The Discrete-Time Framework of the Arbitrage-Free Nelson-Siegel Class of Term Structure Models *

Linlin Niu[a,b] Gengming Zeng[a] †
[a]Wang Yanan Institute for Studies in Economics (WISE), Xiamen University
[b]MOE Key Laboratory of Econometrics, Xiamen University

March 2012

Abstract

We derive the discrete-time arbitrage-free Nelson-Siegel class of term structure models with an exact solution and proof of uniqueness. We design a fast and reliable estimation procedure based on reduced-dimension optimization with multistep embedded regressions. After an analytical illustration, we also show empirically that arbitrage-free restrictions have a bounded advantage for in-sample fit and out-of-sample forecast, compared to its reduced-form counterpart. However, the arbitrage-free model is a powerful tool for analysing risk premia associated with Level, Slope and Curvature factors. Our empirical results have interesting implications for both the US bond yield conundrum of 2004-05 and the recent financial crisis.

Keywords: Yield curve, term structure of interest rates, factor models, state-space models
JEL Classification: G1, E4, C5

*Linlin Niu acknowledges support from the Natural Science Foundation of China (Grant No. 70903053). Gengming Zeng acknowledges the visiting PhD fellowship at University of Antwerp, and thanks seminar participants at the Youth Forum of WISE, Xiamen University.

†Linlin Niu, Rm A301, Economics Building, WISE, Xiamen University, 361005, Xiamen, China. E-mail: linlin.niu@gmail.com. Gengming Zeng, PhD candidate, WISE, Xiamen University, 361005, Xiamen, China. E-mail: gengming.zeng@gmail.com.

1 Introduction

Interest rate term structure or yield curve is of fundamental importance to financial markets, monetary policy and fiscal policy. Substantial research effort has been devoted to modeling the dynamics of the yield curve. The modeling approaches can be classified into two categories: the no-arbitrage approach based on financial theory versus the statistical approach with a reduced form model.

The first approach of modeling the yield curve based on the assumption of no arbitrage was first developed in continuous time by Vasicek (1977) and Cox, Ingersoll and Ross (1985). The no-arbitrage assumption is important in an efficient market, and provides tractability and consistency in bond pricing. Various specific models have since been developed in the no-arbitrage framework. For their ease of derivation and computation, the type of affine term structure model (denoted as ATSM) with closed form bond pricing solutions has gained in popularity. Duffie and Kan (1996) derive a three-factor ATSM which encompasses various already existing ATSMs. Dai and Singleton (2000) make a thorough specification analysis of the three-factor ATSMs. These arbitrage-free models in finance are usually derived in continuous time, until Ang and Piazzesi (2003) popularize the discrete-time model with an application to macrofinance issues. The discrete-time framework has become a workhorse in macrofinance studies in the past decade. However, many of these models exhibit poor empirical performence, especially when forecasting future yields (Duffee, 2002). Confirming the numerical challenges in estimating ATSMs in the literature, Hamilton and Wu (2010) establish that three popular canonical representations of Gaussian ATSMs are unidentified, and propose to estimate their reduced-form representation.

Among the second reduced-form statistical approach where models are usually set up in discrete time, the three-factor Nelson-Siegel interpolation (Nelson and Siegel, 1987) is extremely popular due to its goodness-of-fit, parsimony, and the implied conforming behavior of long-term yields. Diebold and Li (2006) extend this model to a dynamic Nelson-Siegel model (denoted as DNS) and find that (1) it is simple and stable to estimate, (2) it is quite flexible and fits well, and (3) it forecasts well. Effectively the DNS representation is a dynamic model with three Nelson-Siegel latent factors, where the extracted latent factors are labeled Level, Slope and Curvature, and are close to their empirical counterparts or the first three factors extracted from the principal component analysis. However, the Nelson-Siegel model lacks the theoretical foundation to rule out riskless arbitrage opportunities which are important in efficient financial markets (Filipovic, 1999 and Diebold, Piazzesi and Rudebusch, 2005). The reduced form models are typically set up in discrete time.

Christensen, Diebold and Rudebusch (2011) merge the affine arbitrage-free assumption and the Nelson-Siegel interpolation by deriving the affine arbitrage-free Nelson-Siegel (AFNS) term structure model in continuous time The AFNS model addresses both the theoretical weakness of the reduced-form Nelson-Siegel model and the empirical problem of ATSMs. In this class of models, the three Nelson-Siegel yield factors and their traditional loadings are derived under risk-neutral dynamics of the underlying state vector. Compared with the reduced-form DNS model, the arbitrage-free restrictions impose additional constant terms to the yield equations in the AFNS model. Models of this framework not only enjoy the parsimony and flexibility of the reduced-form Nelson-Siegel model, but also rule out arbitrage opportunities in the assumed setting. Christensen, Diebold and Rudebusch (2009) also characterize a more general model with two additional Nelson-Siegel factors in continuous time.

The AFNS class of models would be very useful in a discrete-time setting, if solutions are available. Favero, Niu and Sala (2012) and Gurkaynak and Wright (2010) discussthe approximated discrete-time solution by discretizing the continuous-time solution. However, this approximation-based solution is incompatible with the arbitrage-free assumption. In order to benefit from this emerging model in discrete time, we derive exact solutions for both the AFNS model and the more general model.

Moreover, while Christensen, Diebold and Rudebusch (2009, 2011) prove the existence of the solutions without establishing their uniqueness, we prove both the existence and uniqueness of the solutions. Uniqueness is important for identification as it rules out the possibility of observationally equivalent yields resulting from alternative solutions.

A big challenge in the application of these affine models concerns estimation and inference. Although maximum likelihood estimation (MLE) methods are suitable for such state-space models with latent variables, it becomes difficult to find a global optimum with a high dimension of parameters, typical in multifactor term structure models, and further inference also becomes problematic. While Markov Chain Monte Carlo (MCMC) method is tractible and robust, it is computationally costly. We develop a simple and fast procedure for estimating the AFNS model with reduced-dimension optimization and a multistep embedded regression. This procedure greatly eases the cost of estimation and leads to sufficiently robust results.

We estimate the discrete-time AFNS model with US yield curve data of the last three decades. Compared with the DNS model, our model does better within sample and produces comparable out-of-sample forecasts. We show that the advantage of no-arbitrage restrictions in this model is not in the dynamic prediction of factors, which is dwarfed by dynamic forecast errors in the factors and limited by the upper bound resulting from the inherent parameters.

However, compared to the DNS model, the AFNS model is powerful in separating risk premia and expectations on future short rates. Our empirical results have interesting implications for both the US bond yield conundrum of 2004-05 and the recent financial crisis. The model credits the bond yield conundrum to unusually low risk premia and signals a warning with sharply rising risk premia before the financial crisis.

This paper is organized as follows. Section 2 reviews the general solution to affine term structure model in discrete time and compare it with the reduced-form dynamic Nelson-Siegel model to motivate the solution to the affine arbitrage-free Nelson-Siegel (AFNS) class of models. Section 3 derives the solutions to both the AFNS model and a more general five-factor model with proof of the existence and uniqueness of the solutions. Section 4 develops our simple and fast estimation method. Section 5 describes the data, estimates the model, compares its in-sample and out-of-sample performance with the DNS model and also analyzes yield risk premia. Section 6 concludes.

2 ATSMs in Discrete Time and the DNS Model

In a finance application, when no-arbitrage restrictions are imposed on the yield curve, continuous-time models are commonly used in finance application due to their mathematical ease of derivation and concise representation with stochastic calculus. ATSMs in discrete time with no-arbitrage restrictions have been gaining popularity in macrofinance studies since Ang and Piazzesi (2003) applied a discrete-time essentially affine arbitrage-free term structure model to incorporate macroeconomic factors. The discrete-time framework is also convenient in making econometric analysis, as theories and methods in time series analysis are mostly under the discrete-time framework. In reduced-form term structure models, the discrete-time framework is commonly used for analysis, among which the DNS model in Diebold and Li (2006) is extremely popular among practitioners and central bankers due to its simplicity and effectiveness. However, the DNS model does not rule out arbitrage opportunities, and hence lacks theoretical rigor when applied to an efficient market.

We first discuss the general solution to ATSMs in discrete time under constant volatility, and review the DNS model. Then we make a comparison to identify how to adapt the DNS model to the no-arbitrage restrictions in discrete time.

Although models featuring stochastic volatility are likely to better model the heterogeneity of yields, in a macrofinance application, three factors under constant volatility are sufficient to capture the bulk of variations in the yield curve. Since the DNS model is also under constant volatility, we are spared discussion of the stochastic volatility case. We illustrate the solutions of the essentially affine arbitrage-free term structure model in discrete time closely following the set-up in Ang and Piazzesi (2003).

2.1 The general solution to ATSMs in discrete time under constant volatility

There are several building blocks in ATSMs.

1. Short rate equation.

Denoted by

$$r_t = \delta_0 + \delta_1' X_t \tag{1}$$

where δ_0 is a scalar and δ_1 a $K \times 1$ vector.

2. State dynamics under the physical measure.

Transition equation for X_t follows a VAR(1):

$$X_t = \mu + \Phi X_{t-1} + v_t, , v_t \sim N(0, \Omega). \tag{2}$$

3. State dynamics under the risk-neutral $\mathbb{Q}$ measure.

Transition equation for X_t under the risk-neutral measure also follows VAR(1):

$$X_t = \mu^Q + \Phi^Q X_{t-1} + \tilde{v}_t. \tag{3}$$

4. Prices of risk.

Denoted by Λ_t, investors need to be compensated to be induced to hold long-term bonds. Risk price is associated with the sources of risk of innovations v_t. Under constant volatility, the essentially affine model assumes that the risk price, Λ_t, takes an affine form of the states, i.e.,

$$\Lambda_t = \lambda_0 + \lambda_1 X_t \tag{4}$$

where Λ_t and λ_0 are $K \times 1$ vectors, and λ_1 is a $K \times K$ matrix. As special cases, if investors are risk-neutral, then $\lambda_0 = 0$ and $\lambda_1 = 0$, hence $\Lambda_t = 0$, and there is no risk adjustment; if $\lambda_0 \neq 0$ and $\lambda_1 = 0$, then the price of risk is constant.

5. Pricing kernel.

   No-arbitrage opportunity between bonds with different maturities implies that there is a discount factor, m, linking the bond price of maturity n of this period with that of a bond of maturity $n-1$ of the next period, denoted

$$P_t^{(n)} = E_t\left[m_{t+1}P_{t+1}^{(n-1)}\right], \tag{5}$$

   where the stochastic discount factor m_{t+1} is a function of the short rate and the risk perceived by the market, or

$$m_{t+1} = \exp\left(-r_t - \frac{1}{2}\Lambda_t'\Omega\Lambda_t - \Lambda_t' v_{t+1}\right). \tag{6}$$

Using these building blocks, no-arbitrage recursive relations can be derived such that the price of bonds are exponential affine functions of state variables, and yields are affine on the states. We can state that

$$P_t^{(n)} = E_t\left[\exp\left(A_n + B_n' X_t\right)\right] = E_t\left[\exp\left(-n y_{t,t+n}\right)\right]. \tag{7}$$

Affine functions of the state variables for yields are

$$y_{t,t+n} = a_n + b_n' X_t = \frac{-1}{n}\left(A_n + B_n' X_t\right) \tag{8}$$

where the coefficients follow the difference Equations:[1]

$$\begin{aligned} A_{n+1} &= A_n + B_n'\left(\mu - \Omega\lambda_0\right) + \frac{1}{2}B_n'\Omega B_n + A_1 &(9)\\ B_{n+1}' &= B_n'\left(\Phi - \Omega\lambda_1\right) + B_1' \\ &= B_1'\sum_{k=0}^{n}(\Phi^Q - \Omega\lambda_1)^k &(10)\end{aligned}$$

with $a_1 = \delta_0 = -A_1$ and $b_1 = \delta_1 = -B_1$.

Under the risk-neutral measure,

$$P_t^{(n)} = E_t^Q\left[\exp\left(-r_t\right)P_{t+1}^{(n-1)}\right],$$

[1] Ang and Piazzesi (2003) have a detailed derivation of this solution. They define $\Sigma\Sigma' = \Omega$ in the difference equations above, and identify the matrix Σ whereas we are only interested in Ω. Therefor, λ_0 and λ_1 have therefore different meaning and scale to theirs. The pricing kernel we specify is $m_{t+1} = \exp\left(-r_t - \frac{1}{2}\Lambda_t'\Omega\Lambda_t - \Lambda_t' v_{t+1}\right)$, where $v_{t+1} \sim N(0, \Omega)$; they assume $m_{t+1} = \exp\left(-r_t - \frac{1}{2}\Lambda_t'\Lambda_t - \Lambda_t'\varepsilon_{t+1}\right)$, where $v_{t+1} = \Sigma'\varepsilon_{t+1}$ with $\varepsilon_{t+1} \sim N(0, I)$.

where E_t^Q denotes the expectation under the risk-neutral probability measure, and the dynamics of state vector X_t are characterized by the risk-neutral vector of constants μ^Q and the autoregressive matrix Φ^Q. Under the specification of risk price, the following relationship holds:

$$\mu^Q = \mu - \Omega\lambda_0 \tag{11}$$

$$\Phi^Q = \Phi - \Omega\lambda_1 \tag{12}$$

It is then evident that the difference equations of A_{n+1} and B_{n+1} can be derived and expressed with the risk-neutral parameters as

$$A_{n+1} = A_n + B_n'\mu^Q + \frac{1}{2}B_n'\Omega B_n + A_1 \tag{13}$$

$$\begin{aligned} B_{n+1}' &= B_n'\Phi^Q + B_1' \\ &= B_1'\sum_{k=0}^{n}(\Phi^Q)^k. \end{aligned} \tag{14}$$

It should be noted that Equations (13) and (14) hold in general under the risk-neutral dynamics, Equation (3), even without specifying the particular form of risk prices. This is the key to understanding how to derive the solutions to the arbitrage-free Nelson-Siegel class of models.

2.2 Dynamic Nelson-Siegel models

In the Nelson-Siegel model (Nelson and Siegel, 1987), factors $X_t = [L_t, S_t, C_t]$ are labeled as Level, Slope and Curvature, respectively. The model becomes dynamic in Diebold and Li (2006) where they model X_t as a VAR(1) process as in Equation (2). The factor loadings in the measurement equations of yields are functions of maturity n and a shape parameter λ, or

$$y_{t,t+n} = L_t + \left(\frac{1 - e^{-\lambda n}}{\lambda n}\right) S_t + \left(\frac{1 - e^{-\lambda n}}{\lambda n} - e^{-\lambda n}\right) C_t. \tag{15}$$

To cast Equation (15) in the form of the affine function of Equation (8), we have

$$A_n = 0 \tag{16}$$

$$B_n = [-n, -\frac{1 - e^{-\lambda n}}{\lambda}, ne^{-\lambda n} - \frac{1 - e^{-\lambda n}}{\lambda}]'. \tag{17}$$

2.3 Compatibility of DNS Models and ATSMs

The DNS model is known to be incompatible with the arbitrage-free conditions, because $A_n = 0$ for all n. Examining Equation (9) or Equation (13) for A_n, it is evident that, due to the quadratic term $\frac{1}{2}B_n'\Omega B_n$, A_n cannot be simultanously zero for all maturities under the no-arbitrage assumption.

In order to make the DNS model compatible with the arbitrage-free assumption, it must comply with two requirements:

1) Let A_n follow the difference equation characterized by Equations (9) or (13).

2) Find the underlying parameters in the difference Equations (10) or (14) for B_n, such that the Nelson-Siegel factor loadings characterized in Equation (17) for B_n holds. That is to say, the underlying parameters should be a function of λ in the Nelson-Siegel factor loadings.

Since the difference equations under the risk-neutral measure are easy to work with, i.e., without making specific assumptions on the risk prices, the trick is to find a suitable Φ^Q to equate Equations (14) and (17) such that

$$B'_n = B'_1 \sum_{k=0}^{n-1} (\Phi^Q)^k = [-n, -\frac{1-e^{-\lambda n}}{\lambda}, ne^{-\lambda n} - \frac{1-e^{-\lambda n}}{\lambda}]. \tag{18}$$

This equation is key in understanding the solution and proof to the arbitrage-free Nelson-Siegel class of models.

2.4 Continuous-time solution and discretization

Christensen, Diebold and Rudebusch (2011) derive the continuous-time solution of the risk-neutral process to equate the difference equation for B_n to the Nelson-Siegel factor loadings. The key solution is that, for the underlying continuous-time diffusion process,

$$dX_t = K^Q(t)[\theta^Q(t) - X_t]dt + \Sigma(t)D(X_t, t)dW_t^Q, \tag{19}$$

where the transition matrix, K^Q, needs to have the following specification:

$$K^Q = \begin{pmatrix} 0 & 0 & 0 \\ 0 & \lambda & -\lambda \\ 0 & 0 & \lambda \end{pmatrix}. \tag{20}$$

Christensen, Diebold and Rudebusch (2009) derive the solution to a more general model with five factors where two of them form another set of slope and curvature factors determined by a second shape parameter. This generalization is useful when a yield curve with very long maturity, such as 15-30 years, needs to be modeled as a second hump is necessary to capture the shape of the very long end.

These models then combine the parsimony of the DNS models and the theoretical rigor of ATSMs. However, there is some discussion about having the model perform in discrete time, with attempts to discretize the solution in a discrete-time VAR, such as in Favero, Niu and Sala (2012) and Guerkaynak and Wright (2010). They assume when dt is small, the process can be approximated by a discrete-time VAR process as in Equation (3), when the transition matrix reaches the limit, or

$$\Phi^Q = I_{3\times 3} - K^Q = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1-\lambda & \lambda \\ 0 & 0 & 1-\lambda \end{pmatrix}. \tag{21}$$

However, this is only an approximation in discrete time. When inserting Equation (21) into the general solution for the difference Equation (14), one can easily verify that as maturity n increases, the resulting B_n will diverge from the Nelson-Siegel factor loadings such that Equation (18), the necessary condition for yields with Nelson-Siegel factor loadings to be arbitrage-free, does not hold.

In order to find the exact solution in discrete time, we need to derive the solution in the discrete-time setting, instead of using the discretized approximation of the continuous-time solution.

3 Affine Arbitrage-free Nelson-Siegel Class of Models in Discrete Time

In this section, we first characterize and prove the solution for the AFNS model. Then we generalize it to the affine arbitrage-free generalized Nelson-Siegel (AFGNS) model with a second set of slope and curvature factors governed by a different shape parameter. For both models, we prove both the sufficient and the necessary conditions for a yield curve with Nelson-Siegel factor loadings to be arbitrage free. While the sufficient conditions guarantee the existence of a solution, as proved by Christensen, Diebold and Rudebusch (2009, 2011), we further prove the necessary conditions for the solution to be uniquely characterized by a specific form of transition matrix in the risk-neutral state dynamics. The uniqueness property of this solution is indispensable in empirical estimation and identification, as it rules out alternative solutions which lead to observationally equivalent yields.

3.1 Discrete-time AFNS model

These propositions establish the sufficient and necessary conditions for the DNS model to be arbitrage free.

