lambda\tau_1}}{\lambda\tau_1} & \frac{1-e^{-\lambda\tau_1}}{\lambda\tau_1} - e^{-\lambda\tau_1} \\ 1 & \frac{1-e^{-\lambda\tau_2}}{\lambda\tau_2} & \frac{1-e^{-\lambda\tau_2}}{\lambda\tau_2} - e^{-\lambda\tau_2} \\ \vdots & \vdots & \vdots \\ 1 & \frac{1-e^{-\lambda\tau_N}}{\lambda\tau_N} & \frac{1-e^{-\lambda\tau_N}}{\lambda\tau_N} - e^{-\lambda\tau_N} \end{pmatrix} \begin{pmatrix} X_t^1 \\ X_t^2 \\ X_t^3 \end{pmatrix} - \begin{pmatrix} \frac{A(\tau_1)}{\tau_1} \\ \frac{A(\tau_2)}{\tau_2} \\ \vdots \\ \frac{A(\tau_N)}{\tau_N} \end{pmatrix} + \begin{pmatrix} \epsilon_t(\tau_1) \\ \epsilon_t(\tau_2) \\ \vdots \\ \epsilon_t(\tau_N) \end{pmatrix} \tag{4.34}$$

4.4 Dynamic Generalized Nelson-Siegel Model(DGNS)

DNS models and AFNS models are developed from the Nelson-Siegel model. Hence, all the models are three-factor models that included one level factor, one slope factor and one curvature factor. From Figure 4.1, slope and curvature factors decrease to zero rapidly with maturity. When fitting long-maturity yields, slope and curvature factors are not available, only level factor can be used to describe yields. To solve this problem, Svensson [1995] introduce a four-factor Nelson-Siegel model with an extra curvature factor (Nelson-Siegel Svensson model) and Christensen et al. [2009] create a five-factor Nelson-Siegel model with one level factor, two slope factors and two curvature factors.

The dynamic five-factor Nelson-Siegel model is labelled as the dynamic generalized Nelson-Siegel model (DGNS). The yield function of DGNS is

$$\begin{aligned} y_t(\tau) = L_t &+ S_t^1(\frac{1-e^{-\lambda_1\tau}}{\lambda_1\tau}) + S_t^2(\frac{1-e^{-\lambda_2\tau}}{\lambda_2\tau}) + C_t^1(\frac{1-e^{-\lambda_1\tau}}{\lambda_1\tau} - e^{-\lambda_1\tau}) \\ &+ C_t^2(\frac{1-e^{-\lambda_2\tau}}{\lambda_2\tau} - e^{-\lambda_2\tau}) \end{aligned} \tag{4.35}$$

where the dynamic factors $(L_t, S_t^1, S_t^2, C_t^1, C_t^2)$ are interpreted as one level factor, two slope factors and two curvature factors, and λ_1, λ_2 are described as decay parameters. (Note that Christensen et al. [2009] restricted $\lambda_1 > \lambda_2$.)

Same with the DNS models, the DGNS model is a state-space model and in my thesis, I only consider the independent case. Because the correlated case has 54 parameters and it is too complicated to estimate. In the independent-factor DGNS model, all five state factors follow independent

AR(1) processes. The state transition equation is

$$\begin{pmatrix} L_t-\mu_L \\ S_t^1-\mu_S^1 \\ S_t^2-\mu_S^2 \\ C_t^1-\mu_C^1 \\ C_t^2-\mu_C^2 \end{pmatrix} = \begin{pmatrix} a_{11} & 0 & 0 & 0 & 0 \\ 0 & a_{22} & 0 & 0 & 0 \\ 0 & 0 & a_{33} & 0 & 0 \\ 0 & 0 & 0 & a_{44} & 0 \\ 0 & 0 & 0 & 0 & a_{55} \end{pmatrix} \begin{pmatrix} L_{t-1}-\mu_L \\ S_{t-1}^1-\mu_S^1 \\ S_{t-1}^2-\mu_S^2 \\ C_{t-1}^1-\mu_C^1 \\ C_{t-1}^2-\mu_C^2 \end{pmatrix} + \begin{pmatrix} \eta_t(L) \\ \eta_t(S^1) \\ \eta_t(S^2) \\ \eta_t(C^1) \\ \eta_t(C^2) \end{pmatrix} \tag{4.36}$$

where the state error matrix is

$$Q = \begin{pmatrix} q_{11}^2 & 0 & 0 & 0 & 0 \\ 0 & q_{22}^2 & 0 & 0 & 0 \\ 0 & 0 & q_{33}^2 & 0 & 0 \\ 0 & 0 & 0 & q_{44}^2 & 0 \\ 0 & 0 & 0 & 0 & q_{55}^2 \end{pmatrix} \tag{4.37}$$

And the measurement equation is

$$\begin{pmatrix} y_t(\tau_1) \\ y_t(\tau_2) \\ \vdots \\ y_t(\tau_N) \end{pmatrix} = \begin{pmatrix} 1 & \frac{1-e^{-\lambda_1\tau_1}}{\lambda_1\tau_1} & \frac{1-e^{-\lambda_2\tau_1}}{\lambda_2\tau_1} & \frac{1-e^{-\lambda_1\tau_1}}{\lambda_1\tau_1}-e^{-\lambda_1\tau_1} & \frac{1-e^{-\lambda_2\tau_1}}{\lambda_2\tau_1}-e^{-\lambda_2\tau_1} \\ 1 & \frac{1-e^{-\lambda_1\tau_2}}{\lambda_1\tau_2} & \frac{1-e^{-\lambda_2\tau_2}}{\lambda_2\tau_2} & \frac{1-e^{-\lambda_1\tau_2}}{\lambda_1\tau_2}-e^{-\lambda_1\tau_2} & \frac{1-e^{-\lambda_2\tau_2}}{\lambda_2\tau_2}-e^{-\lambda_2\tau_2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \frac{1-e^{-\lambda_1\tau_N}}{\lambda_1\tau_N} & \frac{1-e^{-\lambda_2\tau_N}}{\lambda_2\tau_N} & \frac{1-e^{-\lambda_1\tau_N}}{\lambda_1\tau_N}-e^{-\lambda_1\tau_N} & \frac{1-e^{-\lambda_2\tau_N}}{\lambda_2\tau_N}-e^{-\lambda_2\tau_N} \end{pmatrix}$$

$$\begin{pmatrix} L_t \\ S_t^1 \\ S_t^2 \\ C_t^1 \\ C_t^2 \end{pmatrix} + \begin{pmatrix} \epsilon_t(\tau_1) \\ \epsilon_t(\tau_2) \\ \vdots \\ \epsilon_t(\tau_N) \end{pmatrix} \tag{4.38}$$

where the measurement errors $\epsilon_t(\tau_i)$ are i.i.d. white noise.

All the necessary models are introduced in this section. I apply the models for estimations and offer an empirical study on Swedish data in next section.

5 Empirical Study

This section mainly includes the empirical study based on five underlying models in Section 4 and estimated methods in Section 3. The empirical study has four parts, the data introduction, the estimation framework of dynamic Nelson-Siegel models, the results of estimated models and the out-of-sample forecasts.

