An Arbitrage-Free Generalized Nelson-Siegel Term Structure Model

Jens H. E. Christensen
Federal Reserve Bank of San Francisco

Francis X. Diebold
University of Pennsylvania and NBER

Glenn D. Rudebusch
Federal Reserve Bank of San Francisco

An Arbitrage-Free Generalized Nelson-Siegel
Term Structure Model[†]

Jens H. E. Christensen
Federal Reserve Bank of San Francisco
jens.christensen@sf.frb.org

Francis X. Diebold
University of Pennsylvania and NBER
fdiebold@sas.upenn.edu

Glenn D. Rudebusch
Federal Reserve Bank of San Francisco
glenn.rudebusch@sf.frb.org

Abstract

The Svensson generalization of the popular Nelson-Siegel term structure model is widely used by practitioners and central banks. Unfortunately, like the original Nelson-Siegel specification, this generalization, in its dynamic form, does not enforce arbitrage-free consistency over time. Indeed, we show that the factor loadings of the Svensson generalization cannot be obtained in a standard finance arbitrage-free affine term structure representation. Therefore, we introduce a closely related generalized Nelson-Siegel model on which the no-arbitrage condition can be imposed. We estimate this new arbitrage-free generalized Nelson-Siegel model and demonstrate its tractability and good in-sample fit.

1 Introduction

To investigate yield-curve dynamics, researchers have produced a vast literature with a wide variety of models. Many of these models assume that at observed bond prices there are no remaining unexploited opportunities for riskless arbitrage. This theoretical assumption is consistent with the observation that bonds of various maturities all trade simultaneously in deep and liquid markets. Rational traders in such markets should enforce a consistency in the yields of various bonds across different maturities—the yield curve at any point in time—and the expected path of those yields over time—the dynamic evolution of the yield curve. Indeed, the assumption that there are no remaining arbitrage opportunities is central to the enormous finance literature devoted to the empirical analysis of bond pricing. Unfortunately, as noted by Duffee (2002), the associated arbitrage-free (AF) models demonstrate disappointing empirical performance, especially with regard to out-of-sample forecasting. In addition, the estimation of these models is problematic, in large part because of the existence of numerous model likelihood maxima that have essentially identical fit to the data but very different implications for economic behavior (Kim and Orphanides, 2005).[1]

In contrast to the popular finance arbitrage-free models, many other researchers have employed models that are empirically appealing but not well grounded in theory. Most notably, the Nelson-Siegel model provides a remarkably good fit to the cross section of yields in many countries and has become a widely used specification among financial market practitioners and central banks. Moreover, Diebold and Li (2006) develop a dynamic version of this model and show that it corresponds exactly to a modern factor model, with yields that are affine in three latent factors, which have a standard interpretation of level, slope, and curvature. Such a dynamic Nelson-Siegel (DNS) model is easy to estimate and forecasts the yield curve quite well. Despite its good empirical performance, however, the DNS model does not impose the presumably desirable theoretical restriction of absence of arbitrage (e.g., Filipović, 1999, and Diebold, Piazzesi, and Rudebusch, 2005).

In Christensen, Diebold, and Rudebusch (2007), henceforth CDR, we show how to reconcile the Nelson-Siegel model with the absence of arbitrage by deriving an affine AF model that maintains the Nelson-Siegel factor loading structure for the yield curve. This hybrid model, which we label the AFNS model, combines the best of both yield-curve modeling traditions. Although it maintains the theoretical restrictions of the affine AF modeling tradition, the Nelson-Siegel structure helps identify the latent yield-curve factors, so the AFNS model can be easily and robustly estimated. Furthermore, our results show that the AFNS model exhibits superior empirical forecasting performance.

In this paper, we consider some important generalizations of the Nelson-Siegel model that are also widely used in central banks and industry (e.g., De Pooter, 2007). Foremost among these is the Svensson (1995) extension to the Nelson-Siegel curve, which is used at the Federal Reserve Board (see Gürkaynak, Sack, and Wright, 2007), the European Central Bank (see Coroneo, Nyholm, and Vidova-Koleva, 2008), and many other central banks (see Bank for International Settlements,

---

[1] A further failing is that the affine arbitrage-free finance models offer little insight into the economic nature of the underlying forces that drive movements in interest rates. This issue has been addressed by a burgeoning macro-finance literature, which is described in Rudebusch and Wu (2007, 2008).

2005). The Svensson extension adds a second curvature term, which allows for a better fit at long maturities. Following Diebold and Li (2006), we first introduce a dynamic version of this model, the DNSS model, which corresponds to a modern four-factor term structure model. Unfortunately, we show that it is not possible to obtain an arbitrage-free "approximation" to this model in the sense of obtaining analytically identical factor loadings for the four factors. Intuitively, such an approximation requires that each curvature factor must be paired with a slope factor that has the same mean-reversion rate. This pairing is simply not possible for the Svensson extension, which has one slope factor and two curvature factors. Therefore, to obtain an arbitrage-free generalization of the Nelson-Siegel curve, we add a second slope factor to pair with the second curvature factor. The simple dynamic version of this model is denoted as the DGNS model, which is a generalized version of the DNS model. We also show that the result in CDR can be extended to obtain an arbitrage-free approximation to that five-factor model, which we refer to as the AFGNS model.

Finally, we show that this new AFGNS model of the yield curve not only displays theoretical consistency but also retains the important properties of empirical tractability and fit. We estimate the independent-factor versions of the four-factor and five-factor non-AF DNSS and DGNS models and the independent-factor version of the five-factor arbitrage-free AFGNS model. We compare the results to those obtained by CDR for the DNS and AFNS models and find remarkably good in-sample fit for the AFGNS model.

The remainder of the paper is structured as follows. Section 2 briefly describes the Nelson-Siegel model and its arbitrage-free equivalent as derived in CDR. Section 3 contains the description of the arbitrage-free generalized Nelson-Siegel model. Section 4 describes the five specific models that we analyze, while Section 5 describes the data, estimation method, and estimation results. Section 6 concludes the paper, and an appendix contains some additional technical details.

2   Nelson-Siegel term structure models

In this section, we review the DNS and AFNS models that maintain the Nelson-Siegel factor loading structure.

2.1   The dynamic Nelson-Siegel model

The Nelson-Siegel model fits the yield curve at any point in time with the simple functional form[2]

$$y(\tau) = \beta_0 + \beta_1\Big(\frac{1-e^{-\lambda\tau}}{\lambda\tau}\Big) + \beta_2\Big(\frac{1-e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau}\Big), \tag{1}$$

where $y(\tau)$ is the zero-coupon yield with τ denoting the time to maturity, and β_0, β_1, β_2, and λ are model parameters.

As many have noted, this model is able to provide a good fit to the cross section of yields at a given point in time, and this is a key reason for its popularity with financial market practitioners. Still, to understand the evolution of the bond market over time, a dynamic representation is required. Diebold and Li (2006) supply such a representation by replacing the parameters with

---

[2]This is equation (2) in Nelson and Siegel (1987).

time-varying factors.

$$y_t(\tau) = L_t + S_t\Big(\frac{1 - e^{-\lambda\tau}}{\lambda\tau}\Big) + C_t\Big(\frac{1 - e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau}\Big). \tag{2}$$

Given their associated Nelson-Siegel factor loadings, Diebold and Li show that L_t, S_t, and C_t can be interpreted as level, slope, and curvature factors. Furthermore, once the model is viewed as a factor model, a dynamic structure can be postulated for the three factors, which yields a fully dynamic version of the Nelson-Siegel model, henceforth referred to as the DNS model.

Despite its good empirical performance, however, the DNS model does not impose absence of arbitrage (e.g., Filipović, 1999, and Diebold, Piazzesi, and Rudebusch, 2005). This problem was solved in CDR, where we derived the affine arbitrage-free class of dynamic Nelson-Siegel term structure models, referred to as the AFNS model in the remainder of this paper.

2.2   The arbitrage-free Nelson-Siegel model

The derivation in CDR of the class of AFNS models starts from the standard continuous-time affine arbitrage-free term structure model. In this framework, we consider a three-factor model with a constant volatility matrix, that is, in the terminology of the canonical characterization of affine term structure models provided by Dai and Singleton (2000), we start with the $A_0(3)$ class of term structure models. Within the $A_0(3)$ class, CDR prove the following proposition.

Proposition 1:

Assume that the instantaneous risk-free rate is defined by

$$r_t = X_t^1 + X_t^2.$$

In addition, assume that the state variables $X_t = (X_t^1, X_t^2, X_t^3)$ are described by the following system of stochastic differential equations (SDEs)under the risk-neutral Q-measure:

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

Then, zero-coupon bond prices are given by

$$P(t,T) = E_t^Q\Big[\exp\Big(-\int_t^T r_u du\Big)\Big] = \exp\big(B^1(t,T)X_t^1 + B^2(t,T)X_t^2 + B^3(t,T)X_t^3 + C(t,T)\big),$$

where $B^1(t,T)$, $B^2(t,T)$, $B^3(t,T)$, and $C(t,T)$ are the unique solutions to the following system of ordinary differential equations (ODEs):

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

and

$$\frac{dC(t,T)}{dt} = -B(t,T)'K^Q\theta^Q - \frac{1}{2}\sum_{j=1}^{3}\left(\Sigma'B(t,T)B(t,T)'\Sigma\right)_{j,j}, \tag{4}$$

with boundary conditions $B^1(T,T) = B^2(T,T) = B^3(T,T) = C(T,T) = 0$. The unique solution for this system of ODEs is:

$$\begin{aligned} B^1(t,T) &= -(T-t), \\ B^2(t,T) &= -\frac{1-e^{-\lambda(T-t)}}{\lambda}, \\ B^3(t,T) &= (T-t)e^{-\lambda(T-t)} - \frac{1-e^{-\lambda(T-t)}}{\lambda}, \end{aligned}$$

and

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

Finally, zero-coupon bond yields are given by

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

Proof: See CDR.

This proposition defines the class of AFNS models. In this class of models, the factor loadings exactly match the Nelson-Siegel ones, but there is an unavoidable additional term in the yield function $-\frac{C(t,T)}{T-t}$, which depends only on the maturity of the bond. The nature of this "yield-adjustment" term is crucial in assessing differences between the AFNS and DNS models, and this yield-adjustment term will have the following form:[3]

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

---

[3]As explained in CDR, this form of the yield-adjustment term is obtained by fixing the mean parameters of the state variables under the Q-measure at zero, i.e., $\theta^Q = 0$, which implies no loss of generality.

