THE AFFINE ARBITRAGE-FREE CLASS OF
NELSON-SIEGEL TERM STRUCTURE MODELS

Jens H. E. Christensen
Francis X. Diebold
Glenn D. Rudebusch

The Affine Arbitrage-Free Class of Nelson-Siegel Term Structure Models
Jens H. E. Christensen, Francis X. Diebold, and Glenn D. Rudebusch
NBER Working Paper No. 13611
November 2007
JEL No. G1,G12

ABSTRACT

We derive the class of arbitrage-free affine dynamic term structure models that approximate the widely-used Nelson-Siegel yield-curve specification. Our theoretical analysis relates this new class of models to the canonical representation of the three-factor arbitrage-free affine model. Our empirical analysis shows that imposing the Nelson-Siegel structure on this canonical representation greatly improves its empirical tractability; furthermore, we find that improvements in predictive performance are achieved from the imposition of absence of arbitrage.

Jens H. E. Christensen
Federal Reserve Bank of San Francisco
Economic Research, MS 1130
101 Market Street
San Francisco, CA 94105-9967
jens.christensen@sf.frb.org

Glenn D. Rudebusch
Federal Reserve Bank of San Francisco
Economic Research, MS 1130
101 Market Street
San Francisco, CA 94105-9967
glenn.rudebusch@sf.frb.org

Francis X. Diebold
Department of Economics
University of Pennsylvania
3718 Locust Walk
Philadelphia, PA 19104-6297
and NBER
fdiebold@sas.upenn.edu

1 Introduction

Understanding the dynamic evolution of interest rates and the yield curve is important for many diverse tasks, such as pricing long-lived assets and their financial derivatives, managing financial risk, allocating portfolios, conducting monetary policy, purchasing capital goods, and structuring fiscal debt. To investigate yield-curve dynamics, researchers have produced a vast literature and a wide variety of models. However, many of these models have tended to be either theoretically rigorous but empirically disappointing or empirically appealing but not well grounded in theory. In this paper, we introduce a hybrid model of the yield curve that displays theoretical consistency as well as empirical tractability and fit.

Since nominal bonds trade in deep and well-organized markets, the theoretical restrictions that rule out opportunities for riskless arbitrage across maturities and over time hold a powerful appeal, and they provide the foundation for a large finance literature on arbitrage-free (AF) models that started with Vasiček (1977) and Cox, Ingersoll, and Ross (1985). These models specify the risk-neutral evolution of the underlying yield-curve factors as well as the dynamics of risk premiums. Following Duffie and Kan (1996), the affine versions of these models are particularly popular because yields are convenient linear functions of underlying latent factors (state variables that are unobserved by the econometrician) with parameters, or "factor loadings," that can be calculated from a simple system of differential equations.

Unfortunately, the canonical affine AF models can exhibit poor empirical time series performance, especially when forecasting future yields (Duffee, 2002). In addition, the estimation of these models is known to be 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). These empirical problems appear to reflect an underlying model over-parameterization, and as a solution, many researchers (e.g., Duffee, 2002, and Dai and Singleton, 2002) simply restrict to zero those parameters with small t-statistics in a first round of estimation. The resulting more parsimonious structure is typically somewhat easier to estimate and has a more robust economic interpretation (fewer troublesome likelihood maxima). However, these additional restrictions on model structure are arbitrary from both a theoretical and a statistical perspective. Furthermore, their arbitrary application, along with the computational burden of estimation, effectively precludes thorough simulation studies of the finite-sample properties of the estimators of the canonical affine model, thus, complicating model validation. In part to overcome such problems, this paper considers the application of a new, arguably less arbitrary, structure to the affine AF class of models.

Our new AF model structure is based on the workhorse yield-curve representation introduced by Nelson and Siegel (1987). The Nelson-Siegel model is a flexible curve that provides a remarkably good fit to the cross section of yields in many countries, and it is very popular among financial market practitioners and central banks (e.g., Svensson, 1995, Bank for International Settlements, 2005, and Gürkaynak, Sack, and Wright, 2006). 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 Diebold

and Li (2006) show that it forecasts the yield curve quite well. Unfortunately, despite its good empirical performance, the DNS model does not impose the desirable theoretical restrictions that rule out opportunities for riskless arbitrage (e.g., Filipović, 1999, and Diebold, Piazzesi, and Rudebusch, 2005).

In this paper, we show how to reconcile the Nelson-Siegel model with the absence of arbitrage by deriving the class of AFDNS models, which are affine AF term structure models that maintain the Nelson-Siegel factor-loading structure. These models combine the best of both yield-curve modeling traditions. They maintain the AF theoretical restrictions of the canonical affine models but can be easily and robustly estimated because the Nelson-Siegel structure helps identify the latent yield-curve factors. In particular, empirical implementation of the AFDNS models is facilitated by the fact that zero-coupon bond prices have analytical solutions, which we provide.

After deriving the new class of AFDNS models, we examine their in-sample fit and out-of-sample forecast performance relative to standard DNS models. For both the DNS and the AFDNS models, we estimate parsimonious and flexible versions (with independent factors and more richly parameterized correlated factors, respectively). We find that the flexible versions of both models are preferred for in-sample fit; however, the parsimonious versions exhibit significantly better out-of-sample forecast performance. Finally, and most importantly, we find that the parsimonious AFDNS model outperforms its DNS counterpart in forecasting, which supports the imposition of the AF restrictions.

We proceed as follows. Section 2 introduces the DNS model and derives the main theoretical result of the paper, which defines the AFDNS class of models. Section 3 derives the relationship between the AFDNS class of models and the canonical representation of affine AF models introduced by Dai and Singleton (2000). For the four specific DNS and AFDNS models used in our empirical analysis, Section 4 describes the estimation method, data, and in-sample fit, while Section 5 examines out-of-sample forecast performance. Section 6 concludes, and appendices contain additional technical details.

2 Nelson-Siegel term structure models

In this section, we review the DNS model and introduce the AFDNS class of arbitrage-free affine term structure models that maintain the Nelson-Siegel factor loading structure.

2.1 The dynamic Nelson-Siegel model

The original Nelson-Siegel model fits the yield curve with the simple functional form

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

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

As noted earlier, this representation is commonly used by financial market practitioners to fit the yield curve at a point in time. Although for some purposes such a static representation

appears useful, a dynamic version is required to understand the evolution of the bond market over time. Therefore, Diebold and Li (2006) reinterpret the β coefficients as time-varying factors L_t, S_t, and C_t, so

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

Given their Nelson-Siegel factor loadings, these factors can be interpreted as level, slope, and curvature. Diebold and Li assume an autoregressive structure for these three factors, which yields the DNS model—a fully dynamic Nelson-Siegel specification.

Empirically, the DNS model is very tractable and provides a good fit to the data; however, as a theoretical matter, it does not require that the dynamic evolution of yields and the yield curve at any point in time cohere such that arbitrage opportunities are precluded. Indeed, the results of Filipović (1999) imply that *whatever* stochastic dynamics are chosen for the DNS factors, it is impossible to rule out arbitrage at the bond prices implicit in the resulting Nelson-Siegel yield curve. Hence, the discounted prices of zero-coupon bonds in the DNS model are not semimartingale processes under the pricing or Q-measure. The next subsection shows how to remedy this theoretical weakness.

2.2 The AFDNS model

Our derivation of the AFDNS model starts from the standard continuous-time affine AF structure (Duffie and Kan, 1996).[1] To represent an affine diffusion process, define a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t), Q)$, where the filtration $(\mathcal{F}_t) = \{\mathcal{F}_t : t \geq 0\}$ satisfies the usual conditions (Williams, 1997). The state variable X_t is assumed to be a Markov process defined on a set $M \subset \mathbf{R}^n$ that solves the following stochastic differential equation (SDE)[2]

$$dX_t = K^Q(t)[\theta^Q(t) - X_t]dt + \Sigma(t)D(X_t, t)dW_t^Q, \tag{3}$$

where W^Q is a standard Brownian motion in $\mathbf{R}^n$, the information of which is contained in the filtration $(\mathcal{F}_t)$. The drift terms $\theta^Q : [0, T] \to \mathbf{R}^n$ and $K^Q : [0, T] \to \mathbf{R}^{n \times n}$ are bounded, continuous functions.[3] Similarly, the volatility matrix $\Sigma : [0, T] \to \mathbf{R}^{n \times n}$ is assumed to be a bounded, continuous function, while $D : M \times [0, T] \to \mathbf{R}^{n \times n}$ is assumed to have the following diagonal structure

$$\begin{pmatrix} \sqrt{\gamma^1(t) + \delta_1^1(t)X_t^1 + \ldots + \delta_n^1(t)X_t^n} & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & \sqrt{\gamma^n(t) + \delta_1^n(t)X_t^1 + \ldots + \delta_n^n(t)X_t^n} \end{pmatrix}.$$

---

[1] Krippner (2006) derives a special case of the AFDNS model with constant risk premiums.

[2] The affine property applies to bond prices; therefore, affine models only impose structure on the factor dynamics under the pricing measure.

[3] Stationarity of the state variables is ensured if all the eigenvalues of $K^Q(t)$ are positive (if complex, the real component should be positive), see Ahn, Dittmar, and Gallant (2002). However, stationarity is not a necessary requirement for the process to be well defined.

To simplify the notation, $\gamma(t)$ and $\delta(t)$ are defined as

$$\gamma(t) = \begin{pmatrix} \gamma^1(t) \\ \vdots \\ \gamma^n(t) \end{pmatrix} \quad \text{and} \quad \delta(t) = \begin{pmatrix} \delta_1^1(t) & \cdots & \delta_n^1(t) \\ \vdots & \ddots & \vdots \\ \delta_1^n(t) & \cdots & \delta_n^n(t) \end{pmatrix},$$

where $\gamma : [0,T] \to \mathbf{R}^n$ and $\delta : [0,T] \to \mathbf{R}^{n \times n}$ are bounded, continuous functions. Given this notation, the SDE of the state variables can be written as

$$dX_t = K^Q(t)[\theta^Q(t) - X_t]dt + \Sigma(t) \begin{pmatrix} \sqrt{\gamma^1(t) + \delta^1(t)X_t} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sqrt{\gamma^n(t) + \delta^n(t)X_t} \end{pmatrix} dW_t^Q,$$

where $\delta^i(t)$ denotes the ith row of the $\delta(t)$-matrix. Finally, the instantaneous risk-free rate is assumed to be an affine function of the state variables

$$r_t = \rho_0(t) + \rho_1(t)'X_t,$$

where $\rho_0 : [0,T] \to \mathbf{R}$ and $\rho_1 : [0,T] \to \mathbf{R}^n$ are bounded, continuous functions.