5.1 Data

The estimates in my thesis are based on monthly Swedish treasury bill and government bond yields. Treasury bills are short-term debt tools with maturities of less than one year. The treasury bills in Sweden include four durations: one month, three months, six months and one year. However, the one-year treasury bill is discontinued in October 2010. So I do not collect it into my data set. Swedish government bonds are long-term debt tools with four maturities: two years, five years, seven years and ten years. Treasury bills and government bonds are issued by the Swedish National Debt Office.

The final data are zero-coupon yields from January 1996 to December 2011 and at seven maturities: one month, three months, six months, two years, five years, seven years and ten years. The data are directly downloaded from the Swedish Central Bank website.

Notice that, in Diebold et al. [2006], Christensen et al. [2011] and other papers, they estimate and test dynamic Nelson-Siegel models by using the monthly US Treasury bill and bond yields with a wide range of maturities and get good performances. The data are zero-coupon yields with sixteen maturities: 3, 6, 9, 12, 18, 24, 36, 48, 60, 84, 96, 108, 120, 180, 240, and 360 months. Compared with the US data, I do not have so many observations at different maturities for my thesis. But it is also fascinating to investigate the performances of the models under this situation.

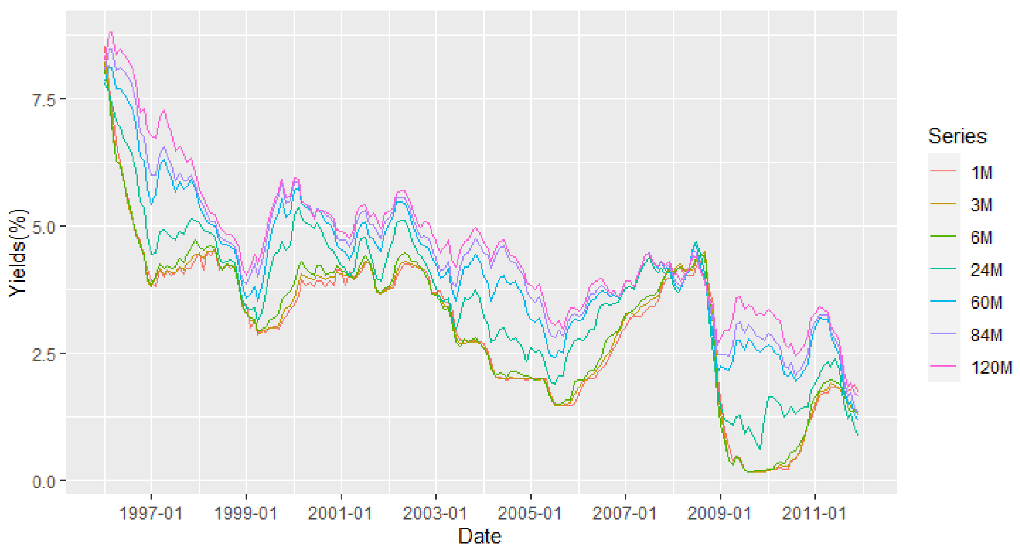

Figure 5.1: The time series of the data from January 1996 to December 2011.

From Figure 5.1, I observe the Swedish yields have a general descending trend from January 1996 to December 2011. Usually, the yields at long maturity are higher than the yields at short maturity. In this figure, the 10-year yield curve (the pink line) is over other yield curves. Under this curve, there are the 7-year yield curve, the 5-year yield curve and so on. But the difference between the yields is not obvious when the yields are at short maturity (less than one year). Another interesting thing is that the yields at seven different maturities become quite close and go down sharply from 2008 to 2009 because of the global financial crisis in 2008. Especially, the crisis has a serious influence on the short-term bonds, which made the bond yields close to zero in nearly two years (2009-2010). But the long-term bonds are not as deeply affected as the short-term bonds.

Table 5.1 shows the descriptive statistics for monthly data at seven maturities. The mean of yields increases when the maturity increases. The standard deviations of yields are acceptable and the differences between minimum and maximum of the yields at all maturities looks huge. The empirical level, slope and curvature factors are also included. Following Diebold and Li [2006], the empirical level factor is interpreted as the yield at the longest maturity because the level of the yield curve is the long-term factor. (In my thesis, it is the 10-year yield.) The empirical slope factor is interpreted as the difference between the three-month yield and ten-year yield. Finally, the empirical curvature factor is interpreted as twice the 2-year yield minus the sum of the 3-month and 10-year yields.

Table 5.1: Descriptive Statistics for the yield curve

Maturity	Mean	Standard Deviation	Minimum	Maximum
1M	3.0240	1.4934	0.1520	8.5241
3M	3.0318	1.4982	0.1518	8.2227
6M	3.0903	1.5007	0.1657	8.0418
24M	3.5945	1.4635	0.6105	7.8177
60M	4.1602	1.3789	1.1596	8.1376
84M	4.3688	1.3974	1.2789	8.4619
120M(Level)	4.5896	1.4157	1.6769	8.7943
Slope	-1.9973	1.1235	-4.2131	0.7512
Curvature	-1.1687	1.7592	-6.0278	2.5035

5.2 Estimation Framework

Five different dynamic Nelson-Siegel models are explored in this part. My estimation focuses on the DNS model and AFNS model with two versions (independent-factor and correlated-factor), and the independent-factor DGNS model. In this part, I introduce the framework to estimate and forecast Swedish data with dynamic Nelson-Siegel models.

5.2.1 DNS Model

For the DNS models, the state-space representations can be written as follows:

$$X_t = (I - A)\mu + AX_{t-1} + \eta_t \tag{5.1}$$
$$Y_t = BX_t + \epsilon_t \tag{5.2}$$

where $X_t = (L_t, S_t, C_t)$, I is the identity matrix, μ is a vector of factor means, the matrices A and B are constant. B is the factor-loading matrix for the decay parameter λ. The error structure is

$$\begin{pmatrix} \eta_t \\ \epsilon_t \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} Q & 0 \\ 0 & H \end{pmatrix}\right], \tag{5.3}$$

According to (4.5) and (4.7), the matrix Q in the independent-factor DNS model is diagonal and in the correlated-factor DNS model is not diagonal. Following Diebold and Li [2006], the matrix H is diagonal.

Now the Kalman filter is applied to estimate the parameters. In the independent-factor DNS model, there are 17 parameters, and in the correlated-factor DNS model, there are 26 parameters. Following Diebold et al. [2006], $X_0 = \mu$ and $\Sigma_0 = V$, where V solves $V = AVA' + Q$. The prediction step of time $t - 1$ is:

$$X_{t|t-1} = (I - A)\mu + AX_{t-1} \tag{5.4}$$
$$\Sigma_{t|t-1} = A\Sigma_{t-1}A' + Q \tag{5.5}$$

The update step of time t is:

$$X_{tt} = X_{t|t-1} + \Sigma_{t|t-1}B'F_t^{-1}v_t \tag{5.6}$$
$$\Sigma_{tt} = (I - \Sigma_{t|t-1}B'F_t^{-1}B)\Sigma_{t|t-1} \tag{5.7}$$

where measurement residual $v_t = y_t - A - BX_{t|t-1}$, $F_t = cov(v_t) = B\Sigma_{t|t-1}B' + H$, $H = diag(\sigma_\epsilon^2(\tau_1), ..., \sigma_\epsilon^2(\tau_N))$.