Proposition 1: Sufficient conditions and the existence of a solution for the AFNS model.

Assumption 1.1: The instantaneous risk-free rate is defined by

$$r_t = \delta_0 + \delta_1' X_t,$$

where $\delta_1 = \begin{bmatrix} 1 & \frac{1-e^{-\lambda}}{\lambda} & \frac{1-e^{-\lambda}}{\lambda} - e^{-\lambda} \end{bmatrix}'$.

Assumption 1.2: The vector of state variables X_t follows a VAR process under the risk-neutral $\mathbb{Q}$ measure

$$X_t = \mu^Q + \Phi^Q X_{t-1} + v_t^Q$$

with the transition matrix taking the specific form

$$\Phi^Q = \begin{bmatrix} 1 & 0 & 0 \\ 0 & e^{-\lambda} & \lambda e^{-\lambda} \\ 0 & 0 & e^{-\lambda} \end{bmatrix}.$$

Assumption 1.1 and *Assumption 1.2* lead to the following implications.

Implication 1.1

The zero-coupon bond prices are given by

$$P_t^n = exp(A_n + B_n' X_t)$$

where B_n is the Nelson-Siegel factor loading,

$$B_n = [-n, -\frac{1-e^{-\lambda n}}{\lambda}, ne^{-\lambda n} - \frac{1-e^{-\lambda n}}{\lambda}]',$$

and A_n satisfies the following difference equation for n >1,

$$A_n = A_{n-1} + B_{n-1}' \mu^Q + \frac{1}{2} B_{n-1}' \Omega B_{n-1}$$

with $B_1 = -\delta_1 = -\begin{bmatrix} 1 & \frac{1-e^{-\lambda}}{\lambda} & \frac{1-e^{-\lambda}}{\lambda} - e^{-\lambda} \end{bmatrix}'$.

Proof of Proposition 1:

Let $B_1 = -\delta_1 = \begin{bmatrix} 1 & \frac{1-e^{-\lambda}}{\lambda} & \frac{1-e^{-\lambda}}{\lambda} - \lambda e^{-\lambda} \end{bmatrix}'$ for $n > 1$. From the difference Equations (13) and (14) for A_n and B_n, respectively, under the risk-neutral measure in the discrete-time ATSM we have

$$\begin{aligned} A_n &= A_{n-1} + B_{n-1}' \mu^Q + \frac{1}{2} B_{n-1}' \Omega B_{n-1} \\ B_n' &= -\delta_1' \sum_{k=0}^{n-1} (\Phi^Q)^k. \end{aligned}$$

Since

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

$$\begin{aligned} \sum_{k=0}^{n-1} (\Phi^Q)^k &= \begin{bmatrix} n & 0 & 0 \\ 0 & \sum_{k=0}^{n-1} e^{-k\lambda} & \sum_{k=0}^{n-1} k\lambda e^{-k\lambda} \\ 0 & 0 & \sum_{k=0}^{n-1} e^{-k\lambda} \end{bmatrix} \\ &= \begin{bmatrix} n & 0 & 0 \\ 0 & \frac{1-e^{-n\lambda}}{1-e^{-\lambda}} & \lambda[\frac{e^{-\lambda}-e^{-n\lambda}}{(1-e^{-\lambda})^2} - \frac{(n-1)e^{-n\lambda}}{1-e^{-\lambda}}] \\ 0 & 0 & \frac{1-e^{-n\lambda}}{1-e^{-\lambda}} \end{bmatrix}, \end{aligned}$$

it follows that

$$\begin{aligned} B_n &= -\left\{ \sum_{k=0}^{n-1} (\Phi^Q)^k \right\}' \delta_1 \\ &= -\begin{bmatrix} n & 0 & 0 \\ 0 & \frac{1-e^{-n\lambda}}{1-e^{-\lambda}} & 0 \\ 0 & \lambda[\frac{e^{-\lambda}-e^{-n\lambda}}{(1-e^{-\lambda})^2} - \frac{(n-1)e^{-n\lambda}}{1-e^{-\lambda}}] & \frac{1-e^{-n\lambda}}{1-e^{-\lambda}} \end{bmatrix} \begin{bmatrix} 1 \\ \frac{1-e^{-\lambda}}{\lambda} \\ \frac{1-e^{-\lambda}}{\lambda} - e^{-\lambda} \end{bmatrix} \\ &= -\begin{bmatrix} n \\ \frac{1-e^{-\lambda n}}{\lambda} \\ \frac{e^{-\lambda}-e^{-n\lambda}}{1-e^{-\lambda}} - (n-1)e^{-n\lambda} + \frac{1-e^{-n\lambda}}{\lambda} - \frac{e^{-\lambda}-e^{-(n+1)\lambda}}{1-e^{-\lambda}} \end{bmatrix} \\ &= [-n, -\frac{1-e^{-\lambda n}}{\lambda}, ne^{-\lambda n} - \frac{1-e^{-\lambda n}}{\lambda}]'. \end{aligned}$$

■

Proposition 2: Necessary conditions and uniqueness of the solution for the AFNS model.

Assumption 2.1: The DNS model with a constant adjustment term, a_n, fits the yield curve completely, we may write

$$y_{t,t+n} = a_n + L_t + \left(\frac{1-e^{-\lambda n}}{\lambda n}\right) S_t + \left(\frac{1-e^{-\lambda n}}{\lambda n} - e^{-\lambda n}\right) C_t.$$

Assumption 2.2: The Nelson-Siegel latent factors follow a VAR(1) process under the risk-neutral $\mathbb{Q}$ measure,

$$X_t = \mu^Q + \Phi^Q X_{t-1} + v_t^Q.$$

There are two implications.

Implication 2.1 The risk-free rate needs to follow the affine process

$$r_t = \delta_0 + \delta_1 X_t$$

where $\delta_1 = \left[\begin{array}{ccc} 1 & \frac{1-e^{-\lambda}}{\lambda} & \frac{1-e^{-\lambda}}{\lambda} - e^{-\lambda} \end{array}\right]'$ and $\delta_0 = a_1$.

Implication 2.2 The risk-neutral dynamic coefficient matrix Φ^Q of these three latent factors needs to satisfy

$$\Phi^Q = \begin{pmatrix} 1 & 0 & 0 \\ 0 & e^{-\lambda} & \lambda e^{-\lambda} \\ 0 & 0 & e^{-\lambda} \end{pmatrix}.$$

In *Proposition 2*, *Implication 2.1* and *Implication 2.2* are exactly the same two assumptions as in *Proposition 1*.

Proof of Proposition 2:

When $n = 1$, *Implication 2.1* is easily satisfied by *Assumption 2.1.*

By *Assumption 2.2* and *Implication 2.1*, we know that this model is an affine term structure model which implies this affine function of yields,

$$y_{t,t+n} = \frac{-1}{n}\left(A_n + B_n' X_t\right) \tag{22}$$

where

$$A_{n+1} = A_n + B_n'\mu^Q + \frac{1}{2}B_n'\Omega B_n + A_1 \tag{23}$$

$$B_{n+1}' = B_n'\Phi^Q + B_1'. \tag{24}$$

Hence, by comparing Equation (22) with *Assumption 2.1*, we have

$$\begin{aligned} B_n &= [-n, -\frac{1-e^{-\lambda n}}{\lambda}, ne^{-\lambda n} - \frac{1-e^{-\lambda n}}{\lambda}]' \\ A_n &= -na_n. \end{aligned} \tag{25}$$

We insert Equation (25) into both sides of Equation (24) and take the transpose, giving

$$\begin{pmatrix} n+1 \\ \frac{1-e^{-\lambda(n+1)}}{\lambda} \\ \frac{1-e^{-\lambda(n+1)}}{\lambda} - (n+1)e^{-\lambda(n+1)} \end{pmatrix} = \begin{pmatrix} \phi_{11} & \phi_{21} & \phi_{31} \\ \phi_{12} & \phi_{22} & \phi_{32} \\ \phi_{13} & \phi_{23} & \phi_{33} \end{pmatrix} \begin{pmatrix} n \\ \frac{1-e^{-\lambda n}}{\lambda} \\ \frac{1-e^{-\lambda n}}{\lambda} - ne^{-\lambda n} \end{pmatrix} + \begin{pmatrix} 1 \\ \frac{1-e^{-\lambda}}{\lambda} \\ \frac{1-e^{-\lambda}}{\lambda} - e^{-\lambda} \end{pmatrix}.$$

We can rearrange it such that

$$\begin{pmatrix} n \\ \frac{e^{-\lambda}-e^{-\lambda(n+1)}}{\lambda} \\ \frac{e^{-\lambda}-e^{-\lambda(n+1)}}{\lambda} - (n+1)e^{-\lambda(n+1)} + e^{-\lambda} \end{pmatrix} = \begin{pmatrix} \phi_{11} & \phi_{21} & \phi_{31} \\ \phi_{12} & \phi_{22} & \phi_{32} \\ \phi_{13} & \phi_{23} & \phi_{33} \end{pmatrix} \begin{pmatrix} n \\ \frac{1-e^{-\lambda n}}{\lambda} \\ \frac{1-e^{-\lambda n}}{\lambda} - ne^{-\lambda n} \end{pmatrix}$$

Next, we solve for the elements of Φ^Q by comparing the coefficients, equation by equation. The first equation implies that

$$\phi_{11} = 1, \;\; \phi_{21} = 0 \;\text{ and }\; \phi_{31} = 0.$$

The second equation implies that

$$\begin{aligned} \phi_{12} &= 0; \\ \frac{e^{-\lambda}}{\lambda} - \frac{e^{-\lambda}e^{-\lambda n}}{\lambda} &= \phi_{22}\frac{1}{\lambda} + \phi_{32}\frac{1}{\lambda} - (\phi_{22}+\phi_{32})\frac{e^{-\lambda n}}{\lambda} - \phi_{32}ne^{-\lambda n} \end{aligned}$$

where the second line requires the following to hold:

$$\phi_{32} = 0 \text{ and } \phi_{22} = e^{-\lambda}.$$

The third equation implies that

$$\begin{aligned} \phi_{13} &= 0 \\ \frac{e^{-\lambda}}{\lambda} + e^{-\lambda} - e^{-\lambda}(1+\lambda)\frac{e^{-\lambda n}}{\lambda} - e^{-\lambda}ne^{-\lambda n} &= \phi_{23}\frac{1}{\lambda} + \phi_{33}\frac{1}{\lambda} - (\phi_{23}+\phi_{33})\frac{e^{-\lambda n}}{\lambda} - \phi_{33}ne^{-\lambda n} \end{aligned}$$

where the second line requires

$$\phi_{33} = e^{-\lambda} \text{ and } \phi_{23} = \lambda e^{-\lambda}.$$

To summarize, we have

$$\Phi^Q = \begin{pmatrix} 1 & 0 & 0 \\ 0 & e^{-\lambda} & \lambda e^{-\lambda} \\ 0 & 0 & e^{-\lambda} \end{pmatrix}.$$

■

3.2 Discrete-time AFGNS model

Christensen, Diebold and Rudebusch (2009) generalize the AFNS model to allow for a second set of slope and curvature factors to be determined by another shape parameter. This AFGNS model adds an additional slope factor to the dynamic Nelson-Siegel-Svensson (DNSS) model (Svensson, 1995) which adds a second curvature factor to the DNS model. This is represented by

$$\begin{aligned} y_{t,t+n} &= a_n + L_t + \left(\frac{1-e^{-\lambda_1 n}}{\lambda_1 n}\right) S_{1,t} + \left(\frac{1-e^{-\lambda_2 n}}{\lambda_2 n}\right) S_{2,t} \\ &+ \left(\frac{1-e^{-\lambda_1 n}}{\lambda_1 n} - e^{-\lambda_1 n}\right) C_{1,t} + \left(\frac{1-e^{-\lambda_2 n}}{\lambda_2 n} - e^{-\lambda_2 n}\right) C_{2,t}. \end{aligned}$$

Christensen, Diebold and Rudebusch (2009) prove the existence of a solution.

We characterize the model in discrete time and prove both the existence and the uniqueness of the solution. The proof closely resembles that for the AFNS model, but slightly more complex. For conciseness, the proof appears in Appendix 1. The crucial element of the solution is that the dynamic coefficient matrix, Φ^Q, takes the following form:

$$\Phi^Q = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & e^{-\lambda_1} & 0 & \lambda_1 e^{-\lambda_1} & 0 \\ 0 & 0 & e^{-\lambda_2} & 0 & \lambda_2 e^{-\lambda_2} \\ 0 & 0 & 0 & e^{-\lambda_1} & 0 \\ 0 & 0 & 0 & 0 & e^{-\lambda_2} \end{bmatrix}.$$

3.3 Specification analysis

As discussed in Dai and Singleton (2000), there often exists an over-identification problem in affine term structure models. For example, some parallel transformation to the drift, μ^Q, in the state dynamics and intercept, δ_0, in the short rate equation may result in the same distribution of the yields. Hence, we need to perform specification analysis to define the canonical representation for identification.

In the AFNS model, we let $\mu^Q = [\mu_L^Q, 0, 0]'$ and $\delta_0 = 0$. Equivalently, we can restrict $\mu^Q = 0$ but free up δ_0. In our analysis we choose the former, with which the difference equations of coefficients in the measurement equations become

$$A_{n+1} = A_n + n\mu_L^Q + \frac{1}{2}B_n'\Omega B_n, \text{ with } A_1 = 0 \tag{26}$$

$$B_n = [-n, -\frac{1-e^{-\lambda n}}{\lambda}, ne^{-\lambda n} - \frac{1-e^{-\lambda n}}{\lambda}]'. \tag{27}$$

Likewise, for the AFGNS model, similar restrictions on specification can be imposed.

From the specification analysis, it is clear that the AFNS model only has one additional parameter, μ_L^Q, compared to the DNS model. This guarantees the maximal parsimony of the AFNS model.

3.4 Risk price

The AFNS and AFGNS models are silent on risk price. No specific form for the price of risks is proposed. However, since the models specify the risk-neutral transition matrix and do not directly impose any restrictions on the physical dynamics, risk price parameters can be backed out with respect to the chosen specification. For example, if we believe the risk price has an affine form as in Equation (4), with the relationship connecting risk-neutral and physical state dynamics, or Equations (11) and (12), the parameters of the risk price can be derived as

$$\lambda_0 = \Omega^{-1}(\mu - \mu^Q) \tag{28}$$

$$\lambda_1 = \Omega^{-1}(\Phi - \Phi^Q). \tag{29}$$

Since the transition matrix Φ^Q is highly sparse and restricted, it amounts to the imposition of restrictions on risk price parameter λ_1, indirectly.

3.5 Risk premia

Although the AFNS and AFGNS models do not make a direct inference on the risk price without an explicit assumption on the form of risk price, the model is still applicable and useful in separating risk premia from the expectation on future short rates, and in analyzing the specific effects of the yield factors on risk premia.

If we define the risk premium of the bond price as the yield component to compensate risk, then it can be derived as the difference between the theoretical yield, $y_{t,t+n}$, and the expected yield with zero risk compensation, $\tilde{y}_{t,t+n}$, when setting $\lambda_0 = 0$ and $\lambda_1 = 0$. That is, we define risk premium as

$$RP_{t,t+n} = y_{t,t+n} - \tilde{y}_{t,t+n}.$$

For a yield without the premium we have

$$\tilde{y}_{t,t+n} = -\frac{1}{n}(\tilde{A}_n + \tilde{B}_n' X_t)$$

where

$$
\begin{aligned}
\tilde{B}'_{n+1} &= \tilde{B}'_n \Phi - \delta'_1 && (30)\\
\tilde{A}_{n+1} &= \tilde{A}_n + \tilde{B}'_n \mu + \frac{1}{2} \tilde{B}'_n \Omega \tilde{B}_n - \delta_0. && (31)
\end{aligned}
$$

These coefficients can be derived solely based on the parameters of the state equation under the physical measure. Then the risk premium can be calculated as

$$
\begin{aligned}
RP_{t,t+n} &= y_{t,t+n} - \tilde{y}_{t,t+n}\\
&= -\frac{1}{n}(A_n + B'_n X_t) + \frac{1}{n}(\tilde{A}_n + \tilde{B}'_n X_t)\\
&= \frac{1}{n}\left[\left(\tilde{A}_n - A_n\right) + \left(\tilde{B}'_n - B'_n\right) X_t\right]. && (32)
\end{aligned}
$$

4 Estimation Methods

In estimation and application, we focus on the three factor no-arbitrage Nelson-Siegel model, as it is more popular and widely applied than the five factor generalized model in empirical applications.

The econometric representation of the AFNS is

$$
\begin{aligned}
y_{t,t+n} &= -\frac{1}{n}(A_n + B'_n X_t) + \varepsilon_t, \varepsilon_t \sim N(0, \sigma^2) && (33)\\
X_t &= \mu + \Phi X_{t-1} + v_t, v_t \sim N(0, \Omega) && (34)
\end{aligned}
$$

with A_n and B_n as defined in Equations (26) and (27).

With latent variables in the state-space representation, the standard method of estimation is a Maximum Likelihood Estimation (MLE) with Kalman filter. It is necessary to use numerical methods to find values of the parameter set in order to maximize the likelihood of the model. However, with high dimensions of the parameter set, such as the 21 parameters in this AFNS model, numerical optimization algorithms often encounter local maxima or non-convergence problems. The standard errors of parameters are often estimated without accuracy.

In this paper, we propose a simple and fast procedure which combines a **r**educed-dimension **o**ptimization with **m**ultistep **e**mbedded **r**egressions, or ROMER. We first demonstrate the appropriateness of this method by simulation. In the real data application following this section, we will also estimate the model with the Bayesian Markov chain Monte Carlo (MCMC) method which leads to efficient inference with respect to selected priors. By comparing the results of the two methods, we find that our simpler procedure works very well and is particularly suitable for quick inference or forecast purposes to reduce computational burden.

4.1 Reduced-dimension optimization with multistep embedded regressions

The ROMER procedure incorporates the two-step OLS regression employed in the DNS model estimation as explained in Diebold and Li (2006), where

$$y_{t,t+n} = b_n' X_t + \varepsilon_t, \qquad \varepsilon_t \sim N(0, \sigma^2) \tag{35}$$

$$X_t = \mu + \Phi X_{t-1} + v_t, \quad v_t \sim N(0, \Omega) \tag{36}$$

with $b_n = -\frac{1}{n} B_n$ as the Nelson-Siegel factor loadings. Given $\lambda = 0.0609$, which maximizes the curvature loading at 30 months maturity, the coefficients of measurement equations, b_n, can be computed for yields of any maturity, $y_{t,t+n}$.

In the first step, the three factors in X_t are extracted from the yield curve. Denote b as the $N \times 3$ matrix stacking all b_n' and y_t as the $N \times 1$ vector of N yields available at time t. With sufficient number of $y_{t,t+n}$ at each time t, X_t can be extracted by an OLS regression treating b as a regressor. Hence we have

$$\hat{X}_t = (b'b)^{-1} b' y_t \tag{37}$$

$$= X_t + (b'b)^{-1} b' \varepsilon_t. \tag{38}$$

In the second step, the VAR coefficients μ and Φ can be estimated by an OLS regression of the state equation, using the extracted $\hat{X}_t$. Due to the presence of error term $(b'b)^{-1} b' \varepsilon_t$, the procedure is not fully efficient or consistent. However, when the number of yields, N, is sufficiently large, the error term is be very small and negligible.