Given a general volatility matrix

$$
\Sigma = \left(
\begin{array}{ccc}
\sigma_{11} & \sigma_{12} & \sigma_{13} \\
\sigma_{21} & \sigma_{22} & \sigma_{23} \\
\sigma_{31} & \sigma_{32} & \sigma_{33}
\end{array}
\right),
$$

the yield-adjustment term can be derived in analytical form as

$$
\begin{aligned}
\frac{C(t,T)}{T-t} &= \frac{1}{2}\frac{1}{T-t}\int_t^T \sum_{j=1}^3 \left(\Sigma' B(s,T)B(s,T)'\Sigma\right)_{j,j} ds \\
&= \overline{A}\frac{(T-t)^2}{6} + \overline{B}\Big[\frac{1}{2\lambda^2} - \frac{1}{\lambda^3}\frac{1-e^{-\lambda(T-t)}}{T-t} + \frac{1}{4\lambda^3}\frac{1-e^{-2\lambda(T-t)}}{T-t}\Big] \\
&+ \overline{C}\Big[\frac{1}{2\lambda^2} + \frac{1}{\lambda^2}e^{-\lambda(T-t)} - \frac{1}{4\lambda}(T-t)e^{-2\lambda(T-t)} - \frac{3}{4\lambda^2}e^{-2\lambda(T-t)} - \frac{2}{\lambda^3}\frac{1-e^{-\lambda(T-t)}}{T-t} + \frac{5}{8\lambda^3}\frac{1-e^{-2\lambda(T-t)}}{T-t}\Big] \\
&+ \overline{D}\Big[\frac{1}{2\lambda}(T-t) + \frac{1}{\lambda^2}e^{-\lambda(T-t)} - \frac{1}{\lambda^3}\frac{1-e^{-\lambda(T-t)}}{T-t}\Big] \\
&+ \overline{E}\Big[\frac{3}{\lambda^2}e^{-\lambda(T-t)} + \frac{1}{2\lambda}(T-t) + \frac{1}{\lambda}(T-t)e^{-\lambda(T-t)} - \frac{3}{\lambda^3}\frac{1-e^{-\lambda(T-t)}}{T-t}\Big] \\
&+ \overline{F}\Big[\frac{1}{\lambda^2} + \frac{1}{\lambda^2}e^{-\lambda(T-t)} - \frac{1}{2\lambda^2}e^{-2\lambda(T-t)} - \frac{3}{\lambda^3}\frac{1-e^{-\lambda(T-t)}}{T-t} + \frac{3}{4\lambda^3}\frac{1-e^{-2\lambda(T-t)}}{T-t}\Big],
\end{aligned}
$$

where

- $\overline{A} = \sigma_{11}^2 + \sigma_{12}^2 + \sigma_{13}^2$,

- $\overline{B} = \sigma_{21}^2 + \sigma_{22}^2 + \sigma_{23}^2$,

- $\overline{C} = \sigma_{31}^2 + \sigma_{32}^2 + \sigma_{33}^2$,

- $\overline{D} = \sigma_{11}\sigma_{21} + \sigma_{12}\sigma_{22} + \sigma_{13}\sigma_{23}$,

- $\overline{E} = \sigma_{11}\sigma_{31} + \sigma_{12}\sigma_{32} + \sigma_{13}\sigma_{33}$,

- $\overline{F} = \sigma_{21}\sigma_{31} + \sigma_{22}\sigma_{32} + \sigma_{23}\sigma_{33}$.

This result has two implications. First, the fact that zero-coupon bond yields in the AFNS class of models are given by an analytical formula greatly facilitates empirical implementation of these models. Second, the nine underlying volatility parameters are not identified. Indeed, only the six terms $\overline{A}$, $\overline{B}$, $\overline{C}$, $\overline{D}$, $\overline{E}$, and $\overline{F}$ can be identified; thus, the maximally flexible AFNS specification that can be identified has a triangular volatility matrix given by[4]

$$
\Sigma = \left(
\begin{array}{ccc}
\sigma_{11} & 0 & 0 \\
\sigma_{21} & \sigma_{22} & 0 \\
\sigma_{31} & \sigma_{32} & \sigma_{33}
\end{array}
\right).
$$

3 Extensions of the Nelson-Siegel model

The main in-sample problem with the regular Nelson-Siegel model is that, for reasonable choices of λ (which are empirically in the range from 0.5 to 1 for U.S. Treasury yield data), the factor

---

[4]The choice of upper or lower triangular is irrelevant for the fit of the model.

(a) Factor loadings in the DNSS model.   (b) Factor loadings in the DGNS model.
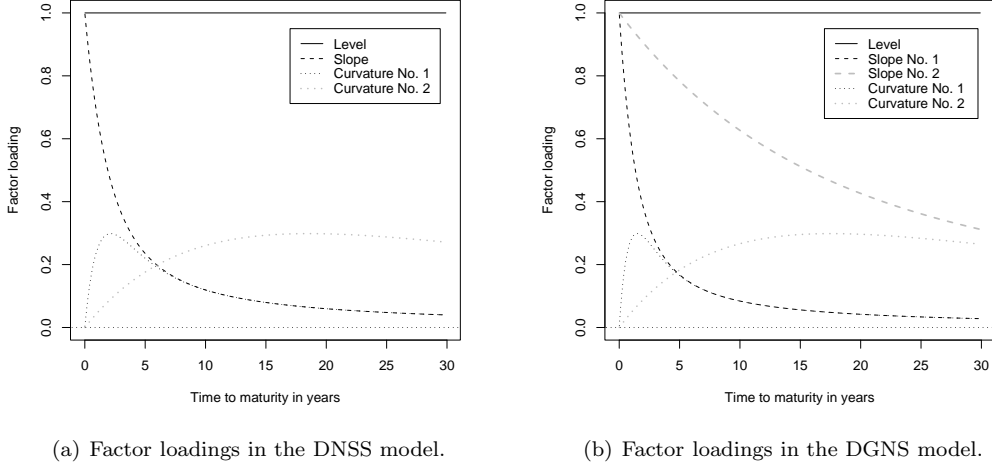
Figure 1: **Factor Loadings in the Yield Functions of the DNSS and DGNS Models.**
The left-hand figure shows the factor loadings of the four state variables in the yield function of the DNSS model with λ_1 and λ_2 equal to 0.8379 and 0.09653, respectively. The right-hand figure shows the factor loadings of the five state variables in the yield function of the DGNS model with λ_1 and λ_2 equal to 1.190 and 0.1021, respectively. These λ_i values equal the estimated values obtained below, and they require maturity to be measured in years.

loading for the slope and the curvature factor decay rapidly to zero as a function of maturity. Thus, the model only has the level factor to fit yields with maturities of ten years or longer. In empirical estimation this limitation shows up as a lack of fit of the long-term yields, as described in CDR.

To address this problem in fitting the cross section of yields, Svensson (1995) introduced an extended version of the Nelson-Siegel yield curve with an additional curvature factor,

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

Just as Diebold and Li (2006) replaced the three β coefficients with dynamic factors in the regular Nelson-Siegel model, we can replace the four β coefficients in the Svensson model with dynamic processes (L_t, S_t, C_t^1, C_t^2) interpreted as a level, a slope, and two curvature factors, respectively. Thus, the dynamic version of the Svensson yield curve, which we label the DNSS model, is given by

$$y_t(\tau) = L_t + S_t\Big(\frac{1 - e^{-\lambda_1\tau}}{\lambda_1\tau}\Big) + C_t^1\Big(\frac{1 - e^{-\lambda_1\tau}}{\lambda_1\tau} - e^{-\lambda_1\tau}\Big) + C_t^2\Big(\frac{1 - e^{-\lambda_2\tau}}{\lambda_2\tau} - e^{-\lambda_2\tau}\Big).$$

The factor loadings of the four state variables in the yield function of the DNSS model are illustrated in Figure 1(a) with λ_1 and λ_2 set equal to our estimates described in Section 5.

The critique raised by Filipović (1999) against the dynamic version of the Nelson-Siegel model also applies to the dynamic version of the Svensson model introduced in this paper. Thus, this model is not consistent with the concept of absence of arbitrage. Ideally, we would like to repeat the work in CDR and derive an arbitrage-free approximation to the Svensson model described earlier. However, from the mechanics of Proposition 1 for the arbitrage-free approximation of the

regular Nelson-Siegel model, it is clear that we can only obtain the specific Nelson-Siegel factor loading structure for the slope and curvature factors under two specific conditions. First, each pair of slope and curvature factors must have identical own mean-reversion rates. Second, the impact of deviations in the curvature factor from its mean on the slope factor must be scaled with a factor equal to that own mean-reversion rate (λ). Thus, it is impossible in an arbitrage-free model to generate the factor loading structure of two curvature factors with only one slope factor. Consequently, it is impossible to create an arbitrage-free version of the Svensson extension to the Nelson-Siegel model that has factor loadings analytically identical to the ones in the DNSS model.

However, this discussion suggests that we can create a generalized AF Nelson-Siegel model by including a fifth factor in the form of a second slope factor. The yield function of this model takes the form

$$y_t(\tau) = L_t + S_t^1\Big(\frac{1-e^{-\lambda_1\tau}}{\lambda_1\tau}\Big) + S_t^2\Big(\frac{1-e^{-\lambda_2\tau}}{\lambda_2\tau}\Big) + C_t^1\Big(\frac{1-e^{-\lambda_1\tau}}{\lambda_1\tau} - e^{-\lambda_1\tau}\Big) + C_t^2\Big(\frac{1-e^{-\lambda_2\tau}}{\lambda_2\tau} - e^{-\lambda_2\tau}\Big).$$

This dynamic generalized Nelson-Siegel model, which we denote as the DGNS model, is a five-factor model with one level factor, two slope factors, and two curvature factors. (Note that we impose the restriction that $\lambda_1 > \lambda_2$, which is nonbinding due to symmetry.[5]) The factor loadings of the five state variables in the yield function of the DGNS model are illustrated in Figure 1(b) with λ_1 and λ_2 set equal to our estimates in Section 5.

A straightforward extension of Proposition 1 delivers the arbitrage-free approximation of this model, which we denote as the AFGNS model.

Proposition 2:

Assume that the instantaneous risk-free rate is defined by

$$r_t = X_t^1 + X_t^2 + X_t^3.$$

In addition, assume that the state variables $X_t = (X_t^1, X_t^2, X_t^3, X_t^4, X_t^5)$ are described by the following system of SDEs under the risk-neutral Q-measure:

$$
\begin{pmatrix} dX_t^1 \\ dX_t^2 \\ dX_t^3 \\ dX_t^4 \\ dX_t^5 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & \lambda_1 & 0 & -\lambda_1 & 0 \\ 0 & 0 & \lambda_2 & 0 & -\lambda_2 \\ 0 & 0 & 0 & \lambda_1 & 0 \\ 0 & 0 & 0 & 0 & \lambda_2 \end{pmatrix} \Bigg[\begin{pmatrix} \theta_1^Q \\ \theta_2^Q \\ \theta_3^Q \\ \theta_4^Q \\ \theta_5^Q \end{pmatrix} - \begin{pmatrix} X_t^1 \\ X_t^2 \\ X_t^3 \\ X_t^4 \\ X_t^5 \end{pmatrix} \Bigg] dt + \Sigma \begin{pmatrix} dW_t^{1,Q} \\ dW_t^{2,Q} \\ dW_t^{3,Q} \\ dW_t^{4,Q} \\ dW_t^{5,Q} \end{pmatrix}, \quad \lambda_1 > \lambda_2 > 0.
$$

Then, zero-coupon bond prices are given by

$$
\begin{aligned}
P(t,T) &= E_t^Q\Big[\exp\Big(-\int_t^T r_u du\Big)\Big] \\
&= \exp\big(B^1(t,T)X_t^1 + B^2(t,T)X_t^2 + B^3(t,T)X_t^3 + B^4(t,T)X_t^4 + B^5(t,T)X_t^5 + C(t,T)\big),
\end{aligned}
$$

---

[5]Björk and Christensen (1999) introduce a related extension of the Nelson-Siegel model with one level factor, two slope factors, and a single curvature factor with the restriction that $\lambda_1 = 2\lambda_2$.