Given all parameters from the Kalman filter, the log-likelihood function can be classified, the prediction error decomposition of the function is

$$logL = -\frac{TN}{2}log(2\pi) - \sum_{t=1}^{T}(\frac{v_t'F_t^{-1}v_t + log|F_t|}{2}) \tag{5.8}$$

where N is the number of observed yields. The maximum log-likelihood estimates are obtained by optimizing the parameters and the decay parameter λ. In this thesis, $N = 192$, and λ for initialling Kalman filter is 0.6915, which is chosen by the mean of the time series of empirical λ parameters. The empirical λ parameters are the results of fixing the basic Nelson-Siegel model with the data.

Here I do not choose the common decay parameter $\lambda = 0.0609$ that is recommended by Diebold et al. [2006], because I think that for different term structure data, the decay parameter is different. Instead, I use the mean of the empirical parameters as the constant initial λ parameter.

5.2.2 AFNS Model

Following Christensen et al. [2011], the state-space representations of the AFNS models can be written as:

$$X_t = (I - \exp(-K^P \Delta t))\theta^P + \exp(-K^P \Delta t)X_{t-1} + \eta_t \tag{5.9}$$

$$Y_t = BX_t - A + \epsilon_t \tag{5.10}$$

where Δt is the time difference between the data. The error structure is same as (5.3), only the matrix Q_t in AFNS models is different.

$$Q_t = \int_0^{\Delta t} e^{-K^P s}\Sigma\Sigma' e^{-(K^P)' s} ds \tag{5.11}$$

Notice that the matrix Q is equal to the conditional covariance matrix.

The Kalman filter-based maximum-likelihood method for AFNS models is similar with the method for DNS models. But I initialize the Kalman filter with the unconditional mean and covariance matrix in AFNS models, $X_0 = \theta^P$ and $\Sigma_0 = \int_0^\infty e^{-K^P s}\Sigma\Sigma' e^{-(K^P)' s} ds$. Hence the prediction step is

$$X_{t|t-1} = E^P[X_t|Y_{t-1}] = (I - \exp(-K^P \Delta t))\theta^P + \exp(-K^P \Delta t)X_{t-1} \tag{5.12}$$

$$\Sigma_{t|t-1} = \exp(-K^P \Delta t)\Sigma_{t-1}[\exp(-K^P \Delta t)]' + Q_t \tag{5.13}$$

Other steps can be found in the DNS model estimation. To keep the covariance stationary under the P-measure, Christensen et al. restrict the eigenvalues of A to be less than 1 in DNS and DGNS models and the real component of each eigenvalue of K^P to be positive.

5.2.3 DGNS Model

The independent-factor DGNS model has the same representations and the error structure as the independent-factor DNS model, the only difference is the size of matrices, e.g. the matrix A and the state error matrix Q are 5×5 matrices, and the factor $X_t = (L_t, S_t^1, S_t^2, C_t^1, C_t^2)$. There are 24 parameters in this model. To avoid collinearity between the factors, the difference of initial λ_1, λ_2 parameters can be set as big as possible. Here I use the optimized λ for the independent-factor DNS model as the first decay parameter, $\lambda_1 = 0.7144$, and set $\lambda_2 = 0.0144$.

5.2.4 Out-of-sample Forecast Method

The out-of-sample forecast is a common and important way to test the performance of financial estimators and models. To check dynamic Nelson-Siegel models, I splice the observations in two parts with forecasting period h. I use the first part to get the forecast yields with h months ahead $\hat{Y}_{t+h}(\tau)$, where τ is the maturity and t is the time, and compare forecast yields with observed yields. Following Christensen et al. [2011], the forecast error is defined as $\hat{e}_{t+h}(\tau) = Y_{t+h}(\tau) - \hat{Y}_{t+h}(\tau)$, and forecast performances are decided by the root mean squared forecast error (RMSFE).

5.3 Estimation Results

This section is the summary of all the estimation results involved in my thesis. The results are illustrated separately by different models.

5.3.1 DNS Model Estimation

The results of the estimates for DNS models are shown below.

Table 5.2: Independent-factor Estimates

	A matrix			Mean	q matrix		
A	L_{t-1}	S_{t-1}	C_{t-1}	μ	q_L	q_S	q_C
L_t	0.9979	0	0	4.1201	0.1988	0	0
S_t	0	0.9777	0	-2.5407	0	0.2920	0
C_t	0	0	0.9342	-1.4222	0	0	0.5947

The table shows the estimated A matrix, the μ vector, and the estimated q matrix. The estimated λ is 0.7144 for maturities measured in months. The maximized log-likelihood is 863.6072.

Table 5.3: Correlated-factor Estimates

	A matrix			Mean	q matrix		
A	L_{t-1}	S_{t-1}	C_{t-1}	μ	q_L	q_S	q_C
L_t	1.0271	0.0270	-0.0186	4.4539	0.1997	0	0
S_t	-0.0434	0.8930	0.1007	-2.4504	-0.1661	0.1747	0
C_t	-0.1289	-0.0072	1.0052	-1.3399	-0.1315	-0.0098	0.6725

The table shows the estimated A matrix, the μ vector, and the estimated q matrix. The estimated λ is 0.6913 for maturities measured in months. The maximized log-likelihood is 950.7484.

In 5.2 and 5.3, from the estimated transition matrix, it is obvious that the level factor is the most persistent factor in both models and the curvature factor is least persistent in the independent-factor model, whereas the slope factor is least persistent in the correlated-factor model. The persistence of the slope factor in the independent model decreases in the correlated model and the persistence of the curvature factor in the independent model increases in the correlated model. These two differences are close.

Observe that most of the off-diagonal elements excepted $A_{S_t,C_{t-1}}$ and $A_{C_t,L_{t-1}}$ in the correlated-factor model are at same level. Christensen et al. [2011] suggest that the key non-zero off-diagonal element for the AFNS model is $A_{S_t,C_{t-1}}$, which is 0.1007 in my estimation. This estimate may investigate that the AFNS model is suitable for the data in my thesis.

Comparing the empirical factor means with the estimated factor means in the DNS models, the deviation is quite significant, which may show that the dynamic models are necessary. Notice that there are few things to describe the observed data, hence the means are important parameters to contrast.

For convenience, I plot the three empirical factors against the estimated DNS factors separately as shown in Figures 5.2, 5.3 and 5.4, where red solid lines represent empirical factors, green solid lines stand for independent factors, and blue dashed lines are correlated factors. There is not much difference between the independent-factor DNS model and the correlated-factor DNS model as shown in figures.

Figure 5.2: The empirical and estimated DNS level factors.