Duffie and Kan (1996) prove that zero-coupon bond prices in this framework are exponential-affine functions of the state variables

$$P(t,T) = E_t^Q \left[\exp\left(-\int_t^T r_u du \right) \right] = \exp\left(B(t,T)'X_t + C(t,T) \right),$$

where $B(t,T)$ and $C(t,T)$ are the solutions to the following system of ordinary differential equations (ODEs)

$$\frac{dB(t,T)}{dt} = \rho_1 + (K^Q)'B(t,T) - \frac{1}{2}\sum_{j=1}^{n}(\Sigma'B(t,T)B(t,T)'\Sigma)_{j,j}(\delta^j)', \quad B(T,T) = 0 \quad (4)$$

$$\frac{dC(t,T)}{dt} = \rho_0 - B(t,T)'K^Q\theta^Q - \frac{1}{2}\sum_{j=1}^{n}(\Sigma'B(t,T)B(t,T)'\Sigma)_{j,j}\gamma^j, \quad C(T,T) = 0, \quad (5)$$

and the possible time-dependence of the parameters is suppressed in the notation. These pricing functions imply that the zero-coupon yields are given by

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

Therefore, for a three-factor affine model with $X_t = (X_t^1, X_t^2, X_t^3)$, the closest approximation to the Nelson-Siegel yield function is a yield function of the form

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

4

so the ODEs for the $B(t,T)$ functions have these solutions:

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

As described in the following proposition, there exists a unique class of affine AF models that satisfy these ODEs.

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 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

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

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 ODEs:

$$
\begin{pmatrix} \frac{dB^1(t,T)}{dt} \\ \frac{dB^2(t,T)}{dt} \\ \frac{dB^3(t,T)}{dt} \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & -\lambda & \lambda \end{pmatrix} \begin{pmatrix} B^1(t,T) \\ B^2(t,T) \\ B^3(t,T) \end{pmatrix} \tag{6}
$$

and

$$
\frac{dC(t,T)}{dt} = -B(t,T)'K^Q\theta^Q - \frac{1}{2}\sum_{j=1}^{3} \left(\Sigma' B(t,T)B(t,T)'\Sigma \right)_{j,j}, \tag{7}
$$

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 Appendix A.

This proposition defines the AFDNS model and has several interesting implications. First, the three state variables are Gaussian Ornstein-Uhlenbeck processes with a constant volatility matrix Σ.[4] The instantaneous interest rate is the sum of level and slope factors (X_t^1 and X_t^2), while the curvature factor (X_t^3) is a truly latent factor in the sense that its sole role is as a stochastic time-varying mean for the slope factor under the Q-measure. Second, Proposition 1 only imposes structure on the dynamics of the AFDNS model under the Q-measure and is silent about the dynamics under the P-measure. Still, the observation that curvature is a truly latent factor generally accords with the empirical literature where it has been difficult to find sensible interpretations of curvature under the P-measure (Diebold, Rudebusch, and Aruoba, 2006). Similarly, the level factor is a unit-root process under the Q-measure, which accords with the usual finding that one or more of the interest rate factors are close to being nonstationary processes under the P-measure.[5] Third, Proposition 1 provides insight into the nature of the parameter λ. Nelson and Siegel (1987) allowed λ to vary at each point in time, but subsequent dynamic implementations of the Nelson-Siegel model tended to fix λ over the sample (e.g., Diebold

---

[4]Proposition 1 can be extended to include jumps in the state variables. As long as the jump arrival intensity is state-independent, the Nelson-Siegel factor loading structure in the yield function is maintained since only $C(t,T)$ is affected by the inclusion of such jumps. See Duffie, Pan, and Singleton (2000) for the needed modification of the ODEs for $C(t,T)$ in this case.

[5]With the unit root in the level factor, as maturity increases, $-\frac{C(t,T)}{T-t} \to -\infty$, which implies that, strictly speaking, this model is not arbitrage-free. However, if we modify the mean-reversion matrix K^Q to

$$
K^Q(\varepsilon) = \begin{pmatrix} \varepsilon & 0 & 0 \\ 0 & \lambda & -\lambda \\ 0 & 0 & \lambda \end{pmatrix}
$$

and consider a converging sequence $\varepsilon_n > 0$, $\varepsilon_n \downarrow 0$, then there is a converging sequence of AF models with a limit given by the result in Proposition 1. Thus, by choosing $\varepsilon > 0$ sufficiently small, we can obtain an AF model that is indistinguishable from the AFDNS model in Proposition 1.

and Li, 2006). In the AFDNS model, λ is indeed a constant, namely, the mean-reversion rate of the curvature and slope factors as well as the scale by which a deviation of the curvature factor from its mean affects the mean of the slope factor. Fourth, relative to the Nelson-Siegel model, the AFDNS model contains an additional maturity-dependent term $-\frac{C(t,T)}{T-t}$ in the function for the zero-coupon bond yields. The nature of this "yield-adjustment" term is crucial in assessing differences between the AFDNS and DNS models, and we now turn to a theoretical analysis of this term.

2.3 The AFDNS yield-adjustment term

The only parameters in the system of ODEs for the AFDNS $B(t,T)$ functions are ρ_1 and K^Q, i.e., the factor loadings of r_t and the mean-reversion structure for the state variables under the Q-measure. The drift term θ^Q and the volatility matrix Σ do not appear in the ODEs but in the yield-adjustment term $-\frac{C(t,T)}{T-t}$. Therefore, in the AFDNS model, the choice of the volatility matrix Σ affects both the P-dynamics and the yield function through the yield-adjustment term. In contrast, the DNS model is silent about the real-world dynamics of the state variables, so the choice of P-dynamics is irrelevant for the yield function.

As discussed in the next section, we identify the AFDNS models by fixing the mean levels of the state variables under the Q-measure at 0, i.e., $\theta^Q = 0$. This implies that the yield-adjustment term will have the following form:

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

Given a general volatility matrix

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

the yield-adjustment term can be derived in analytical form (see Appendix B) 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 AFDNS class of models are given by an analytical formula will greatly facilitate 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 AFDNS specification that can be identified has a triangular volatility matrix given by[6]

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

In Section 4, we quantify the yield-adjustment term and examine how it affects the empirical performance of two specific AFDNS models relative to their corresponding DNS models. These models are introduced next.

2.4 Four specific Dynamic Nelson-Siegel models

In general, the DNS and AFDNS models are silent about the P-dynamics, so there are an infinite number of possible specifications that could be used to match the data. However, for continuity with the existing literature, our econometric analysis focuses on two specific versions of the DNS model that have been estimated in recent studies, and, for consistency, we also examine the two corresponding versions of the AFDNS model.

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

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

---

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

The *correlated-factor DNS model* has factor P-dynamics described by a first-order vector autoregression (VAR(1))

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

as in Diebold, Rudebusch, and Aruoba (2006). The innovations $\eta_t(L)$, $\eta_t(S)$, and $\eta_t(C)$ are allowed to be correlated with a conditional covariance matrix given by $Q = qq'$, where the Cholesky factor q of the covariance matrix Q is

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

In both of these DNS models, 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 i.i.d. white noise.

The corresponding AFDNS models are 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. In order 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,

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

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 Q-dynamic structure described in Proposition 1. Therefore, we focus on two AFDNS models that correspond to the specific two DNS models above.

In the *independent-factor AFDNS model*, all three factors are assumed to be independent under the P-measure

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

This model is the AF equivalent of our first DNS model.

In the *correlated-factor AFDNS model*, the three shocks may be correlated, and there may be full interaction among the factors as they adjust to the steady state

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

This is the most flexible version of the AFDNS models where all parameters are identified.

For both AFDNS models, 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.

3 The AFDNS subclass of canonical affine AF models

Before proceeding to an empirical analysis of the various DNS and AFDNS models, we first answer a key theoretical question: What, precisely, are the restrictions that the AFDNS model imposes on the canonical representation of three-factor affine AF models—the $A_0(3)$ representation (with three state variables and zero square-root processes) as defined in Dai and Singleton (2000). Denoting the state variables by Y_t, the canonical $A_0(3)$ model is given by

$$\begin{aligned} r_t &= \delta_0^Y + (\delta_1^Y)' Y_t \\ dY_t &= K_Y^P[\theta_Y^P - Y_t]dt + \Sigma_Y dW_t^P \\ dY_t &= K_Y^Q[\theta_Y^Q - Y_t]dt + \Sigma_Y dW_t^Q, \end{aligned}$$

with $\delta_0^Y \in \mathbf{R}$, $\delta_1^Y, \theta_Y^P, \theta_Y^Q \in \mathbf{R}^3$, and $K_Y^P, K_Y^Q, \Sigma_Y \in \mathbf{R}^{3\times3}$. If the essentially affine risk premium specification $\Gamma_t = \gamma_Y^0 + \gamma_Y^1 Y_t$ is imposed on the model, the drift terms under the P-measure (K_Y^P, θ_Y^P) can be chosen independently of the drift terms under the Q-measure (K_Y^Q, θ_Y^Q).

Because the latent state variables may rotate without changing the probability distribution of bond yields, not all parameters in the above model can be identified. Dai and Singleton (2000) impose identifying restrictions under the P-measure but not under the Q-measure. Specifically,

they set the mean $\theta_Y^P = 0$, the volatility matrix Σ_Y equal to the identity matrix, and the mean-reversion matrix K_Y^P equal to a triangular matrix.[7] Thus, their canonical representation has P-dynamics given by

$$
\begin{pmatrix} dY_t^1 \\ dY_t^2 \\ dY_t^3 \end{pmatrix} = - \begin{pmatrix} \kappa_{11}^{Y,P} & \kappa_{12}^{Y,P} & \kappa_{13}^{Y,P} \\ 0 & \kappa_{22}^{Y,P} & \kappa_{23}^{Y,P} \\ 0 & 0 & \kappa_{33}^{Y,P} \end{pmatrix} \begin{pmatrix} Y_t^1 \\ Y_t^2 \\ Y_t^3 \end{pmatrix} dt + \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} dW_t^{1,P} \\ dW_t^{2,P} \\ dW_t^{3,P} \end{pmatrix},
$$

and Q-dynamics given by

$$
\begin{pmatrix} dY_t^1 \\ dY_t^2 \\ dY_t^3 \end{pmatrix} = \begin{pmatrix} \kappa_{11}^{Y,Q} & \kappa_{12}^{Y,Q} & \kappa_{13}^{Y,Q} \\ \kappa_{21}^{Y,Q} & \kappa_{22}^{Y,Q} & \kappa_{23}^{Y,Q} \\ \kappa_{31}^{Y,Q} & \kappa_{32}^{Y,Q} & \kappa_{33}^{Y,Q} \end{pmatrix} \left[\begin{pmatrix} \theta_1^{Y,Q} \\ \theta_2^{Y,Q} \\ \theta_3^{Y,Q} \end{pmatrix} - \begin{pmatrix} Y_t^1 \\ Y_t^2 \\ Y_t^3 \end{pmatrix} \right] dt + \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} dW_t^{1,Q} \\ dW_t^{2,Q} \\ dW_t^{3,Q} \end{pmatrix}.
$$

The instantaneous risk-free rate is given by

$$
r_t = \delta_0^Y + \delta_{1,1}^Y Y_t^1 + \delta_{1,2}^Y Y_t^2 + \delta_{1,3}^Y Y_t^3.
$$

Appealing to mathematical symmetry and the latent nature of the state variables, it seems that the canonical identifying restrictions could have been imposed on the Q-dynamics instead of the P-dynamics, namely fixing $\theta_Y^Q = 0$ and requiring K_Y^Q to be triangular. However, from an economic point of view, these two identifications may not be equivalent, because the yield function being fit to the observed yields is determined solely by the dynamics under the Q-measure so imposing restrictions on the Q-measure drift terms could limit the ability of the model to fit observed yields. The following proposition provides a simple condition under which the two types of identifying restrictions are indeed equivalent.[8]

Proposition 2:

Consider an $A_0(3)$ model with dynamics given by the canonical representation described in Dai and Singleton (2000)

$$
\begin{aligned}
dY_t &= -K_Y^P Y_t dt + I dW_t^P \\
dY_t &= K_Y^Q [\theta_Y^Q - Y_t] dt + I dW_t^Q,
\end{aligned}
$$

with the risk-free rate given by $r_t = \delta_0^Y + (\delta_1^Y)' Y_t$. If the eigenvalues of the 3×3 matrix K_Y^Q are all real, then there exists an equivalent representation of this model with dynamics given by

$$
\begin{aligned}
dZ_t &= K_Z^P [\theta_Z^P - Z_t] dt + I d\widehat{W}_t^P \\
dZ_t &= -K_Z^Q Z_t dt + I d\widehat{W}_t^Q,
\end{aligned}
$$

---

[7]Without loss of generality, we will take it to be upper triangular in the following.