In the state-space representation of Equations (33) and (34), the only difference with the AFNS model is that there is a constant term, $a_n = -\frac{1}{n} A_n$, in the measurement equation, which is a function of underlying parameters. However, even if λ is known, since A_n is determined also by Ω, a parameter in the state equation, and μ_L^Q, a parameter in the latent risk-neutral dynamics, the two-step regressions in Diebold and Li (2006) becomes invalid.

We show in the following that, given λ and μ_L^Q in the measurement equations, a multistep embedded OLS regression procedure leads to a robust estimation of parameters (μ, Φ, Ω) in the state equation. Parameters only involved in the measurement equations, $(\lambda, \mu_L^Q, \sigma)$, can be estimated by minimizing the log-likelihood of the measurement equations. We call this the ROMER procedure.

Compared to the standard MLE procedure, the optimization in ROMER is of reduced dimensions, significantly reduced from twenty-one to three. It is not only simple and fast, but also robust and reliable, which will be proved by simulation and comparison with MCMC estimation results.

Before further elaboration, we first divide the parameter set into two subsets, $\Theta_1 = (\lambda, \mu_L^Q, \sigma)$ and $\Theta_2 = (\mu, \Phi, \Omega)$. While the former is involved only in the measurement equations, the latter is the parameter set of the state equations which also appear in the adjustment term of the measurement equations.

4.1.1 Multistep embedded regressions

Given Θ_1, specifically λ and μ_L^Q, we use a multistep OLS procedure to infer the parameters in state equations, taking into consideration the adjustment terms in the measurement equations.

The first stage: Diebold-Li two-step regressions with $A = 0$. As in the estimation of Diebold and Li (2006), we perform a linear regression on the measurement equations to obtain an estimate of the latent factors, X_t, denoted as $\hat{X}_t$. We denote the regression residual as $\hat{\varepsilon}_t$ and its variance as $\hat{\sigma}_\varepsilon^2$. Then we regress $\hat{X}_t$ on a constant and $\hat{X}_{t-1}$ to obtain the parameters of the state equations, denoted as $\hat{\mu}$, $\hat{\Phi}$ and $\hat{\Omega}$.

If $a_n = 0$, such that Equation (35) holds as in the Diebold and Li model, then $\hat{X}_t$ is an unbiased estimator of X_t, although with some error terms inherited from the measurement equations. However, with no-arbitrage restrictions, $a_n = 0$ cannot hold for all n. Thus $\hat{X}_t$ is biased. To see this, substitute measurement Equation (33) for y_t into the OLS regression of the DNS model, Equation (37),

$$\hat{X}_t = (b'b)^{-1}b'y_t \tag{39}$$

$$= (b'b)^{-1}b'a + X_t + (b'b)^{-1}b'\varepsilon_t. \tag{40}$$

It is clear that $\hat{X}_t$ is biased with a constant term $(b'b)^{-1}b'a$, where a is an $N \times 1$ vector of a_n. This term mainly affects the inference on $\hat{\mu}$, not on $\hat{\Phi}$ and $\hat{\Omega}$, to the extent that $(b'b)^{-1}b'\varepsilon_t$ is negligible with a large N.

To correct for this first-step bias in the measurement equations, we substitute $\hat{\Omega}$ into Equation (26) to obtain an estimate for A_n, denoted as $\hat{A}_n$, such that

$$\begin{aligned}\hat{A}_{n+1} &= \hat{A}_n + n\mu_L^Q + \frac{1}{2}B_n'\hat{\Omega}B_n, \text{with } A_1 = 0,\ \forall n > 1 \\ \hat{a}_n &= -\frac{1}{n}\hat{A}_n.\end{aligned}$$

Since $\hat{\Omega}$ is the same whether or not the bias term $(b'b)^{-1}b'a$ is present, the computation of $\hat{A}_n$ using $\hat{\Omega}$ is valid.

The second stage: two-step regressions with the adjustment term A. Subtract $\hat{A}_n$ from y_t, and redo the two-step regressions to extract X_t, denoted as $\tilde{X}_t$, such that

$$\tilde{X}_t = (b'b)^{-1}b'(y_t - \hat{a}) \tag{41}$$

$$= \hat{X}_t - (b'b)^{-1}b'\hat{a} \tag{42}$$

$$= X_t + (b'b)^{-1}b'\varepsilon_t + (b'b)^{-1}b'(a - \hat{a}). \tag{43}$$

As long as $\hat{a}$ is close to the true value, the bias will be effectively controlled for. Then we can re-run the OLS regression for the VAR state Equation (36) with this $\tilde{X}_t$ to obtain estimates of parameters in the state equation, denoted as $(\tilde{\mu}, \tilde{\Phi}, \tilde{\Omega})$. To the extent that $\hat{a}$ is close to a, the term $(b'b)^{-1}b'(a - \hat{a})$ is negligible. This is because the term is effectively the regression coefficient of $(a - \hat{a})$ onto b, i.e., decomposing $(a - \hat{a})$ into Level, Slope and Curvature according to the Nelson-Siegel factor loadings. When $(a - \hat{a})$ is flat around zero, the decomposed factors are close to zero. We will give a numerical illustration with simulated data in a later section to show the difference between $\hat{a}$ and a.

4.1.2 Optimization on $\Theta_1 = (\lambda, \mu_L^Q, \sigma)$

We show in the above that, conditional on $\Theta_1 = (\lambda, \mu_L^Q, \sigma)$, specifically on λ and μ_L^Q, $\Theta_2 = (\mu, \Phi, \Omega)$ and X_t can be estimated with the two-stage regressions described above, with the latter step adjusting for the constant term a_n in the measurement equations. Hence $(\tilde{\mu}, \tilde{\Phi}, \tilde{\Omega}, \tilde{X}_{1:T})$ and coefficients $\tilde{a}$ and b in the measurement equations can be regarded as a function of Θ_1 and the yield data, y.

We can then use MLE on the measurement equations to estimate Θ_1. The likelihood function is

$$\begin{aligned}\mathcal{L}(\Theta_1) &= \prod_{t=1}^{T} f(y_t|\Theta_1) \\ &= (\frac{1}{2\pi\sigma^2})^{\frac{NT}{2}} \exp(-\frac{\Sigma_{t=1}^{T} y_t - \tilde{a}(\lambda, \mu_L^Q) - b(\lambda)\tilde{X}_t(\lambda, \mu_L^Q)]'[y_t - \tilde{a}(\lambda, \mu_L^Q) - b(\lambda)\tilde{X}_t(\lambda, \mu_L^Q)}{2\sigma^2}).\end{aligned}$$

Taking the log,

$$\begin{aligned}\log(L(\Theta_1)) \propto\ & -\frac{NT}{2}\log(\sigma^2) - \\ & \frac{1}{2\sigma^2}\sum_t [y_t - \tilde{a}(\lambda, \mu_L^Q) - b(\lambda)\tilde{X}_t(\lambda, \mu_L^Q)]'[y_t - \tilde{a}(\lambda, \mu_L^Q) - b(\lambda)\tilde{X}_t(\lambda, \mu_L^Q)].\end{aligned}$$

Maximizing the log likelihood function, we obtain the estimate $\hat{\Theta}_1$. Since the parameter set is only of three dimensions, standard optimization procedures can be applied easily to find reliable estimates and to compute standard errors.

Correponding to $\hat{\Theta}_1$, we can obtain a conditional estimate of Θ_2 using the embedded multistep regressions and denote it as $\tilde{\Theta}_2(\hat{\Theta}_1)$. Standard errors of the estimate on Θ_2 come from two sources: the variance conditional on $\hat{\Theta}_1$, which can be obtained from the embedded OLS regression, and the variance resulting from the variance of $\hat{\Theta}_1$, $var(\hat{\Theta}_1)$,

$$Var(\Theta_2) = Var_{|\hat{\Theta}_1}(\tilde{\Theta}_2) + (\partial\tilde{\Theta}_2/\partial\hat{\Theta}_1) * Var(\hat{\Theta}_1) * (\partial\tilde{\Theta}_2/\partial\hat{\Theta}_1)^{\intercal}$$

In fact, a DNS model with free λ can also be estimated by the ROMER procedure, with the optimization only dealing with λ and σ. It is much simpler as no adjustment term needs to be corrected. So in applying ROMER to a DNS model with free λ, only the original Diebold-Li two-step procedure is needed in the embedded step.

We concede that, due to the error-in-regression problem inherited in the multistep embedded regressions, our optimization procedure is essentially an approximated MLE, which is not theoretically efficient or unbiased. The validity of the procedure depends on the assumption that the number of yields, N, is large so that the error term, $(b'b)^{-1}b'\varepsilon_t$, is negligible. This is the same implicit assumption justifying the Diebold-Li two-step procedure for the DNS model.

In what follows, we will estimate the model with this procedure using simulated data to verify its performance. In the section of empirical estimation with real data, we will also compare results from this procedure with results obtained using the MCMC method.

4.2 Simulation

To verify the effectiveness and robustness of this estimation method, we first apply it to simulated data generated by a known parameter set. Table 1 shows the parameter values for this simulation.

[Table 1. Parameter values for simulation]

The parameter values are somewhat arbitrary, but reflect the essential empirical features of Nelson-Siegel factors extracted from yield data, in that there is descending persistence of the three yield factors. As in Diebold and Li (2006), λ is calibrated to maximize the curvature at 30 months as in Diebold and Li (2006). The variance-covariance matrix is set to a reasonable scale according to the data. The point here is to test the procedure by estimating synthetic data with reasonable features of yields and to verify its estimation property compared to the known parameter values.

We choose to simulate yields of the following maturities: 3, 6, 9, 12, 15, 18, 21, 24, 30, 36, 48, 60, 72, 84, 96, 108 and 120 months, a total of 17 yields. This maturity structure is similar to the US Treasury yield curve data used in Diebold and Li (2006). Assuming the observation is monthly, we simulate yield factors and yields according to the above parameter set for 360 months, i.e., 30 years. We simulate this data set with the same parameter set, for 5000 times. With each simulated sample, we apply our method to estimate the parameters. We calculate the means and standard errors for the point estimates of the parameters from the 5000 simulated samples. Table 2 shows the estimation results. For the variance-covariance matrix Ω, we report only the lower triangular part for convenience.

[Table 2. Parameter estimates from simulated yield data]

Compared to the true parameter set in Table 1, Table 2 shows that the MLE estimation of the three parameters in Θ_1 are highly significant and precise. The parameters in the state equations are also well estimated. Nonzero parameters in Φ and Ω that crucially determine the factor dynamics are also precisely estimated. The estimate of μ is insignificant, as it is so small in scale compared to the variation of regressors. The small scale and insignificance in the estimates are confirmed in the empirical estimation of the DNS or continuous-time AFNS models in Christensen, Diebold and Rudebusch (2011).

Another way to assess the estimation precision is by looking at the estimated factors in comparison to the real factors simulated. As an illustration, Figure 1 shows the comparison between estimated and true factors from one of the 5000 simulated samples. The real latent factors and the estimated latent factors, denoted by $\{L, S, C\}$in solid lines and $\{L\_, S\_, C\_\}$ in circled points, respectively, are so close that that it is hard to distinguish them visually from each other.

We also use the simulated samples to compute the distribution of the estimated adjustment term, $\tilde{A}_n$, in comparison to its true value. The results are shown in Figure 2. The true value of A_n is shown in solid line, the estimated mean of $\tilde{A}_n$ in dotted line, and the confidence interval of ± 2 standard deviations is in dashed lines. The figure confirms that the estimate is very close to the true value.

[Figure 1. Comparison of the estimated and simulated true latent factors]

[Figure 2. Adjustment term from simulated data and estimation]

4.3 Bayesian estimation

For quick estimation, forecast and inference, the ROMER procedure is recommended for its simplicity and speed. However, for efficient estimation, the MCMC method is ideal in providing robust and consistent estimation even when the state-space is of high dimension. The resulting posterior distributions of parameters are also handy for making further inferences.

We detail the MCMC procedure in Appendix 2 for further reference. We provide an estimation comparison to the ROMER estimation results with real data in the next section, and also make further inference and analysis on risk premia based on the results.

4.4 Estimation of the AFGNS model

The methods of both the ROMER and the Bayesian MCMC procedures are applicable to the five-factor arbitrage-free generalized Nelson-Siegel model. However, caution needs to be paid to the maturity structure of the data set. In the AFGNS model, there are two shape parameters, λ_1 and λ_2, each governing a curvature shape at a different horizon. The first one peaks at a relatively short maturity and the second one peaks at a relatively long maturity. The second curvature, before reaching its peak, behaves like a slope factor. When yield data do not have enough observations for very long maturities such as twenty or even thirty years, the second hump is not easily identifiable from the first set of slope and curvature factors. The resulting collinearity leads to imprecise estimation with a large bias and with regression errors.

5 Data and Estimation

5.1 Data

We use US Treasury zero-coupon-equivalent yield data with maturities of 3, 6, 9, 12, 18, 24, 30, 36, 48, 60, 72, 84, 96, 108 and 120 months, a total of 15 yields from August 1971 through September 2010. Among them, the 3- and 6-month yields are converted from the 3- and 6-month Treasury bill rates on a discount basis, from the Federal Reserve statistical release H.15 of selected interest rates. The yields from 9 to 120 months are from the research data of the Federal Reserve Board, released with the paper of Gurkaynak, Sack and Wright (2007), which originally contains yields of maturities in integer years. We add the 9-, 18- and 30-month yields interpolated according to the parameters provided in the file. Both data have daily frequency, updated constantly. We use month-end data for the empirical application. A monthly frequency is often used for macrofinance analysis, where such discrete-time models are widely used. Figure 3 shows the dynamics of this yield curve.

[Figure 3. US yield curve from 1971:8 to 2010:9]

5.2 Estimation

We first estimate the model with the whole sample using the ROMER procedure discussed in Section 3. The result is verified against the result of the MCMC estimation to show its reliability and robustness in real data estimation. We then use the ROMER procedure to estimate both the AFNS and DNS models for comparison of in-sample fit and out-of-sample forecast. The MCMC results of posterior distributions of parameters are used for analysis on the price of risk and risk premia.

Since the original data are annualized yields in percentage terms as shown in Figure 3, for our estimation with monthly frequency, we divide the data by 1200 to adjust them to monthly yields in decimal units. This is important in no-arbitrage models. As the no-arbitrage restrictions across equations are derived under the proper unit and have nonlinear form, linear transformation of the yields and factors proportionally in the estimation lead to a violation of the no-arbitrage assumption.

5.2.1 Estimation results from the ROMER procedure

We estimate the whole sample with the ROMER procedure.

In Figure 4, we plot the fitted residuals between the discrete-time AFNS model and the real yield curve, which shows that errors are fairly small and that this model captures the overall yield dynamics very well.

[Figure 4. Fitting residuals of the discrete-time AFNS model]

Table 3 reports the estimation result of the whole sample with the ROMER procedure. Standard errors are shown in brackets below the point estimate of each parameter. Bold-faced parameters are significant at the 95 percent confidence level.

The estimate of λ is 0.0715, implying a peak of curvature at the maturity of about twenty-five months, i.e., roughly two years. Parameter σ measures the average fitting error of the model. With an estimate of 6.58×10^{-5}, which is equivalent to 7.90 basis points for annualized yields, it implies that the model fits the data very well. The Level factor, L_t, is very persistent, with an autoregressive coefficient of 0.99. Interestingly, entries in the second and third rows of Φ are significant, indicating interactions between factors in the dynamics. The estimate of the variance-covariance matrix shows that the innovations of the Level and Slope factors have a strong negative correlation with each other, which is consistent with findings in Christensen, Diebold and Rudebusch (2011).

[Table 3. Estimates with the ROMER procedure]

In Figure 5, we show the comparison between the yield factors in the AFNS model and the corresponding empirical proxies of Level $(y_t^{(3)} + y_t^{(24)} + y_t^{(120)})/3$, Slope $(y_t^{(3)} - y_t^{(120)})$ and Curvature $(2y_t^{(24)} - y_t^{(3)} - y_t^{(120)})$. The figure confirms that the three factors in the AFNS model also closely correspond to the empirical proxies.

Figure 6 displays the yield adjustment term in the measurement equations and its confidence interval of two standard deviations. We obtain the confidence interval by simulating the adjustment term according to the distributions of parameter estimates. It shows that the yield adjustment increases with maturity, which means risk compensation increases as the risk of holding long term bonds increases. Compared to the DNS model, the AFNS model adds an adjustment term which is determined by the variance-covariance of the state innovations to account for the risk compensation for future uncertainty.

[Figure 5. Comparison of the three latent factors with empirical proxies]

[Figure 6. Adjustment term from the AFNS model]

5.2.2 Estimation results from MCMC

We demonstrate with simulated data in Section 3 that the ROMER procedure is easy to implement with high precision in the estimates. For the real data application, we now estimate the AFNS model with the alternative Bayesian Markov chain Monte Carlo (MCMC) method to provide a second check. The MCMC results are also useful in making further inferences on risk price and risk premia, for which the confidence intervals are easily computed based on the posterior distribution of parameters.

Table 4 presents the parameter estimates of the whole sample with the MCMC method. For each parameter we report the posterior median and a 95 percent confidence interval (CI).

[Table 4. Parameter estimates with the MCMC method]

Comparing these new results with the estimation results in Table 3, we find that the results are quite similiar. The point estimates are fairly close to each other; in particular, the point estimates of the VAR coefficient matrix are almost identical. The point estimates of the ROMER procedure always fall into the 95 percent confidence interval of the MCMC estimates, and the medians of the MCMC estimates always fall within the two standard deviations of the ROMER procedure estimates.

However, the MCMC results reveal an interesting skewness in the distribution of λ, which has a median at 0.0691 and the 95 percent confidence interval of $[0.0466, 0.0719]$. Since the higher λ is, the earlier the peak of curvature is and the quicker the loadings decay towards longer maturity. The concentration of λ to the right side of the median indicates a high probability that the peak is around two years and that loadings on slope and curvature decay relatively fast as maturities lengthen.

With this comparison using real data, we are reassured about the estimation property of the ROMER procedure, which has a particular advantage in its simplicity and speed. Running codes on a normal laptop computer with an Intel Core Processor of 3.3GHz, the ROMER procedure needs one second to deliver results, while the MCMC procedure of 10,000 draws needs about two days. It is of value to do further indirect inferences with the MCMC estimation, as we will show in the next section of inference on risk price and risk premia.

However, for real time applications with a time constraint, the ROMER procedure dominates. We will use the ROMER method for comparing the AFNS model with the DNS model on in-sample fit and out-of-sample forecast. We will then use the MCMC method's posterior samples to make indirect inferences on risk price and risk premia implied by the AFNS model.