where $B^1(t,T)$, $B^2(t,T)$, $B^3(t,T)$, $B^4(t,T)$, $B^5(t,T)$, and $C(t,T)$ are the unique solutions to the following system of ODEs:

$$\begin{pmatrix} \frac{dB^1(t,T)}{dt} \\ \frac{dB^2(t,T)}{dt} \\ \frac{dB^3(t,T)}{dt} \\ \frac{dB^4(t,T)}{dt} \\ \frac{dB^5(t,T)}{dt} \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & \lambda_1 & 0 & 0 & 0 \\ 0 & 0 & \lambda_2 & 0 & 0 \\ 0 & -\lambda_1 & 0 & \lambda_1 & 0 \\ 0 & 0 & -\lambda_2 & 0 & \lambda_2 \end{pmatrix} \begin{pmatrix} B^1(t,T) \\ B^2(t,T) \\ B^3(t,T) \\ B^4(t,T) \\ B^5(t,T) \end{pmatrix} \qquad (5)$$

and

$$\frac{dC(t,T)}{dt} = -B(t,T)'K^Q\theta^Q - \frac{1}{2}\sum_{j=1}^{5}\left(\Sigma'B(t,T)B(t,T)'\Sigma\right)_{j,j}, \qquad (6)$$

with boundary conditions $B^1(T,T) = B^2(T,T) = B^3(T,T) = B^4(T,T) = B^5(T,T) = C(T,T) = 0$. The unique solution for this system of ODEs is:

$$\begin{aligned} B^1(t,T) &= -(T-t), \\ B^2(t,T) &= -\frac{1-e^{-\lambda_1(T-t)}}{\lambda_1}, \\ B^3(t,T) &= -\frac{1-e^{-\lambda_2(T-t)}}{\lambda_2}, \\ B^4(t,T) &= (T-t)e^{-\lambda_1(T-t)} - \frac{1-e^{-\lambda_1(T-t)}}{\lambda_1}, \\ B^5(t,T) &= (T-t)e^{-\lambda_2(T-t)} - \frac{1-e^{-\lambda_2(T-t)}}{\lambda_2}, \end{aligned}$$

and

$$\begin{aligned} C(t,T) &= (K^Q\theta^Q)_2\int_t^T B^2(s,T)ds + (K^Q\theta^Q)_3\int_t^T B^3(s,T)ds + (K^Q\theta^Q)_4\int_t^T B^4(s,T)ds + (K^Q\theta^Q)_5\int_t^T B^5(s,T)ds \\ &\quad + \frac{1}{2}\sum_{j=1}^{5}\int_t^T \left(\Sigma'B(s,T)B(s,T)'\Sigma\right)_{j,j}ds. \end{aligned}$$

Finally, zero-coupon bond yields are given by

$$y(t,T) = X_t^1 + \frac{1-e^{-\lambda_1(T-t)}}{\lambda_1(T-t)}X_t^2 + \frac{1-e^{-\lambda_2(T-t)}}{\lambda_2(T-t)}X_t^3 + \left[\frac{1-e^{-\lambda_1(T-t)}}{\lambda_1(T-t)} - e^{-\lambda_1(T-t)}\right]X_t^4 + \left[\frac{1-e^{-\lambda_2(T-t)}}{\lambda_2(T-t)} - e^{-\lambda_2(T-t)}\right]X_t^5 - \frac{C(t,T)}{T-t}.$$

Proof: Straightforward extension of CDR.

Similar to the AFNS class of models, the yield-adjustment term will have the following form:[6]

$$-\frac{C(t,T)}{T-t} = -\frac{1}{2}\frac{1}{T-t}\sum_{j=1}^{5}\int_t^T \left(\Sigma'B(s,T)B(s,T)'\Sigma\right)_{j,j}ds.$$

---

[6]The analytical formula for the yield-adjustment term in the AFGNS model is provided in Appendix A. As was the case for Proposition 1, Proposition 2 is also silent about the P-dynamics of the state variables, so to identify the model, we follow CDR and fix the mean under the Q-measure at zero, i.e., $\theta^Q = 0$.

Following arguments similar to the ones provided for the AFNS class of models in the previous section, the maximally flexible specification of the volatility matrix that can be identified in estimation is given by a triangular matrix

$$\Sigma = \begin{pmatrix} \sigma_{11} & 0 & 0 & 0 & 0 \\ \sigma_{21} & \sigma_{22} & 0 & 0 & 0 \\ \sigma_{31} & \sigma_{32} & \sigma_{33} & 0 & 0 \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_{44} & 0 \\ \sigma_{51} & \sigma_{52} & \sigma_{53} & \sigma_{54} & \sigma_{55} \end{pmatrix}.$$

4 Five specific Nelson-Siegel models

In general, all the models considered in this paper are silent about the P-dynamics, an infinite number of possible specifications could be used to match the data. However, for continuity with the existing literature, our econometric analysis focuses on five independent-factor versions of the models we have described. These models include the DNS and AFNS models from CDR and the generalized DNSS, DGNS, and AFGNS models introduced in Section 3.

In the *independent-factor DNS model*, all three state variables are assumed to be independent first-order autoregressions, as in Diebold and Li (2006). Using their notation, the state equation is given by

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

where the error terms $\eta_t(L)$, $\eta_t(S)$, and $\eta_t(C)$ have a conditional covariance matrix given by

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

In this model, the measurement equation takes the form

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

where the measurement errors $\varepsilon_t(\tau_i)$ are assumed to be independently and identically distributed (i.i.d.) white noise.

The corresponding AFNS model is formulated in continuous time and the relationship between the real-world dynamics under the P-measure and the risk-neutral dynamics under the Q-measure is given by the measure change

$$dW_t^Q = dW_t^P + \Gamma_t dt,$$

where Γ_t represents the risk premium specification. To preserve affine dynamics under the P-measure, we limit our focus to essentially affine risk premium specifications (see Duffee, 2002). Thus, Γ_t will take the form

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

With this specification, the SDE for the state variables under the P-measure,

$$
dX_t = K^P[\theta^P - X_t]dt + \Sigma dW_t^P, \tag{7}
$$

remains affine. Due to the flexible specification of Γ_t, we are free to choose any mean vector θ^P and mean-reversion matrix K^P under the P-measure and still preserve the required Q-dynamic structure described in Proposition 1. Therefore, we focus on the *independent-factor AFNS model*, which corresponds to the specific DNS model from earlier in this section and assumes all three factors are independent under the P-measure

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

In this case, the measurement equation takes the form

$$
\begin{pmatrix} y_t(\tau_1) \\ y_t(\tau_2) \\ \vdots \\ y_t(\tau_N) \end{pmatrix} = \begin{pmatrix} 1 & \frac{1-e^{-\lambda\tau_1}}{\lambda\tau_1} & \frac{1-e^{-\lambda\tau_1}}{\lambda\tau_1} - e^{-\lambda\tau_1} \\ 1 & \frac{1-e^{-\lambda\tau_2}}{\lambda\tau_2} & \frac{1-e^{-\lambda\tau_2}}{\lambda\tau_2} - e^{-\lambda\tau_2} \\ \vdots & \vdots & \vdots \\ 1 & \frac{1-e^{-\lambda\tau_N}}{\lambda\tau_N} & \frac{1-e^{-\lambda\tau_N}}{\lambda\tau_N} - e^{-\lambda\tau_N} \end{pmatrix} \begin{pmatrix} X_t^1 \\ X_t^2 \\ X_t^3 \end{pmatrix} - \begin{pmatrix} \frac{C(\tau_1)}{\tau^1} \\ \frac{C(\tau_2)}{\tau^2} \\ \vdots \\ \frac{C(\tau_N)}{\tau^N} \end{pmatrix} + \begin{pmatrix} \varepsilon_t(\tau_1) \\ \varepsilon_t(\tau_2) \\ \vdots \\ \varepsilon_t(\tau_N) \end{pmatrix},
$$

where, again, the measurement errors $\varepsilon_t(\tau_i)$ are assumed to be i.i.d. white noise.

We now turn to the three generalized Nelson-Siegel models. In the *independent-factor DNSS model*, all four state variables are assumed to be independent first-order autoregressions, as in Diebold and Li (2006). Using their notation, the state equation is given by

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

where the error terms $\eta_t(L)$, $\eta_t(S)$, $\eta_t(C^1)$, and $\eta_t(C^2)$ have a conditional covariance matrix given

10

by

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

In the DNSS model, the measurement equation takes the form

$$\begin{pmatrix} y_t(\tau_1) \\ y_t(\tau_2) \\ \vdots \\ y_t(\tau_N) \end{pmatrix} = \begin{pmatrix} 1 & \frac{1-e^{-\lambda_1\tau_1}}{\lambda_1\tau_1} & \frac{1-e^{-\lambda_1\tau_1}}{\lambda_1\tau_1}-e^{-\lambda_1\tau_1} & \frac{1-e^{-\lambda_2\tau_1}}{\lambda_2\tau_1}-e^{-\lambda_2\tau_1} \\ 1 & \frac{1-e^{-\lambda_1\tau_2}}{\lambda_1\tau_2} & \frac{1-e^{-\lambda_1\tau_2}}{\lambda_1\tau_2}-e^{-\lambda_1\tau_2} & \frac{1-e^{-\lambda_2\tau_2}}{\lambda_2\tau_2}-e^{-\lambda_2\tau_2} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & \frac{1-e^{-\lambda_1\tau_N}}{\lambda_1\tau_N} & \frac{1-e^{-\lambda_1\tau_N}}{\lambda_1\tau_N}-e^{-\lambda_1\tau_N} & \frac{1-e^{-\lambda_2\tau_N}}{\lambda_2\tau_N}-e^{-\lambda_2\tau_N} \end{pmatrix} \begin{pmatrix} L_t \\ S_t \\ C_t^1 \\ C_t^2 \end{pmatrix} + \begin{pmatrix} \varepsilon_t(\tau_1) \\ \varepsilon_t(\tau_2) \\ \vdots \\ \varepsilon_t(\tau_N) \end{pmatrix},$$

where the measurement errors $\varepsilon_t(\tau_i)$ are assumed to be i.i.d. white noise.

In the *independent-factor DGNS model*, all five state variables are assumed to be independent first-order autoregressions, and the state equation is given by

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

where the error terms $\eta_t(L)$, $\eta_t(S^1)$, $\eta_t(S^2)$, $\eta_t(C^1)$, and $\eta_t(C^2)$ have a conditional covariance matrix given by

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

In the DGNS model, the measurement equation takes the form

$$\begin{pmatrix} y_t(\tau_1) \\ y_t(\tau_2) \\ \vdots \\ y_t(\tau_N) \end{pmatrix} = \begin{pmatrix} 1 & \frac{1-e^{-\lambda_1\tau_1}}{\lambda_1\tau_1} & \frac{1-e^{-\lambda_2\tau_1}}{\lambda_2\tau_1} & \frac{1-e^{-\lambda_1\tau_1}}{\lambda_1\tau_1}-e^{-\lambda_1\tau_1} & \frac{1-e^{-\lambda_2\tau_1}}{\lambda_2\tau_1}-e^{-\lambda_2\tau_1} \\ 1 & \frac{1-e^{-\lambda_1\tau_2}}{\lambda_1\tau_2} & \frac{1-e^{-\lambda_2\tau_2}}{\lambda_2\tau_2} & \frac{1-e^{-\lambda_1\tau_2}}{\lambda_1\tau_2}-e^{-\lambda_1\tau_2} & \frac{1-e^{-\lambda_2\tau_2}}{\lambda_2\tau_2}-e^{-\lambda_2\tau_2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \frac{1-e^{-\lambda_1\tau_N}}{\lambda_1\tau_N} & \frac{1-e^{-\lambda_2\tau_N}}{\lambda_2\tau_N} & \frac{1-e^{-\lambda_1\tau_N}}{\lambda_1\tau_N}-e^{-\lambda_1\tau_N} & \frac{1-e^{-\lambda_2\tau_N}}{\lambda_2\tau_N}-e^{-\lambda_2\tau_N} \end{pmatrix} \begin{pmatrix} L_t \\ S_t^1 \\ S_t^2 \\ C_t^1 \\ C_t^2 \end{pmatrix} + \begin{pmatrix} \varepsilon_t(\tau_1) \\ \varepsilon_t(\tau_2) \\ \vdots \\ \varepsilon_t(\tau_N) \end{pmatrix},$$

where the measurement errors $\varepsilon_t(\tau_i)$ are assumed to be i.i.d. white noise.