[8]"Equivalent" representations generate the same distributions for the short-rate process and bond yields and, in estimation, have identical fit (Dai and Singleton, 2000).

where K_Z^Q is an upper-triangular 3×3 matrix, and the risk-free rate is given by $r_t = \delta_0^Z + (\delta_1^Z)'Z_t$.

Proof: See Appendix C.[9]

Proposition 2 shows that if the eigenvalues of the mean-reversion matrix under the Q-measure are real, then it does not matter under which measure we impose the identifying restrictions. This is useful because Q-measure identification is imposed in Proposition 1 for the AFDNS model. The AFDNS mean-reversion matrix under the Q-measure is

$$
\begin{pmatrix}
0 & 0 & 0 \\
0 & \lambda & -\lambda \\
0 & 0 & \lambda
\end{pmatrix}.
$$

The eigenvalues of this matrix, 0 and λ, are both real numbers; thus, by construction, any canonical representation equivalent of the AFDNS model must satisfy the requirement in Proposition 2.[10] Hence there is no loss of generality when we fix the AFDNS model mean under the Q-measure at 0 and leave the mean under the P-measure, θ^P, to be estimated.

Proposition 2 also facilitates the identification of the restrictions on the canonical affine representation that are required to obtain the class of AFDNS models. Indeed, we can, without loss of generality, limit our focus to $A_0(3)$ models where the identifying restrictions have been imposed under the Q-measure, namely

$$
\begin{pmatrix} dZ_t^1 \\ dZ_t^2 \\ dZ_t^3 \end{pmatrix} = - \begin{pmatrix} \kappa_{11}^{Z,Q} & \kappa_{12}^{Z,Q} & \kappa_{13}^{Z,Q} \\ 0 & \kappa_{22}^{Z,Q} & \kappa_{23}^{Z,Q} \\ 0 & 0 & \kappa_{33}^{Z,Q} \end{pmatrix} \begin{pmatrix} Z_t^1 \\ Z_t^2 \\ Z_t^3 \end{pmatrix} dt + \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} dW_t^{1,Q} \\ dW_t^{2,Q} \\ dW_t^{3,Q} \end{pmatrix},
$$

while the P-dynamics are given by

$$
\begin{pmatrix} dZ_t^1 \\ dZ_t^2 \\ dZ_t^3 \end{pmatrix} = \begin{pmatrix} \kappa_{11}^{Z,P} & \kappa_{12}^{Z,P} & \kappa_{13}^{Z,P} \\ \kappa_{21}^{Z,P} & \kappa_{22}^{Z,P} & \kappa_{23}^{Z,P} \\ \kappa_{31}^{Z,P} & \kappa_{32}^{Z,P} & \kappa_{33}^{Z,P} \end{pmatrix} \left[\begin{pmatrix} \theta_1^{Z,P} \\ \theta_2^{Z,P} \\ \theta_3^{Z,P} \end{pmatrix} - \begin{pmatrix} Z_t^1 \\ Z_t^2 \\ Z_t^3 \end{pmatrix} \right] dt + \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} dW_t^{1,P} \\ dW_t^{2,P} \\ dW_t^{3,P} \end{pmatrix}.
$$

The procedure through which the restrictions are identified is based on so-called *affine invariant transformations*. Appendix C describes such transformations, and Appendix D derives the restrictions associated with the AFDNS model, which are summarized in Table 1.

Table 1 shows that for the correlated-factor AFDNS model, there are three key parameter restrictions on the canonical affine model (which itself has a total of 22 free parameters). First,

---

[9]We are grateful to Peter Feldhütter for helpful discussions about this proof, which obviously generalizes to the $A_0(n)$ model.

[10]This follows from the fact that the relationship between the mean-reversion matrix of a model and its transformed equivalent is $K^Z = AK^Y A^{-1}$, where Y refers to the original model, Z refers to the transformed model, and A is a nonsingular square matrix defining the transformation (see appendix C for more details). The eigenvalues of the transformed mean-reversion matrix are given by the solutions to the characteristic equation

$$
|K^Z - \gamma I| = |A^{-1}K^Y A - \gamma A^{-1}A| = |A^{-1}| \cdot |K^Y - \gamma I| \cdot |A| = |K^Y - \gamma I| = 0.
$$

Hence K^Z will have the same eigenvalues as K^Y.

AFDNS Model	δ_0^Z, δ_1^Z	κ_Z^Q	κ_Z^P	θ_Z^P	No. restrictions
Independent-factor	$\delta_0^Z = 0,$ $\delta_{1,3}^Z = 0$	$\kappa_{1,1}^{Z,Q} = \kappa_{1,2}^{Z,Q} = \kappa_{1,3}^{Z,Q} = 0,$ $\kappa_{2,2}^{Z,Q} = \kappa_{3,3}^{Z,Q}$	κ_Z^P is diagonal	No restriction	12
Correlated-factor	$\delta_0^Z = 0$	$\kappa_{1,1}^{Z,Q} = 0,$ $\kappa_{2,2}^{Z,Q} = \kappa_{3,3}^{Z,Q}$	No restriction	No restriction	3

Table 1: **AFDNS Model Parameter Restrictions on the Canonical Representation**
These are the restrictions on the $A_0(3)$ model needed to obtain the independent-factor and correlated-factor AFDNS specifications.

$\delta_0^Z = 0$, so there is no constant in the equation for the instantaneous risk-free rate. There is no need for this constant because, with the second restriction $\kappa_{1,1}^{Z,Q} = 0$, the first factor must be a unit-root process under the Q-measure, which also implies that this factor can be identified as the level factor. Finally, $\kappa_{2,2}^{Z,Q} = \kappa_{3,3}^{Z,Q}$, so the own mean-reversion rates of the second and third factors under the Q-measure must be identical. The independent-factor AFDNS model maintains these three parameter restrictions and adds nine others under both the P- and Q-measures. (For both specifications, there is a further modest restriction described in Appendix D: $\kappa_{2,3}^{Z,Q}$ must have the opposite sign of $\kappa_{2,2}^{Z,Q}$ and $\kappa_{3,3}^{Z,Q}$, but its absolute numerical size can vary freely.)

The Nelson-Siegel parameter restrictions on the canonical affine AF model greatly facilitate estimation.[11] They allow a closed-form solution and, as described in the next section, eliminate in an appealing way the surfeit of troublesome likelihood maxima in estimation.[12]

4 Estimation of the DNS and AFDNS models

Here we describe estimation methods and results for the DNS and AFDNS models.

4.1 Estimation methods

The Kalman filter is an efficient and consistent estimator for both the DNS and AFDNS models. For the DNS models, the state equation is

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

where $X_t = (L_t, S_t, C_t)$, 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 stationarity of the state variables, which is ensured by imposing that the eigenvalues of A are smaller than 1.

---

[11]Note that in the AFDNS model, the connection between the P-dynamics and the yield function is explicitly tied to the yield adjustment term through the specification of the volatility matrix, while in the canonical representation it is blurred by an interplay between the specifications of δ_1^Z and K_Z^Q.

[12]This contrasts with the common practice, mentioned earlier, of zeroing out an arbitrary set of individual coefficients.

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

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

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

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

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 ensured 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[13]

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

Finally, the AFDNS measurement equation is given by

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

For both types of models, the error structure is

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

where H is a diagonal matrix

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

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

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

[13]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_{t-1}	μ		q	1	2	3
L_t	0.9827	0	0	0.06958		1	0.002485	0	0
	(0.0128)			(0.0137)			(0.000153)		
S_t	0	0.9778	0	-0.02487		2	0	0.003329	0
		(0.0166)		(0.0151)				(0.000194)	
C_t	0	0	0.9189	-0.01075		3	0	0	0.007471
			(0.0284)	(0.00786)					(0.000396)

Table 2: **Parameter Estimates of the Independent-Factor DNS Model.**
The left-hand panel contains the estimated A matrix and μ vector. The right-hand panel contains the estimated q matrix. Estimated standard deviations of the parameter estimates are given in parentheses. The associated estimated λ is 0.06040 (when yield maturities are measured in months) with a standard deviation of 0.00100. The maximized log-likelihood value is 16332.94.

A	L_{t-1}	S_{t-1}	C_{t-1}	μ		q	1	2	3
L_t	0.9874	0.0050	-0.0097	0.0723		1	0.002457	0	0
	(0.0165)	(0.0183)	(0.0157)	(0.0145)			(0.000147)		
S_t	0.0066	0.9332	0.0819	-0.0294		2	-0.002227	0.002265	0
	(0.0228)	(0.0229)	(0.0202)	(0.0159)			(0.000255)	(0.000110)	
C_t	0.0152	0.0401	0.9011	-0.0120		3	0.002752	0.000618	0.006554
	(0.0526)	(0.0418)	(0.0377)	(0.0126)			(0.000706)	(0.000610)	(0.000441)

Table 3: **Parameter Estimates of the Correlated-Factor DNS Model.**
The left-hand panel contains the estimated A matrix and μ vector. The right-hand panel contains the estimated q matrix. Estimated standard deviations of the parameter estimates are given in parentheses. The associated estimated λ is 0.06248 (when yield maturities are measured in months) with a standard deviation of 0.00109. The maximum log-likelihood value is 16415.36.

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.

4.2 DNS model estimation results

In this subsection, we present estimation results for the two versions of the DNS model. These specifications, along with the two AFDNS specifications described in the next subsection, are estimated using monthly data on U.S. Treasury security yields from January 1987 to December 2002. The data are end-of-month, unsmoothed Fama-Bliss (1987) zero-coupon yields at the following 16 maturities: 3, 6, 9, 12,18, 24, 36, 48, 60, 84, 96, 108, 120, 180, 240, and 360 months.

The estimates of the DNS models with independent and correlated factors are shown in Tables 2 and 3, respectively. In both models, the level factor is the most persistent factor, while the curvature factor has the fastest rate of mean-reversion. Interestingly, for the correlated factors DNS model, the only significant off-diagonal element (the 0.0819) in the estimated A-matrix is $A_{S_t,C_{t-1}}$, which is the key non-zero off-diagonal element required in Proposition 1 for the AFDNS specification.