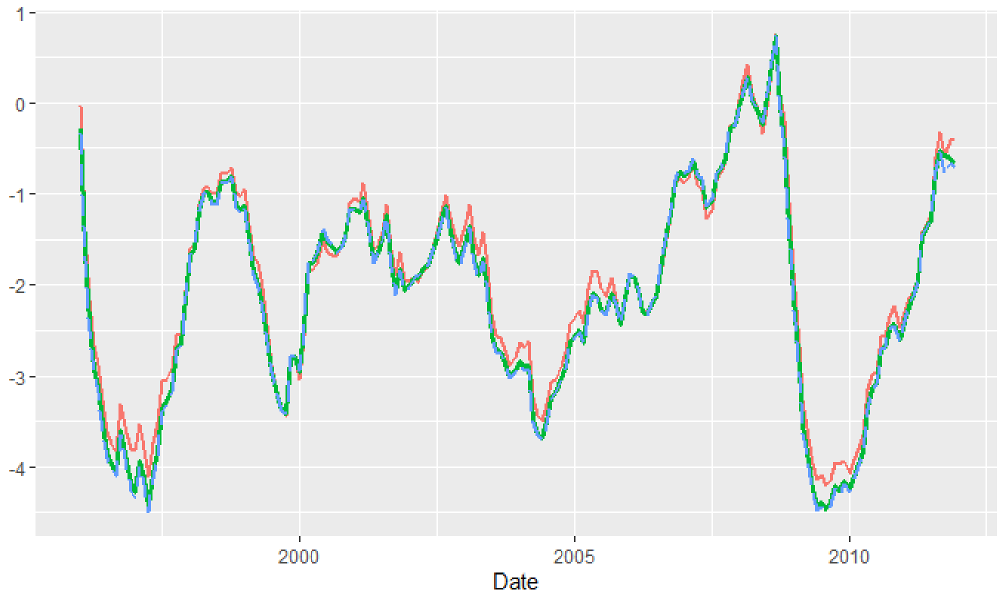

Figure 5.3: The empirical and estimated DNS slope factors.

Figure 5.4: The empirical and estimated DNS curvature factors.

Following Diebold et al. [2006], the state transition covariance matrix is given by $Q = qq'$, using the state error estimates in 5.2 and 5.3. The conditional covariance matrices are

$$Q^{DNS}_{indep} = \begin{pmatrix} 0.0395 & 0 & 0 \\ 0 & 0.0853 & 0 \\ 0 & 0 & 0.3537 \end{pmatrix} \tag{5.14}$$

$$Q^{DNS}_{corr} = \begin{pmatrix} 0.0399 & -0.0332 & -0.0263 \\ -0.0332 & 0.0581 & 0.0201 \\ -0.0263 & 0.0201 & 0.4696 \end{pmatrix} \tag{5.15}$$

From the matrices above, the volatilities of the level factor in both models are similar. The variance of the slope factor decreases and the variance of curvature factor increases in the correlated DNS model. The correlation for shocks between the level and slope factors is -0.6895, the correlation for shocks between the level and curvature factors is -0.1921, and the correlation for shocks between the slope and curvature factors is 0.1217.

Based on the structures of the independent-factor model and the correlated-factor model, the two models are nested. I followed Christensen et al. [2011] to test the independent factor restrictions by the standard likelihood ratio test. The null hypothesis is the independent-factor model. The likelihood ratio is given by

$$LR = 2[logL(\theta_{corr}) - logL(\theta_{indep})] \sim \chi^2(9) \tag{5.16}$$

I obtain $LR = 174.2824$ and the associated p-value less than 0.0001. Hence, the restrictions established in the independent-factor model are rejected.

Table 5.4 is the summary of in-sample fit that shows the means and root mean squared errors for residuals of DNS and AFNS models. Notice that although the correlated DNS model has more parameters and is more flexible than the independent DNS model, there is no obvious difference in the in-sample fit.

Finally, yield estimates of DNS models with observed yields at seven different maturities separately are drawn as Figure A.1. The legends of this figure are the same as figures of factors. Figure A.1 prove that the DNS models can fit the observed data perfectly.

Table 5.4: Summary Statistics of In-Sample Fit For DNS and AFNS Models

Maturity	Independent DNS		Correlated DNS		Independent AFNS		Correlated AFNS	
in months	Mean	RMSE	Mean	RMSE	Mean	RMSE	Mean	RMSE
1	0.0361	0.0946	0.0367	0.0934	0.0039	0.0907	-0.0010	0.0878
3	-0.0001	0.0006	-0.0001	0.0001	-6e-07	0.0014	-3e-07	0.0019
6	-0.0114	0.0649	-0.0116	0.0629	0.0197	0.0632	0.0238	0.0604
24	0.0587	0.1381	0.0599	0.1405	0.0279	0.1472	0.0112	0.2052
60	0.0050	0.0549	0.0078	0.0602	-0.1822	1.1164	-0.2329	1.4526
84	-0.0184	0.0798	-0.0166	0.0831	0.0001	1.7108	0.0008	2.2256
120	-0.0004	0.0140	-0.0004	0.0027	0.7670	2.3204	0.9268	3.0391

5.3.2 AFNS Model Estimation

So far I have evaluated the simple dynamic Nelson-Siegel model with independent-factor and correlated-factor cases. Now the estimation results of the independent-factor arbitrage-free Nelson-Siegel model and the correlated-factor arbitrage-free Nelson-Siegel model are shown in this part.

The estimates for independent-factor AFNS model are displayed in Table 5.5, and the estimates for correlated-factor AFNS model are shown in Table 5.6

Table 5.5: Independent-factor AFNS Model Estimates

	K^P matrix			Mean	Σ matrix		
K^P	$K^P_{.,1}$	$K^P_{.,2}$	$K^P_{.,1}$	θ^P	$\Sigma_{.,1}$	$\Sigma_{.,2}$	$\Sigma_{.,3}$
$K^P_{1,.}$	0.0081	0	0	5.2524	0.6897	0	0
$K^P_{2,.}$	0	0.0911	0	-2.5457	0	0.9559	0
$K^P_{3,.}$	0	0	0.0510	-2.0331	0	0	2.4386

The table shows the estimated K^P matrix, the θ^P vector, and the estimated Σ matrix. The estimated λ is 0.5196 for maturities measured in months. The maximized log-likelihood is 315.0290.

Table 5.6: Correlated-factor AFNS Model Estimates

	K^P matrix			Mean	Σ matrix		
K^P	$K^P_{.,1}$	$K^P_{.,2}$	$K^P_{.,1}$	θ^P	$\Sigma_{.,1}$	$\Sigma_{.,2}$	$\Sigma_{.,3}$
$K^P_{1,.}$	0.1236	-0.3347	0.3226	5.2100	0.6157	0	0
$K^P_{2,.}$	-0.2236	0.4007	-0.6015	-2.4938	0	0.8074	0
$K^P_{3,.}$	0.7017	0.0392	0.5886	-1.9977	-0.5496	0.5599	2.1989

The table shows the estimated K^P matrix, the θ^P vector, and the estimated Σ matrix. The estimated λ is 0.4977 for maturities measured in months. The maximized log-likelihood is 241.0708.

First, I compare the independent DNS model with the independent AFNS model. Both two models have 17 parameters, hence the log-likelihoods can be contrasted straightforwardly. Because the log-likelihood value of the estimated independent AFNS model (315.0290) is much lower than the log-likelihood value of the estimated independent DNS model (863.6072), the independent AFNS model shows weaker in-sample performance in Table 5.4. The continuous-time matrix K^P in AFNS model can not be compared directly with the one-month conditional matrix A in DNS model, so I translate the K^P matrix into the one-month conditional matrix, then I compare the mean-reversion matrix (5.17) with the matrix A in Table 5.2, the level factor and the curvature factor in the AFNS model show more persistence than in the DNS model and the slope factor becomes less persistent in the AFNS model.

$$exp\left(-K^P\frac{1}{12}\right) = \begin{pmatrix} 0.9993 & 0 & 0 \\ 0 & 0.9924 & 0 \\ 0 & 0 & 0.9957 \end{pmatrix} \tag{5.17}$$

Then I use same method to compare two correlated models. The lower log-likelihood value of AFNS model shows weaker in-sample performance. The one-month conditional matrix is

$$exp\left(-K^P\frac{1}{12}\right)=\begin{pmatrix}0.9898 & 0.0283 & -0.0265\\ 0.0012 & 0.9366 & 0.0008\\ -0.0008 & -0.0010 & 0.9430\end{pmatrix} \tag{5.18}$$

In the correlated-factor case, the level factor and the curvature factor in the AFNS model show less persistence than in the DNS model and the slope factor becomes more persistent in the AFNS model. Here the result is different from the result of the independent-factor case.