Finally, as for the AFNS model, the AFGNS model is formulated in continuous time and the relationship between the real-world dynamics under the P-measure and the risk-neutral dynamics under the Q-measure is given by the measure change

$$dW_t^Q = dW_t^P + \Gamma_t dt,$$

11

where Γ_t represents the risk premium specification. Again, to preserve affine dynamics under the P-measure, we limit our focus to essentially affine risk premium specifications (see Duffee, 2002). Thus, Γ_t takes the form

$$
\Gamma_t = \begin{pmatrix} \gamma_1^0 \\ \gamma_2^0 \\ \gamma_3^0 \\ \gamma_4^0 \\ \gamma_5^0 \end{pmatrix} + \begin{pmatrix} \gamma_{11}^1 & \gamma_{12}^1 & \gamma_{13}^1 & \gamma_{14}^1 & \gamma_{15}^1 \\ \gamma_{21}^1 & \gamma_{22}^1 & \gamma_{23}^1 & \gamma_{24}^1 & \gamma_{25}^1 \\ \gamma_{31}^1 & \gamma_{32}^1 & \gamma_{33}^1 & \gamma_{34}^1 & \gamma_{35}^1 \\ \gamma_{41}^1 & \gamma_{42}^1 & \gamma_{43}^1 & \gamma_{44}^1 & \gamma_{45}^1 \\ \gamma_{51}^1 & \gamma_{52}^1 & \gamma_{53}^1 & \gamma_{54}^1 & \gamma_{55}^1 \end{pmatrix} \begin{pmatrix} X_t^1 \\ X_t^2 \\ X_t^3 \\ X_t^4 \\ X_t^5 \end{pmatrix}.
$$

With this specification, the SDE for the state variables under the P-measure,

$$
dX_t = K^P[\theta^P - X_t]dt + \Sigma dW_t^P, \tag{8}
$$

remains affine. Due to the flexible specification of Γ_t, we are free to choose any mean vector θ^P and mean-reversion matrix K^P under the P-measure and still preserve the required structure for the Q-dynamics described in Proposition 2. Therefore, we focus on the AFGNS model that corresponds to the specific DGNS model we have described earlier. In this *independent-factor AFGNS model*, all five factors are assumed to be independent under the P-measure

$$
\begin{pmatrix} dX_t^1 \\ dX_t^2 \\ dX_t^3 \\ dX_t^4 \\ dX_t^5 \end{pmatrix} = \begin{pmatrix} \kappa_{11}^P & 0 & 0 & 0 & 0 \\ 0 & \kappa_{22}^P & 0 & 0 & 0 \\ 0 & 0 & \kappa_{33}^P & 0 & 0 \\ 0 & 0 & 0 & \kappa_{44}^P & 0 \\ 0 & 0 & 0 & 0 & \kappa_{55}^P \end{pmatrix} \left[\begin{pmatrix} \theta_1^P \\ \theta_2^P \\ \theta_3^P \\ \theta_4^P \\ \theta_5^P \end{pmatrix} - \begin{pmatrix} X_t^1 \\ X_t^2 \\ X_t^3 \\ X_t^4 \\ X_t^5 \end{pmatrix} \right] dt + \begin{pmatrix} \sigma_1 & 0 & 0 & 0 & 0 \\ 0 & \sigma_2 & 0 & 0 & 0 \\ 0 & 0 & \sigma_3 & 0 & 0 \\ 0 & 0 & 0 & \sigma_4 & 0 \\ 0 & 0 & 0 & 0 & \sigma_5 \end{pmatrix} \begin{pmatrix} dW_t^{1,P} \\ dW_t^{2,P} \\ dW_t^{3,P} \\ dW_t^{4,P} \\ dW_t^{5,P} \end{pmatrix}.
$$

For the AFGNS model, the measurement equation takes the form

$$
\begin{pmatrix} y_t(\tau_1) \\ y_t(\tau_2) \\ \vdots \\ y_t(\tau_N) \end{pmatrix} = \begin{pmatrix} 1 & \frac{1-e^{-\lambda_1\tau_1}}{\lambda_1\tau_1} & \frac{1-e^{-\lambda_2\tau_1}}{\lambda_2\tau_1} & \frac{1-e^{-\lambda_1\tau_1}}{\lambda_1\tau_1} - e^{-\lambda_1\tau_1} & \frac{1-e^{-\lambda_2\tau_1}}{\lambda_2\tau_1} - e^{-\lambda_2\tau_1} \\ 1 & \frac{1-e^{-\lambda_1\tau_2}}{\lambda_1\tau_2} & \frac{1-e^{-\lambda_2\tau_2}}{\lambda_2\tau_2} & \frac{1-e^{-\lambda_1\tau_2}}{\lambda_1\tau_2} - e^{-\lambda_1\tau_2} & \frac{1-e^{-\lambda_2\tau_2}}{\lambda_2\tau_2} - e^{-\lambda_2\tau_2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \frac{1-e^{-\lambda_1\tau_N}}{\lambda_1\tau_N} & \frac{1-e^{-\lambda_2\tau_N}}{\lambda_2\tau_N} & \frac{1-e^{-\lambda_1\tau_N}}{\lambda_1\tau_N} - e^{-\lambda_1\tau_N} & \frac{1-e^{-\lambda_2\tau_N}}{\lambda_2\tau_N} - e^{-\lambda_2\tau_N} \end{pmatrix} \begin{pmatrix} X_t^1 \\ X_t^2 \\ X_t^3 \\ X_t^4 \\ X_t^5 \end{pmatrix} - \begin{pmatrix} \frac{C(\tau_1)}{\tau_1} \\ \frac{C(\tau_2)}{\tau_2} \\ \vdots \\ \frac{C(\tau_N)}{\tau_N} \end{pmatrix} + \begin{pmatrix} \varepsilon_t(\tau_1) \\ \varepsilon_t(\tau_2) \\ \vdots \\ \varepsilon_t(\tau_N) \end{pmatrix},
$$

where, again, the measurement errors $\varepsilon_t(\tau_i)$ are assumed to be i.i.d. white noise.

5   Estimation of the models

In this section, we will first describe the interest rate data to be used and the estimation method. Next, we examine estimation results and in-sample fit for the DNS, AFNS, DNSS, DGNS, and AFGNS models.

5.1   Data

Our data are monthly unsmoothed Fama-Bliss zero-coupon yields covering the period from January 1987 to December 2002 and described in CDR. The maturities range from three months to 30

Maturity	Mean	St.dev.	Skewness	Kurtosis
3	0.0509	0.0174	-0.0598	2.8199
6	0.0522	0.0175	-0.1400	2.7892
9	0.0533	0.0176	-0.1681	2.7474
12	0.0548	0.0177	-0.1960	2.7663
18	0.0570	0.0173	-0.1951	2.7605
24	0.0581	0.0166	-0.1797	2.7415
36	0.0606	0.0155	-0.1160	2.6952
48	0.0626	0.0148	-0.0829	2.5919
60	0.0636	0.0144	-0.0196	2.4418
84	0.0660	0.0138	0.0465	2.2071
96	0.0670	0.0136	0.0610	2.1290
108	0.0674	0.0136	0.0638	2.0617
120	0.0674	0.0135	0.0618	1.9843
180	0.0716	0.0123	0.2130	1.8874
240	0.0725	0.0113	0.0760	1.7757
360	0.0677	0.0121	0.0589	1.7428

Table 1: **Summary Statistics for U.S. Treasury Yields.**
The summary statistics for our sample of monthly observed unsmoothed Fama-Bliss zero-coupon Treasury bond yields, which covers the period from January 1987 to December 2002.
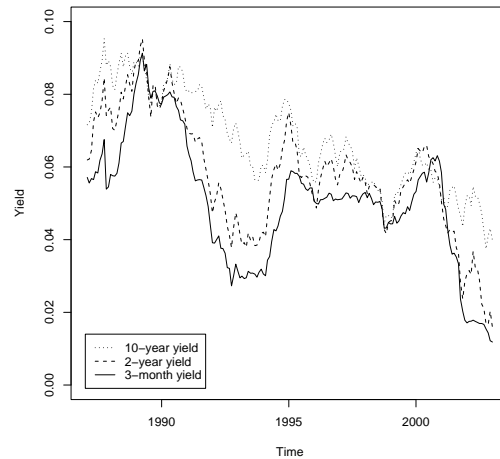

Figure 2: **Time series of U.S. Treasury Yields.**
Illustration of the observed Treasury zero-coupon bond yields covering the period from January 1987 to December 2002. The yields shown have three-month, two-year, and ten-year maturities.

years. Summary statistics are provided in Table 1, while Figure 2 illustrates the time series for the three-month, two-year, and ten-year yields.

5.2   Estimation methods

The Kalman filter is an efficient and consistent estimator for all five models. For the DNS, DNSS, and DGNS models, the state equation is

$$X_t = (I - A)\mu + AX_{t-1} + \eta_t, \quad \eta_t \sim N(0, Q),$$

where $X_t = (L_t, S_t, C_t)$, $X_t = (L_t, S_t, C_t^1, C_t^2)$, and $X_t = (L_t, S_t^1, S_t^2, C_t^1, C_t^2)$, respectively, while the measurement equation is given by

$$y_t = BX_t + \varepsilon_t.$$

Following Diebold, Rudebusch, and Aruoba (2006), we start the algorithm at the unconditional mean and variance of the state variables. This assumes the state variables are stationary, which is imposed with the constraint that the eigenvalues of A are smaller than 1.

For the continuous-time AFNS and AFGNS models, the conditional mean vector and the conditional covariance matrix are given by

$$
\begin{aligned}
E^P[X_T|\mathcal{F}_t] &= (I - \exp(-K^P \Delta t))\theta^P + \exp(-K^P \Delta t)X_t, \\
V^P[X_T|\mathcal{F}_t] &= \int_0^{\Delta t} e^{-K^P s} \Sigma\Sigma' e^{-(K^P)' s} ds,
\end{aligned}
$$

where $\Delta t = T - t$. By discretizing the continuous dynamics under the P-measure, we obtain the state equation

$$X_i = (I - \exp(-K^P \Delta t_i))\theta^P + \exp(-K^P \Delta t_i)X_{i-1} + \eta_t,$$

where $\Delta t_i = t_i - t_{i-1}$ is the time between observations. The conditional covariance matrix for the shock terms is given by

$$Q = \int_0^{\Delta t_i} e^{-K^P s} \Sigma\Sigma' e^{-(K^P)' s} ds.$$

Stationarity of the system under the P-measure is imposed by restricting the real component of each eigenvalue of K^P to be positive. The Kalman filter for these models is also started at the unconditional mean and covariance[7]

$$\widehat{X}_0 = \theta^P \quad \text{and} \quad \widehat{\Sigma}_0 = \int_0^\infty e^{-K^P s} \Sigma\Sigma' e^{-(K^P)' s} ds.$$

Finally, the AFNS and AFGNS measurement equation is given by

$$y_t = A + BX_t + \varepsilon_t.$$

For all five models, the error structure is

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

---

[7]In the estimation, $\int_0^\infty e^{-K^P s} \Sigma\Sigma' e^{-(K^P)' s} ds$ is approximated by $\int_0^{10} e^{-K^P s} \Sigma\Sigma' e^{-(K^P)' s} ds$.

A	L_{t-1}	S_{t-1}	C^1_{t-1}	C^2_{t-1}	μ	q
L_t	0.9839	0	0	0	0.04907	0.001835
	(0.0145)				(0.0112)	(0.000280)
S_t	0	0.9889	0	0	-0.006021	0.002728
		(0.0126)			(0.0208)	(0.000216)
C^1_t	0	0	0.9565	0	0.003424	0.007988
			(0.0221)		(0.0169)	(0.000448)
C^2_t	0	0	0	0.9864	0.06082	0.006355
				(0.0146)	(0.0422)	(0.000682)

Table 2: **Estimated Dynamic Parameters in the DNSS Model.**
This table reports the estimated A matrix and μ vector along with the estimated parameters of the Q matrix in the independent-factor DNSS model for the sample period from January 1987 to December 2002. The maximum log likelihood value is 16658.40. The estimated value of λ_1 is 0.8379 (0.0117), while the estimated value of λ_2 is 0.09653 (0.0163). The numbers in parentheses are the estimated standard deviations of the parameter estimates.

where H is a diagonal matrix

$$
H = \begin{pmatrix} \sigma^2(\tau_1) & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sigma^2(\tau_N) \end{pmatrix}.
$$

The linear least-squares optimality of the Kalman filter requires that the transition and measurement errors be orthogonal to the initial state, i.e.,

$$
E[f_0 \eta_t'] = 0, \quad E[f_0 \varepsilon_t'] = 0.
$$

Finally, parameter standard deviations are calculated as

$$
\Sigma(\widehat{\psi}) = \frac{1}{T} \Big[\frac{1}{T} \sum_{t=1}^{T} \frac{\partial \log l_t(\widehat{\psi})}{\partial \psi} \frac{\partial \log l_t(\widehat{\psi})}{\partial \psi}' \Big]^{-1},
$$

where $\widehat{\psi}$ denotes the estimated model parameter set.