Volatility parameters will be most easily compared by using the one-month conditional covari-

ance matrices for the independent-factor model

$$Q_{indep}^{DNS} = qq' = \begin{pmatrix} 6.17 \times 10^{-6} & 0 & 0 \\ 0 & 1.11 \times 10^{-5} & 0 \\ 0 & 0 & 5.58 \times 10^{-5} \end{pmatrix} \tag{9}$$

and the correlated-factor model

$$Q_{corr}^{DNS} = qq' = \begin{pmatrix} 6.03 \times 10^{-6} & -5.47 \times 10^{-6} & 6.76 \times 10^{-6} \\ -5.47 \times 10^{-6} & 1.01 \times 10^{-5} & -4.73 \times 10^{-6} \\ 6.76 \times 10^{-6} & -4.73 \times 10^{-6} & 5.09 \times 10^{-5} \end{pmatrix}. \tag{10}$$

Across the two models, the variances of each state variable are similar, with the level factor the least volatile and the curvature factor the most volatile. The covariance estimates obtained in the correlated-factor DNS model translate into a correlation of -0.701 for innovations to the level and slope factor, a correlation of 0.385 for innovations to the level and curvature factor, and a correlation of -0.208 for innovations to the slope and curvature factor.

The two DNS models are nested, so we can test the independent-factor restricted parameter set $\theta_{indep.}$ versus the correlated-factor unrestricted parameter set $\theta_{corr.}$ with a likelihood ratio test

$$LR = 2[\log L(\theta_{corr}) - \log L(\theta_{indep})] = 164.8 \sim \chi^2(q),$$

where q, the number of parameter restrictions, equals nine. The associated p-value is less than .0001, so the restrictions imposed in the independent-factor DNS model are not supported by the data.[14] Still, the increased flexibility of the correlated-factor DNS model provides little advantage in fitting the observed yields.[15] Table 4 reports summary statistics for the fitted errors for each of the four models considered in this study. For the two DNS models, the differences in RMSEs at any maturity are not large (less than 0.58 basis points), and there is no consistent advantage for the correlated factors model. Interestingly, both models have difficulty fitting yields beyond the 10-year maturity, which suggests that a maturity-dependent yield adjustment term, as in the AFDNS models that we turn to next, could improve fit.

4.3 AFDNS model estimation results

As many have noted, estimation of the canonical affine $A_0(3)$ term structure model is very difficult and time-consuming and effectively prevents the kind of repetitive re-estimation required in a comprehensive simulation study or out-of-sample forecast exercise, which we pursue with the AFDNS model in the next section.[16] By comparison, the estimation of the AFDNS model is

---

[14]This rejection reflects an elevated negative correlation between the innovations to the level and slope factor and a significant positive correlation through the mean-reversion matrix between changes in the slope factor and deviations of the curvature factor from its mean.

[15]The similarity in fit is not too surprising, since there is no direct connection in these DNS models between the yield function and the assumed P-dynamics of the state variables. Indeed, across the two models, the level, slope, and curvature factors are very highly correlated.

[16]For example, Rudebusch, Swanson, and Wu (2006) report being unable to replicate the published estimates of a no-arbitrage model even though they use the same data and programs that generated the model's parameter

Maturity in months	DNS indep.-factor		DNS corr.-factor		AFDNS indep.-factor		AFDNS corr.-factor	
	Mean	RMSE	Mean	RMSE	Mean	RMSE	Mean	RMSE
3	-1.64	12.26	-1.84	11.96	-2.85	18.54	-2.47	11.53
6	-0.24	1.09	-0.29	1.34	-1.19	7.12	-0.04	0.75
9	-0.54	7.13	-0.51	6.92	-1.24	3.44	-0.35	6.86
12	4.04	11.19	4.11	10.86	3.58	9.60	3.69	10.11
18	7.22	10.76	7.28	10.42	7.15	10.44	5.49	8.31
24	1.18	5.83	1.19	5.29	1.37	5.94	-1.20	4.37
36	-0.07	1.51	-0.19	2.09	0.31	1.98	-1.10	3.16
48	-0.67	3.92	-0.85	4.03	-0.39	3.72	0.94	4.14
60	-5.33	7.13	-5.51	7.31	-5.27	6.82	-1.99	5.20
84	-1.22	4.25	-1.30	4.25	-1.50	4.29	0.90	3.83
96	1.31	2.10	1.29	2.02	1.02	2.11	1.05	1.83
108	0.03	2.94	0.07	3.11	-0.11	3.02	-3.24	5.28
120	-5.11	8.51	-5.01	8.53	-4.96	8.23	-11.67	14.02
180	24.11	29.44	24.40	29.66	27.86	32.66	3.76	16.50
240	25.61	34.99	26.00	35.33	35.95	42.61	4.20	23.93
360	-29.62	37.61	-29.12	37.18	1.37	22.04	-0.81	23.02

Table 4: **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.

straightforward and robust in large part because the role of each latent factor is not left unidentified as in the maximally flexible $A_0(3)$ model. Even though the factors are latent in the AFDNS model, with the Nelson-Siegel factor loading structure, they can be clearly identified as level, slope, and curvature. This identification eliminates the troublesome local maxima reported by Kim and Orphanides (2005), i.e. maxima with likelihood values very close to the global maximum but with very different interpretations of the three factors and their dynamics.[17]

The estimated parameters of the independent-factor AFDNS model are reported in Table 5. The factor means are close to those of the DNS model. To compare the mean-reversion parameters, we translate the continuous-time matrix in Table 5 into the one-month conditional mean-reversion matrix

$$\exp\left(-K^P\frac{1}{12}\right) = \begin{pmatrix} 0.994 & 0 & 0 \\ 0 & 0.983 & 0 \\ 0 & 0 & 0.903 \end{pmatrix}. \tag{11}$$

We also convert the volatility matrix into a one-month conditional covariance matrix

$$Q_{indep}^{AFDNS} = \int_0^{\frac{1}{12}} e^{-K^P s}\Sigma\Sigma'^{-(K^P)'s}ds = \begin{pmatrix} 2.15 \times 10^{-6} & 0 & 0 \\ 0 & 9.97 \times 10^{-6} & 0 \\ 0 & 0 & 5.28 \times 10^{-5} \end{pmatrix}. \tag{12}$$

These too appear little different from the ones reported for the independent-factor DNS model.

estimates.

[17]Other strategies to facilitate estimation include adding survey information (Kim and Orphanides, 2005) or assuming the latent yield-curve factors are observable (Ang, Piazzesi, and Wei, 2006).

K^P	$K^P_{\cdot,1}$	$K^P_{\cdot,2}$	$K^P_{\cdot,3}$	μ^P		Σ	$\Sigma_{\cdot,1}$	$\Sigma_{\cdot,2}$	$\Sigma_{\cdot,3}$
$K^P_{1,\cdot}$	0.06734	0	0	0.07243		$\Sigma_{1,\cdot}$	0.005095	0	0
	(0.0623)			(0.0148)			(0.000139)		
$K^P_{2,\cdot}$	0	0.2083	0	-0.02825		$\Sigma_{2,\cdot}$	0	0.01103	0
		(0.179)		(0.0173)				(0.000580)	
$K^P_{3,\cdot}$	0	0	1.230	-0.009266		$\Sigma_{3,\cdot}$	0	0	0.02647
			(0.423)	(0.00609)					(0.00137)

Table 5: **Parameter Estimates of the Independent-Factor AFDNS Model.**
The left-hand panel contains the estimated K^P matrix and μ vector. The right-hand panel contains the estimated Σ matrix. Estimated standard deviations of the parameter estimates are given in parentheses. The associated estimated λ is 0.5971 with a standard deviation of 0.0115. The maximum log-likelihood value is 16279.55.

K^P	$K^P_{\cdot,1}$	$K^P_{\cdot,2}$	$K^P_{\cdot,3}$	μ^P		Σ	$\Sigma_{\cdot,1}$	$\Sigma_{\cdot,2}$	$\Sigma_{\cdot,3}$
$K^P_{1,\cdot}$	4.729	8.046	-9.730	0.07982		$\Sigma_{1,\cdot}$	0.01542	0	0
	(1.14)	(1.19)	(1.23)	(0.00738)			(0.000364)		
$K^P_{2,\cdot}$	-0.8584	-0.3617	0.5775	-0.03798		$\Sigma_{2,\cdot}$	-0.003763	0.01088	0
	(1.12)	(1.96)	(2.38)	(0.0200)			(0.00480)	(0.00174)	
$K^P_{3,\cdot}$	-32.89	-59.34	72.49	-0.02636		$\Sigma_{3,\cdot}$	-0.1615	-0.05981	0.01457
	(9.68)	(1.01)	(1.02)	(0.0189)			(0.00658)	(0.0102)	(0.0430)

Table 6: **Parameter Estimates of the Correlated-Factor AFDNS Model.**
The left-hand panel contains the estimated K^P matrix and μ vector. The right-hand panel contains the estimated Σ matrix. Estimated standard deviations of the parameter estimates are given in parentheses. The associated estimated λ is 0.8219 with a standard deviation of 0.0122. The maximum log-likelihood value is 16492.00.

Still, although the two independent-factor models are non-nested, they contain the same number of parameters, and the lower log-likelihood value obtained for the AFDNS model (16279 vs. 16332) suggests a slightly weaker in-sample performance for that model, which appears consistent with the RMSEs in Table 4.

Similar fit to the data by the two models is not too surprising because they make identical assumptions about the P-dynamics, so the only difference between the two models is the inclusion of the yield-adjustment term in the AFDNS model yield function. For the independent-factor AFDNS model, this term is given by

$$
\begin{aligned}
-\frac{C(t,T)}{T-t} &= -\frac{\sigma_{11}^2}{2}\frac{1}{T-t}\int_t^T B^1(s,T)^2 ds - \frac{\sigma_{22}^2}{2}\frac{1}{T-t}\int_t^T B^2(s,T)^2 ds - \frac{\sigma_{33}^2}{2}\frac{1}{T-t}\int_t^T B^3(s,T)^2 ds \\
&= -\sigma_{11}^2\frac{(T-t)^2}{6} - \sigma_{22}^2\Big[\frac{1}{2\lambda^2} - \frac{1}{\lambda^3}\frac{1-e^{-\lambda(T-t)}}{T-t} + \frac{1}{4\lambda^3}\frac{1-e^{-2\lambda(T-t)}}{T-t}\Big] \\
&\quad - \sigma_{33}^2\Big[\frac{1}{2\lambda^2} + \frac{1}{\lambda^2}e^{-\lambda(T-t)} - \frac{1}{4\lambda}(T-t)e^{-2\lambda(T-t)} - \frac{3}{4\lambda^2}e^{-2\lambda(T-t)} - \frac{2}{\lambda^3}\frac{1-e^{-\lambda(T-t)}}{T-t} + \frac{5}{8\lambda^3}\frac{1-e^{-2\lambda(T-t)}}{T-t}\Big].
\end{aligned}
$$

The estimated yield-adjustment term and its three components associated with the variances of the three state variables are shown in Figure 1. All three components are negative, regardless of the size of the volatility parameters. In general, the rather simple functional form of the yield-adjustment term suggests that the lack of improvement in fit of this model is not too surprising.