5.3 Model comparison with DNS

By adding adjustment terms to the measurement equations, the AFNS model modifies the DNS model and adding theoretical rigor to exclude arbitrage opportunities under state dynamics. The AFNS model is still highly parsimonous, with only one additional parameters than the DNS model.

The question is this: can the AFNS model improve the in-sample fit and the out-of-sample forecast performance compared to the DNS model? If the answer is yes, then to what extent?

We answer the question in two steps. First, we make an analytical comparison of the two models in fitting and forecasting yield data, under the assumption that the AFNS model is the true data generating process with known parameter values. We look at how much is lost by fitting the yield curve to a DNS model. We establish an upper bound of difference in in-sample and out-of-sample root mean squared errors (RMSE), and discuss the impact factors when testing the difference. Our analysis is useful to understand why it is hard to find statistically favorable evidence to support a no-arbitrage restricted model. In the second step, we compare the models using real data. The results, although subject to both model uncertainty and parameter uncertainty, echo our analytical findings.

5.3.1 Prediction difference between AFNS and DNS models

Illustration with an ideal case: fitting an AFNS model with a DNS model. We assume, in an ideal scenario, that the AFNS model is the true data generating process as described in Equations (33) and (34), and that the parameters are also perfectly known. The measurement equation can be written compactly as

$$y_t = a + bX_t + \varepsilon_t, \varepsilon_t \sim N(0, I\sigma_\varepsilon^2) \tag{44}$$

where a is the $N \times 1$ vector containing $-\frac{A_n}{n}$, and b is the $N \times 3$ matrix containing the Nelson-Siegel loadings, $-\frac{B_n}{n}$, for a total number of N yields.

For a large N, X_t can be consistently estimated from the measurement equations with OLS, such that

$$\begin{aligned}\hat{X}_t &= (b'b)^{-1}b'(y_t - a) \\ &= (b'b)^{-1}b'(bX_t + \varepsilon_t) \\ &= X_t + (b'b)^{-1}b'_t\varepsilon_t\end{aligned}$$

where $(b'b)^{-1}b'_t\varepsilon_t$ is the regression error.

The h-steps ahead forecast of X_{t+h} is

$$\hat{X}_{t+h|t} = \sum_{i=0}^{h-1}\Phi^i\mu + \Phi^h\hat{X}_t \tag{45}$$

with prediction error

$$\begin{aligned}\hat{v}_{t,t+h} &\equiv X_{t+h} - \hat{X}_{t+h|t} \\ &= \sum_{i=0}^{h-1}\Phi^i v_{t+h-i} - \Phi^h(b'b)^{-1}b'\varepsilon_t.\end{aligned}$$

The forecast error for yields h-steps ahead is then

$$\begin{aligned}\hat{\varepsilon}_{t,t+h} &= y_{t+h} - a - b\hat{X}_{t+h|t} \\ &= a + bX_{t+h} + \varepsilon_{t+h} - a - b\hat{X}_{t+h|t} \\ &= b\cdot\hat{v}_{t,t+h} + \varepsilon_{t+h}\end{aligned}$$

for $h = 0, 1, 2, ...$

Now suppose we fit the yields generated by the AFNS model to the DNS model, by simply ignoring the constant term, a, such that

$$y_t = bZ_t + u_t.$$

The OLS estimation of Z_t is

$$\begin{aligned}\hat{Z}_t &= (b'b)^{-1}b'y_t \\ &= (b'b)^{-1}b'(a + bX_t + \varepsilon_t) \\ &= X_t + (b'b)^{-1}b'(a + \varepsilon_t) \\ &= \hat{X}_t + (b'b)^{-1}b'a.\end{aligned}$$

The constant term, $(b'b)^{-1}b'a$, is the difference between $\hat{X}_t$ and $\hat{Z}_t$. Understanding the difference and ignoring parameter uncertainty, the best forecast of Z_{t+h} is only different from the forecast of X_{t+h} by the constant term $(b'b)^{-1}b'a$, or

$$\hat{Z}_{t+h|t} = \hat{X}_{t+h|t} + (b'b)^{-1}b'a.$$

The resulting forecast error for yields h-steps ahead under the DNS model is then

$$\begin{aligned} \hat{u}_{t,t+h} &= y_{t+h} - b\hat{Z}_{t+h|t} \\ &= a + bX_{t+h} + \varepsilon_{t+h} - b\left(\hat{X}_{t+h|t} + (b'b)^{-1}b'a\right) \\ &= b \cdot \hat{v}_{t,t+h} + \varepsilon_{t+h} + Ma \end{aligned}$$

where $M \equiv I - b(b'b)^{-1}b'$, for $h = 0, 1, 2, ...$ It is clear that the difference between the forecast of yields for the AFNS and DNS models is Ma, or

$$\begin{aligned} \hat{y}_{t+h|t,AFNS} - \hat{y}_{t+h|t,DNS} &= a + b\hat{X}_{t+h|t} - b\hat{Z}_{t+h|t} \\ &= Ma. \end{aligned} \tag{46}$$

This term can be regarded as a residual term of a regressing on b in the regression

$$a_{(N\times1)} = b_{(N\times3)} \cdot \theta_{(3\times1)} + \iota_{(N\times1)}$$

where the OLS estimate of θ is $\hat{\theta} = (b'b)^{-1}b'a$. The fitted residual $\hat{\iota} = Ma$. Since we know that a is an upward sloping curve with a reasonable range of parameters, and the Nelson-Siegel factor loadings, b, is able to fit a curve with such a shape, the resulting residuals are very small.

Under this comparison with perfect information on parameter values, the following results hold:

1. The difference in mean squared error (MSE) between the DNS and AFNS models has a positive mean, but is subject to variations due to forecast errors of factors and measurement errors of yields.

Since AFNS is the true model in our scenario, it is expected that the in-sample and forecast RMSE from the AFNS model is smaller than the twisted DNS model. However, to what extent is the DNS model at a disadvantage? To see this, let us look at the difference in the MSEs:

$$
\begin{aligned}
& MSE_{DNS} - MSE_{AFNS} \\
= \;& diag\left[\frac{1}{T}\sum_{t=1}^{T}\hat{u}_{t,t+h}\hat{u}'_{t,t+h} - \frac{1}{T}\sum_{t=1}^{T}\hat{\varepsilon}_{t,t+h}\hat{\varepsilon}'_{t,t+h}\right] \\
= \;& diag\left[\frac{1}{T}\sum_{t=1}^{T}\left(b\cdot v_{t,t+h} + \varepsilon_{t+h} + Ma\right)\left(b\cdot v_{t,t+h} + \varepsilon_{t+h} + Ma\right)' \right.\\
& \left. -\frac{1}{T}\sum_{t=1}^{T}\left(b\cdot v_{t,t+h} + \varepsilon_{t+h}\right)\left(b\cdot v_{t,t+h} + \varepsilon_{t+h}\right)'\right] \\
= \;& diag\left[\frac{1}{T}\sum_{t=1}^{T}\left[(b\cdot v_{t,t+h} + \varepsilon_{t+h})a'M + Ma(b\cdot v_{t,t+h} + \varepsilon_{t+h})' + Maa'M\right]\right] \\
= \;& diag(Maa'M) + \frac{2}{T}diag\left(Ma\sum_{t=1}^{T}v'_{t,t+h}b'\right) + \frac{2}{T}diag\left(Ma\sum_{t=1}^{T}\varepsilon'_{t+h}\right).
\end{aligned}
$$

This result shows that the MSE of the AFNS model is, indeed, smaller in mean than the MSE of the DNS model at a constant level, $diag(Maa'M)$, wether in- or out-of-sample, and across forecast horizons. However, the difference is random with its variations affected by the measurement errors, ε_{t+h}, and forecast errors of $v_{t,t+h}$. These features also hold for the differences of mean absolute error (MAE) and RMSE, but the derivation is more involved.

2. The upper bound of the difference between in-sample RMSE and forecast RMSE of the AFNS and DNS models is $|Ma|$.

From Equation (46), we have

$$
\begin{aligned}
& |\sqrt{E(\hat{y}_{t+h|t,AFNS} - y_{t+h})^2} - \sqrt{E(\hat{y}_{t+h|t,DNS} - y_{t+h})^2}| \\
\leq \;& \sqrt{E(\hat{y}_{t+h|t,AFNS} - \hat{y}_{t+h|t,DNS})^2} \\
= \;& |Ma|. \qquad (47)
\end{aligned}
$$

This analysis shows that, being the true model, the AFNS model does fit and forecast better in the sense of a smaller MAE or RMSE on average, but the advantage is bounded by a small term and subject to randomness due to measurement and forecast errors. As the forecast horizon increases, both the MAE and RMSE increase with accumulated forecast errors, but the bound is constant, so the advantage of the AFNS model is diminishing.

Effects on fitting and forecasting the yield curve with the DNS model. To understand intuitively the effects of fitting and predicting an AFNS model with a DNS model specification, we use the estimates from the whole sample with the ROMER procedure for an illustration. We plot in Figure 7 the terms of a and Ma together, where Ma is the fitted residual of the DNS model with respect to the AFNS model. Figure 7 shows that, although the adjustment term, a, can reach about 100 basis points at the 10-year maturity, Ma is at a maximum 9 basis points and on average only about 3 basis points. At the three turning points where Ma changes signs, the differences are all zero such that statistical tests based on MAE or RMSE find it extremely difficult to distinguish between these two models, even when the sample length, T, is very large.

[Figure 7. Fitting an AFNS with DNS model: the fitted residual for a]

Next, we set λ free, so that the data optimally finds the ideal λ^* for the DNS model to minimize the residual squares. This will result in a fitted residual for the constant term M^*a $= [I - b^*(b^{*\prime}b^*)^{-1}b^{*\prime}]a$, where b^* is the Nelson-Siegel factor loadings determined by λ^*. The estimated $\lambda^* = 0.0570$ is smaller than the value from the true model, the AFNS model. This means a slower speed of decay for Slope and Curvature, and that curvature peaks later than in the AFNS model. Since λ^* is set optimally to fit the whole yield curve, not just the adjustment term, there will still be an unfitted residual term M^*a.

Figure 8 displays the factor loadings of Slope and Curvature of the AFNS model and the loadings resulting from the fitted DNS model. It is clear that the peak of Curvature in the AFNS model is earlier than in the DNS model.

[Figure 8. Fitting an AFNS model with DNS model: the factor loadings]

Complication with real application. In a real data application, more uncertainty is involved. Even if the AFNS model is the true model, parameters need to be estimated from the data, and parameter uncertainty from regression errors increases the variation of the difference between the two models. Besides, in reality, we do not know the true model. Model uncertainty will further complicate the comparison. The difference bound, $|Ma|$, will hold only roughly, but serve as a useful reference.

In what follows, we compare the two models with US data, using the measures of MAE and RMSE. The results are in line with the above analysis.

5.3.2 In-sample fit comparison

We estimate the models under two scenarios.

1. Where λ is free. Both the AFNS and DNS models are estimated with the ROMER procedure. While the AFNS model involves optimization of three parameters, λ, μ_L^Q and σ, the DNS model involves optimization of only two parameters, λ and σ, without the step of correcting for the adjustment term.

2. Where $\lambda = 0.0609$, the value calibrated by Diebold and Li (2006). At fixed λ, the first stage regression of the discrete-time AFNS model coincides with the Diebold-Li two-step regressions. Then the difference comes from the second stage regression from a correction of the constant adjustment term in the measurement equations. This experiment reveals the advantage of imposing no-arbitrage restrictions on the DNS model.

We use the MAE and the RMSE to evaluate the adequacy of in-sample fit. The errors are converted to percentage terms.

Table 5 presents the performance for in-sample fit of the discrete-time AFNS model and the DNS model, measured by MAE and RMSE. For each measure, the first two columns show the results under free λ, and the next two columns show the results under fixed λ. For each pair of comparison pair, we mark the better (smaller) MAE and RMSE in bold face.

[Table 5. In-sample fit comparison (1971.8-2010.9)]

Table 6 shows the subsample result of average in-sample fit across the yield curve with free λ. Here we divide the sample into four intervals, each of about ten years. We estimate the model separately for each subsample.

[Table 6. In-sample fit comparison of subsamples]

From Tables 6 and 7, the following observations are in order.

1) Both models fit the data well, with MAE and RMSE under ten basis points for almost the whole yield curve.

2) The discrete-time AFNS model dominates the DNS model for a majority of yields and for the average yield curve both in the whole sample and in the subsamples. Compared to the DNS model, the discrete-time AFNS model achieves a 15-20 percent reduction for both measures, which

amounts to an approximate one basis point improvement. During the 1990s, the reduction in RMSE and MAE reaches 30-40 percent, or about two basis points.

3) Similar to the previous findings from Nelson and Siegel (1987) and others, we find that the results are not very sensitive to the choice of λ, whether it is estimated freely or fixed ex ante.

5.3.3 Out-of-sample forecast comparison

The DNS model of Diebold and Li (2006) proves to be a useful tool for yield curve forecast. Christensen, Diebold and Rudebusch (2011) show favorable forecast performance of the AFNS model, especially the independent factor AFNS model, compared with the corresponding DNS specifications. However, in their comparison, the AFNS model is in continuous time, while the DNS model is in discrete time. Both models are estimated using MLE with a Kalman filter which may be subject to problems of local optimum and non-convergence due to high-dimensionality of the parameter space. The possible effects of different frameworks and the estimation method are hard to control for in the interpretation of the results. We estimate both models with in discrete time with the ROMER procedure.

Forecast models. We obtain $h-$steps ahead forecasts for the states by iterating the one-step model forward:[2]

$$\hat{X}_{t+h|t} = \sum_{i=0}^{h-1} \hat{\Phi}^i \hat{\mu} + \hat{\Phi}^h \hat{X}_t. \tag{48}$$

Yield forecasts based on model specifications are computed as follows. For the AFNS model, the optimal yield forecast for a maturity-τ yield made at time t for time $t+h$ is

$$\hat{y}_{t+h,t+h+n|t} = -\frac{A(\tau)}{\tau} + \hat{L}_{t+h|t} + \hat{S}_{t+h|t}\left(\frac{1-e^{-\lambda n}}{\lambda n}\right) + \hat{C}_{t+h|t}\left(\frac{1-e^{-\lambda n}}{\lambda n} - e^{-\lambda n}\right). \tag{49}$$

For the DNS model, the forecast has a similar expression but without the constant term,

$$\hat{y}_{t+h,t+h+n|t} = \hat{L}_{t+h|t} + \hat{S}_{t+h|t}\left(\frac{1-e^{-\lambda n}}{\lambda n}\right) + \hat{C}_{t+h|t}\left(\frac{1-e^{-\lambda n}}{\lambda n} - e^{-\lambda n}\right). \tag{50}$$

[2] The alternative would be to obtain forecasts by projecting h-steps ahead: $\hat{X}_{t+h|t} = \hat{\mu}_h + \hat{\Phi}_h X_t$. Given the nature of no-arbitrage models, only the iterated forecast can be computed for them. For this reason, we employ iterated forecasts for both models.

Forecast strategy. We do a rolling estimation for both models with a fixed window length at each point in time. We consider three forecasting horizons (denoted by h): 1 month, 6 months and 12 months. We choose two different window lengths for the rolling estimation. The first window length is ten years, i.e., a sample size of 120 months, using the sample period 1971:8-1981:7. For the 1-month ahead forecasting horizon, we conduct forecasts for all dates within the period 1981:8-2010:9, a total of 350 periods; for the 6-month ahead forecast, we end up with a total of 345 forecasts; and for the 12-month ahead forecast, a total of 339 forecasts. The second sample size is twenty years, using the sample period 1971:8-1991:7. For the 1-month ahead forecasting horizon, we conduct our exercise for all dates in the period 1991:8-2010:9, a total of 230 periods; for the 6-month ahead forecast, we end up with a total of 225 forecasts; and for the 12 month ahead forecast, a total of 219 forecasts.

Forecast measures. We choose the forecast root mean squared error (FRMSE) as a measure of forecast performance, and convert it to percentage terms.

Table 7 presents the results of the 10-year rolling window. For each maturity of yield, we report the 1-, 6- and 12-month ahead FRMSE for the two models: discrete-time AFNS and DNS. It is obvious that the differences between the AFNS and the DNS models are small and similar in scale, at maximally about three to four basis points, across different specifications and forecast horizons.

[Table 7. Forecast comparison with FRMSE]

Table 8 presents the results of the 20-year rolling window. Since the forecast period does not include the 1980s, a highly volatile period for yields, the FRMSEs are, in general, much smaller than the previous results. However, the fact that the difference between the two models across model specifications and forecast horizons is small, still holds. Although, in the case of independent factor models, the AFNS model tends to do better across maturities than the DNS model, in general, the independent-factor specification is not superior to the correlated-factor one.

[Table 8. Forecast comparison with FRMSE]

To summarize, although the AFNS model does perform better generally within sample, compared with the DNS model, its out-of-sample performance is not superior. The differences between them, whether in-sample or out-of-sample, are fairly small and are not uniformly positive or negative. In other words, the DNS model fares well for forecasts. If purely forecasting the level of yields, the no-arbitrage restrictions are not particularly helpful compared with its simple reduced-form counterpart, the DNS model. We provide an analytical illustration, under the assumption that the AFNS model is the true model, where the differences in fitting and forecasting yields are bounded by a small constant term which will be dwarfed by the forecast errors in factors as the forecast horizon increases. This result, as a specific example, supports the view of Duffee (2011), that cross-sectional relations among yields in a linear factor model are not helpful for forecasts, because cross-sectional properties of yields are easy to infer with high precision; dynamic restrictions are useful, but can be imposed without relying on the no-arbitrage structure.

5.4 Analysis on risk price and risk premia

The fact that no-arbitrage restrictions do not help for forecasts, at least in terms of the Nelson-Siegel class of models, is not a verdict of incompetence on this type of models. Again, at least in the forecast comparison of the AFNS and DNS models, the AFNS model does not fare worse. It is fair to say that it is safe to use the DNS model in forecasts, with its merit lying in its simplicity. However, the usefulness of no-arbitrage models in pricing derivatives are not substitutable by reduced-form models. In finance studies, the important issues of the risk price of factors and risk premia decomposition, can not be investigated in a reduced-form model.

Although the Nelson-Siegel factors of Level, Slope and Curvature have been extremely popular among investors, researchers and policy makers in yield curve interpolation, it is only possible to analyze their related risk price and risk premia within the no-arbitrage framework.

5.4.1 Price of risk

It is well known from empirical work that many popular ATSMs have identification problems, especially in the parameters of risk price. Hamilton and Wu (2010) establish that three popular canonical representations of Gaussian ATSMs are unidentified. The common approach is to set zero restrictions by assumption or to set some insignificant parameters from the initial estimation step to zero. There is no robust and systematic approach in dealing with this problem. The AFNS class of models addresses this problem indirectly by reverse engineering the restrictions from the Nelson-Siegel factor loadings which have been proved to make an effective interpolation of the yield curve.