5.3   DNSS model estimation results

Table 2 presents the estimated mean-reversion matrix A and the estimated vector of mean parameters μ, along with the estimated parameters of the conditional covariance matrix Q obtained for the DNSS model. The results reveal that the slope factor is the most persistent factor. Also, the relatively large standard deviations of the estimated mean parameters suggest some difficulty in pinning down their value under the P-measure, which is likely related to the fairly high persistence of the state variables (e.g., Kim and Orphanides, 2005). The λ_1 parameter is estimated at 0.838, which implies a factor loading for the first curvature factor that peaks near the two-year maturity. The estimated value of λ_2 is 0.097, so the factor loading of the second curvature factor reaches its maximum near the 19-year maturity. (These are illustrated in Figure 1(a).) Clearly, the two
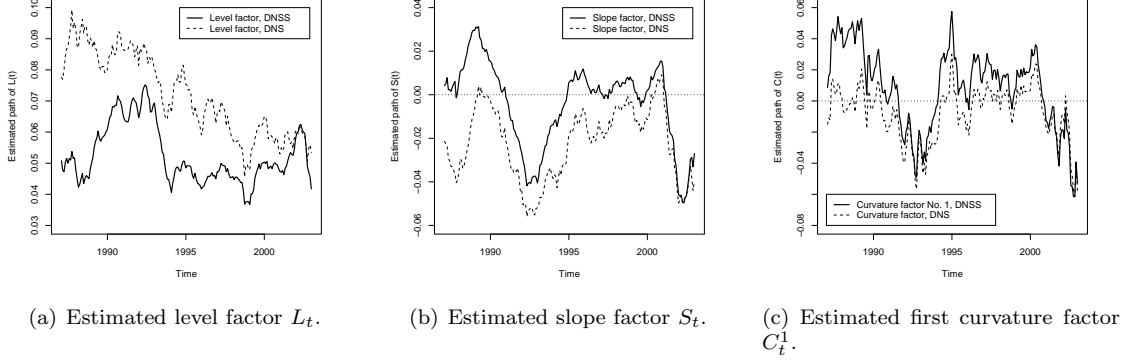
(a) Estimated level factor L_t.   (b) Estimated slope factor S_t.   (c) Estimated first curvature factor C_t^1.

Figure 3: **Level, Slope, and First Curvature Factors in the DNSS Model.**
The estimated paths of the level, slope, and first curvature factors from the independent-factor DNSS model are shown. For comparison, the corresponding paths from the independent-factor DNS model are included.

curvature factors take on very different roles in the fit of the model.

Volatility parameters across the various models are most easily compared by focusing on the one-month conditional covariance matrix that they generate. For the independent-factor DNSS model, the estimated matrix is given by

$$
Q_{indep}^{DNSS} = qq' = \begin{pmatrix}
3.37 \times 10^{-6} & 0 & 0 & 0 \\
0 & 7.44 \times 10^{-6} & 0 & 0 \\
0 & 0 & 6.38 \times 10^{-5} & 0 \\
0 & 0 & 0 & 4.04 \times 10^{-5}
\end{pmatrix}. \tag{9}
$$

The level factor has the smallest volatility, and the two curvature factors are the most volatile, similar to the CDR results for the DNS model.

In Figure 3, we compare the estimated level, slope, and first curvature factors in the DNSS model to the corresponding factors estimated by CDR for the independent-factor DNS model. The correlations for these three factors across the two models are 0.553, 0.844, and 0.899, respectively. Thus, only the level factor changes notably when the second curvature factor is added to the model. Intuitively, without the second curvature factor, only the level factor is able to fit the long-term yields. However, the second curvature factor can fit yields with maturities in the 10- to 30-year range, so when it is included, the level factor is allowed to fit other areas of the yield curve.

Figure 4 shows the second curvature factor. The purpose of this factor is to improve the fit of long-term yields, and there is a clear relationship between it and the ten-year yield (with a correlation coefficient of 0.793). The second curvature factor also inherits the downward trend observed in long-term yields over this sample period, while the DNSS level factor starts to look more stationary.

Table 3 reports summary statistics for the fitted errors of all five models. With its additional flexibility, the DNSS model does show some improvement in fit over the DNS model, especially
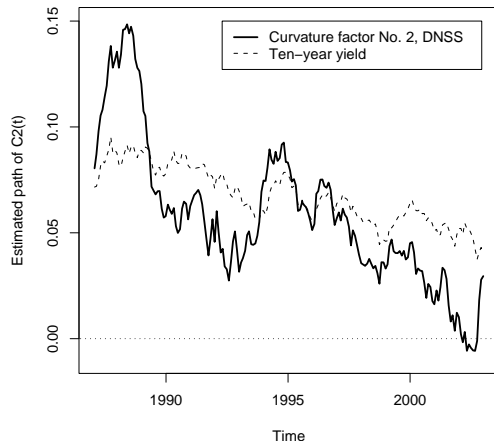
16

Figure 4: **Second Curvature Factor in the DNSS Model.**
The estimated path of the second curvature factor from the independent-factor DNSS model is shown with the ten-year yield for comparison.

in the maturity range from three months to eight years. There is also a slightly better DNSS model fit with long-term yields, which is consistent with the second curvature factor operating at long maturities. However, Figure 5, which displays the fit of all five models on four specific dates,[8] shows that at times the DNSS model still does not fit the long end of the yield curve very well. Indeed, since the factor loading of the second curvature factor is practically flat in the 10- to 30-year maturity range, it can only provide a level difference between the shorter end of the yield curve and the very long end of the curve, but it cannot fit deviations between the 10-, 15-, 20-, and 30-year yields.

The fitted errors reported in Table 3 for the DNSS model can be compared loosely to the errors reported by Gürkaynak, Sack, and Wright (2007), who use the Svensson yield curve to fit bond yields. Importantly, they fit the curve separately for each business day with no regard for the time series behavior of the extracted factors, which show dramatic variation over time. Their estimation will always produce a better fit on any given day than ours, but the fit of the DNSS model is quite comparable to theirs over the maturity range from six months to nine years.

5.4   DGNS model estimation results

Table 4 presents the estimated mean-reversion matrix A and the estimated vector of mean parameters μ along with the estimated parameters of the conditional covariance matrix Q for the independent-factor DGNS model. Relative to the independent-factor DNSS model reported in the previous section, the level factor and the two curvature factors preserve their relatively high rate of persistence after the inclusion of the second slope factor. However, for the two slope factors, we see a significant change in the estimated mean-reversion rates after this addition. Overall, though,

---

[8]These four dates provide examples of the variety of yield curve shapes observed over this sample period and were selected by De Pooter (2007).

Maturity in months	DNS indep.-factor		AFNS indep.-factor		DNSS indep.-factor		DGNS indep.-factor		AFGNS indep.-factor	
	Mean	RMSE	Mean	RMSE	Mean	RMSE	Mean	RMSE	Mean	RMSE
3	-1.64	12.26	-2.85	18.54	2.53	10.65	2.36	9.07	0.03	9.52
6	-0.24	1.09	-1.19	7.12	0.01	0.60	-0.06	1.05	0.01	0.86
9	-0.54	7.13	-1.24	3.44	-2.73	6.82	-2.64	6.15	-1.58	5.94
12	4.04	11.19	3.58	9.60	0.53	8.16	0.77	6.84	1.99	7.62
18	7.22	10.76	7.15	10.44	3.19	5.87	3.60	5.56	4.12	6.11
24	1.18	5.83	1.37	5.94	-1.82	4.11	-1.44	3.61	-1.76	3.80
36	-0.07	1.51	0.31	1.98	0.07	2.68	0.03	2.57	-0.62	2.65
48	-0.67	3.92	-0.39	3.72	1.69	3.78	1.20	3.12	1.56	3.47
60	-5.33	7.13	-5.27	6.82	-2.32	5.24	-2.99	5.15	-1.56	4.71
84	-1.22	4.25	-1.50	4.29	-0.26	4.04	-0.36	3.73	0.65	3.92
96	1.31	2.10	1.02	2.11	0.47	0.85	0.99	1.80	0.31	0.77
108	0.03	2.94	-0.11	3.02	-2.67	4.49	-1.41	3.27	-4.56	6.08
120	-5.11	8.51	-4.96	8.23	-9.51	12.13	-7.46	9.73	-13.60	15.47
180	24.11	29.44	27.86	32.66	16.37	24.94	21.97	28.16	-0.04	12.03
240	25.61	34.99	35.95	42.61	23.12	34.62	30.72	36.43	1.51	6.67
360	-29.62	37.61	1.37	22.04	-8.65	24.45	-0.96	6.81	-2.65	24.62
Mean	1.19	11.29	3.82	11.41	1.25	9.59	2.77	8.32	-1.01	7.14
Median	-0.16	7.13	0.10	6.97	0.04	5.56	-0.02	5.36	-0.01	6.01

Table 3: **Summary Statistics of In-Sample Fit.**
The means and the root mean squared errors for 16 different maturities. All numbers are measured in basis points.

A	L_{t-1}	S^1_{t-1}	S^2_{t-1}	C^1_{t-1}	C^2_{t-1}	μ	q
L_t	0.9758	0	0	0	0	0.05140	0.001998
	(0.0239)					(0.0104)	(0.000268)
S^1_t	0	0.9235	0	0	0	-0.007039	0.004309
		(0.0295)				(0.00718)	(0.000371)
S^2_t	0	0	0.9306	0	0	0.0006993	0.003462
			(0.0341)			(0.00686)	(0.000363)
C^1_t	0	0	0	0.9543	0	-0.0006114	0.005807
				(0.0223)		(0.0109)	(0.000405)
C^2_t	0	0	0	0	0.9782	0.05536	0.005223
					(0.0194)	(0.0207)	(0.000756)

Table 4: **Estimated Dynamic Parameters in the DGNS Model.**
This table reports the estimated A matrix and μ vector along with the estimated parameters of the Q matrix in the DGNS model with independent factors for the sample period from January 1987 to December 2002. The maximum log likelihood value is 16816.08. The estimated value of λ_1 is 1.190 (0.0350), while the estimated value of λ_2 is 0.1021 (0.00863). The numbers in parentheses are the estimated standard deviations of the parameter estimates.

all the factors have become less persistent than what we observed in the DNSS model.