Greater flexibility is allowed in the correlated-factor AFDNS model, and the estimated parameters of this model are reported in Table 6. Since this model nests the independent-factor version,
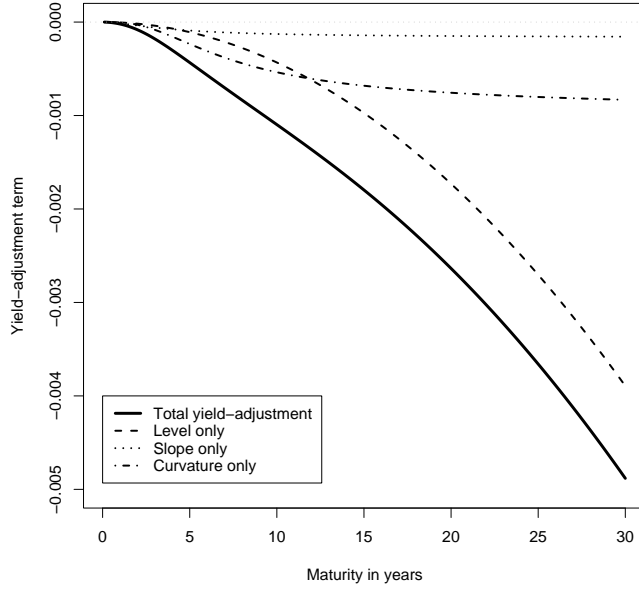
Figure 1: **The Yield-Adjustment Term for the Independent-Factor AFDNS Model.**
The yield-adjustment term $-\frac{C(\tau)}{\tau}$ and its three components.

a standard likelihood ratio test can be performed,

$$LR = 2[\log L(\theta_{corr}) - \log L(\theta_{indep})] = 424.9 \sim \chi^2(q),$$

where q, the number of parameter restrictions, equals nine. The associated p-value is again minuscule, so the independent factor restrictions are not supported by the data in sample.

The greater flexibility is apparent in the complexity of the yield-adjustment term for this model:

$$
\begin{aligned}
-\frac{C(t,T)}{T-t} =\ & -\sigma_{11}^2 \frac{(T-t)^2}{6} - (\sigma_{21}^2 + \sigma_{22}^2)\Big[\frac{1}{2\lambda^2} - \frac{1}{\lambda^3}\frac{1-e^{-\lambda(T-t)}}{T-t} + \frac{1}{4\lambda^3}\frac{1-e^{-2\lambda(T-t)}}{T-t}\Big] \\
& - (\sigma_{31}^2 + \sigma_{32}^2 + \sigma_{33}^2) \\
& \times \Big[\frac{1}{2\lambda^2} + \frac{1}{\lambda^2}e^{-\lambda(T-t)} - \frac{1}{4\lambda}(T-t)e^{-2\lambda(T-t)} - \frac{3}{4\lambda^2}e^{-2\lambda(T-t)} - \frac{2}{\lambda^3}\frac{1-e^{-\lambda(T-t)}}{T-t} + \frac{5}{8\lambda^3}\frac{1-e^{-2\lambda(T-t)}}{T-t}\Big] \\
& - \sigma_{11}\sigma_{21}\Big[\frac{1}{2\lambda}(T-t) + \frac{1}{\lambda^2}e^{-\lambda(T-t)} - \frac{1}{\lambda^3}\frac{1-e^{-\lambda(T-t)}}{T-t}\Big] \\
& - \sigma_{11}\sigma_{31}\Big[\frac{3}{\lambda^2}e^{-\lambda(T-t)} + \frac{1}{2\lambda}(T-t) + \frac{1}{\lambda}(T-t)e^{-\lambda(T-t)} - \frac{3}{\lambda^3}\frac{1-e^{-\lambda(T-t)}}{T-t}\Big] \\
& - (\sigma_{21}\sigma_{31} + \sigma_{22}\sigma_{32})\Big[\frac{1}{\lambda^2} + \frac{1}{\lambda^2}e^{-\lambda(T-t)} - \frac{1}{2\lambda^2}e^{-2\lambda(T-t)} - \frac{3}{\lambda^3}\frac{1-e^{-\lambda(T-t)}}{T-t} + \frac{3}{4\lambda^3}\frac{1-e^{-2\lambda(T-t)}}{T-t}\Big].
\end{aligned}
$$

Figure 2 displays this yield-adjustment term and its various components. This term has an interesting hump with a peak in the 15- to 20-year maturity range, which appears to improve the fit of those long-term yields in particular, but also of yields with fairly short maturities. This
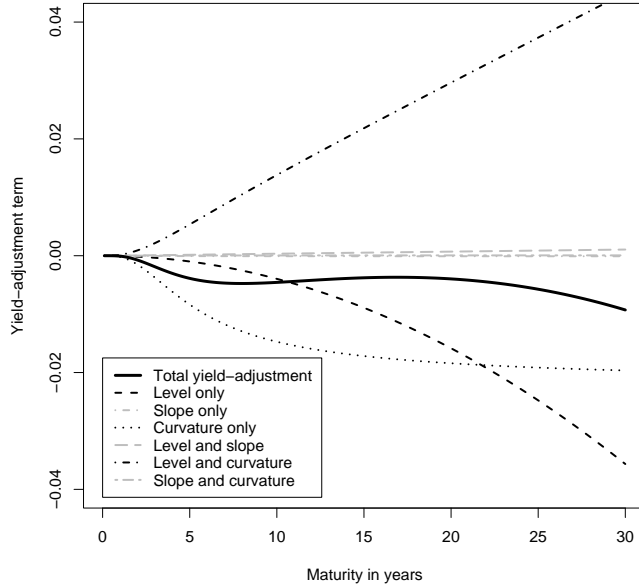
19

Figure 2: **The Yield-Adjustment Term for the Correlated-Factor AFDNS Model.**
The yield adjustment term $-\frac{C(\tau)}{\tau}$ and its six components.

added flexibility allows the level factor to become less persistent, as is evident in the estimated one-month conditional mean-reversion matrix

$$\exp\left(-K^P\frac{1}{12}\right) = \begin{pmatrix} 0.915 & -0.170 & 0.124 \\ 0.0499 & 0.992 & -0.00222 \\ 0.451 & 0.765 & 0.0556 \end{pmatrix}. \tag{13}$$

It appears that to the extent long-term yields are fit through the yield-adjustment term, the level factor becomes less persistent because it blends with slope and curvature in an effort to provide an improved fit for maturities up to nine years.

The one-month conditional covariance matrix is given by

$$Q_{corr}^{AFDNS} = \int_0^{\frac{1}{12}} e^{-K^P s} \Sigma\Sigma'^{-(K^P)'s} ds = \begin{pmatrix} 7.44 \times 10^{-6} & -6.37 \times 10^{-6} & -8.35 \times 10^{-6} \\ -6.37 \times 10^{-6} & 1.09 \times 10^{-5} & 2.80 \times 10^{-6} \\ -8.35 \times 10^{-6} & 2.80 \times 10^{-6} & 2.04 \times 10^{-4} \end{pmatrix}. \tag{14}$$

The conditional variances in the diagonal are about the same for the level and slope factors as those obtained in the correlated-factor DNS model, but the conditional variance for curvature is much larger. In terms of covariances, the negative correlation between the innovations to level and slope is maintained. For the correlations between shocks to curvature and shocks to level and slope, the signs have changed relative to the unconstrained correlated-factor DNS model. This suggests that the off-diagonal elements of Σ are heavily influenced by the required shape of the

yield-adjustment term rather than the dynamics of the state variables. This interpretation will be supported by our out-of-sample forecast exercise in the next section.

5 Forecast performance

In this section, we investigate whether the in-sample superiority of the flexible correlated-factor models carries over to out-of-sample forecast accuracy. We first describe the recursive estimation and forecasting procedure employed, and then we proceed to the results.

5.1 Construction of out-of-sample forecasts

We construct one-, six-, and twelve-month-ahead forecasts from the four DNS and AFDNS models for six yields with maturities of 3 months and 1, 3, 5, 10, and 30 years. We use a recursive procedure. For the first set of forecasts, the model is estimated from January 1987 to December 1996; then, one month of data is added, the models are reestimated, and another set of forecasts is constructed. The largest estimation sample for the one-month-ahead forecasts ends in November 2002 (72 forecasts in all). For the six- and 12-month horizons, the largest samples end in June 2002 and December 2001 (67 and 61 forecasts), respectively. This recursive estimation strategy is greatly facilitated for the AF model by the addition of the Nelson-Siegel factor loadings. For the usual method of estimating the canonical $A_0(3)$ model, each additional month requires reexamination of the zero exclusion restrictions, which is prohibitively time consuming.

For the DNS models, the period-t forecast of the τ-maturity yield h periods ahead is simply the conditional expectation

$$\widehat{y}_{t+h}^{DNS}(\tau) \equiv E_t^P[y_{t+h}(\tau)] = E_t^P[L_{t+h}] + E_t^P[S_{t+h}]\Big(\frac{1 - e^{-\lambda\tau}}{\lambda\tau}\Big) + E_t^P[C_{t+h}]\Big(\frac{1 - e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau}\Big).$$

Given parameter estimates for A and μ from a sample that ends in period t, the discrete-time state equation for the DNS model can be written

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

where $X_t = (L_t, S_t, C_t)$. Recursive iteration (and i.i.d. innovations) imply that the conditional expectation of the state variables in period $t + h$ are

$$E_t^P[X_{t+h}] = \Big(\sum_{i=0}^{h-1} A^i\Big)(I - A)\mu + A^h X_t,$$

so it is straightforward to calculate forecasted yields.

For the AFDNS models, the forecast of the τ-maturity yield in period $t+h$ based on information available at time t is simply the conditional expectation

$$\widehat{y}_{t+h}^{AFDNS}(\tau) \equiv E_t^P[y_{t+h}(\tau)] = E_t^P[X_{t+h}^1] + E_t^P[X_{t+h}^2]\Big(\frac{1 - e^{-\lambda\tau}}{\lambda\tau}\Big) + E_t^P[X_{t+h}^3]\Big(\frac{1 - e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau}\Big) - \frac{C(\tau)}{\tau}.$$

In this case, the requisite conditional expectations are given by

$$E_0^P[X_t] = (I - \exp(-K^P t))\theta^P + \exp(-K^P t)X_0,$$

where $X_t = (X_t^1, X_t^2, X_t^3)$. Thus, with estimates for K^P, θ^P, λ, and Σ along with the optimally filtered paths of the three factors, it is easy to calculate future factor expected values and yields.

5.2 Evaluation of out-of-sample forecasts

Out-of-sample forecast accuracy has been a key metric to evaluate the adequacy of AF yield-curve models.[18] The forecast performances of the four models are compared using the root mean squared error (RMSE) of the forecast error $\varepsilon_t(\tau, h) = \hat{y}_{t+h}(\tau) - y_{t+h}(\tau)$, for $\tau = 3, 12, 36, 60, 120, 360$, and $h = 1, 6, 12$ (in months). These RMSEs are shown in Table 7. For each of the 18 combinations of yield maturity and forecast horizon, the most accurate model's RMSE is underlined. The results are quite striking. In 14 of the 18 combinations, the most accurate model is the independent-factor AFDNS model. In particular, the in-sample advantage of the correlated-factor AFDNS model disappears out of sample. Evidently, the correlated-factor AFDNS model is prone to in-sample overfitting, due to its complex yield-adjustment term and rich P-dynamics. Furthermore, the cases in which the independent-factor AFDNS model is not the most accurate all pertain to shorter-maturity yields. Specifically, it is only for the 3-month yield, that the correlated-factor models have lower RMSEs. This advantage likely reflects idiosyncratic fluctuations in short-term Treasury bill yields from institutional factors that are unrelated to yields on longer-maturity Treasuries, as described by Duffee (1996). The more flexible models appear to have a slight advantage in fitting these idiosyncratic movements.