Under a popular assumption of essentially affine term structure model in discrete time, which is widely used in macrofinance study, the risk price is an affine function of the state variables as described in Equation (4). Risk price parameters can be derived as

$$\lambda_0 = \Omega^{-1}(\mu - \mu^Q) \quad (51)$$

$$\lambda_1 = \Omega^{-1}(\Phi - \Phi^Q). \quad (52)$$

With the estimates of parameters in physical dynamics and in risk-neutral dynamics, specifically μ_L^Q and λ which fully determine μ^Q and Φ^Q, we can infer the market price of risk. Estimates from the ROMER and the MCMC procedures are both reliable. In order to use the ROMER estimates, one needs to make a simulation based on the parameter estimates, while the posterior samples of the MCMC method are readily available to make further inferences. So we may directly compute the posterior distribution of the risk price with the MCMC results.

Table 9 shows the median and 95 percent quantile of the posterior distribution of parameters λ_0 and λ_1. It can be seen that only three coefficients of time-varying risk parameter λ_1 are significant, where the Slope factor affects the risk price of Level positively, and the Level factor affects the risk price of Slope and Curvature with different signs. All parameters in the constant risk price, λ_0, are not significant.

[Table 9. Risk price parameters of λ_0 and λ_1]

We then compute the risk price, $\Lambda_t = \lambda_0 + \lambda_1 X_t$, and report the historical dynamics of the market price of risk for Level, Slope and Curvature in Figure 9. The solid lines in the middle are medians of risk prices associated with the three factors. Dotted lines denote the 95 percent quantiles of the distributions. The figure reveals that the market price of risk for Level and Slope are mostly negative and significant, while the risk price for Curvature is usually around zero and insignificant. These results are intuitive. The negative risk price of Level implies that when Level increases, the prices of bonds tend to be lower for compensating level risk, which results in risk premia in yields. When Slope increases, where Slope is defined as the short rate minus the long rate, the risk price for holding bonds also decreases, which leads to positive risk premia.

[Figure 9. Risk prices of the three factors]

5.4.2 Risk premia

The discussion on risk price is under the assumption on the particular functional form of risk price, as in Equation (4). The interpretation for risk price is, thus, only valid under this specification. But the AFNS model is actually silent on the specification form of risk price, and different risk price formulae may result in observationally equivalent yields and risk premia decomposition. To avoid possible specification errors and identification problems, it is useful to look at the risk premia decomposition, as derived in Equations (32), which is uniquely determined by the estimated model.

Figure 10 shows the factor loadings of the term structure of risk premia, where each curve shows the yield risk premia resulting from the underlying factor across maturities. We see that Level has positive factor loadings for risk premia across the yield curve. Slope has negative factor loadings across the yield and the loadings decrease with maturity. Curvature has apositive humped-shape loadings.

The positive loadings of Level on risk premia imply that when Level increases, the prices of bonds tend to be lower for compensating level risk. The longer the maturity, the higher the risk premia for holding long term bonds. Since the Nelson-Siegel Slope is close to the empirical concept of the difference between short- and long-term yields, i.e., $y_3 - y_{120}$, which is usually negative, the negative loadings of Slope on risk premia imply positive risk premia in yields associated with Slope. This risk premia is positively correlated with term premia which increases with maturity. The positive and hump-shaped loadings of Curvature on risk premia implies that medium-term yields have a significant positive response to the Curvature factor, which partly explains the hump-shaped loadings of Curvature on yields.

[Figure 10. Factor loadings of risk premia]

Combining the effects of factors on risk premia, we find the resulting total risk premia of yields for different maturities. We further decompose yields into two parts: one from the weighted average of the expected future short rate, i.e., expected yield under risk-neutral probability, the other from time-varying risk premia due to risk compensation on the underlying risk factors. Figure 11 plots the 10-year yield with its decomposition into the two components, with the estimation of the whole sample in Figure 11.a and from a subsample of the last ten years in Figure 11.b. For each sample, we first plot the yield decomposition all together, and then plot separately the residual term, which is left unexplained by the model.

[Figure 11. Decomposition of 10-year yield into expectation and risk premia]

As can be seen from these figures, the risk premia of 10-year yields are positive and time-varying, which reflects that investors in the bond market are risk-averse and require larger risk compensation for holding longer term bonds, but that the required compensation varies with the economic conditions.

Figure 11 also reveals an interesting implication from the AFNS model for yield dynamics in the past decade. During 2004 and 2005, when the Federal Reserve continually increased the short rate by a total of 150 basis points, the long-term yields stayed unexpectedly low, which does not comply with past empirical regularity and the expectation hypothesis. Greenspan dubbed this phenomenon the bond yield conundrum. Rudebusch, Swanson and Wu (2006) find commonly used essentially affine term structure models in macrofinance studies cannot capture this behavior, because the long-term yields have a large residual term that is unexplained from the fitted data during this period. From mid-2004 to the end of 2005, the BRS model (Bernanke, Reinhart and Sack, 2005) with five observable macroeconomic variables cannot fit the 10-year yield at all, leaving a residual of about 50 basis points; the Rudebusch-Wu model (Rudebusch and Wu, 2008) with Level, Slope and two macro variables also underestimates the 10-year yields with the residual around 40 basis points.

In Figure 11, we show the residual terms of the 10-year yield, which cannot be explained by the model. For the overlapping period of both samples, the residual terms agree with each other. With the three Nelson-Siegel factors, the 10-year yield has been very well fitted, with the largest errors occuring during the volatile period of the early 1980s and during the recent financial crisis. The residual term is mostly within 20 basis points during these unusual periods. During the bond yield conundrum period of mid-2004 to the end of 2005, there does not appear to be "conundrum" at all in terms of the residual term, as the residuals have been well within a few basis points. These results demonstrate that the AFNS model is able to explain the low level of long-term yields from its decomposition of risk premia and the expectation on the movement of the short-term interest rate. The upper panel of each figure shows the decomposition of the 10-year yield into risk premia and expectation term. While it is true that the expectation component denoted in dotted line rises sharply following the increasing short-rate during 2004-2005, the dashed line depicting the risk premia drops dramatically in due time, resulting in a persistently low long-term yield. The conundrum is thus caused by an ever falling and unusually low risk premia at the time.

Moreover, when we examine a more recent period, starting from 2007, although the expectation component drops dramatically as the Federal Reserve reduces the short-term interest rate, the long-term yields remain relatively stable. The model implies that risk premia picks up rapidly for the 10-year yield from 2007 up to the financial crisis. The risk premia even picks up earlier than when the expectation component began to drop in 2008. This phenomenon could serve as warning sign for the changing market sentiment about underlying risks.

6 Conclusions

Christensen, Diebold and Rudebusch (2009, 2011) derive the class of arbitrage-free Nelson-Siegel term structure model in continuous time with proof of the sufficient conditions for a Nelson-Siegel class yield curve model to be arbitrage free. The model is appealing as it combines the parsimony of the Nelson-Siegel model and the rigor of ATSMs.

For this class of model to be applied in a discrete-time framework, we derive the exact solution, in discrete time, for an arbitrage-free Nelson-Siegel class of term structure model, which features a dynamic coefficient matrix in exponential form under the risk-neutral measure. As done in Christensen, Diebold and Rudebusch (2009, 2011), we prove the existence of a solution. Moreover, we also prove the uniqueness of the solutions to rule out alternative solutions which may result in equivalent observation of yields.

To enhance the applicability of the model, we develop a simple and fast optimization procedure with imbedded multi-step regressions. The dimension of optimization is dramatically reduced from twenty-one to three in the AFNS model. The robustness and reliability of the procedure is verified with simulated data and results from the MCMC estimation with real data.

We show that the discrete-time AFNS model, although improving in-sample fit, does not improve out-of-sample forecast, compared to the reduced-form DNS model. We provide an analytical discussion on why the no-arbitrage restricted Nelson-Siegel model cannot outperform the DNS model for forecasts, although it is powerful in analyzing risk price and risk premia of yields within the popular interpolation of Level, Slope and Curvature. Estimation of the risk premia reveals that the model is useful in analyzing yield curve movement. The AFNS model provides a natural explanation to the bond yield conundrum of 2004-2005 where the risk premia fell dramatically, which counter-balanced the increasing expectation of short yield. The model also provides a warning sign of increasing risk premia ahead of the recent financial crisis.

Appendix 1. Proof of solution to the AFGNS model

For the affine arbitrage-free generalized Nelson-Siegel model which takes the following representation in yields,

$$\begin{aligned} y_{t,t+n} &= a_n + L_t + \left(\frac{1-e^{-\lambda_1 n}}{\lambda_1 n}\right) S_{1,t} + \left(\frac{1-e^{-\lambda_2 n}}{\lambda_2 n}\right) S_{2,t} \\ &+ \left(\frac{1-e^{-\lambda_1 n}}{\lambda_1 n} - e^{-\lambda_1 n}\right) C_{1,t} + \left(\frac{1-e^{-\lambda_2 n}}{\lambda_2 n} - e^{-\lambda_2 n}\right) C_{2,t}. \end{aligned}$$

We prove in the following the solution in discrete time.

Proposition A1: Sufficient conditions and the existence of a solution for the AFGNS model.

Assumption A1.1: The instantaneous risk-free rate is defined by

$$r_t = \delta_0 + \delta_1' X_t,$$

where $\delta_1 = \left[\begin{array}{ccccc} 1 & \frac{1-e^{-\lambda_1}}{\lambda_1} & \frac{1-e^{-\lambda_2}}{\lambda_2} & \frac{1-e^{-\lambda_1}}{\lambda_1} - e^{-\lambda_1} & \frac{1-e^{-\lambda_2}}{\lambda_2} - e^{-\lambda_2} \end{array}\right]'$.

Assumption A1.2: The state variables in X_t follow a VAR process under the risk-neutral $\mathbb{Q}$ measure such that

$$X_t = \mu^Q + \Phi^Q X_{t-1} + v_t^Q$$

with

$$\Phi^Q = \left[\begin{array}{ccccc} 1 & 0 & 0 & 0 & 0 \\ 0 & e^{-\lambda_1} & 0 & \lambda_1 e^{-\lambda_1} & 0 \\ 0 & 0 & e^{-\lambda_2} & 0 & \lambda_2 e^{-\lambda_2} \\ 0 & 0 & 0 & e^{-\lambda_1} & 0 \\ 0 & 0 & 0 & 0 & e^{-\lambda_2} \end{array}\right].$$

There are two implications.

Implication A1.1: Zero-coupon bond prices are given by

$$P_t^n = exp(A_n + B_n' X_t)$$

where B_n are Nelson-Siegel factor loadings with a second set of slope and curvature loadings determined by a different shape parameter. That is

$$B_n = \left[\begin{array}{ccccc} -n & -\frac{1-e^{-\lambda_1 n}}{\lambda_1} & -\frac{1-e^{-\lambda_2 n}}{\lambda_2} & ne^{-\lambda_1 n} - \frac{1-e^{-\lambda_1 n}}{\lambda_1} & ne^{-\lambda_2 n} - \frac{1-e^{-\lambda_2 n}}{\lambda_2} \end{array}\right]$$

and A_n satisfies the difference equation written as

$$A_n = A_{n-1} + B_{n-1}' \mu^Q + \frac{1}{2} B_{n-1}' \Omega B_{n-1} for n > 1.$$

Proof: By the ordinary difference equation (ODE) of factor loadings B_n in Equation (14), we have

$$\begin{bmatrix} B^1_{n+1} \\ B^2_{n+1} \\ B^3_{n+1} \\ B^4_{n+1} \\ B^5_{n+1} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & e^{-\lambda_1} & 0 & 0 & 0 \\ 0 & 0 & e^{-\lambda_2} & 0 & 0 \\ 0 & \lambda_1 e^{-\lambda_1} & 0 & e^{-\lambda_1} & 0 \\ 0 & 0 & \lambda_2 e^{-\lambda_2} & 0 & e^{-\lambda_2} \end{bmatrix} \begin{bmatrix} B^1_n \\ B^2_n \\ B^3_n \\ B^4_n \\ B^5_n \end{bmatrix} - \begin{bmatrix} 1 \\ \frac{1-e^{-\lambda_1}}{\lambda_1} \\ \frac{1-e^{-\lambda_2}}{\lambda_2} \\ \frac{1-e^{-\lambda_1}}{\lambda_1} - e^{-\lambda_1} \\ \frac{1-e^{-\lambda_2}}{\lambda_2} - e^{-\lambda_2} \end{bmatrix}.$$

We can divide this into two sets of independent differential equations:

$$\begin{bmatrix} B^1_{n+1} \\ B^2_{n+1} \\ B^4_{n+1} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & e^{-\lambda_1} & 0 \\ 0 & \lambda_1 e^{-\lambda_1} & e^{-\lambda_1} \end{bmatrix} \begin{bmatrix} B^1_n \\ B^2_n \\ B^4_n \end{bmatrix} - \begin{bmatrix} 1 \\ \frac{1-e^{-\lambda_1}}{\lambda_1} \\ \frac{1-e^{-\lambda_1}}{\lambda_1} - e^{-\lambda_1} \end{bmatrix}$$

and

$$\begin{bmatrix} B^1_{n+1} \\ B^3_{n+1} \\ B^5_{n+1} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & e^{-\lambda_2} & 0 \\ 0 & \lambda_2 e^{-\lambda_2} & e^{-\lambda_2} \end{bmatrix} \begin{bmatrix} B^1_n \\ B^3_n \\ B^5_n \end{bmatrix} - \begin{bmatrix} 1 \\ \frac{1-e^{-\lambda_2}}{\lambda_2} \\ \frac{1-e^{-\lambda_2}}{\lambda_2} - e^{-\lambda_2} \end{bmatrix}.$$

Then, using the proof of proposition 1 for the AFNS model in the paper, we know that

$$B_n = \begin{bmatrix} -n & -\frac{1-e^{-\lambda_1 n}}{\lambda_1} & -\frac{1-e^{-\lambda_2 n}}{\lambda_2} & ne^{-\lambda_1 n} - \frac{1-e^{-\lambda_1 n}}{\lambda_1} & ne^{-\lambda_2 n} - \frac{1-e^{-\lambda_2 n}}{\lambda_2} \end{bmatrix}.$$

■

Proposition A2: Necessary conditions and uniqueness of the solution for the AFGNS model.

Assumption A2.1: The following dynamic generalized Nelson-Siegel model with constant term can fit the yield curve completely, where $\lambda_1 \neq \lambda_2$, that is

$$\begin{aligned} y_{t,t+n} &= a_n + L_t + \left(\frac{1-e^{-\lambda_1 n}}{\lambda_1 n}\right) S_{1,t} + \left(\frac{1-e^{-\lambda_2 n}}{\lambda_2 n}\right) S_{2,t} \\ &\quad + \left(\frac{1-e^{-\lambda_1 n}}{\lambda_1 n} - e^{-\lambda_1 n}\right) C_{1,t} + \left(\frac{1-e^{-\lambda_2 n}}{\lambda_2 n} - e^{-\lambda_2 n}\right) C_{2,t}. \end{aligned}$$

Assumption A2.2: The general Nelson-Siegel latent factors follow a VAR(1) process under the risk-neutral $\mathbb{Q}$ measure such that

$$X_t = \mu^Q + \Phi^Q X_{t-1} + v^Q_t.$$

There are two implications.

Implication A2.1: The risk-free rate follows the affine process,

$$r_t = \delta_0 + \delta_1 X_t$$

where $\delta_1 = \left[\begin{array}{ccccc} 1 & \frac{1-e^{-\lambda_1}}{\lambda_1} & \frac{1-e^{-\lambda_2}}{\lambda_2} & \frac{1-e^{-\lambda_1}}{\lambda_1} - e^{-\lambda_1} & \frac{1-e^{-\lambda_2}}{\lambda_2} - e^{-\lambda_2} \end{array} \right]'$ and $\delta_0 = a_1$.

Implication A2.2: The risk-neutral dynamics for the latent factors satisfy

$$\Phi^Q = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & e^{-\lambda_1} & 0 & \lambda_1 e^{-\lambda_1} & 0 \\ 0 & 0 & e^{-\lambda_2} & 0 & \lambda_2 e^{-\lambda_2} \\ 0 & 0 & 0 & e^{-\lambda_1} & 0 \\ 0 & 0 & 0 & 0 & e^{-\lambda_2} \end{bmatrix}.$$

Proof: When $n = 1$, *Implication A2.1* is guaranteed by *Assumption A2.1*.