For the estimated mean parameters we find little change after adding the second slope factor to the model. If anything, it seems like the uncertainty about these parameters has declined notably. This ties in well with the fact that the factors have become less persistent, which allows the estimation to determine their means more precisely.
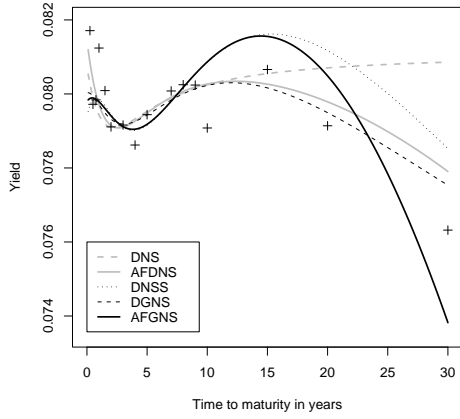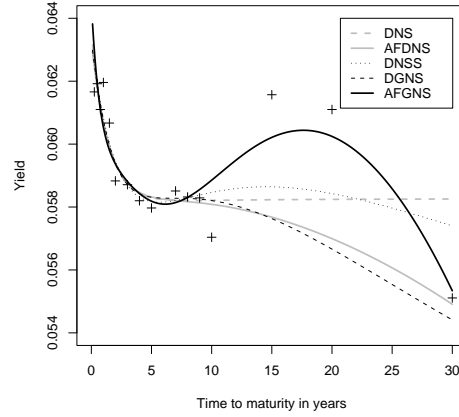
(a) June 30, 1989.
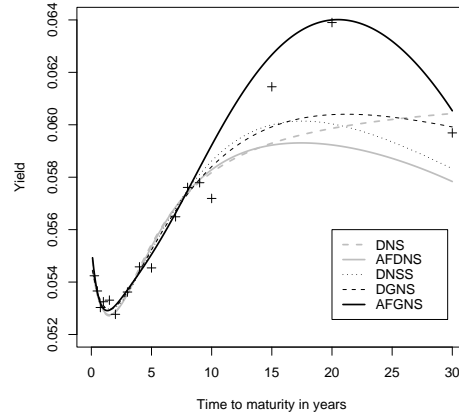
(b) November 30, 1995.
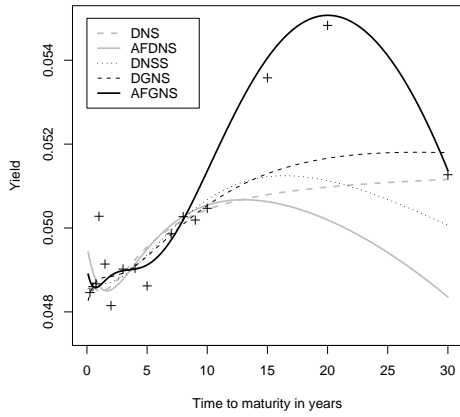

(c) August 31, 1998.

(d) September 29, 2000.

Figure 5: **Fitted Yield Curves for Four Specific Dates.**
The observed yields on four specific dates (June 30, 1989, November 30, 1995, August 31, 1998, September 29, 2000) are indicated with plus signs. The fitted yield curves on these same dates are from the independent-factor DNS, AFNS, DNSS, DGNS, and AFGNS models estimated over the full sample from January 1987 to December 2002.

For the independent-factor DGNS model, the estimated q-parameters translate into a one-month conditional covariance matrix given by

$$Q_{indep}^{DGNS} = qq' = \begin{pmatrix} 3.99 \times 10^{-6} & 0 & 0 & 0 & 0 \\ 0 & 1.86 \times 10^{-5} & 0 & 0 & 0 \\ 0 & 0 & 1.20 \times 10^{-5} & 0 & 0 \\ 0 & 0 & 0 & 3.37 \times 10^{-5} & 0 \\ 0 & 0 & 0 & 0 & 2.73 \times 10^{-5} \end{pmatrix}. \quad (10)$$

This matrix shows that for the level factor and the two curvature factors the estimated volatilities
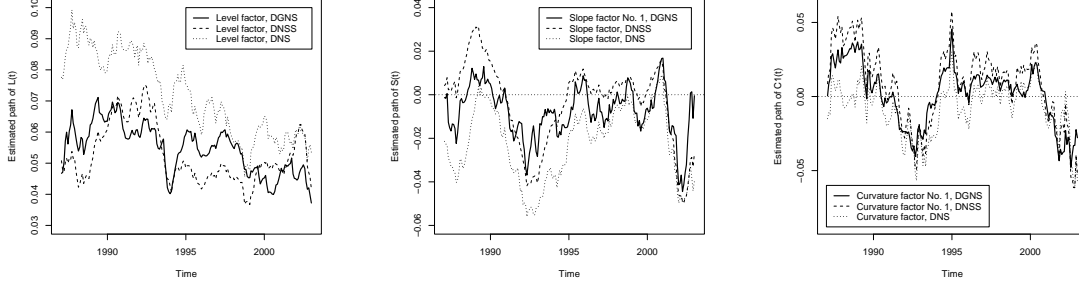
(a) The estimated level factor L_t.　(b) The estimated first slope factor S_t^1.　(c) The estimated first curvature factor C_t^1.

Figure 6: **Estimated Paths of the Level, First Slope, and First Curvature Factor in the DGNS Model.**
The estimated level, first slope, and first curvature factors from the Kalman filter estimation of the DGNS model with independent factors. For ease of comparison the estimated paths from the independent-factor DNS and DNSS models have been included. In all three cases the data used are unsmoothed Fama-Bliss yields covering the period from January 1987 to December 2002.

are now smaller than the ones reported in Equation (9) for the DNSS model. In contrast, the estimated volatilities of the two slope factors are notably higher than the one reported for the single slope factor in the DNSS model.

The estimated values of λ_1 and λ_2, which are 1.19 and 0.102, respectively, are also quite interesting. The high value of λ_1 compared to the estimate of 0.838 we obtained for the DNSS model implies that the factor loading of the first slope and curvature factor decay to zero at a more rapid pace. Thus, as illustrated in Figure 1(b), these two factors have a limited impact on yields beyond the five-year maturity. However, that lack of impact is made up for by the second slope factor. Due to the maintained low estimate of λ_2, this factor has a loading that decays very slowly. Therefore, this factor can affect the important intermediate range of maturities from 5 to 15 years of maturity.

In Figure 6, we compare the level, the first slope, and the first curvature factors in the DGNS model to the corresponding factors obtained for the DNS model (from CDR) and the DNSS model (described earlier in this section). The correlations of these three factors across the DNS and DGNS models are 0.730, 0.804, and 0.793, respectively. For the DNS and DNSS models, the correlations are 0.549, 0.821, and 0.949, respectively. Thus, while the level factor is affected by the addition of a second curvature factor, as in the DNSS model, the impact of a second slope factor, as in the DGNS model, is more limited. Also, the first slope and curvature factors have very similar sample paths across all three models. Given the fairly large estimated values of λ_1 in all three models, the factor loadings of these two factors decay towards zero relatively rapidly as a function of maturity, so their roles in fitting the shorter end of the yield curve are well defined.

Figure 7 shows the second slope and curvature factors from the DGNS model. There is a clear correlation between the curvature factor and the ten-year yield, as in the DNSS model. The second slope factor appears to be a stationary process with a fairly high rate of mean-reversion, but its intuition is not obvious.
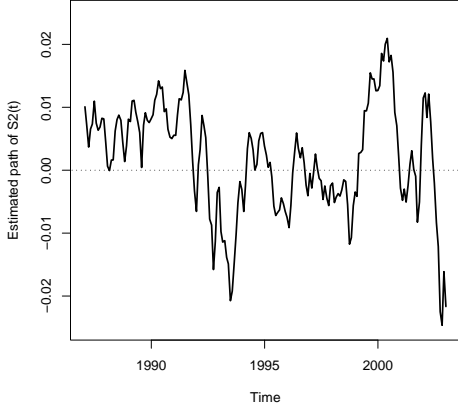
20

(a) The estimated second slope factor S_t^2.
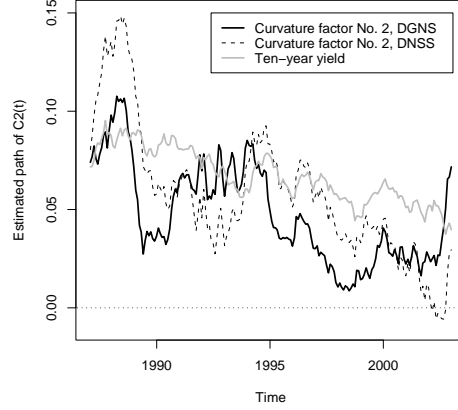
(b) The estimated second curvature factor C_t^2.

Figure 7: **Second Slope and Second Curvature Factors in the DGNS Model.**
The estimated paths of the second slope and curvature factors of the independent-factor DGNS model. The estimated path of the second curvature factor from the independent-factor DNSS model has been included for comparison.

K^P	$K_{\cdot,1}$	$K_{\cdot,2}$	$K_{\cdot,3}$	$K_{\cdot,4}$	$K_{\cdot,5}$	θ^P	Σ
$K_{1,\cdot}$	1.012	0	0	0	0	0.1165	0.01057
	(0.716)					(0.00651)	(0.000262)
$K_{2,\cdot}$	0	0.2685	0	0	0	-0.04551	0.01975
		(0.497)				(0.0493)	(0.00255)
$K_{3,\cdot}$	0	0	0.3812	0	0	-0.02912	0.01773
			(0.603)			(0.0322)	(0.00225)
$K_{4,\cdot}$	0	0	0	1.409	0	-0.02398	0.05049
				(0.970)		(0.0227)	(0.00304)
$K_{5,\cdot}$	0	0	0	0	0.8940	-0.09662	0.04304109
					(0.927)	(0.0338)	(0.00305)

Table 5: **Estimated Dynamic Parameters in the AFGNS Model.**
This table reports the estimated K^P matrix and θ^P mean vector along with the estimated parameters of the Σ volatility matrix in the AFGNS model with independent factors for the sample period from January 1987 to December 2002. The maximum log likelihood value is 16982.52. λ_1 is estimated at 1.005 (0.0246), and λ_2 is estimated at 0.2343 (0.00922). The numbers in parentheses are the estimated standard deviations of the parameter estimates.

If we focus on the fit of the DGNS model in Table 3, we see fairly uniform improvement in the fit in the maturity range from three months to ten years and a dramatic improvement in the fit of the 30-year yield. The improved fit for the long yield in the DGNS model relative to the DNSS model reflects the presence of the second slope factor and is also visible in Figure 5. However, there is still no improvement for the 15- or 20-year yields, a deficiency that can perhaps be alleviated by imposing the AF restrictions.

5.5 AFGNS model estimation results

Table 5 presents the estimated parameters for the mean-reversion matrix K^P, the mean vector θ^P, and the volatility matrix Σ for the AFGNS model with independent factors. To compare the estimated mean-reversion parameters in this model to the results reported for the previous models, we calculate the one-month conditional discrete-time mean-reversion matrix, which is given by

$$
\exp\left(-\frac{1}{12}K^P\right) = \begin{pmatrix} 0.9191 & 0 & 0 & 0 & 0 \\ 0 & 0.9779 & 0 & 0 & 0 \\ 0 & 0 & 0.9687 & 0 & 0 \\ 0 & 0 & 0 & 0.8892 & 0 \\ 0 & 0 & 0 & 0 & 0.9282 \end{pmatrix}. \tag{11}
$$

Compared to the estimated A matrix reported for the DGNS model in Table 4, this shows that by imposing an absence of arbitrage on that model, the level and two curvature factors become notably less persistent, while the two slope factors become more persistent.

Based on the estimated volatility parameters, the one-month conditional covariance matrix in the AFGNS model is given by

$$
Q_{indep}^{AFGNS} = \int_0^{\frac{1}{12}} e^{-K^P s} \Sigma\Sigma' e^{-(K^P)'s} ds = \begin{pmatrix} 8.52 \times 10^{-6} & 0 & 0 & 0 & 0 \\ 0 & 3.17 \times 10^{-5} & 0 & 0 & 0 \\ 0 & 0 & 2.53 \times 10^{-5} & 0 & 0 \\ 0 & 0 & 0 & 0.000188 & 0 \\ 0 & 0 & 0 & 0 & 0.000143 \end{pmatrix}. \tag{12}
$$

Across the board, the volatility of each factor is notably higher in the AFGNS model than in the corresponding non-AF DGNS model.

The estimated AFGNS values of λ_1 and λ_2 are 1.01 and 0.234, compared with the DGNS values of 1.19 and 0.1016. The lower value of λ_1 implies that the first slope and curvature factors decay somewhat slower to zero than in the DGNS model, while the higher value of λ_2 indicates that the model is using the additional yield-adjustment term to get the level of the long-term yields right, which eases the tension on the second curvature factor. This shows up as a much larger estimate for λ_2.