In examining forecast performance, we are interested in two broad comparisons. First, how do the correlated-factor models do against the independent-factor models, and second, how does the imposition of the AF structure affect forecast performance. Table 8 brings these two questions into sharper focus by showing the ratios of the forecast RMSEs of various models. The two columns on the left divide the DNS and AFDNS independent-factor model RMSEs by their respective correlated-factor model RMSEs. These are almost uniformly below one (outside of the 3-month yield noted above), which supports the parsimonious versions of these models. These differences in forecast accuracy are also generally statistically significant. For each maturity and horizon combination, we use the Diebold-Mariano (1995) test to compare model performance.[19] The asterisks in Table 8 denote significant differences in out-of-sample model performance at the 1, 5, and 10 percent levels. For both the DNS and AFDNS models, the preponderance of evidence supports the parsimonious models.[20]

The two columns on the right divide the RMSEs of the AF versions of the independent- and

---

[18]Recent analyses of the forecast performance of AF models include Ang and Piazzesi (2003), Hördahl, Tristani, and Vestin (2005), Mönch (2006), De Pooter, Ravazzolo, and van Dijk (2007).

[19]We implement this test by regressing the differences between the squared forecast errors for two models on an intercept and examining the significance of that intercept using standard errors that are corrected for possibly heteroskedastic and autocorrelated residuals.

[20]We also examined model accuracy using the generalized Diebold-Mariano test proposed by Christensen et al. (2007), which can pool observations across all maturities or horizons simultaneously. This test supported our conclusions from the individual comparisons.

Model	Forecast horizon		
	One month	Six months	Twelve months
3-month yield			
DNS_{indep}	22.93	96.87	173.39
DNS_{corr}	<u>20.43</u>	<u>87.43</u>	166.91
$AFDNS_{indep}$	22.84	91.60	164.97
$AFDNS_{corr}$	20.56	88.67	<u>162.33</u>
1-year yield			
DNS_{indep}	29.41	103.25	170.85
DNS_{corr}	<u>27.06</u>	102.71	173.14
$AFDNS_{indep}$	29.12	<u>98.58</u>	<u>164.01</u>
$AFDNS_{corr}$	33.89	98.87	165.99
3-year yield			
DNS_{indep}	30.64	92.22	135.24
DNS_{corr}	30.59	99.55	145.82
$AFDNS_{indep}$	<u>30.29</u>	<u>87.23</u>	<u>127.78</u>
$AFDNS_{corr}$	36.95	91.00	136.44
5-year yield			
DNS_{indep}	30.77	87.87	122.09
DNS_{corr}	31.23	94.95	132.40
$AFDNS_{indep}$	<u>30.13</u>	<u>82.68</u>	<u>113.83</u>
$AFDNS_{corr}$	32.37	88.46	125.42
10-year yield			
DNS_{indep}	28.35	74.71	105.02
DNS_{corr}	29.06	79.48	112.37
$AFDNS_{indep}$	<u>27.18</u>	<u>67.72</u>	<u>93.36</u>
$AFDNS_{corr}$	35.08	90.42	124.28
30-year yield			
DNS_{indep}	38.42	71.35	96.90
DNS_{corr}	38.73	72.71	99.68
$AFDNS_{indep}$	<u>30.42</u>	<u>48.82</u>	<u>63.50</u>
$AFDNS_{corr}$	38.30	71.35	96.86

Table 7: **Out-of-Sample Forecast RMSE for Four Models.**
For each maturity and horizon, the most accurate model's RMSE is underlined. All numbers are measured in basis points.

Forecast horizon	Ratios of out-of-sample forecast RMSEs			
(in months)	$\dfrac{\text{DNS}_{indep}}{\text{DNS}_{corr}}$	$\dfrac{\text{AFDNS}_{indep}}{\text{AFDNS}_{corr}}$	$\dfrac{\text{AFDNS}_{indep}}{\text{DNS}_{indep}}$	$\dfrac{\text{AFDNS}_{corr}}{\text{DNS}_{corr}}$
	3-month yield			
1	1.12*	1.11**	1.00	1.01
6	1.11	1.03**	0.95*	1.01
12	1.04	1.02	0.95*	0.97
	12-month yield			
1	1.09**	0.86**	0.99	1.25**
6	1.01	1.00	0.95	0.96
12	0.99	0.99	0.96	0.96
	36-month yield			
1	1.00	0.82***	0.99	1.21**
6	0.93**	0.96	0.95	0.91
12	0.93**	0.94	0.94	0.94**
	60-month yield			
1	0.99	0.93*	0.98**	1.04
6	0.93***	0.93	0.94*	0.93**
12	0.92**	0.91	0.93	0.95***
	120-month yield			
1	0.98***	0.77***	0.96***	1.21***
6	0.94***	0.75***	0.91**	1.14***
12	0.93***	0.75***	0.89*	1.11***
	360-month yield			
1	0.99	0.79*	0.79	0.99
6	0.98**	0.68***	0.68***	0.98
12	0.97***	0.66***	0.66***	0.97

Table 8: **RMSE Ratios for Out-of-Sample Forecast Errors.**
The ratios of the RMSEs for two different models are shown for each forecast horizon and yield maturity. The statistical significance of these forecast comparisons (based on tests of equal forecast accuracy using quadratic loss) are denoted by * at the 10% level, ** at the 5% level, and *** at the 1% level.

correlated-factor models by their non-AF counterparts. Here the story is more mixed, but for the independent-factor case, which is arguably the one of interest given the generally poor performance and overparameterization of the correlated-factor models, the AF version dominates. The bottom line is that out-of-sample forecast performance is improved by imposing the AF restrictions—especially at longer horizons and for longer maturities.

6 Concluding Remarks

Asset pricing, portfolio allocation, and risk management are the fundamental tasks in financial asset markets. For fixed income securities, superior yield-curve modeling translates into superior pricing, portfolio returns, and risk management. Accordingly, we have focused on two important and successful yield curve literatures: the Nelson-Siegel empirically based one and the no-arbitrage theoretically based one. Yield-curve models in both of these traditions are impressive successes, albeit for very different reasons. Ironically, both approaches are equally impressive failures, and for the same reasons, swapped. That is, models in the Nelson-Siegel tradition fit and forecast well, but they lack theoretical rigor insofar as they admit arbitrage possibilities. Conversely, models in the arbitrage-free tradition are theoretically rigorous insofar as they enforce absence of arbitrage, but they fit and forecast poorly.

In this paper we have bridged this divide, proposing hybrid Nelson-Siegel-inspired models that simultaneously enforce absence of arbitrage. We analyzed our models theoretically and empirically, relating them to the canonical Dai-Singleton representation of three-factor arbitrage-free affine models and documenting that predictive gains may be achieved by imposing absence of arbitrage, particularly for moderate to long maturities and forecast horizons.

Appendix A: Proof of Proposition 1

Start the analysis by limiting the volatility to be constant. Then the system of ODEs for $B(t,T)$ is given by

$$\frac{dB(t,T)}{dt} = \rho_1 + (K^Q)'B(t,T), \quad B(T,T) = 0.$$

Because

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

it follows from the system of ODEs that

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

or, equivalently, using the boundary conditions

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

Now impose the following structure on $(K^Q)'$ and ρ_1:

$$(K^Q)' = \begin{pmatrix} 0 & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & -\lambda & \lambda \end{pmatrix} \quad \text{and} \quad \rho_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}.$$

It is then easy to show that

$$e^{(K^Q)'(T-t)} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & e^{\lambda(T-t)} & 0 \\ 0 & -\lambda(T-t)e^{\lambda(T-t)} & e^{\lambda(T-t)} \end{pmatrix} \quad \text{and} \quad e^{-(K^Q)'(T-t)} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & e^{-\lambda(T-t)} & 0 \\ 0 & \lambda(T-t)e^{-\lambda(T-t)} & e^{-\lambda(T-t)} \end{pmatrix}.$$

Inserting this in the ODE, we obtain

$$\begin{aligned}
B(t,T) &= -\begin{pmatrix} 1 & 0 & 0 \\ 0 & e^{-\lambda(T-t)} & 0 \\ 0 & \lambda(T-t)e^{-\lambda(T-t)} & e^{-\lambda(T-t)} \end{pmatrix} \int_t^T \begin{pmatrix} 1 & 0 & 0 \\ 0 & e^{\lambda(T-s)} & 0 \\ 0 & -\lambda(T-s)e^{\lambda(T-s)} & e^{\lambda(T-s)} \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} ds \\
&= -\begin{pmatrix} 1 & 0 & 0 \\ 0 & e^{-\lambda(T-t)} & 0 \\ 0 & \lambda(T-t)e^{-\lambda(T-t)} & e^{-\lambda(T-t)} \end{pmatrix} \int_t^T \begin{pmatrix} 1 \\ e^{\lambda(T-s)} \\ -\lambda(T-s)e^{\lambda(T-s)} \end{pmatrix} ds.
\end{aligned}$$

Because

$$\int_t^T ds = T - t,$$

and

$$\int_t^T e^{\lambda(T-s)}ds = \left[\frac{-1}{\lambda}e^{\lambda(T-s)}\right]_t^T = -\frac{1 - e^{\lambda(T-t)}}{\lambda},$$

and

26

$$\int_t^T -\lambda(T-s)e^{\lambda(T-s)}ds = \frac{1}{\lambda}\int_{\lambda(T-t)}^0 xe^x dx = \frac{1}{\lambda}[xe^x]_{\lambda(T-t)}^0 - \frac{1}{\lambda}\int_{\lambda(T-t)}^0 e^x dx = -(T-t)e^{\lambda(T-t)} - \frac{1-e^{\lambda(T-t)}}{\lambda},$$

the system of ODEs can be reduced to

$$B(t,T) = - \begin{pmatrix} 1 & 0 & 0 \\ 0 & e^{-\lambda(T-t)} & 0 \\ 0 & \lambda(T-t)e^{-\lambda(T-t)} & e^{-\lambda(T-t)} \end{pmatrix} \begin{pmatrix} T-t \\ -\frac{1-e^{\lambda(T-t)}}{\lambda} \\ -(T-t)e^{\lambda(T-t)} - \frac{1-e^{\lambda(T-t)}}{\lambda} \end{pmatrix} = \begin{pmatrix} -(T-t) \\ -\frac{1-e^{-\lambda(T-t)}}{\lambda} \\ (T-t)e^{-\lambda(T-t)} - \frac{1-e^{-\lambda(T-t)}}{\lambda} \end{pmatrix},$$