By *Assumption A2.2* and *Implication A2.1*, we know that this is an affine term structure model, so the yields are affine functions of X_t, or

$$y_{t,t+n} = \frac{-1}{n}\left(A_n + B_n' X_t\right)$$

where

$$A_{n+1} = A_n + B_n'\mu^Q + \frac{1}{2}B_n'\Omega B_n + A_1 \quad (53)$$

$$B_{n+1}' = B_n'\Phi^Q + B_1'. \quad (54)$$

Hence by comparing Equation (54) and *Assumption A2.1*, we have

$$B_n = \left[\begin{array}{ccccc} -n & -\frac{1-e^{-\lambda_1 n}}{\lambda_1} & -\frac{1-e^{-\lambda_1 n}}{\lambda_1} & ne^{-\lambda_1 n} - \frac{1-e^{-\lambda_1 n}}{\lambda_1} & ne^{-\lambda_2 n} - \frac{1-e^{-\lambda_2 n}}{\lambda_2} \end{array} \right].$$

We insert it into both sides of (54) and take transpose, we then have

$$\begin{bmatrix} B_{n+1}^1 \\ B_{n+1}^2 \\ B_{n+1}^3 \\ B_{n+1}^4 \\ B_{n+1}^5 \end{bmatrix} = \begin{bmatrix} \phi_{11} & \phi_{21} & \phi_{31} & \phi_{41} & \phi_{51} \\ \phi_{12} & \phi_{22} & \phi_{32} & \phi_{42} & \phi_{52} \\ \phi_{13} & \phi_{23} & \phi_{33} & \phi_{43} & \phi_{53} \\ \phi_{14} & \phi_{24} & \phi_{34} & \phi_{44} & \phi_{54} \\ \phi_{15} & \phi_{25} & \phi_{35} & \phi_{45} & \phi_{55} \end{bmatrix} \begin{bmatrix} B_n^1 \\ B_n^2 \\ B_n^3 \\ B_n^4 \\ B_n^5 \end{bmatrix} - \begin{bmatrix} 1 \\ \frac{1-e^{-\lambda_1}}{\lambda_1} \\ \frac{1-e^{-\lambda_2}}{\lambda_2} \\ \frac{1-e^{-\lambda_1}}{\lambda_1} - \lambda_1 e^{-\lambda_1} \\ \frac{1-e^{-\lambda_2}}{\lambda_2} - \lambda_2 e^{-\lambda_2} \end{bmatrix}.$$

After we rearrange,

$$\begin{bmatrix} n \\ \frac{e^{-\lambda_1} - e^{-\lambda_1(n+1)}}{\lambda_1} \\ \frac{e^{-\lambda_2} - e^{-\lambda_2(n+1)}}{\lambda_2} \\ \frac{e^{-\lambda_1} - e^{-\lambda_1(n+1)}}{\lambda_1} - (n+1)e^{-\lambda_1(n+1)} + e^{-\lambda_1} \\ \frac{e^{-\lambda_2} - e^{-\lambda_2(n+1)}}{\lambda_2} - (n+1)e^{-\lambda_2(n+1)} + e^{-\lambda_2} \end{bmatrix} = \begin{bmatrix} \phi_{11} & \phi_{21} & \phi_{31} & \phi_{41} & \phi_{51} \\ \phi_{12} & \phi_{22} & \phi_{32} & \phi_{42} & \phi_{52} \\ \phi_{13} & \phi_{23} & \phi_{33} & \phi_{43} & \phi_{53} \\ \phi_{14} & \phi_{24} & \phi_{34} & \phi_{44} & \phi_{54} \\ \phi_{15} & \phi_{25} & \phi_{35} & \phi_{45} & \phi_{55} \end{bmatrix} \begin{bmatrix} n \\ \frac{1-e^{-\lambda_1 n}}{\lambda_1} \\ \frac{1-e^{-\lambda_2 n}}{\lambda_2} \\ \frac{1-e^{-\lambda_1 n}}{\lambda_1} - ne^{-\lambda_1 n} \\ \frac{1-e^{-\lambda_2 n}}{\lambda_2} - ne^{-\lambda_2 n} \end{bmatrix}.$$

Next, we solve for the elements of Φ^Q by comparing the coefficients, equation by equation. The first equation implies that

$$\begin{aligned} \phi_{11} &= 1 \\ \phi_{21} &= \phi_{31} = \phi_{41} = \phi_{51} = 0. \end{aligned}$$

The second and third equations imply that

$$\begin{aligned} \phi_{12} &= \phi_{32} = \phi_{42} = \phi_{52} = 0 \\ \phi_{22} &= e^{-\lambda_1} \\ \phi_{13} &= \phi_{33} = \phi_{43} = \phi_{53} = 0 \\ \phi_{33} &= e^{-\lambda_2}. \end{aligned}$$

The fourth equation implies that

$$\phi_{14} = \phi_{34} = \phi_{54} = 0$$

and

$$\begin{aligned} & \frac{e^{-\lambda_1}}{\lambda_1} + e^{-\lambda_1} - e^{-\lambda_1}(1+\lambda_1)\frac{e^{-\lambda_1 n}}{\lambda_1} - e^{-\lambda_1} n e^{-\lambda_1 n} \\ = \; & \phi_{24}\frac{1}{\lambda_1} + \phi_{44}\frac{1}{\lambda_1} - (\phi_{24} + \phi_{44})\frac{e^{-\lambda_1 n}}{\lambda_1} - \phi_{44} n e^{-\lambda_1 n}) \end{aligned}$$

where the second line requires

$$\begin{aligned} \phi_{44} &= e^{-\lambda_1} \\ \phi_{24} &= \lambda_1 e^{-\lambda_1}. \end{aligned}$$

Similarly, for the fifth equation, we have

$$\begin{aligned} \phi_{15} &= \phi_{25} = \phi_{45} = 0 \\ \phi_{55} &= e^{-\lambda_2} \\ \phi_{35} &= \lambda_2 e^{-\lambda_2}. \end{aligned}$$

In summary, we have

$$\Phi^Q = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & e^{-\lambda_1} & 0 & \lambda_1 e^{-\lambda_1} & 0 \\ 0 & 0 & e^{-\lambda_2} & 0 & \lambda_2 e^{-\lambda_2} \\ 0 & 0 & 0 & e^{-\lambda_1} & 0 \\ 0 & 0 & 0 & 0 & e^{-\lambda_2} \end{bmatrix}.$$

■

Appendix 2. Bayesian Estimation

The model can also be estimated by the Bayesian Markov chain Monte Carlo (MCMC) method with a Gibbs sampling algorithm.

To ensure that the estimated results are accurate, we use the MCMC method to estimate this model. Here we adopt the MCMC diagnostic of Geweke (1992) to ensure the estimated results are reliable and that the draws converge to its posterior distribution. Briefly, under the weak conditions necessary for the Gibbs sampler to converge to a sequence of draws from the posterior, we can obtain a familiar central limit theorem for any functions of interest $g(\theta)$, or

$$\sqrt{S}\left\{\hat{g}_S - E\left[g\left(\theta\right)|y\right]\right\} \rightarrow N(0, \sigma_g^2) \tag{55}$$

where $\hat{g}_S$ is a sample average value of $g(\theta)$ in sample interval $[S_0, S]$, or

$$\hat{g}_S = \frac{1}{S}\sum_{s=S_0+1}^{S} g(\theta^{(s)}). \tag{56}$$

For simplicity, we use a weighted sum of each parameter as $g(\theta)$. This weighting allows each parameter to have the same order of magnitude as $g(\theta)$. That is

$$g(\theta) = 10\lambda + 1 \times 10^4 \mu_L^Q + 1 \times 10^4 \sigma + 1 \times 10^3 \sum_i \mu_i + \sum_{i,j} \Phi_{ij} + 1 \times 10^7 \sum_{i,j} \Omega_{ij}.$$

We take a total of $\bar{S} = 8000$ draws, and let an initial $S_0 = 3000$ be the burn-in replications. The remaining 5000 draws, S_1, are divided into sets with the first set of $S_A = 0.2S_1$ draws, a second set of $S_B = 0.6S_1$ draws and a third set of $S_C = 0.2S_1$ draws. Let $\hat{g}_{S_A}$ and $\hat{g}_{S_B}$ be the estimates of $E\left[g\left(\theta\right)|y\right]$ using the first S_A replications after the burn-in and the last S_C replications, respectively, using Equation (56). We define $\frac{\hat{\sigma}_A}{\sqrt{S_A}}$ and $\frac{\hat{\sigma}_b}{\sqrt{S_B}}$ to be the numerical standard errors of these two estimates. Then a central limit theorem, analogous to Equation (55), can be invoked such that

$$CD \to N(0,1)$$

where CD is the convergence diagnostic given by

$$CD = \frac{\hat{g}_{S_A} - \hat{g}_{S_B}}{\frac{\hat{\sigma}_A}{\sqrt{S_A}} + \frac{\hat{\sigma}_b}{\sqrt{S_B}}}.$$

A large value of CD indicates that $\hat{g}_{S_A}$ and $\hat{g}_{S_B}$ are quite different from one another and there is an insufficient number of replications. Here we estimate CD as 0.3739, and its P-value as 0.7085. The difference is insignificant, hence it indicates that our replication numbers are enough, and our draws $[S_0, S_1]$ have converged to the posterior distribution.

A.1 Drawing the latent factors

Conditional on parameter set Θ, we can sample the latent variable X_t with the simulation smoother proposed by DeJong and Shephard (1995). A detailed illustration of the procedure is also given by Koop (2003). In order to adapt our state-space model to the form used for the algorithm, we first transform the variables to remove the means, such that

$$\begin{aligned} y_t^* &= y_t - A - B\theta \\ &= B(X_t - \theta) + [I_N, 0] \begin{bmatrix} \varepsilon_t \\ \eta_t \end{bmatrix}, \varepsilon_t \tilde{\ } N(0, h^{-1} I_N) \\ X_{t+1}^* &= X_{t+1} - \theta \\ &= \Phi(X_t - \theta) + [0, \Sigma] \begin{bmatrix} \varepsilon_t \\ \eta_t \end{bmatrix}, \eta_t \tilde{\ } N(0, h^{-1} I_3) \end{aligned}$$

where $\theta = (I - \Phi)^{-1}\mu$ is the unconditional mean of the latent variables, $A = [A_{n_1}, A_{n_2}, \cdots, A_{n_N}]'$, $B = \left[B'_{n_1}; B'_{n_2}, \cdots, B'_{n_N}\right]$, $h = (\sigma_\varepsilon^2)^{-1}$, and $\Sigma = h^{\frac{1}{2}} chol(\Omega)$, where $chol$ stands for Cholesky factorization of variance-covariance matrix. Hence, y_t^* and X_t^* are the vectors of demeaned yields and latent factors.

In this procedure, a standard Kalman filter is used to filter out X_t^* forward. Then a backward smoothing algorithm is used to obtain draw η_t, hence X_t^*, effectively conditional on all information up to time T. We can then transform X_t^* back to X_t for later inferences on parameters.

A.2 Drawing $\Theta_2 = \{\mu, \Phi, \Omega\}$

Conditional on X_t, we can draw parameters in the state equation, $\Theta_2 = \{\mu, \Phi, \Omega\}$. Although one can make independent assumptions on the distributions of $\{\mu, \Phi\}$ and Ω, for simplicity we employ a conjugate normal-inverted Wishart (N-IW) prior and posterior. Specifically, if we let

K denote the number of latent factors, let $X_{(T-1)\times(K+1)}$ contain in each row $[1, X_t']_{t=1,\ldots,T-1}$, let $Z_{(T-1)\times 3}$ contain $[X_{t+1}']_{t=1,\ldots,T-1}$ in each row, and $\Psi = \begin{bmatrix} \mu' \\ \Phi' \end{bmatrix}_{(K+1)\times K}$, then the state equation can be expressed as

$$Z = X\Psi + U$$

where rows of U are *i.i.d.* $N(0, \Omega)$.

Let the prior distribution of Ω be proportional to $|\Omega|^{-(K+1)/2}$, where K is the dimension of the state variables, and uninformative over Ψ (the so-called Jeffreys prior). Specifically, let $\hat{\Psi}$ denote the OLS estimate of Ψ, and denote

$$\hat{H} = (Z - X\hat{\Psi})'(Z - X\hat{\Psi})$$

and

$$\hat{\psi} = vec(\hat{\Psi}).$$

The corresponding posterior distribution is also NI-W, given by

$$\begin{aligned} P(\psi|\Omega) &\sim N(\hat{\psi}|\Omega \otimes (X'X)^{-1}) \\ P(\Omega) &\sim IW(\hat{H}, \tau, K) \end{aligned}$$

where $IW(\hat{H}, \tau, K)$ denotes an inverted-Wishart distribution with $K \times K$ parameter matrix $\hat{H}$ and degrees of freedom denoted $\tau = T - 1$.

A.3 Drawing $\Theta_1 = \{\lambda, \mu_L^Q, \sigma_\varepsilon^2\}$

Conditional on X_t and Θ_2, we can then infer $\Theta_1 = \{\lambda, \mu_L^Q, \sigma_\varepsilon^2\}$. These three parameters are all involved in the measurement equations in a non-linear form such that

$$y_{t,t+n} = A_n(\lambda, \mu_L^Q) + B_n(\lambda)X_t + \varepsilon_t, \varepsilon_t \sim N(0, \sigma_\varepsilon^2).$$

In fact, here two Gibbs steps are involved: conditional on γ to draw h and conditional on h to draw γ.

Let the prior of $h = (\sigma_\varepsilon^2)^{-1}$ and $\gamma = \begin{pmatrix} \lambda & \mu_L^Q \end{pmatrix}$ be a noninformative distribution, given by

$$p(\gamma, h) \propto \frac{1}{h}.$$

The posterior distributions are

$$
\begin{aligned}
h|y,\gamma &\sim G(s_1^{-2}, v_1) \\
P(\gamma|y,h) &\propto \exp(-\frac{h}{2}\sum_n (y_{t,t+n} - A_n(\lambda,\mu_L^Q) - B_n(\lambda)X_t)'(y_{t,t+n} - A_n(\lambda,\mu_L^Q) - B_n(\lambda)X_t))
\end{aligned}
$$

where

$$
\begin{aligned}
s_1^2 &= \frac{\sum_n [(y_{t,t+n} - A_n(\lambda,\mu_L^Q) - B_n(\lambda)X_t)'(y_{t,t+n} - A_n(\lambda,\mu_L^Q) - B_n(\lambda)X_t)]}{v_1} \\
v_1 &= TN.
\end{aligned}
$$

The conditional posterior density of γ does not follow a distribution that can be directly drawn from. We use a random walk chain Metropolis-Hasting algorithm to draw γ from $P(\gamma|y,h)$.

7 References

1. Ang, A. and M. Piazzesi, 2003, A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables. Journal of Monetary Economics 50, 4, 745-787.

2. Bernanke, B., V. Reinhart, and B. Sack, 2005, Monetary policy alternatives at the zero bound: An empirical assessment. Brookings Papers on Economic Activity 2, 1-78.

3. Christensen, J. H. E., F. X. Diebold and G. D. Rudebusch, 2009, An arbitrage-free generalized Nelson-Siegel term structure model. Econometrics Journal 1, 1-31.

4. Christensen, J. H. E., F. X. Diebold and G. D. Rudebusch, 2011, The affine arbitrage-free class of Nelson-Siegel term structure models. Journal of Econometrics 164, 4-20.

5. Cox, J. C., J. E. Ingersoll and S. A. Ross, 1985, A theory of the term structure of interest rates. Econometrica 53, 385-407.

6. Dai, Q. and K. Singleton, 2000, Specification analysis of affine term structure models. Journal of Finance 55, 5, 1943-78.

7. DeJong, P. and N. Shephard, 1995, The simulation smoother for time series models. Biometrika 82, 339-350.

8. Diebold, F. X. and C. Li, 2006, Forecasting the term structure of government bond yields. Journal of Econometrics 130, 337-364.

9. Diebold, F. X., M. Piazzesi and G. D. Rudebusch, 2005, Modeling bond yields in finance and macroeconomics. American Economic Review 95, 415-420.

10. Duffee, G. R., 2002, Term premia and interest rate forecasts in affine models. Journal of Finance 57, 405-443.

11. Duffee, G. R., 2011, Forecasting with the term structure: The role of no-arbitrage restrictions. Working paper, Johns Hopkins University.

12. Duffie, D. and R. Kan, 1996, A yield-factor model of interest rates. Mathematical Finance 6, 379-406.

13. Favero, C. A., L. Niu and L. Sala, 2012, Term structure forecasting: No-arbitrage restrictions versus large information set. Forthcoming in Journal of Forecasting.

14. Filpovic, D., 1999, A note on the Nelson-Siegel family. Mathematical Finance 9, 349-359.

15. Geweke, J., 1992, Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In Bernardo, J., Berger, J. Dawid, A. and Smith, A. (eds), Bayesian Statistics 4, 641-649. Oxford: Clarendon Press.

16. Gurkaynak, R. S., B. Sack, and J. H. Wright, 2007, The U.S. Treasury yield curve: 1961 to the present. Journal of Monetary Economics 54, 8, 2291-2304.

17. Gurkaynak, R. S. and J. H. Wright, 2010, Macroeconomics and the term structure. CEPR Discussion Paper No. DP8018.

18. Hamilton, J. D. and J. C. Wu, 2010, Identification and estimation of Gaussian affine term structure models. Chicago Booth Research Paper No. 11-28.

19. Koop, G., 2003, Bayesian Econometrics, John Wiley & Sons Ltd.

20. Nelson C. R. and A. F. Siegel, 1987, Parsimonious modeling of yield curves. Journal of Business 60, 473-89.

21. Rudebusch, G., E. Swanson and T. Wu, 2006, The bond yield "conundrum" from a macro-finance perspective. Monetary and Economic Studies 24(S-1), 83-128.

22. Rudebusch, G. and T. Wu, 2008, A Macro-finance model of the term structure, monetary policy, and the economy. Economic Journal 118, 906-926.

23. Svensson, L. E. O., 1995, Estimating forward interest rates with the extended Neslon-Siegel method. Quarterly Review, No. 3, Sveriges Riksbank, 13-26.

24. Vasicek, O., 1977, An equilibrium characterization of the term structure. Journal of Financial Economics 5, 177-188.

Table 1: Parameter values for simulation

λ	μ_L^Q	σ
0.0609	2×10^{-5}	5×10^{-5}

Φ	L_{t-1}	S_{t-1}	C_{t-1}	μ
L_t	0.98	0	0	1×10^{-4}
S_t	-0.1	0.91	0.1	1×10^{-4}
C_t	0	0	0.89	-1×10^{-4}

Ω	1	2	3
1	1×10^{-7}	-0.5×10^{-7}	0
2	-0.5×10^{-7}	1×10^{-7}	0
3	0	0	5×10^{-7}

Table 2. Parameter estimates from simulated yield data

λ	μ_L^Q	σ
0.0609 (0.0008)	**1.99×10^{-4}** (0.067×10^{-4})	**5.00×10^{-4}** (0.005×10^{-4})

Φ	L_{t-1}	S_{t-1}	C_{t-1}	μ
L_t	**0.96** (0.02)	0.00 (0.02)	0.01 (0.01)	**1.54×10^{-4}** (0.62×10^{-4})
S_t	**-0.10** (0.02)	**0.91** (0.01)	**0.10** (0.01)	0.75×10^{-4} (0.58×10^{-4})
C_t	0.00 (0.05)	0.00 (0.03)	**0.86** (0.03)	-1.36×10^{-4} (1.42×10^{-4})

Ω	1	2	3
1	**1.0×10^{-7}** (0.07×10^{-7})		
2	**-0.53×10^{-7}** (0.06×10^{-7})	**1.0×10^{-7}** (0.08×10^{-7})	
3	-0.01×10^{-7} (0.12×10^{-7})	0.00×10^{-7} (0.13×10^{-6})	**5.55×10^{-7}** (0.42×10^{-7})

Table 3. Estimates with the ROMER procedure

λ	μ_L^Q	σ
0.0715 (0.0035)	**1.86×10⁻⁵** (0.05×10⁻⁵)	**6.58×10⁻⁵** (0.06×10⁻⁵)

Φ	L_{t-1}	S_{t-1}	C_{t-1}	μ
L_t	**0.99** (0.01)	**0.03** (0.01)	-0.01 (0.01)	4.68×10^{-5} (3.77×10^{-5})
S_t	**-0.03** (0.01)	**0.92** (0.02)	**0.05** (0.01)	5.49×10^{-5} (6.82×10^{-5})
C_t	**0.07** (0.02)	0.05 (0.03)	**0.88** (0.02)	**-2.32×10^{-4}** (1.18×10^{-4})

Ω	1	2	3
1	**7.28×10^{-8}** (0.70×10^{-8})		
2	**-3.78×10^{-8}** (0.86×10^{-8})	**2.38×10^{-7}** 0.35×10^{-8}	
3	0.15×10^{-8} (1.31×10^{-8})	0.22×10^{-8} (4.52×10^{-8})	**7.17×10^{-7}** (0.88×10^{-7})

Table 4. Parameter estimates with the MCMC method

λ	μ_L^Q	σ
0.0691 [0.0466,0.0719]	**1.87×10^{-5}** [1.53,1.97]×10^{-5}	**6.67×10^{-5}** [6.23,7.88]×10^{-5}

Φ	L_{t-1}	S_{t-1}	C_{t-1}	μ
L_t	**0.99** [0.97,1.01]	0.03 [-0.01,0.05]	-0.01 [-0.03,0.01]	3.56×10^{-5} [-3.50,12.16]×10^{-5}
S_t	**-0.03** [-0.07,0.00]	**0.92** [0.89,0.96]	**0.05** (0.02,0.08]	5.56×10^{-5} [-8.71,18.94]×10^{-5}
C_t	**0.06** [0.01,0.11]	0.04 [-0.01,0.10]	**0.89** [0.84,0.94]	-1.84×10^{-4} [-4.46,0.28]×10^{-4}

Ω	1	2	3
1	**6.39×10^{-8}** [5.50,8.02]×10^{-8}		
2	**-3.36×10^{-8}** [-4.95,-2.10]×10^{-8}	**2.37×10^{-7}** [2.04,2.69]×10^{-7}	
3	2.21×10^{-8} [-5.20,4.58]×10^{-8}	-1.61×10^{-8} [-3.76,5.05]×10^{-8}	**6.03×10^{-7}** [5.19,7.90]×10^{-7}

Table 5. In-sample fit comparison (1971.8-2010.9)

n	MAE				RMSE			
	λ free		λ=0.0609		λ free		λ=0.0609	
	AFNS	DNS	AFNS	DNS	AFNS	DNS	AFNS	DNS
3	**0.084**	0.126	**0.093**	0.123	**0.123**	0.172	**0.137**	0.167
6	**0.062**	0.065	**0.061**	0.064	**0.082**	0.084	**0.082**	0.084
9	**0.094**	0.116	**0.099**	0.117	**0.127**	0.149	**0.134**	0.149
12	**0.063**	0.095	**0.068**	0.095	**0.087**	0.121	**0.095**	0.120
18	**0.035**	0.053	**0.034**	0.052	**0.046**	0.066	**0.046**	0.064
24	0.035	**0.024**	0.031	**0.025**	0.044	**0.030**	0.040	**0.031**
30	0.036	**0.026**	0.037	**0.028**	0.049	**0.038**	0.049	**0.041**
36	**0.038**	0.044	**0.039**	0.047	**0.054**	0.058	**0.055**	0.061
48	**0.045**	0.064	**0.041**	0.067	**0.061**	0.080	**0.056**	0.083
60	**0.047**	0.063	**0.040**	0.065	**0.060**	0.076	**0.053**	0.079
72	**0.041**	0.046	**0.035**	0.047	**0.050**	0.056	**0.045**	0.057
84	0.027	**0.021**	0.027	**0.021**	0.032	**0.027**	0.032	**0.027**
96	0.017	**0.013**	0.017	**0.015**	0.023	**0.019**	0.023	**0.020**
108	**0.033**	0.043	**0.027**	0.046	**0.043**	0.053	**0.036**	0.056
120	**0.061**	0.074	**0.051**	0.078	**0.076**	0.088	**0.066**	0.094
Mean	**0.048**	0.058	**0.047**	0.059	**0.064**	0.075	**0.063**	0.076

Note: “n” stands for number of maturity in months.