Figure 8 displays the estimated paths for the level, first slope, and first curvature factors in the AFGNS model and the earlier models. The correlations for these three factors across the AFGNS and DGNS models are 0.692, 0.668, and 0.952, respectively. Thus, for the level and first slope factors, the imposition of an absence of arbitrage leads to some changes.[9]

Figure 9 illustrates the second slope and curvature factors and the effect of the increase in the estimated value of λ_2. Figure 9(a) shows that there is a notable change in the path of the second slope factor in the AFGNS model relative to the DGNS model, and the two paths show a correlation of only 0.046. There is greater correlation between the AFGNS and DGNS second curvature factors (of 0.696), as depicted in Figure 9(b).

---

[9]Note that, with the inclusion of the yield-adjustment term in the yield function of the AFGNS model, the estimated values of all five factors are rescaled relative to the estimated values obtained in the DGNS model (and reflected in the mean parameter estimates as well).
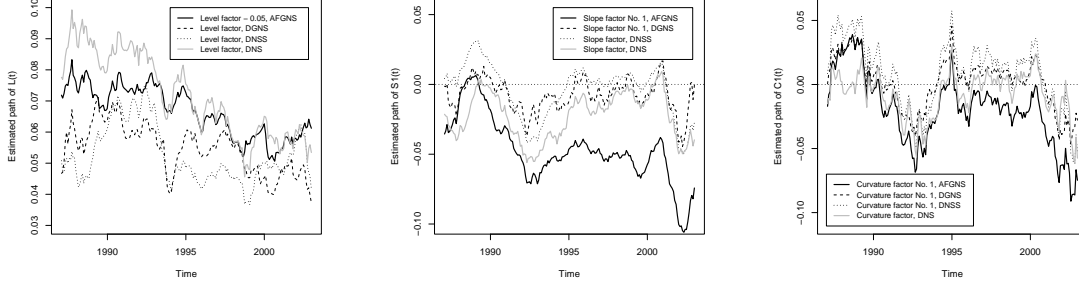
(a) The estimated level factor L_t.

(b) The estimated first slope factor S_t^1.

(c) The estimated first curvature factor C_t^1.

Figure 8: **Level, First Slope, and First Curvature Factors in the AFGNS Model.**
The estimated level, first slope, and first curvature factors for the independent-factor AFGNS model are shown with the estimated paths from the DNS, DNSS, and DGNS models for comparison.
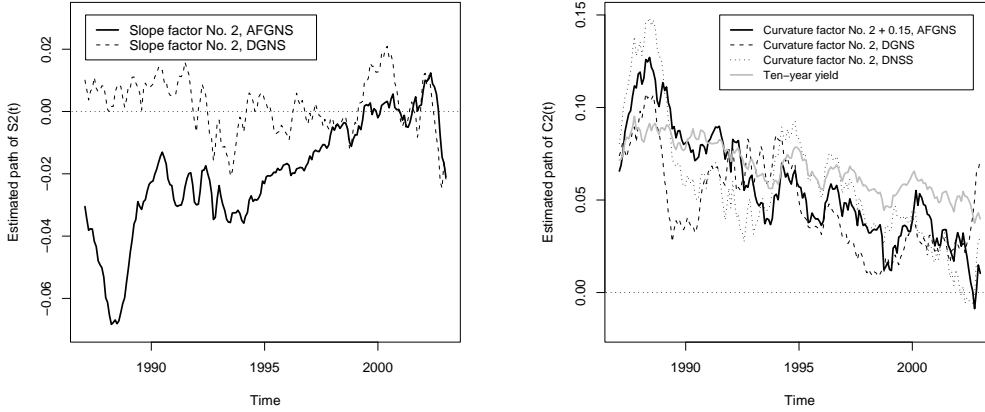

(a) The estimated second slope factor S_t^2.

(b) The estimated second curvature factor C_t^2.

Figure 9: **Second Slope and Second Curvature Factors in the AFGNS Model.**
The estimated second slope and curvature factors for the independent-factor AFGNS model are shown with corresponding estimated paths from the DNSS and DGNS models for comparison.

Focusing on the fit of the AFGNS model in Table 3, it is clear that the AFGNS model provides a more balanced fit across maturities than the DNSS model. Indeed, only the 30-year yield does not really benefit from adding the second slope factor or the AF restrictions. There are also benefits relative to the DGNS model, especially on the four specific dates studied in Figure 5 when the improvement in the fit of the 15- and 20-year yields obtained with the AFGNS model is quite apparent. The increase in the maximum log likelihood value from 16816.08 to 16982.52 from the imposition of the AF restrictions also indicates that the overall fit of the model has been improved notably.

The only difference between the DGNS and the AFGNS models is tied to the yield-adjustment term $-\frac{C(\tau)}{\tau}$, which is a maturity-dependent function that appears in the yield function as a
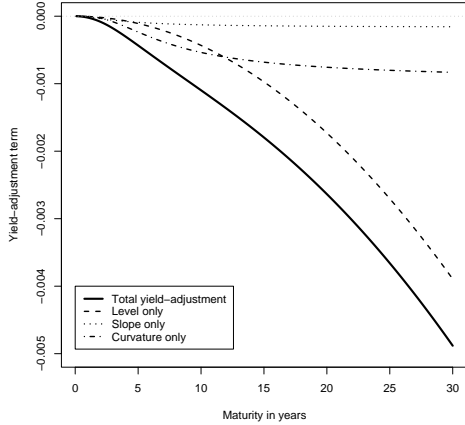
(a) Yield-adjustment in the AFNS model.
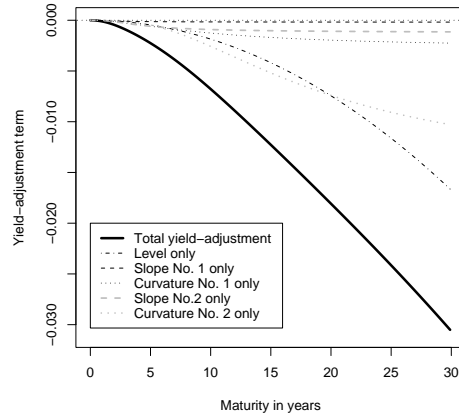
(b) Yield-adjustment in the AFGNS model.

Figure 10: **Yield-Adjustment Term for the AFNS and AFGNS Models.**
The estimated yield-adjustment term $-\frac{C(\tau)}{\tau}$ in the AFNS and AFGNS models and their respective subcomponents.

result of the imposition of absence of arbitrage and is a consequence of convexity effects. Figure 10 displays the AFNS yield-adjustment term from CDR (and its three subcomponents) and the AFGNS yield-adjustment term (and its five subcomponents).[10] These two yield adjustments have similar shapes but a somewhat different scale. In the AFNS model, the yield-adjustment term stays below 50 basis points even at the 30-year maturity, while in the AFGNS model it reaches a full 3 percentage points at that same maturity. The AFGNS model uses the large negative values of the yield adjustment at long maturities to generate the second hump of the yield curve in order to deliver a reasonable fit to the 15- to 30-year yields.

6 Conclusion

The Nelson-Siegel (1987) model and the dynamic version of Diebold and Li (2006) have trouble fitting long-maturity yields. In this paper we address that problem, while *simultaneously* imposing an absence of arbitrage, building on Christensen, Diebold, and Rudebusch (2007). We argue that a suitably-dynamized version of Svensson's (1995) four-factor extension of Nelson-Siegel (DNSS) should improve long-maturity fit, but that there does not exist an arbitrage-free yield-curve model with DNSS factor loadings. We then show that we can address this problem by using a simple five-factor generalization of DNSS, where the fifth factor has a natural interpretation as a second slope factor which *does* achieve freedom from arbitrage. Finally, we show that the estimation of this new arbitrage-free generalized Nelson-Siegel model is tractable and that the AFGNS model provides a good fit to the yield curve.

---

[10]As long as we only consider models with diagonal volatility matrices, the yield-adjustment term will be a negative, monotonically decreasing function of maturity that will eventually converge to $-\infty$ due to the level factor imposed in the Nelson-Siegel model.

Appendix A: Yield-Adjustment Term in the AFGNS Model

Given a general volatility matrix

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} & \sigma_{14} & \sigma_{15} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} & \sigma_{24} & \sigma_{25} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} & \sigma_{34} & \sigma_{35} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_{44} & \sigma_{45} \\ \sigma_{51} & \sigma_{52} & \sigma_{53} & \sigma_{54} & \sigma_{55} \end{pmatrix}$$

the analytical AFGNS yield-adjustment term, via calculations available from the authors, is