Figure 5.5, 5.6, 5.7 are given for comparing the estimated AFNS factors with empirical factors, where red solid lines represent empirical factors, green solid lines stand for independent factors, and blue dashed lines are correlated factors. From the figures, there are obvious biases between three AFNS factors and empirical factors that may cause bad performances on in-sample fit and out-of-sample forecast.

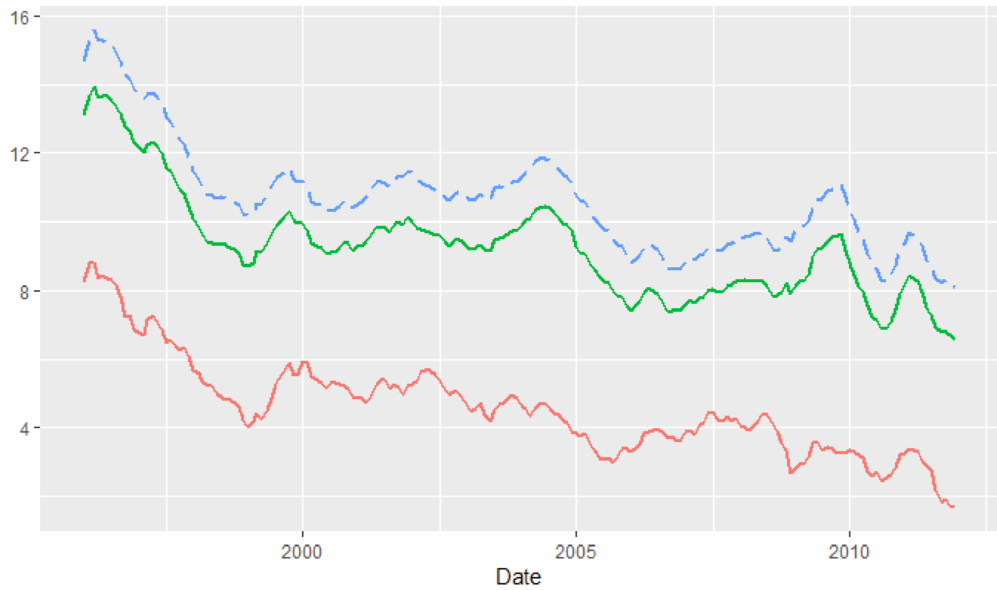

Figure 5.5: The empirical and estimated AFNS level factors.

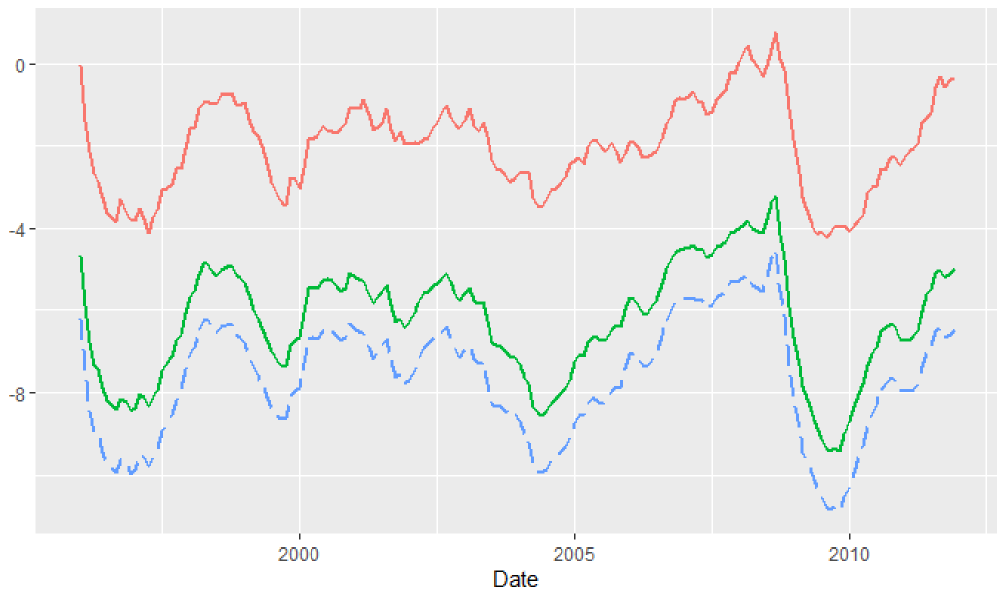

Figure 5.6: The empirical and estimated AFNS slope factors.

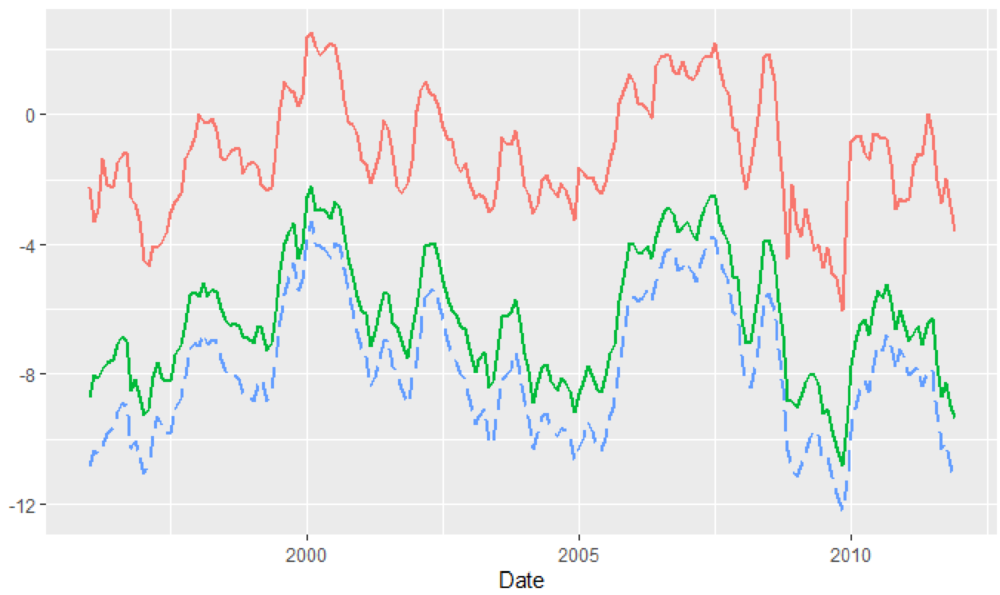

Figure 5.7: The empirical and estimated AFNS curvature factors.