Table 6. In-sample fit comparison of subsamples

	1971.8-1981.7		1981.8-1991.7		1991.8-2001.7		2001.8-2010.9	
	AFNS	DNS	AFNS	DNS	AFNS	DNS	AFNS	DNS
mean of MAE	**0.046**	0.052	**0.049**	0.058	**0.031**	0.054	**0.034**	0.040
mean of RMSE	**0.060**	0.068	**0.060**	0.073	**0.038**	0.061	**0.045**	0.050

Table 7. Forecast comparison with FRMSE

(10-year rolling estimation for forecasts between 1981:1 to 2010:9)

n	1-m ahead				6-m ahead				12-m ahead			
	Corr-factor		Ind-factor		Corr-factor		Ind-factor		Corr-factor		Ind-factor	
	AFNS	DNS	AFNS	DNS	AFNS	DNS	AFNS	DNS	AFNS	DNS	AFNS	DNS
3	**0.415**	0.427	**0.432**	0.440	**1.207**	1.229	**1.272**	1.279	**2.054**	2.082	**2.098**	2.113
6	0.410	**0.408**	0.422	**0.420**	1.233	**1.231**	1.283	**1.276**	2.076	**2.073**	2.101	**2.095**
9	**0.390**	0.391	**0.392**	0.394	1.200	**1.189**	1.235	**1.224**	2.020	**2.004**	2.031	**2.016**
12	0.396	**0.394**	0.400	0.400	1.211	**1.195**	1.238	**1.226**	2.013	**1.991**	2.014	**1.997**
18	0.403	**0.399**	0.410	**0.408**	1.205	**1.191**	1.221	**1.212**	1.967	**1.949**	1.957	**1.945**
24	0.405	**0.403**	**0.412**	0.413	1.181	**1.175**	1.191	**1.190**	1.904	**1.897**	1.889	**1.888**
30	**0.403**	0.405	**0.411**	0.414	**1.150**	1.154	**1.157**	1.164	**1.838**	1.843	**1.824**	1.832
36	**0.400**	0.405	**0.406**	0.414	**1.119**	1.131	**1.124**	1.138	**1.775**	1.791	**1.763**	1.780
48	**0.388**	0.400	**0.394**	0.406	**1.061**	1.084	**1.064**	1.086	**1.667**	1.694	**1.662**	1.687
60	**0.375**	0.388	**0.380**	0.393	**1.013**	1.037	**1.016**	1.038	**1.582**	1.610	**1.584**	1.609
72	**0.364**	0.374	**0.367**	0.377	**0.976**	0.994	**0.978**	0.995	**1.518**	1.538	**1.527**	1.543
84	**0.356**	0.361	**0.359**	0.363	**0.948**	0.956	**0.951**	0.957	**1.470**	1.478	**1.485**	1.489
96	0.352	**0.351**	0.353	0.353	0.928	**0.923**	0.931	**0.926**	1.434	**1.426**	1.454	**1.444**
108	0.351	**0.346**	0.351	**0.348**	0.914	**0.896**	0.917	**0.900**	1.407	**1.383**	1.432	**1.407**
120	0.352	**0.347**	0.352	**0.348**	0.905	**0.873**	0.908	**0.879**	1.388	**1.346**	1.416	**1.376**

Note: "n" stands for number of maturity in months. "Corr-factor" means correlated factors and the state dynamics is a VAR process. "Ind-factor" means independent factors and the state dynamics of each factor is modeled as a AR process .

Table 8. Forecast comparison with FRMSE

(20-year rolling estimation for forecast between 1991:1 to 2010:9)

n	1-m ahead				6-m ahead				12-m ahead			
	Corr-factor		Indept-factor		Corr-factor		Indept-factor		Corr-factor		Indept-factor	
	AFNS	DNS	AFNS	DNS	AFNS	DNS	AFNS	DNS	AFNS	DNS	AFNS	DNS
3	**0.259**	0.274	**0.310**	0.317	**0.903**	0.923	**1.018**	1.026	**1.528**	1.557	**1.596**	1.620
6	0.255	**0.254**	**0.274**	0.278	0.960	**0.959**	**1.038**	1.042	**1.575**	1.576	**1.622**	1.632
9	**0.255**	0.260	**0.263**	0.273	0.969	**0.959**	**1.040**	1.043	1.566	1.553	**1.619**	1.624
12	**0.269**	0.272	**0.272**	0.280	0.990	**0.977**	**1.052**	1.054	1.573	**1.555**	**1.626**	1.630
18	0.300	0.300	**0.301**	0.302	1.007	**0.993**	**1.060**	1.065	1.553	**1.534**	**1.609**	1.618
24	0.324	**0.319**	0.327	**0.326**	1.001	**0.993**	**1.053**	1.064	1.510	**1.500**	**1.569**	1.591
30	0.340	**0.337**	**0.344**	0.346	0.984	0.984	**1.034**	1.054	**1.458**	1.460	**1.520**	1.554
36	**0.347**	0.349	**0.352**	0.358	**0.960**	0.967	**1.008**	1.037	**1.403**	1.417	**1.467**	1.513
48	**0.345**	0.355	**0.348**	0.364	**0.903**	0.923	**0.947**	0.989	**1.299**	1.329	**1.361**	1.423
60	**0.330**	0.343	**0.331**	0.351	**0.846**	0.870	**0.885**	0.930	**1.208**	1.243	**1.266**	1.331
72	**0.312**	0.323	**0.312**	0.329	**0.800**	0.816	**0.830**	0.869	**1.133**	1.162	**1.185**	1.243
84	**0.300**	0.304	**0.297**	0.307	**0.755**	0.766	**0.784**	0.812	**1.073**	1.088	**1.120**	1.162
96	0.291	0.291	**0.288**	0.291	0.723	**0.722**	**0.748**	0.761	1.026	**1.023**	**1.069**	1.090
108	0.288	**0.287**	**0.285**	0.286	0.700	**0.684**	0.721	**0.719**	0.990	**0.966**	1.029	1.029
120	**0.289**	0.292	**0.286**	0.290	0.681	**0.654**	0.701	**0.685**	0.962	**0.917**	1.000	**0.976**

Note: "n" stands for number of maturity in months. "Corr-factor" means correlated factors and the state dynamics is a VAR process. "Ind-factor" means independent factors and the state dynamics of each factor is modeled as a AR process .

Table 9. Risk price parameters of λ_0 and λ_1

	λ_0 (: $\times 10^2$)	λ_1 (: $\times 10^5$)		
L	5.33 [-6.11,17.54]	-1.42 [-3.86,0.88]	**4.97** [2.01,7.87]	-2.09 [-4.54,0.82]
S	3.21 [-2.91,9.29]	**-1.44** [-2.67,-0.22]	0.36 [-1.05,1.76]	-0.84 [-2.10,0.40]
C	-3.21 [-6.90,0.34]	**1.00** [0.28,1.79]	0.53 [-0.31,1.47]	-0.55 [-1.49,0.21]