$$
\begin{aligned}
\frac{C(t,T)}{T-t} &= \frac{1}{2}\frac{1}{T-t}\int_t^T \sum_{j=1}^5 \left(\Sigma' B(s,T)B(s,T)'\Sigma\right)_{j,j} ds \\
&= \overline{A}\frac{(T-t)^2}{6} \\
&\quad +\overline{B}\Big[\frac{1}{2\lambda_1^2} - \frac{1}{\lambda_1^3}\frac{1-e^{-\lambda_1(T-t)}}{T-t} + \frac{1}{4\lambda_1^3}\frac{1-e^{-2\lambda_1(T-t)}}{T-t}\Big] \\
&\quad +\overline{C}\Big[\frac{1}{2\lambda_2^2} - \frac{1}{\lambda_2^3}\frac{1-e^{-\lambda_2(T-t)}}{T-t} + \frac{1}{4\lambda_2^3}\frac{1-e^{-2\lambda_2(T-t)}}{T-t}\Big] \\
&\quad +\overline{D}\Big[\frac{1}{2\lambda_1^2} + \frac{1}{\lambda_1^2}e^{-\lambda_1(T-t)} - \frac{1}{4\lambda_1}(T-t)e^{-2\lambda_1(T-t)} - \frac{3}{4\lambda_1^2}e^{-2\lambda_1(T-t)} + \frac{5}{8\lambda_1^3}\frac{1-e^{-2\lambda_1(T-t)}}{T-t} - \frac{2}{\lambda_1^3}\frac{1-e^{-\lambda_1(T-t)}}{T-t}\Big] \\
&\quad +\overline{E}\Big[\frac{1}{2\lambda_2^2} + \frac{1}{\lambda_2^2}e^{-\lambda_2(T-t)} - \frac{1}{4\lambda_2}(T-t)e^{-2\lambda_2(T-t)} - \frac{3}{4\lambda_2^2}e^{-2\lambda_2(T-t)} + \frac{5}{8\lambda_2^3}\frac{1-e^{-2\lambda_2(T-t)}}{T-t} - \frac{2}{\lambda_2^3}\frac{1-e^{-\lambda_2(T-t)}}{T-t}\Big] \\
&\quad +\overline{F}\Big[\frac{1}{2\lambda_1}(T-t) + \frac{1}{\lambda_1^2}e^{-\lambda_1(T-t)} - \frac{1}{\lambda_1^3}\frac{1-e^{-\lambda_1(T-t)}}{T-t}\Big] \\
&\quad +\overline{G}\Big[\frac{1}{2\lambda_2}(T-t) + \frac{1}{\lambda_2^2}e^{-\lambda_2(T-t)} - \frac{1}{\lambda_2^3}\frac{1-e^{-\lambda_2(T-t)}}{T-t}\Big] \\
&\quad +\overline{H}\Big[\frac{3}{\lambda_1^2}e^{-\lambda_1(T-t)} + \frac{1}{2\lambda_1}(T-t) + \frac{1}{\lambda_1}(T-t)e^{-\lambda_1(T-t)} - \frac{3}{\lambda_1^3}\frac{1-e^{-\lambda_1(T-t)}}{T-t}\Big] \\
&\quad +\overline{I}\Big[\frac{3}{\lambda_2^2}e^{-\lambda_2(T-t)} + \frac{1}{2\lambda_2}(T-t) + \frac{1}{\lambda_2}(T-t)e^{-\lambda_2(T-t)} - \frac{3}{\lambda_2^3}\frac{1-e^{-\lambda_2(T-t)}}{T-t}\Big] \\
&\quad +\overline{J}\Big[\frac{1}{\lambda_1\lambda_2} - \frac{1}{\lambda_1^2\lambda_2}\frac{1-e^{-\lambda_1(T-t)}}{T-t} - \frac{1}{\lambda_1\lambda_2^2}\frac{1-e^{-\lambda_2(T-t)}}{T-t} + \frac{1}{\lambda_1\lambda_2(\lambda_1+\lambda_2)}\frac{1-e^{-(\lambda_1+\lambda_2)(T-t)}}{T-t}\Big] \\
&\quad +\overline{K}\Big[\frac{1}{\lambda_1^2} + \frac{1}{\lambda_1^2}e^{-\lambda_1(T-t)} - \frac{1}{2\lambda_1^2}e^{-2\lambda_1(T-t)} - \frac{3}{\lambda_1^3}\frac{1-e^{-\lambda_1(T-t)}}{T-t} + \frac{3}{4\lambda_1^3}\frac{1-e^{-2\lambda_1(T-t)}}{T-t}\Big] \\
&\quad +\overline{L}\Big[\frac{1}{\lambda_1\lambda_2} + \frac{1}{\lambda_1\lambda_2}e^{-\lambda_2(T-t)} - \frac{1}{\lambda_1(\lambda_1+\lambda_2)}e^{-(\lambda_1+\lambda_2)(T-t)} - \frac{1}{\lambda_1^2\lambda_2}\frac{1-e^{-\lambda_1(T-t)}}{T-t} \\
&\qquad - \frac{2}{\lambda_1\lambda_2^2}\frac{1-e^{-\lambda_2(T-t)}}{T-t} + \frac{\lambda_1+2\lambda_2}{\lambda_1\lambda_2(\lambda_1+\lambda_2)^2}[1 - e^{-(\lambda_1+\lambda_2)(T-t)}]\Big] \\
&\quad +\overline{M}\Big[\frac{1}{\lambda_1\lambda_2} + \frac{1}{\lambda_1\lambda_2}e^{-\lambda_1(T-t)} - \frac{1}{\lambda_2(\lambda_1+\lambda_2)}e^{-(\lambda_1+\lambda_2)(T-t)} - \frac{1}{\lambda_1\lambda_2^2}\frac{1-e^{-\lambda_2(T-t)}}{T-t} \\
&\qquad - \frac{2}{\lambda_1^2\lambda_2}\frac{1-e^{-\lambda_1(T-t)}}{T-t} + \frac{\lambda_2+2\lambda_1}{\lambda_1\lambda_2(\lambda_1+\lambda_2)^2}[1 - e^{-(\lambda_1+\lambda_2)(T-t)}]\Big] \\
&\quad +\overline{N}\Big[\frac{1}{\lambda_2^2} + \frac{1}{\lambda_2^2}e^{-\lambda_2(T-t)} - \frac{1}{2\lambda_2^2}e^{-2\lambda_2(T-t)} - \frac{3}{\lambda_2^3}\frac{1-e^{-\lambda_2(T-t)}}{T-t} + \frac{3}{4\lambda_2^3}\frac{1-e^{-2\lambda_2(T-t)}}{T-t}\Big] \\
&\quad +\overline{O}\Big[\frac{1}{\lambda_1\lambda_2} + \frac{1}{\lambda_1\lambda_2}e^{-\lambda_1(T-t)} + \frac{1}{\lambda_1\lambda_2}e^{-\lambda_2(T-t)} - \Big(\frac{1}{\lambda_1}+\frac{1}{\lambda_2}\Big)\frac{1}{\lambda_1+\lambda_2}e^{-(\lambda_1+\lambda_2)(T-t)} \\
&\qquad - \frac{2}{(\lambda_1+\lambda_2)^2}e^{-(\lambda_1+\lambda_2)(T-t)} - \frac{1}{\lambda_1+\lambda_2}(T-t)e^{-(\lambda_1+\lambda_2)(T-t)} - \frac{2}{\lambda_1^2\lambda_2}\frac{1-e^{-\lambda_1(T-t)}}{T-t} - \frac{2}{\lambda_1\lambda_2^2}\frac{1-e^{-\lambda_2(T-t)}}{T-t} \\
&\qquad + \frac{2}{(\lambda_1+\lambda_2)^3}\frac{1-e^{-(\lambda_1+\lambda_2)(T-t)}}{T-t} + \Big(\frac{1}{\lambda_1}+\frac{1}{\lambda_2}\Big)\frac{1}{(\lambda_1+\lambda_2)^2}\frac{1-e^{-(\lambda_1+\lambda_2)(T-t)}}{T-t} + \frac{1}{\lambda_1\lambda_2(\lambda_1+\lambda_2)}\frac{1-e^{-(\lambda_1+\lambda_2)(T-t)}}{T-t}\Big],
\end{aligned}
$$

where

- $\overline{A} = \sigma_{11}^2 + \sigma_{12}^2 + \sigma_{13}^2 + \sigma_{14}^2 + \sigma_{15}^2,$

- $\overline{B} = \sigma_{21}^2 + \sigma_{22}^2 + \sigma_{23}^2 + \sigma_{24}^2 + \sigma_{25}^2,$

- $\overline{C} = \sigma_{31}^2 + \sigma_{32}^2 + \sigma_{33}^2 + \sigma_{34}^2 + \sigma_{35}^2,$

- $\overline{D} = \sigma_{41}^2 + \sigma_{42}^2 + \sigma_{43}^2 + \sigma_{44}^2 + \sigma_{45}^2,$

- $\overline{E} = \sigma_{51}^2 + \sigma_{52}^2 + \sigma_{53}^2 + \sigma_{54}^2 + \sigma_{55}^2,$

- $\overline{F} = \sigma_{11}\sigma_{21} + \sigma_{12}\sigma_{22} + \sigma_{13}\sigma_{23} + \sigma_{14}\sigma_{24} + \sigma_{15}\sigma_{25},$

- $\overline{G} = \sigma_{11}\sigma_{31} + \sigma_{12}\sigma_{32} + \sigma_{13}\sigma_{33} + \sigma_{14}\sigma_{34} + \sigma_{15}\sigma_{35},$

- $\overline{H} = \sigma_{11}\sigma_{41} + \sigma_{12}\sigma_{42} + \sigma_{13}\sigma_{43} + \sigma_{14}\sigma_{44} + \sigma_{15}\sigma_{45},$

- $\overline{I} = \sigma_{11}\sigma_{51} + \sigma_{12}\sigma_{52} + \sigma_{13}\sigma_{53} + \sigma_{14}\sigma_{54} + \sigma_{15}\sigma_{55},$

- $\overline{J} = \sigma_{21}\sigma_{31} + \sigma_{22}\sigma_{32} + \sigma_{23}\sigma_{33} + \sigma_{24}\sigma_{34} + \sigma_{25}\sigma_{35},$

- $\overline{K} = \sigma_{21}\sigma_{41} + \sigma_{22}\sigma_{42} + \sigma_{23}\sigma_{43} + \sigma_{24}\sigma_{44} + \sigma_{25}\sigma_{45},$

- $\overline{L} = \sigma_{21}\sigma_{51} + \sigma_{22}\sigma_{52} + \sigma_{23}\sigma_{53} + \sigma_{24}\sigma_{54} + \sigma_{25}\sigma_{55},$

- $\overline{M} = \sigma_{31}\sigma_{41} + \sigma_{32}\sigma_{42} + \sigma_{33}\sigma_{43} + \sigma_{34}\sigma_{44} + \sigma_{35}\sigma_{45},$

- $\overline{N} = \sigma_{31}\sigma_{51} + \sigma_{32}\sigma_{52} + \sigma_{33}\sigma_{53} + \sigma_{34}\sigma_{54} + \sigma_{35}\sigma_{55},$

- $\overline{O} = \sigma_{41}\sigma_{51} + \sigma_{42}\sigma_{52} + \sigma_{43}\sigma_{53} + \sigma_{44}\sigma_{54} + \sigma_{45}\sigma_{55}.$

Empirically, we can only identify the 15 terms $(\overline{A}, \overline{B}, \overline{C}, \overline{D}, \overline{E}, \overline{F}, \overline{G}, \overline{H}, \overline{I}, \overline{J}, \overline{K}, \overline{L}, \overline{M}, \overline{N}, \overline{O})$. Thus, not all 25 volatility parameters can be identified. This implies that the maximally flexible specification that is well identified has a volatility matrix given by a triangular volatility matrix[11]

$$
\Sigma = \begin{pmatrix}
\sigma_{11} & 0 & 0 & 0 & 0 \\
\sigma_{21} & \sigma_{22} & 0 & 0 & 0 \\
\sigma_{31} & \sigma_{32} & \sigma_{33} & 0 & 0 \\
\sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_{44} & 0 \\
\sigma_{51} & \sigma_{52} & \sigma_{53} & \sigma_{54} & \sigma_{55}
\end{pmatrix}.
$$

---

[11]Note that it can be either upper or lower triangular. The choice is irrelevant for the fit of the model.

References

Bank for International Settlements, 2005, "Zero-Coupon Yield Curves: Technical Documentation," BIS papers, No. 25.

Björk, Tomas and Bent J. Christensen, 1999, "Interest Rate Dynamics and Consistent Forward Rate Curves," Mathematical Finance, Vol. 9, 323-348.

Christensen, Jens H., Francis X. Diebold, and Glenn D. Rudebusch, 2007, "The Affine Arbitrage-free Class of Nelson-Siegel Term Structure Models," NBER working paper, No. 13611, November.

Coroneo, Laura, Ken Nyholm, and Rositsa Vidova-Koleva, 2008, "How Arbitrage-Free is the Nelson-Siegel Model?," working paper, European Central Bank, No. 874.

Dai, Qiang and Kenneth J. Singleton, 2000, "Specification Analysis of Affine Term Structure Models," Journal of Finance, Vol. 55, 1943-1978.

De Pooter, Michel 2007, "Examining the Nelson-Siegel Class of Term Structure Models, " Tinbergen Institute Discussion Papers No. 43, June.

Diebold, Francis X. and Canlin Li, 2006, "Forecasting the Term Structure of Government Bond Yields," Journal of Econometrics, Vol. 130, 337-364.

Diebold, Francis X., Monika Piazzesi, and Glenn D. Rudebusch, 2005, "Modeling Bond Yields in Finance and Macroeconomics," American Economic Review, Vol, 95, 415-420.

Diebold, Francis X., Glenn D. Rudebusch and S. Boragan Aruoba, 2006, "The Macroeconomy and the Yield Curve: A Dynamic Latent Factor Approach," Journal of Econometrics, Vol. 131, 309-338.

Duffee, Gregory R., 2002, "Term Premia and Interest Rate Forecasts in Affine Models," Journal of Finance, Vol. 57, 405-443.

Filipović, Damir, 1999, "A Note on the Nelson-Siegel Family," Mathematical Finance, Vol. 9, 349-359.

Gürkaynak, Refet S., Brian Sack, and Jonathan H. Wright, 2007, "The U.S. Treasury Yield Curve: 1961 to the Present," Journal of Monetary Economics, Vol. 54, 2291-2304.

Kim, Don H. and Athanasios Orphanides, 2005, "Term Structure Estimation with Survey Data on Interest Rate Forecasts," Finance and Economics Discussion Series, No. 48, Board of Governors of the Federal Reserve System.

Nelson, Charles R. and Andrew F. Siegel, 1987, "Parsimonious Modeling of Yield Curves," Journal of Business, Vol. 60, 473-489.

Rudebusch, Glenn D. and Tao Wu, 2007, "Accounting for a Shift in Term Structure Behavior with No-Arbitrage and Macro-Finance Models," Journal of Money, Credit, and Banking, Vol. 39, 395-422

Rudebusch, Glenn D. and Tao Wu, 2008, "A Macro-Finance Model of the Term Structure, Monetary Policy, and the Economy," forthcoming Economic Journal, Vol. 118.

Svensson, Lars E. O., 1995, "Estimating Forward Interest Rates with the Extended Nelson-Siegel Method," Quarterly Review, No. 3, Sveriges Riksbank, 13-26.