Same with the DNS models, the independent-factor AFNS model and the correlated-factor AFNS model are nested. Applying the equation (5.16), $LR = 147.9164$ and the associated p-value less than 0.0001. The restrictions established in the independent-factor model are rejected. Hence, the independent-factor cases are enough to fit the data set in DNS and AFNS models.

Observing the factor means from Table 5.5 and 5.6, the AFNS models factor means estimates are higher than the estimated factor means in DNS models, especially, the estimated level factors in AFNS models are nearly 20% higher than the level factors in DNS models. In Nelson-Siegel models, the level factor is the long-term factor that influenced yields at long maturities. Hence, higher level factors might indicate that AFNS models are not suitable for long-maturity yields. Meanwhile, from Table 5.4, the residuals and RMSEs are quite big when the maturity equals 120 months. To figure out this problem straightforwardly, I plot yield estimates with observed yields at seven different maturities separately as Figure A.2, the legends of this figure are the same as in previous figures. There are obvious biases between the estimated data and observations when the maturities are five years, seven years and ten years. The AFNS models can fit the yields with maturities less than or equal to 2 years. Hence, AFNS models have poor performance when fitting the long-maturity yields.

Based on the results I get, for my data set, it is no significant difference between DNS models and AFNS models when the maturities are short. For long-maturity yields, DNS models might be a good choice. However, since my data set only has 1344 samples and seven maturities, we need to test more data to get a general conclusion.

5.3.3 DGNS Model Estimation

The estimation of independent-factor DGNS model is interpreted as below.

Table 5.7: Independent-factor DGNS Estimates

	A matrix					Mean
A	L_{t-1}	S^1_{t-1}	S^2_{t-1}	C^1_{t-1}	C^2_{t-1}	μ
L_t	0.8003	0	0	0	0	4.6207
S^1_t	0	0.9506	0	0	0	-1.0735
S^2_t	0	0	0.9940	0	0	-0.1879
C^1_t	0	0	0	0.9438	0	-1.4205
C^2_t	0	0	0	0	0.9879	-0.5323

The table shows the estimated A matrix, the μ vector. The estimated λ_1 is 1.4316 and λ_2 is 0.0530 for maturities measured in months. The maximized log-likelihood is 933.938.

The state transition covariance matrix is

$$Q_{indep}^{DGNS} = \begin{pmatrix} 5.4908 \times 10^{-5} & 0 & 0 & 0 & 0 \\ 0 & 0.1036 & 0 & 0 & 0 \\ 0 & 0 & 0.0694 & 0 & 0 \\ 0 & 0 & 0 & 0.1302 & 0 \\ 0 & 0 & 0 & 0 & 0.4089 \end{pmatrix} \tag{5.19}$$

Following Christensen et al. [2009], I compare the level factor, first slope factor and first curvature factor of the independent DGNS model with the corresponding factors in independent DNS model. Both level and first slope factors are less persistent than DNS factors, while the first curvature factor is more persistent. Estimated λ parameters are 1.4316 and 0.0530. The high value of the first λ describes that the factor loading of the first slope and curvature factors drop to zero faster than other models. But the second slope and curvature factors loading decrease extremely slow because of the small value of second λ, which may have a good fit to long-maturity yields.

Comparing Table 5.4 with Table 5.8, when maturities are 84 months and 120 months, the means and RMSEs of residuals in the independent DGNS model are smaller than in the independent DNS model. The result supports the guess that the DGNS model is suitable for yields at long maturities.

Figure A.3 shows the estimated results of the DGNS model with observations. The red line is the observations and the green line is the estimated yields. Compared with the AFNS model estimates, the DGNS model has distinct improvement on fitting the long-maturity yield curve. However, it is a slight difference between the DGNS model performance and the DNS model performance.

Table 5.8: In-Sample Fit For DGNS Model

Maturity	Independent DGNS	
in months	Mean	RMSE
1	0.0065	0.0502
3	-0.0026	0.0171
6	0.0017	0.0208
24	0.0166	0.1477
60	-0.0057	0.0833
84	-0.0056	0.0692
120	2e-06	4e-05

5.4 Out-of-sample Forecasts

The out-of-sample forecasting RMSFE results are shown in Table 5.9, Table 5.10, and Table 5.11. The three tables include five estimated models at six maturities with 3-,6-,12-month forecast separately.

The bold value represents the smallest RMSFE value of the most accurate model among five models. From the three tables, it is easy to find that DNS models (especially the independent case) have good performances half of the time (nine of eighteen combinations). In seven of eighteen combinations, the most accurate model is the independent-factor DNS model.

When the maturity is less than or equal to two years, the AFNS models perform extremely well, e.g., the three-month yield, but when the maturity is long, the RMSFEs of AFNS models increase fast that show bad forecast performances. Finally, the independent-factor DGNS model is suitable for long-maturity yields with long forecast periods.

Table 5.9: 3-month out-of-sample RMSEs For Five Models

Maturities / Models	3	6	24	60	84	120
Independent DNS	0.0156	0.0090	0.0320	**0.0311**	**0.0339**	**0.0177**
Correlated DNS	0.0170	0.0166	**0.0293**	0.0340	0.0365	0.0195
Independent AFNS	0.0285	0.0168	0.0485	0.1586	0.2378	0.3095
Correlated AFNS	**0.0085**	0.0125	0.0515	0.2131	0.3176	0.4169
Independent DGNS	0.0136	**0.0069**	0.0412	0.0374	0.0352	0.0179

Table 5.10: 6-month out-of-sample RMSEs For Five Models

Maturities / Models	3	6	24	60	84	120
Independent DNS	0.0447	0.0629	**0.1607**	**0.1837**	**0.1827**	**0.1633**
Correlated DNS	0.0894	0.1131	0.2187	0.2185	0.2043	0.1723
Independent AFNS	0.0763	0.0843	0.1944	0.3800	0.4832	0.5784
Correlated AFNS	**0.0155**	**0.0304**	0.1726	0.4252	0.5636	0.6976
Independent DGNS	0.0680	0.0846	0.1906	0.2063	0.1983	0.1742

Table 5.11: 12-month out-of-sample RMSEs For Five Models

Models \ Maturities	3	6	24	60	84	120
Independent DNS	0.1002	0.0974	0.1438	0.2177	0.2308	0.2285
Correlated DNS	0.0852	0.0954	0.1878	0.2371	0.2369	**0.2210**
Independent AFNS	**0.0778**	0.0877	0.1953	0.4513	0.6086	0.7611
Correlated AFNS	0.1837	0.1688	**0.1150**	0.4277	0.6302	0.8309
Independent DGNS	0.0826	**0.0843**	0.1554	**0.2172**	**0.2256**	0.2282

6 Conclusion

In this thesis, I model the Swedish treasury bill and government bond yields from January 1996 to December 2011 with three different kinds of dynamic Nelson-Siegel models. They are the DNS model, the AFNS model and the DGNS model. Especially, for DNS and AFNS models, I estimate with independent-factor and correlated-factor cases.

I can get some information from the estimation results. First, the dynamic Nelson-Siegel model is a reliable method to model the Swedish data. It has stable performances on fitting the yields with all maturities.