which is identical to the claim in Proposition 1. **QED**

Appendix B: The AFDNS yield-adjustment term

In the AFDNS models the yield-adjustment term is in general given by

$$
\begin{aligned}
\frac{C(t,T)}{T-t} &= \frac{1}{2}\frac{1}{T-t}\int_t^T \sum_{j=1}^3 \left(\Sigma' B(s,T)B(s,T)'\Sigma\right)_{j,j} ds \\
&= \frac{1}{2}\frac{1}{T-t}\int_t^T \sum_{j=1}^3 \left[\begin{pmatrix} \sigma_{11} & \sigma_{21} & \sigma_{31} \\ \sigma_{12} & \sigma_{22} & \sigma_{32} \\ \sigma_{13} & \sigma_{23} & \sigma_{33} \end{pmatrix} \begin{pmatrix} B^1(t,T) \\ B^2(t,T) \\ B^3(t,T) \end{pmatrix} \begin{pmatrix} B^1(t,T) & B^2(t,T) & B^3(t,T) \end{pmatrix} \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{pmatrix} \right]_{j,j} \\
&= \frac{\overline{A}}{2}\frac{1}{T-t}\int_t^T B^1(s,T)^2 ds + \frac{\overline{B}}{2}\frac{1}{T-t}\int_t^T B^2(s,T)^2 ds + \frac{\overline{C}}{2}\frac{1}{T-t}\int_t^T B^3(s,T)^2 ds \\
&+ \overline{D}\frac{1}{T-t}\int_t^T B^1(s,T)B^2(s,T)ds + \overline{E}\frac{1}{T-t}\int_t^T B^1(s,T)B^3(s,T)ds + \overline{F}\frac{1}{T-t}\int_t^T B^2(s,T)B^3(s,T)ds,
\end{aligned}
$$

where

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

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

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

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

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

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

To derive the analytical formula for $\frac{C(t,T)}{T-t}$, six integrals need to be solved:

$$I_1 = \frac{\overline{A}}{2}\frac{1}{T-t}\int_t^T B^1(s,T)^2 ds = \frac{\overline{A}}{2}\frac{1}{T-t}\int_t^T (T-s)^2 ds = \frac{\overline{A}}{6}(T-t)^2.$$

$$I_2 = \frac{\overline{B}}{2}\frac{1}{T-t}\int_t^T B^2(s,T)ds = \frac{\overline{B}}{2}\frac{1}{T-t}\int_t^T \left[-\frac{1-e^{-\lambda(T-s)}}{\lambda}\right]^2 ds = \overline{B}\left[\frac{1}{2\lambda^2} - \frac{1}{\lambda^3}\frac{1-e^{-\lambda(T-t)}}{T-t} + \frac{1}{4\lambda^3}\frac{1-e^{-2\lambda(T-t)}}{T-t}\right].$$

27

$$I_3 = \frac{\overline{C}}{2}\frac{1}{T-t}\int_t^T B^3(s,T)ds = \frac{\overline{C}}{2}\frac{1}{T-t}\int_t^T \left\{(T-s)e^{-\lambda(T-s)} - \frac{1-e^{-\lambda(T-s)}}{\lambda}\right\}^2 ds$$

$$= \overline{C}\left[\frac{1}{2\lambda^2} + \frac{1}{\lambda^2}e^{-\lambda(T-t)} - \frac{1}{4\lambda}(T-t)e^{-2\lambda(T-t)} - \frac{3}{4\lambda^2}e^{-2\lambda(T-t)} - \frac{2}{\lambda^3}\frac{1-e^{-\lambda(T-t)}}{T-t} + \frac{5}{8\lambda^3}\frac{1-e^{-2\lambda(T-t)}}{T-t}\right].$$

$$I_4 = \frac{\overline{D}}{T-t}\int_t^T B^1(s,T)B^2(s,T)ds = \frac{\overline{D}}{T-t}\int_t^T \left[-(T-s)\right]\left[-\frac{1-e^{-\lambda(T-s)}}{\lambda}\right]ds = \overline{D}\left[\frac{1}{2\lambda}(T-t) + \frac{1}{\lambda^2}e^{-\lambda(T-t)} - \frac{1}{\lambda^3}\frac{1-e^{-\lambda(T-t)}}{T-t}\right].$$

$$I_5 = \overline{E}\frac{1}{T-t}\int_t^T B^1(s,T)B^3(s,T)ds = \overline{E}\frac{1}{T-t}\int_t^T \left[-(T-s)\right]\left[(T-s)e^{-\lambda(T-s)} - \frac{1-e^{-\lambda(T-s)}}{\lambda}\right]ds$$

$$= \overline{E}\left[\frac{3}{\lambda^2}e^{-\lambda(T-t)} + \frac{1}{2\lambda}(T-t) + \frac{1}{\lambda}(T-t)e^{-\lambda(T-t)} - \frac{3}{\lambda^3}\frac{1-e^{-\lambda(T-t)}}{T-t}\right].$$

$$I_6 = \overline{F}\frac{1}{T-t}\int_t^T B^2(s,T)B^3(s,T)ds = \overline{F}\frac{1}{T-t}\int_t^T \left[-\frac{1-e^{-\lambda(T-s)}}{\lambda}\right]\left[(T-s)e^{-\lambda(T-s)} - \frac{1-e^{-\lambda(T-s)}}{\lambda}\right]ds$$

$$= \overline{F}\left[\frac{1}{\lambda^2} + \frac{1}{\lambda^2}e^{-\lambda(T-t)} - \frac{1}{2\lambda^2}e^{-2\lambda(T-t)} - \frac{3}{\lambda^3}\frac{1-e^{-\lambda(T-t)}}{T-t} + \frac{3}{4\lambda^3}\frac{1-e^{-2\lambda(T-t)}}{T-t}\right].$$

Combining the six integrals, the analytical formula reported in subsection 2.3 is obtained.

Appendix C: Proof of Proposition 2

Before we can turn to the proof of Proposition 2 and the subsequent derivation of the restrictions that need to be imposed on the canonical representation of the $A_0(3)$ class of affine models to arrive at the models equivalent to the AFDNS model, we need to introduce the concept of so-called *affine invariant transformations*.

Consider an arbitrary affine diffusion process represented by

$$dY_t = K_Y^Q[\theta_Y^Q - Y_t]dt + \Sigma_Y dW_t^Q.$$

Now consider the affine transformation $\mathcal{T}_Y : AY_t + \eta$, where A is a nonsingular square matrix of the same dimension as Y_t while η is a vector of constants of the same dimension as Y_t. Denote the transformed process by $X_t = AY_t + \eta$. By Ito's lemma it follows that

$$dX_t = AdY_t = [AK_Y^Q\theta_Y^Q - AK_Y^QY_t]dt + A\Sigma_Y dW_t^Q = AK_Y^QA^{-1}[A\theta_Y^Q - AY_t - \eta + \eta]dt + A\Sigma_Y dW_t^Q$$

$$= AK_Y^QA^{-1}[A\theta_Y^Q + \eta - X_t]dt + A\Sigma_Y dW_t^Q = K_X^Q[\theta_X^Q - X_t]dt + \Sigma_X dW_t^Q.$$

Thus, X_t is itself an affine diffusion process with the following parameter specification:

$$K_X^Q = AK_Y^QA^{-1}, \quad \theta_X^Q = A\theta_Y^Q + \eta, \quad \text{and} \quad \Sigma_X = A\Sigma_Y.$$

A similar result holds for the dynamics under the P-measure.

In terms of the short rate process there exists the following relationship:

$$
\begin{aligned}
r_t &= \delta_0^Y + (\delta_1^Y)'Y_t = \delta_0^Y + (\delta_1^Y)'A^{-1}AY_t = \delta_0^Y + (\delta_1^Y)'A^{-1}[AY_t + \eta - \eta] \\
&= \delta_0^Y - (\delta_1^Y)'A^{-1}\eta + (\delta_1^Y)'A^{-1}X_t.
\end{aligned}
$$

Thus, defining $\delta_0^X = \delta_0^Y - (\delta_1^Y)'A^{-1}\eta$ and $\delta_1^X = (\delta_1^Y)'A^{-1}$, the short rate process is left unchanged and may be represented in either way

$$
r_t = \delta_0^Y + (\delta_1^Y)'Y_t = \delta_0^X + (\delta_1^X)'X_t.
$$

Because both Y_t and X_t are affine latent factor processes that deliver the same distribution for the short rate process r_t, they are equivalent representations of the same fundamental model.

The upshot is that the canonical representation provided by Dai and Singleton (2000) is just one way of representing this model. There are an infinite number of representations of the same model that all share the property that the risk-free short rate process and, by consequence, all bond yields will have the same distribution independent of the choice of representation. Hence $\mathcal{T}_X$ is called an affine invariant transformation.

We now proceed to a proof of Proposition 2. Consider the canonical representation of maximal flexibility under the Q-measure, namely that with identifying restrictions imposed under the P-measure with $\theta_Y^P = 0$, $\Sigma_Y = I$, K_Y^P upper triangular, and K_Y^Q any 3×3 matrix:

$$
\begin{aligned}
dY_t &= -K_Y^P Y_t dt + I dW_t^P \\
dY_t &= K_Y^Q[\theta_Y^Q - Y_t]dt + I dW_t^Q.
\end{aligned}
$$

By Schur's decomposition, the K_Y^Q-matrix can be written as

$$
K_Y^Q = A'[D + N]A,
$$

where A is an orthogonal matrix, D is a diagonal matrix containing the eigenvalues of K_Y^Q, and N is an upper-triangular matrix. Although the Schur decomposition is not unique, the argument below only relies on the existence of such matrices. If the eigenvalues of K_Y^Q are real (as assumed in Proposition 2), then $D + N$ and A are real matrices and, in particular, $T = D + N$ is an upper-triangular matrix.

Now consider the invariant affine transformation given by $\mathcal{T}_A(Y_t) = AY_t$. Given the transformed process $X_t = AY_t$, the factor loadings in the risk-free rate are given by

$$
\delta_0^X = \delta_0^Y, \quad \delta_1^X = (\delta_1^Y)'A^{-1} = (\delta_1^Y)'A',
$$

and the volatility matrix is given by
$$
\Sigma_X = A,
$$

while the P- and Q-dynamics are given by

$$K_X^P = AK_Y^P A^{-1} = AK_Y^P A', \quad \theta_X^P = A\theta_Y^P = 0, \quad K_X^Q = AK_Y^Q A^{-1} = AK_Y^Q A', \quad \theta_X^Q = A\theta_Y^Q.$$

However, because $K_Y^Q = A'TA$ by the Schur decomposition, it follows that

$$K_X^Q = AK_Y^Q A' = A[A'TA]A' = T.$$

Hence we have obtained an upper-triangular mean-reversion matrix under the Q-measure.