Figure 1. Comparison of the estimated and true latent factors

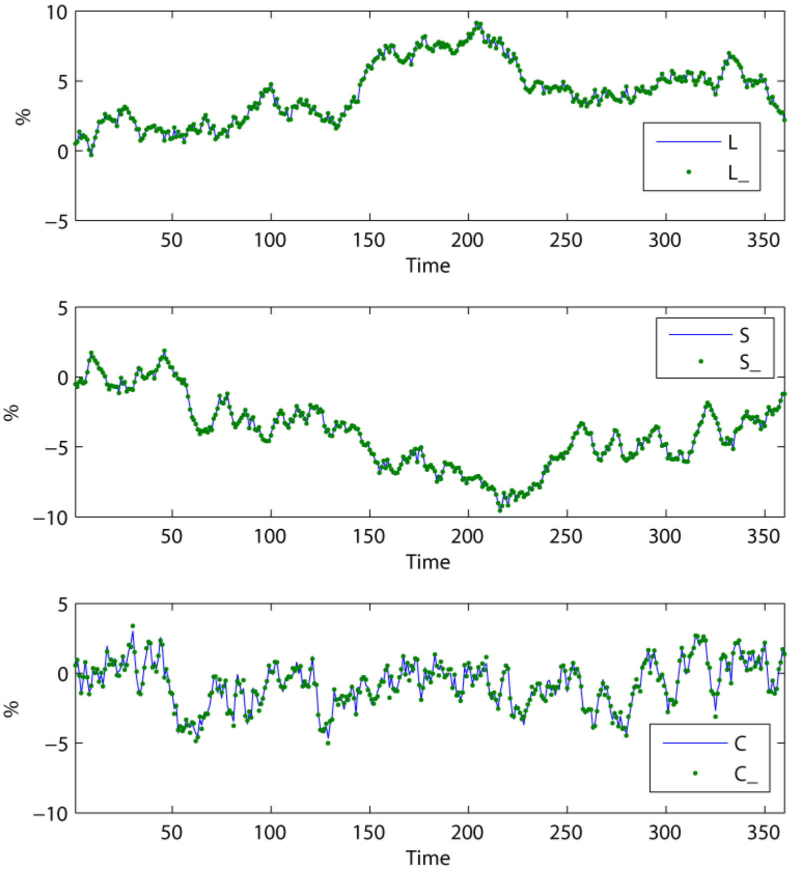


Figure 2. Adjustment term from simulated data and estimation

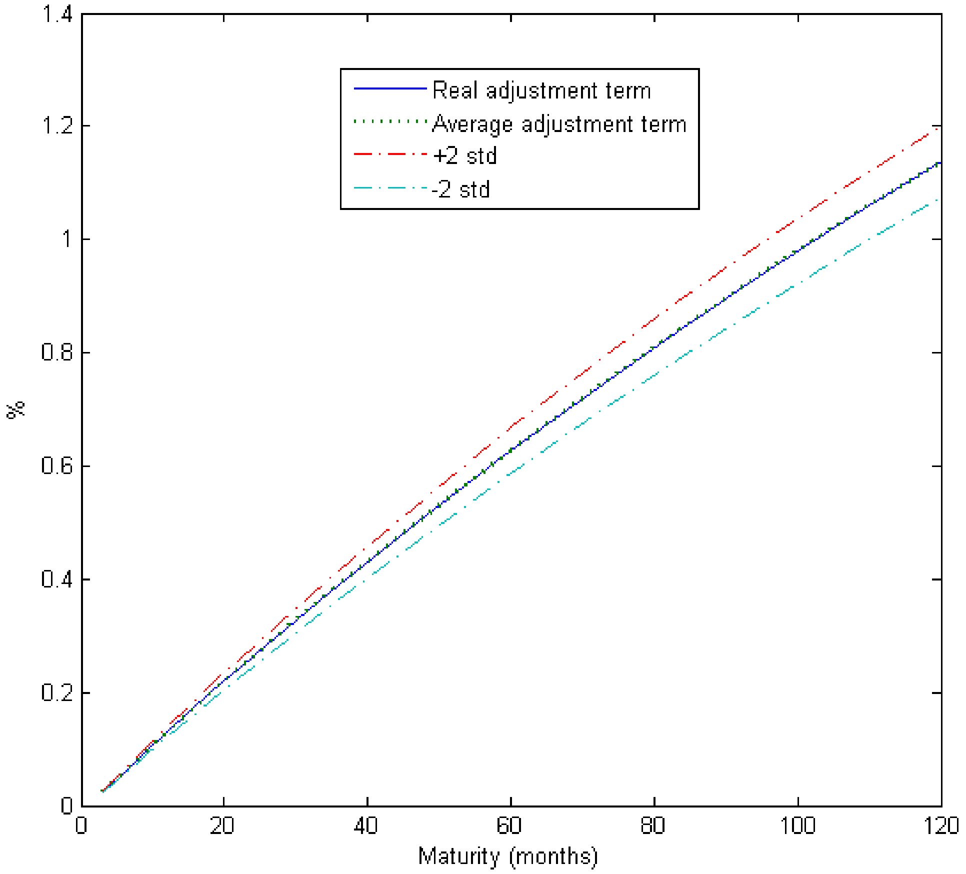


Figure 3. US yield curve from 1971:8 to 2010:9

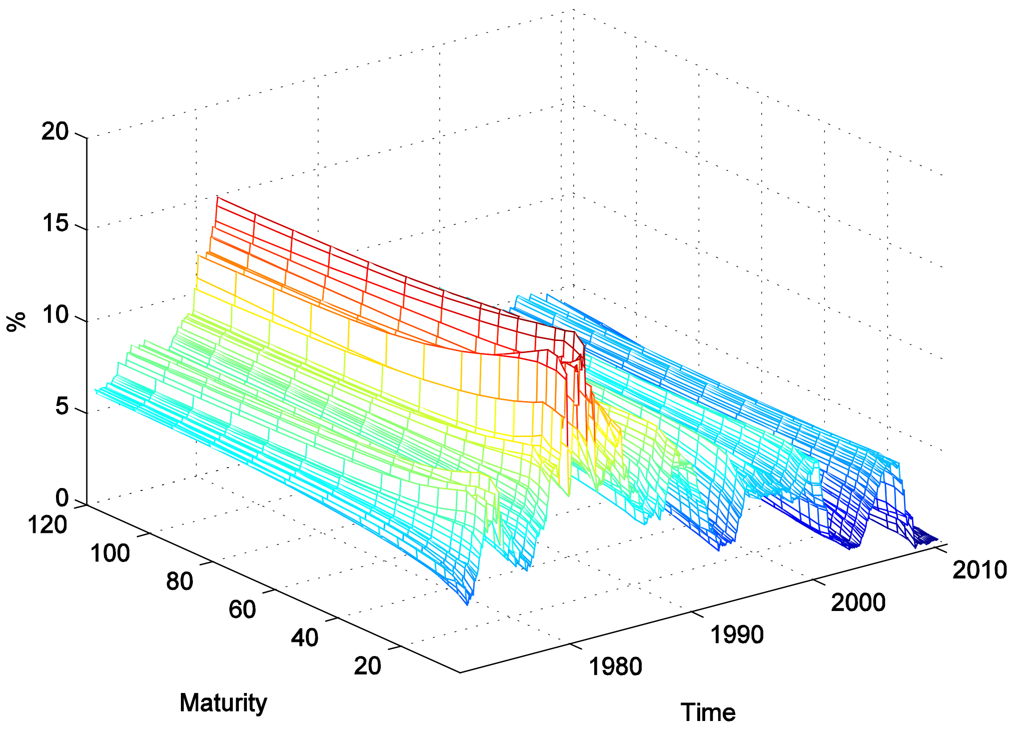


Figure 4. Fitted residuals of the discrete-time AFNS model

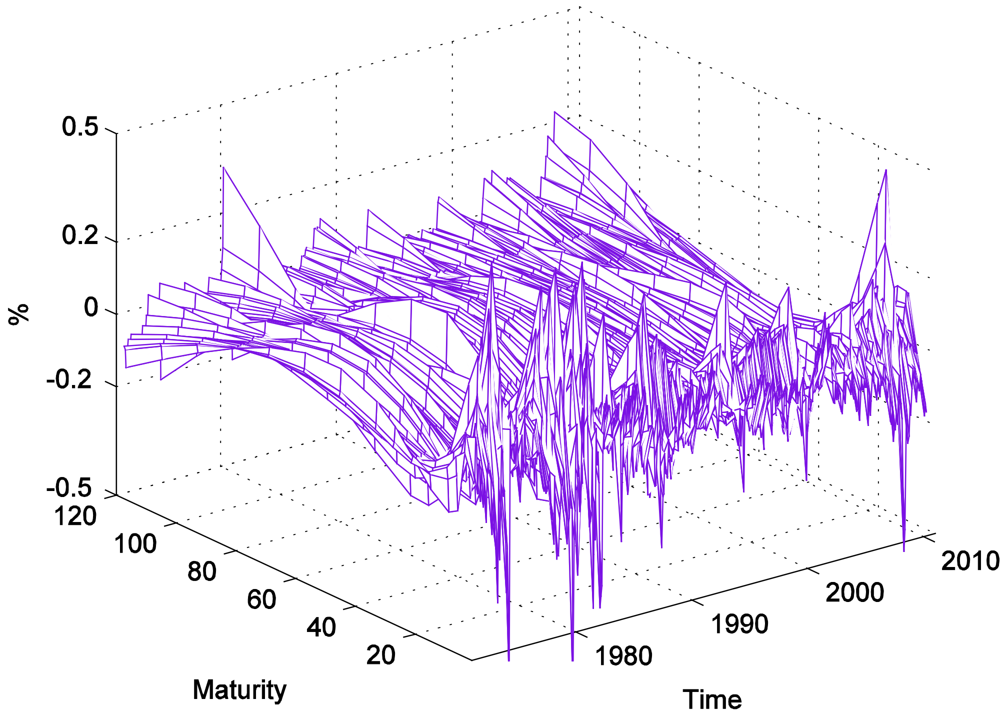


Figure 5. Comparison of the three latent factors with empirical proxies

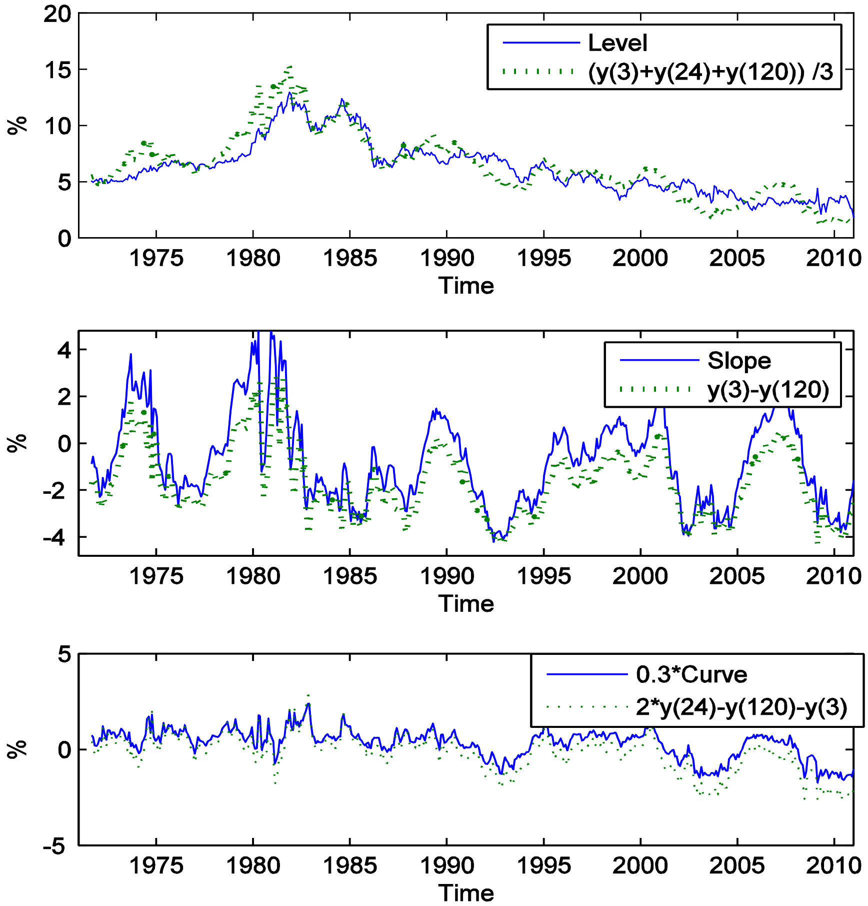


Figure 6. Adjustment term from the AFNS model

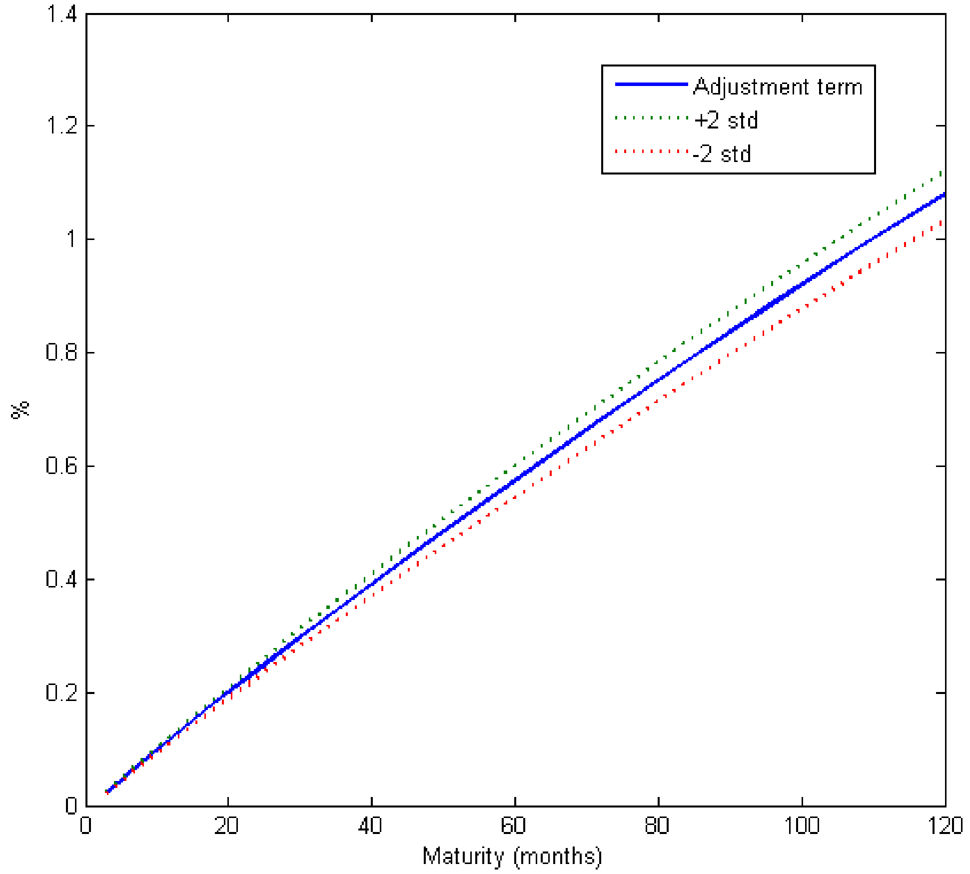


Figure 7. Fitting an AFNS model with a DNS model: the fitted residual for *a*

(λ fixed at 0.0715)

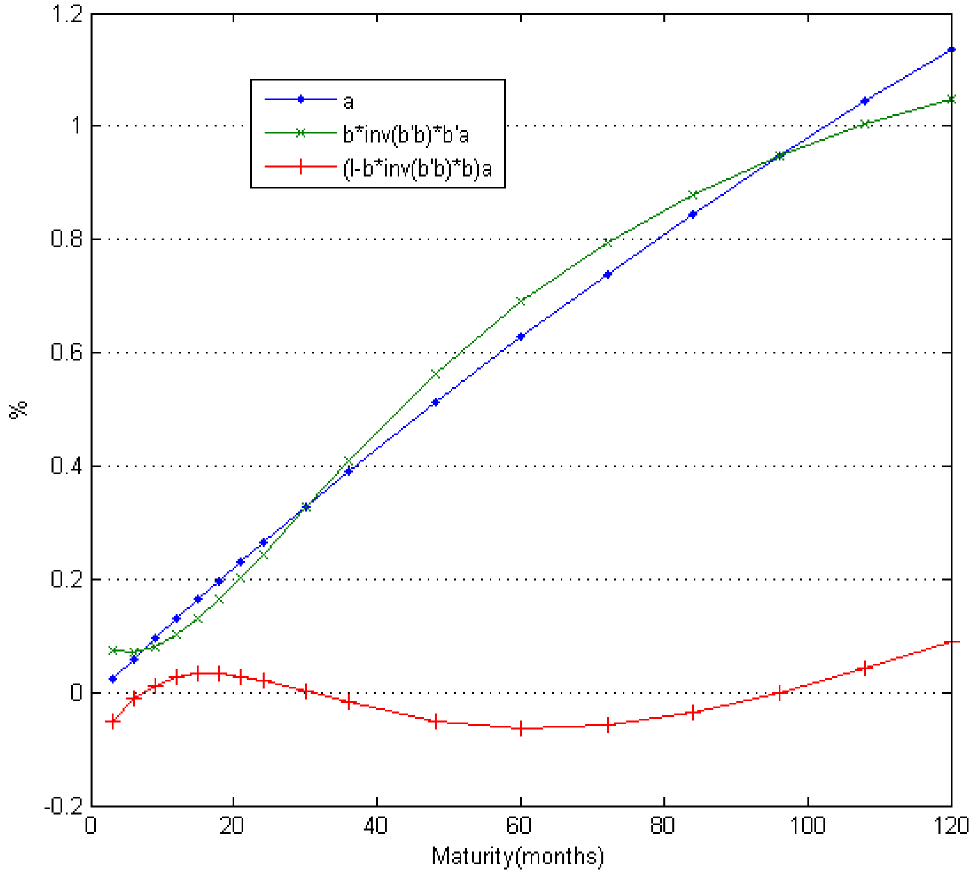


Figure 8. Fitting an AFNS model with a DNS model: the factor loadings

(λ freely estimated for the DNS)

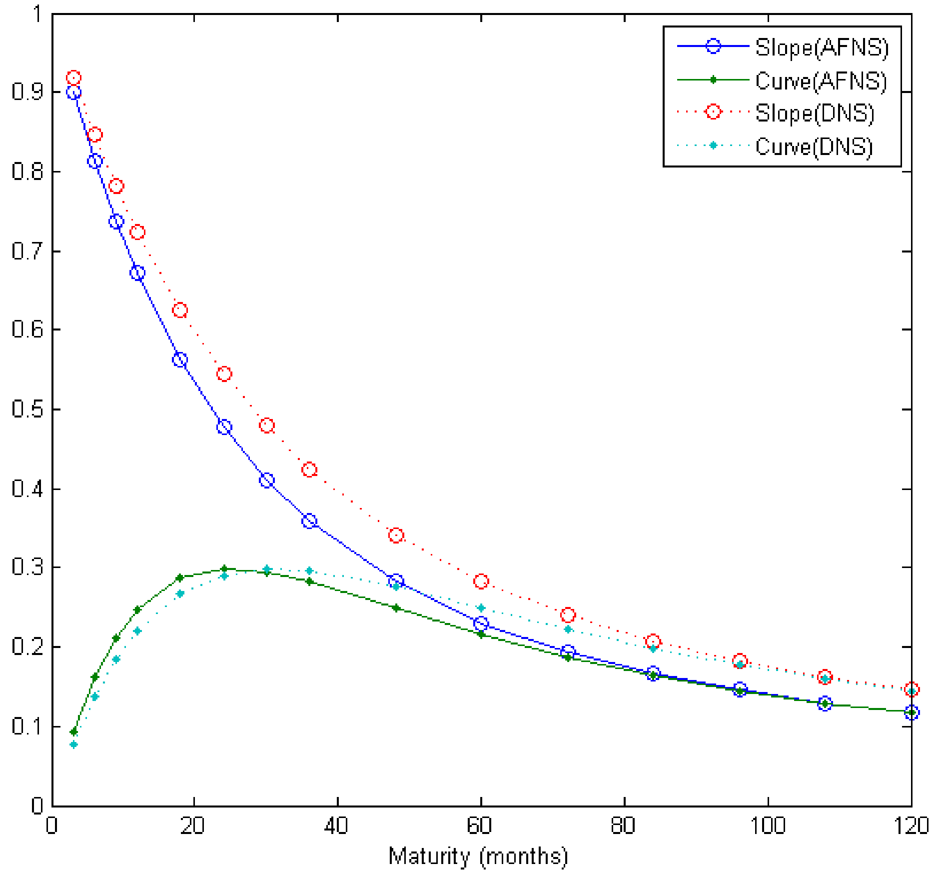


Figure 9. Risk prices of the three factors

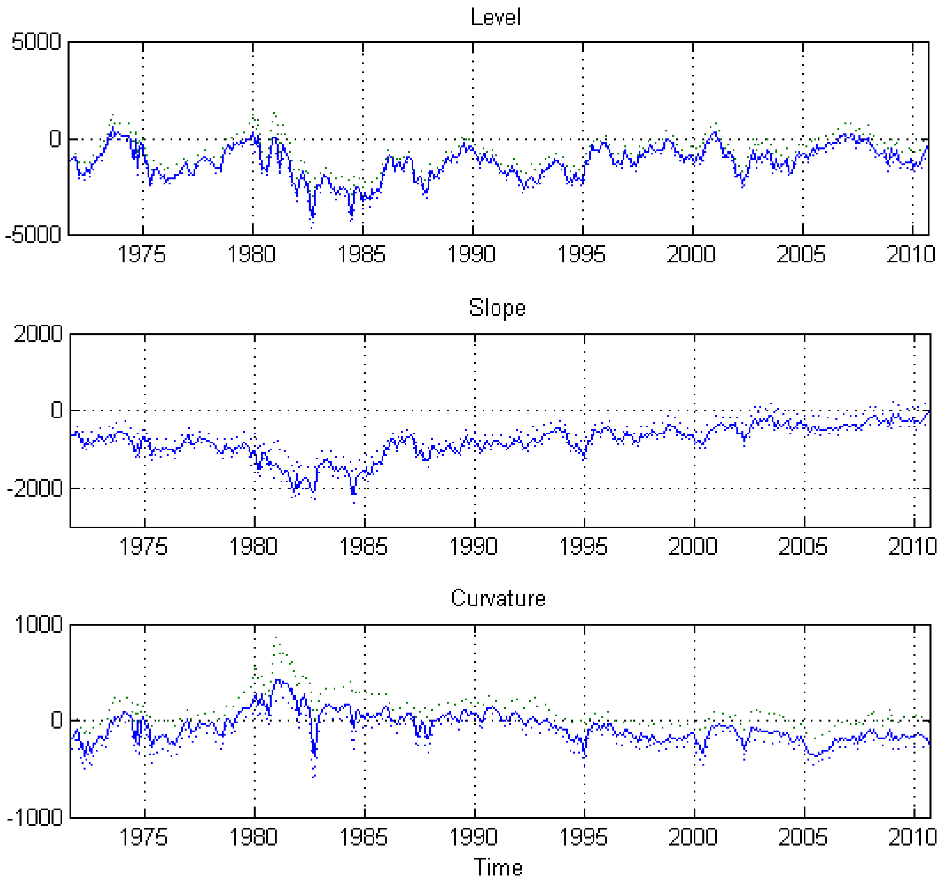


Figure 10. Factor loadings of risk premia

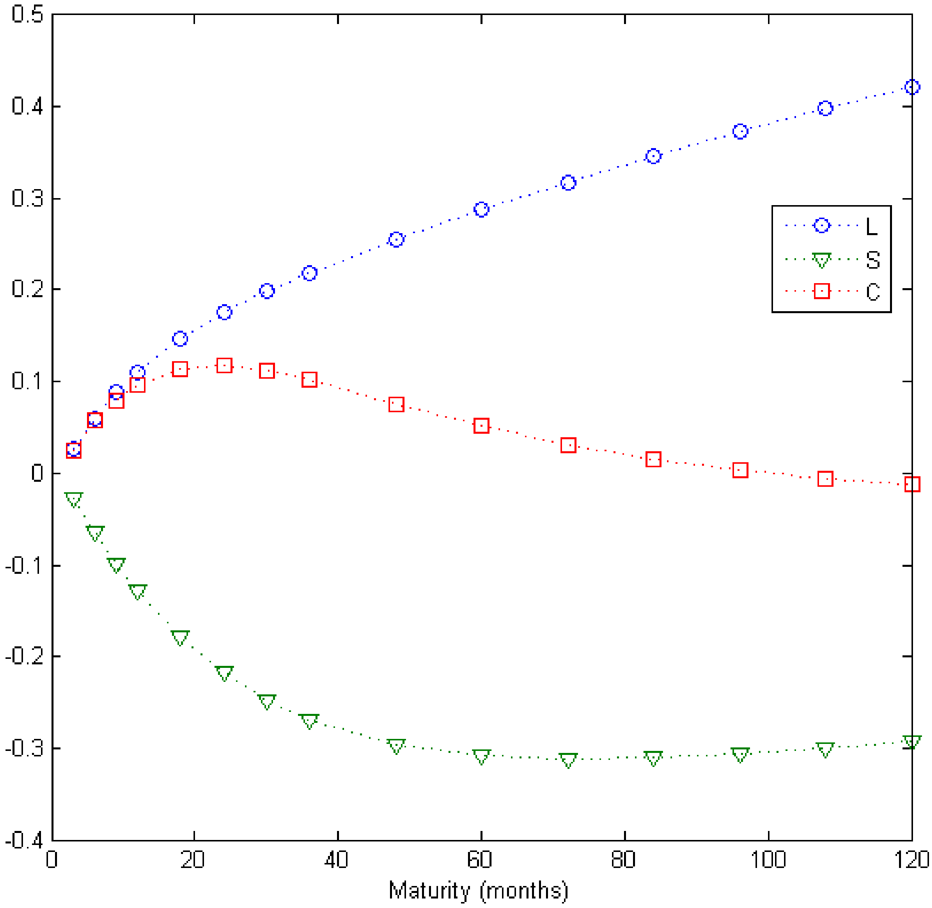
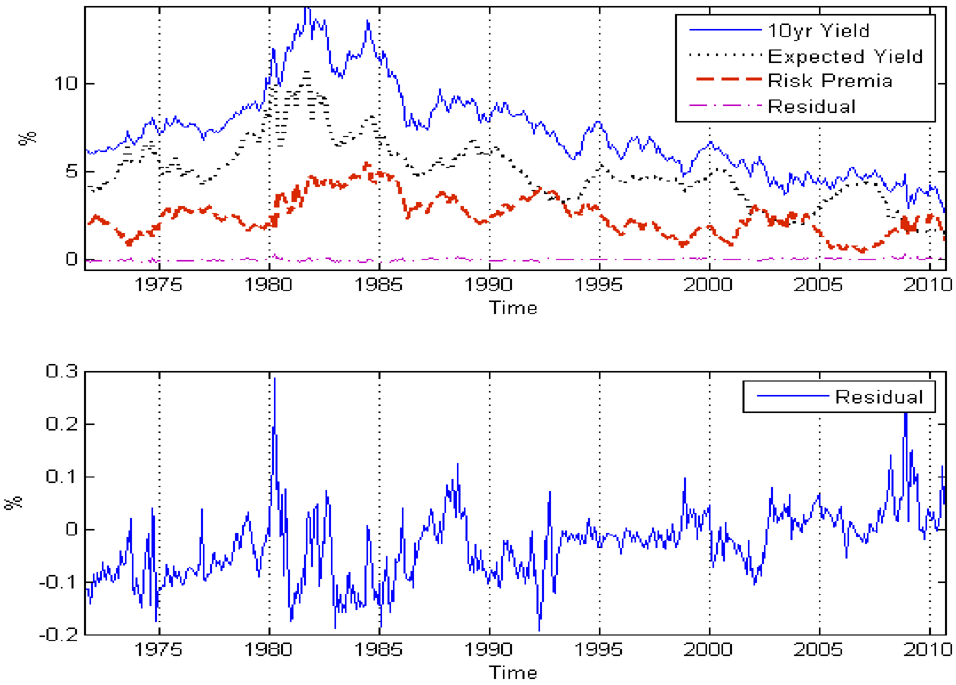


Figure 11. Decomposition of 10-year yield into expectation and risk premia

a. Full sample: 1971.8-2010.9



b. Subsample estimate: 2001.8-2010.9