Second, the arbitrage-free Nelson-Siegel model has obvious advantages and disadvantages. It works well when fitting short-maturity yields, particularly the correlated AFNS model, which has perfect out-of-sample forecast performance on three-month and six-month yields. However, the AFNS model has poor performance on medium-maturity and long-maturity yields, e.g., ten-year yield. This conclusion is same as the summary of Christensen et al. [2011]. Diebold and Rudebusch [2013] suggests that the imposition of free-arbitrage must degrade in-sample fit and the independent-factor AFNS model must fit worse than the correlated-factor AFNS model. But the results of my empirical study are different from these two conclusions. The possible reason might be the different term structures between Swedish data and American data.

Finally, the independent DGNS model can fix on yields with long-maturity, but most of the time, the performances of this five-factor model do not have a significant difference from three-factor DNS model.

Besides the models in this thesis, there are many interesting versions of Nelson-Siegel models worth estimating Swedish government yields. For example, the Markov switching dynamic Nelson-Siegel model and adaptive dynamic Nelson-Siegel model. we can have a try in the future.

Bibliography

Charles R Nelson and Andrew F Siegel. Parsimonious modeling of yield curves. *Journal of Business*, pages 473–489, 1987.

Francis X. Diebold and Canlin Li. Forecasting the term structure of government bond yields. *Journal of Econometrics*, 130(2):337–364, 2006.

Francis X Diebold, Glenn D Rudebusch, and S Boragan Aruoba. The macroeconomy and the yield curve: a dynamic latent factor approach. *Journal of Econometrics*, 131(1-2):309–338, 2006.

Jens HE Christensen, Francis X Diebold, and Glenn D Rudebusch. The affine arbitrage-free class of nelson-siegel term structure models. *Journal of Econometrics*, 164(1):4–20, 2011.

Francis X Diebold and Glenn D Rudebusch. *Yield curve modeling and forecasting: the dynamic Nelson-Siegel approach*. Princeton University Press, 2013.

Lars EO Svensson. Estimating forward interest rates with the extended nelson& siegel method. *Sveriges Riksbank Quarterly Review*, 3(1):13–26, 1995.

Jens HE Christensen, Francis X Diebold, and Glenn D Rudebusch. An arbitrage-free generalized nelson-siegel term structure model, 2009.

Robert Kunst. State space models and the kalman filter. *Vektorautoregressive Methoden*, 2007.

Youngjoo Kim and Hyochoong Bang. Introduction to kalman filter and its applications. *Introduction and Implementations of the Kalman Filter*, 1: 1–16, 2018.

Tomas Björk. *Arbitrage theory in continuous time*. Oxford university press, 2009.

Darrell Duffie and Rui Kan. A yield-factor model of interest rates. *Mathematical finance*, 6(4):379–406, 1996.

Gregory R Duffee. Term premia and interest rate forecasts in affine models. *The Journal of Finance*, 57(1):405–443, 2002.

A Figures

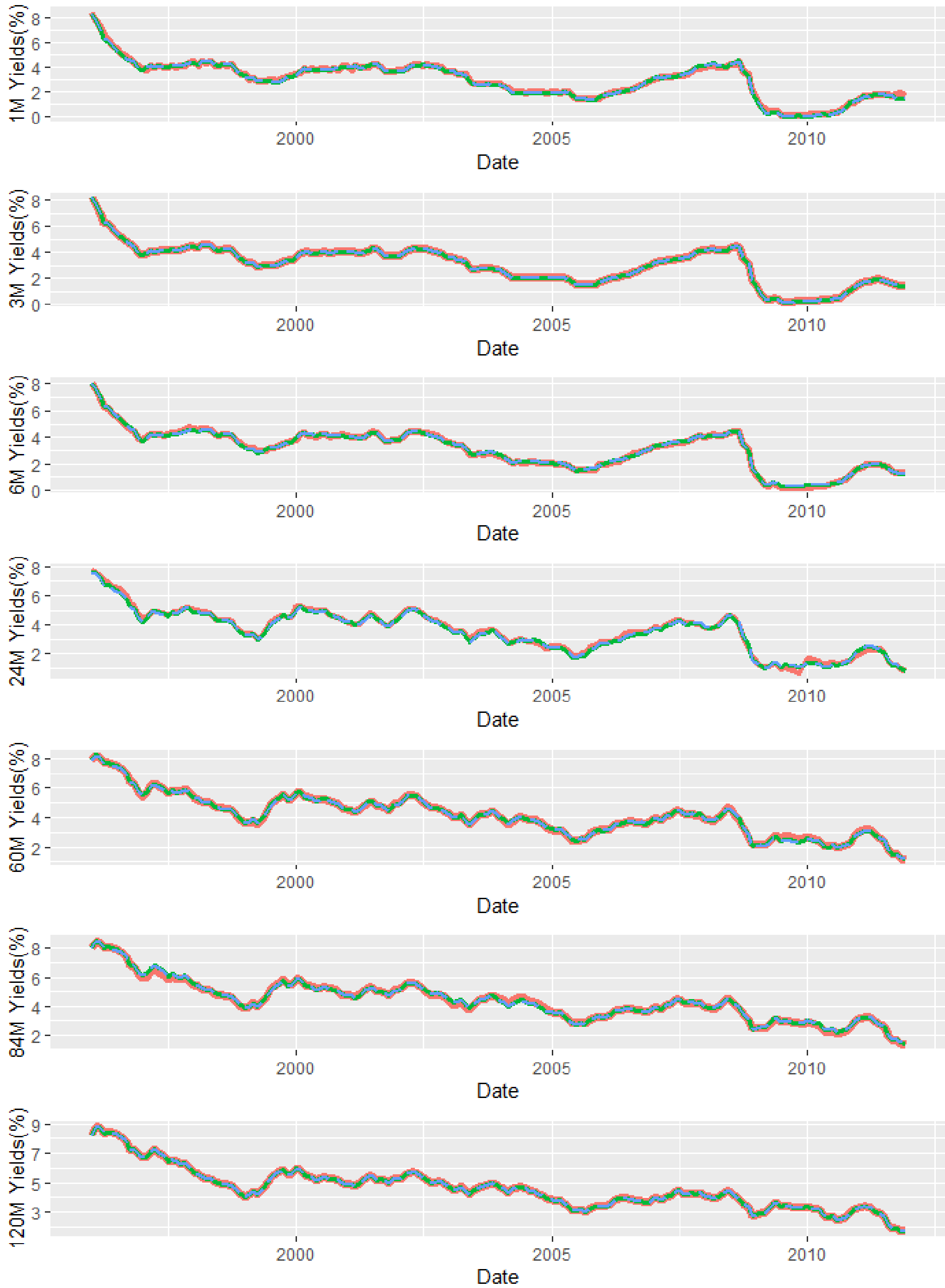

Figure A.1: Estimated DNS results with observations

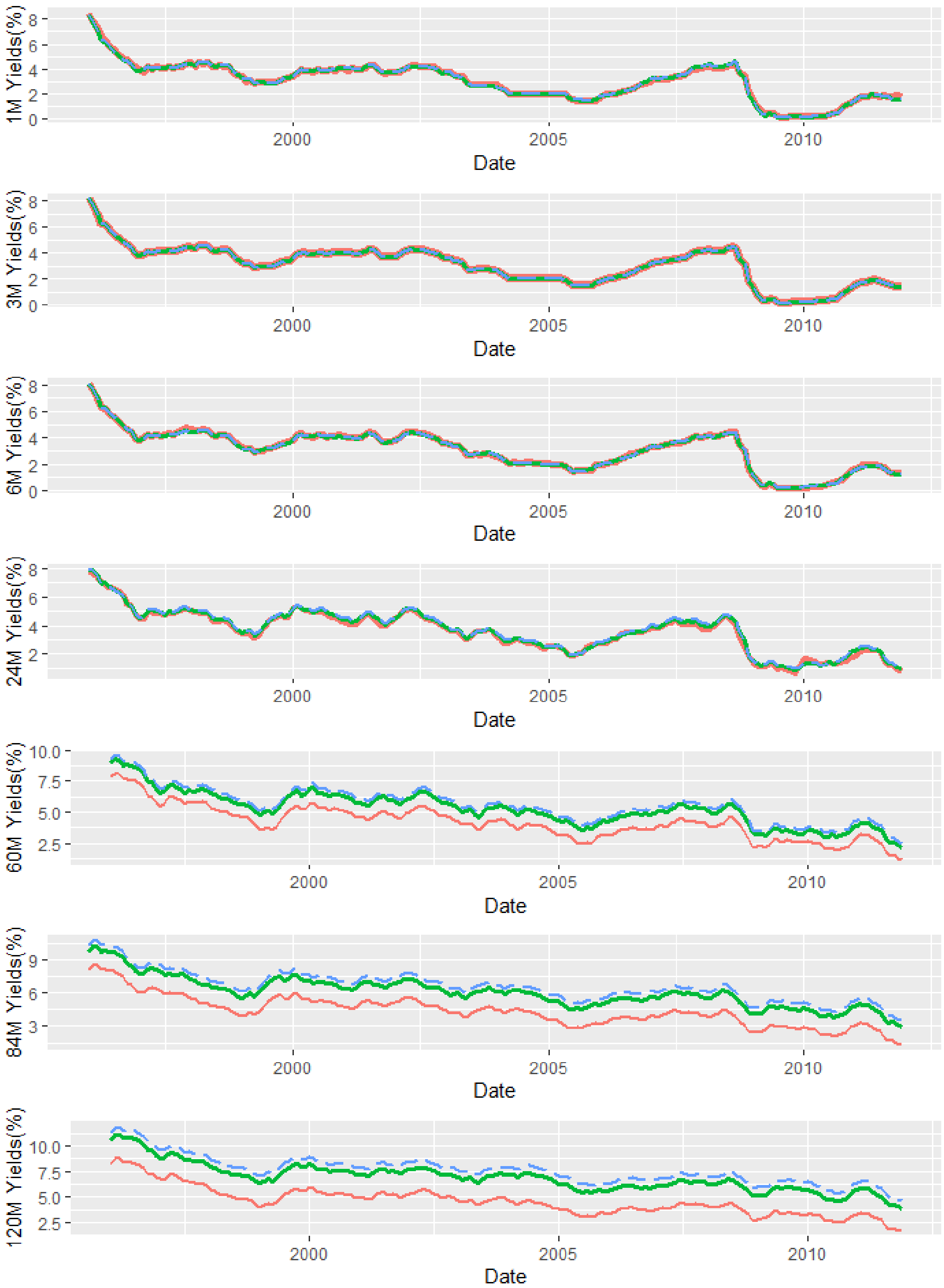

Figure A.2: Estimated AFNS results with observations

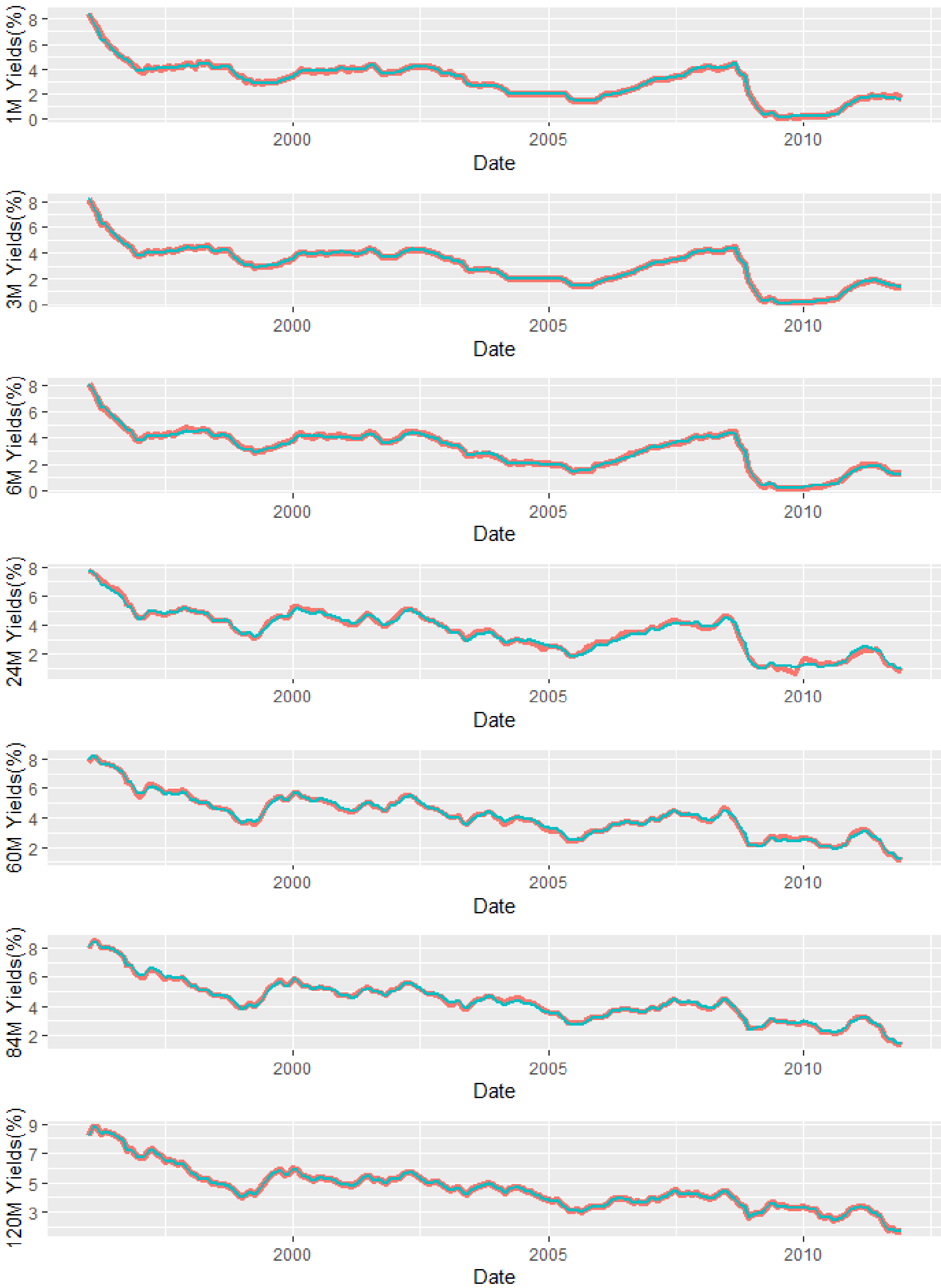

Figure A.3: Estimated DGNS results with observations