The next step in the proof is to transform the volatility matrix $\Sigma_X = A$ into the identity matrix without affecting any of the drift terms. This can be done through a so-called Brownian Motion Rotation.[21] This consists in defining an affine transformation of the Brownian motion

$$T_O(W_t) = OW_t,$$

where O must be an orthogonal matrix. The orthogonality of O maintains the independence of the transformed Brownian motions and thereby makes the transformation invariant. In the current case the rotation is performed with the orthogonal matrix A and the rotated Brownian motion is given by $\widehat{W}_t^P = AW_t^P$ and $\widehat{W}_t^Q = AW_t^Q$. Applied to the X_t-process, it follows that

$$
\begin{aligned}
dX_t &= -K_X^P X_t dt + AdW_t^P = -K_X^P X_t dt + A[A'A]dW_t^P = -K_X^P X_t dt + Id\widehat{W}_t^P, \\
dX_t &= T[\theta_X^Q - X_t]dt + AdW_t^Q = T[\theta_X^Q - X_t]dt + A[A'A]dW_t^Q = T[\theta_X^Q - X_t]dt + Id\widehat{W}_t^Q.
\end{aligned}
$$

The only remaining problem is to eliminate the mean vector under the Q-measure, θ_X^Q. To that end apply the invariant affine transformation $\mathcal{T}_A(X_t) = X_t - \theta_X^Q$. The transformed process $Z_t = X_t - \theta_X^Q$ is characterized by the following vectors and matrices:

$$
\begin{aligned}
\delta_0^Z &= \delta_0^X + (\delta_1^X)'\theta_X^Q = \delta_0^Y + [A\delta_1^Y]'A\theta_Y^Q = \delta_0^Y + (\delta_1^Y)'\theta_Y^Q, \\
\delta_1^Z &= \delta_1^X = A\delta_1^Y. \\
\Sigma_Z &= \Sigma_X = I \\
K_Z^P &= K_X^P = AK_Y^P A' \\
\theta_Z^P &= \theta_X^P - \theta_X^Q = -A\theta_Y^Q \\
K_Z^Q &= K_X^Q = T \\
\theta_Z^Q &= \theta_X^Q - \theta_X^Q = 0,
\end{aligned}
$$

or equivalently,

$$
\begin{aligned}
dZ_t &= AK_Y^P A'[-A\theta_Y^Q - Z_t]dt + Id\widehat{W}_t^P \\
dZ_t &= -TZ_t dt + Id\widehat{W}_t^Q,
\end{aligned}
$$

[21]See Appendix A in Dai and Singleton (2000) for details.

with the instantaneous risk-free rate given by

$$r_t = \delta_0^Y + (\delta_1^Y)'\theta_Y^Q + (\delta_1^Y)'A'Z_t.$$

Hence the dynamics of the Z_t-process are identical to the canonical representation with the identifying restrictions imposed under the Q-measure, and since all the transformations that converted the original Y_t-process into the Z_t-process have been invariant transformations, Y_t and Z_t are equivalent representations of the same underlying model. **QED**

Appendix D: Parameter restrictions imposed in AFDNS

This appendix derives the connection between the AFDNS models and the canonical representation of the $A_0(3)$ class of affine term structure models. In the canonical representation of the subset of $A_0(3)$ affine term structure models considered here, the dynamics under the Q-measure are given by

$$\begin{pmatrix} dZ_t^1 \\ dZ_t^2 \\ dZ_t^3 \end{pmatrix} = - \begin{pmatrix} \kappa_{11}^{Z,Q} & \kappa_{12}^{Z,Q} & \kappa_{13}^{Z,Q} \\ 0 & \kappa_{22}^{Z,Q} & \kappa_{23}^{Z,Q} \\ 0 & 0 & \kappa_{33}^{Z,Q} \end{pmatrix} \begin{pmatrix} Z_t^1 \\ Z_t^2 \\ Z_t^3 \end{pmatrix} dt + \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} dW_t^{1,Q} \\ dW_t^{2,Q} \\ dW_t^{3,Q} \end{pmatrix},$$

and the P-dynamics are given by

$$\begin{pmatrix} dZ_t^1 \\ dZ_t^2 \\ dZ_t^3 \end{pmatrix} = \begin{pmatrix} \kappa_{11}^{Z,P} & \kappa_{12}^{Z,P} & \kappa_{13}^{Z,P} \\ \kappa_{21}^{Z,P} & \kappa_{22}^{Z,P} & \kappa_{23}^{Z,P} \\ \kappa_{31}^{Z,P} & \kappa_{32}^{Z,P} & \kappa_{33}^{Z,P} \end{pmatrix} \left[\begin{pmatrix} \theta_1^{Z,P} \\ \theta_2^{Z,P} \\ \theta_3^{Z,P} \end{pmatrix} - \begin{pmatrix} Z_t^1 \\ Z_t^2 \\ Z_t^3 \end{pmatrix} \right] dt + \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} dW_t^{1,P} \\ dW_t^{2,P} \\ dW_t^{3,P} \end{pmatrix}.$$

Finally, the instantaneous risk-free rate is given by

$$r_t = \delta_0^Z + \delta_{1,1}^Z Z_t^1 + \delta_{1,2}^Z Z_t^2 + \delta_{1,3}^Z Z_t^3.$$

There are 22 parameters in this maximally flexible canonical representation of the $A_3(0)$ class of models. We seek to find the parameter restrictions that need to be imposed on the canonical representation of this maximally flexible $A_0(3)$ model to arrive at a model equivalent to the affine AFDNS models considered in this paper. We first consider the independent-factor case, and then we examine correlated factors.

(1) The AFDNS model with independent factors

The independent-factor AFDNS model has P-dynamics given by

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

and the Q-dynamics are given by Proposition 1 as

$$
\begin{pmatrix} dX_t^1 \\ dX_t^2 \\ dX_t^3 \end{pmatrix} = - \begin{pmatrix} 0 & 0 & 0 \\ 0 & \lambda & -\lambda \\ 0 & 0 & \lambda \end{pmatrix} \begin{pmatrix} X_t^1 \\ X_t^2 \\ X_t^3 \end{pmatrix} dt + \begin{pmatrix} \sigma_{11}^X & 0 & 0 \\ 0 & \sigma_{22}^X & 0 \\ 0 & 0 & \sigma_{33}^X \end{pmatrix} \begin{pmatrix} dW_t^{1,Q} \\ dW_t^{2,Q} \\ dW_t^{3,Q} \end{pmatrix}.
$$

Finally, the short rate process is $r_t = X_t^1 + X_t^2$. This model has a total of 10 parameters. Thus, the task is to determine the 12 parameter restrictions that need to be imposed on the canonical $A_0(3)$ model to arrive at this model.

It is easy to verify that the affine invariant transformation $\mathcal{T}_A(Z_t) = AZ_t + \eta$ with

$$
A = \begin{pmatrix} \sigma_{11}^X & 0 & 0 \\ 0 & \sigma_{22}^X & 0 \\ 0 & 0 & \sigma_{33}^X \end{pmatrix} \qquad \eta = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}
$$

will convert the canonical representation into the independent-factor AFDNS model. For the mean-reversion matrices, the relationship between the two representations is

$$
\begin{aligned}
K_X^P &= AK_Z^P A^{-1} &\Longleftrightarrow&& K_Z^P &= A^{-1}K_X^P A \\
K_X^Q &= AK_Z^Q A^{-1} &\Longleftrightarrow&& K_Z^Q &= A^{-1}K_X^Q A.
\end{aligned}
$$

The equivalent mean-reversion matrix under the Q-measure is then given by

$$
K_Z^Q = \begin{pmatrix} \frac{1}{\sigma_{11}^X} & 0 & 0 \\ 0 & \frac{1}{\sigma_{22}^X} & 0 \\ 0 & 0 & \frac{1}{\sigma_{33}^X} \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 \\ 0 & \lambda & -\lambda \\ 0 & 0 & \lambda \end{pmatrix} \begin{pmatrix} \sigma_{11}^X & 0 & 0 \\ 0 & \sigma_{22}^X & 0 \\ 0 & 0 & \sigma_{33}^X \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & \lambda & -\lambda\frac{\sigma_{33}^X}{\sigma_{22}^X} \\ 0 & 0 & \lambda \end{pmatrix}.
$$

Thus, four restrictions need to be imposed on the upper triangular mean-reversion matrix K_Z^Q:

$$
K_{11}^{Z,Q} = 0, \quad K_{12}^{Z,Q} = 0, \quad K_{13}^{Z,Q} = 0 \quad \text{and} \quad K_{33}^{Z,Q} = K_{22}^Z.
$$

Furthermore, notice that $K_{23}^{Z,Q}$ will always have the opposite sign of $K_{22}^{Z,Q}$ and $K_{33}^{Z,Q}$, but its absolute size can vary independently of these two parameters. Since K_X^P, A, and A^{-1} are all diagonal matrices, K_Z^P is a diagonal matrix, too. This gives another six restrictions.

Finally, we can study the factor loadings in the affine function for the short rate process. In all AFDNS models, $r_t = X_t^1 + X_t^2$, which is equivalent to fixing

$$
\delta_0^X = 0, \quad \delta_1^X = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}.
$$

From the relation $(\delta_1^X)' = (\delta_1^Z)'A^{-1}$ it follows that

$$(\delta_1^Z)' = (\delta_1^X)'A = \begin{pmatrix} 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} \sigma_{11}^X & 0 & 0 \\ 0 & \sigma_{22}^X & 0 \\ 0 & 0 & \sigma_{33}^X \end{pmatrix} = \begin{pmatrix} \sigma_{11}^X & \sigma_{22}^X & 0 \end{pmatrix}.$$

For the constant term it holds that

$$\delta_0^X = \delta_0^Z - (\delta_1^Z)'A^{-1}\eta \iff \delta_0^Z = \delta_0^X = 0.$$

Thus, we have obtained two additional parameter restrictions

$$\delta_0^Z = 0 \quad \text{and} \quad \delta_{1,3}^Z = 0.$$

(2) The AFDNS model with correlated factors

In the correlated-factor AFDNS model, the P-dynamics are given by

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

and the Q-dynamics are given by Proposition 1 as

$$\begin{pmatrix} dX_t^1 \\ dX_t^2 \\ dX_t^3 \end{pmatrix} = - \begin{pmatrix} 0 & 0 & 0 \\ 0 & \lambda & -\lambda \\ 0 & 0 & \lambda \end{pmatrix} \begin{pmatrix} X_t^1 \\ X_t^2 \\ X_t^3 \end{pmatrix} dt + \begin{pmatrix} \sigma_{11}^X & \sigma_{12}^X & \sigma_{13}^X \\ 0 & \sigma_{22}^X & \sigma_{23}^X \\ 0 & 0 & \sigma_{33}^X \end{pmatrix} \begin{pmatrix} dW_t^{1,Q} \\ dW_t^{2,Q} \\ dW_t^{3,Q} \end{pmatrix}.$$

This model has a total of 19 parameters. Thus, there are three parameter restrictions to be determined as compared to the maximally flexible canonical representation of the $A_0(3)$ class.

It is easy to verify that the affine invariant transformation $\mathcal{T}_A(Z_t) = AZ_t + \eta$ with

$$A = \begin{pmatrix} \sigma_{11}^X & \sigma_{12}^X & \sigma_{13}^X \\ 0 & \sigma_{22}^X & \sigma_{23}^X \\ 0 & 0 & \sigma_{33}^X \end{pmatrix} \quad \text{and} \quad \eta = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

will convert the canonical representation into the correlated-factor AFDNS model. For the mean-reversion matrices the relationship between the two representations is

$$K_X^P = AK_Z^P A^{-1} \quad \iff \quad K_Z^P = A^{-1}K_X^P A$$
$$K_X^Q = AK_Z^Q A^{-1} \quad \iff \quad K_Z^Q = A^{-1}K_X^Q A.$$

The equivalent mean-reversion matrix under the Q-measure is then given by

$$
K_Z^Q = \begin{pmatrix} \frac{1}{\sigma_{11}^X} & -\frac{\sigma_{12}^X}{\sigma_{11}^X \sigma_{22}^X} & -\left(\frac{\sigma_{13}^X}{\sigma_{11}^X \sigma_{33}^X} - \frac{\sigma_{12}^X \sigma_{23}^X}{\sigma_{11}^X \sigma_{22}^X \sigma_{33}^X}\right) \\ 0 & \frac{1}{\sigma_{22}^X} & -\frac{\sigma_{23}^X}{\sigma_{22}^X \sigma_{33}^X} \\ 0 & 0 & \frac{1}{\sigma_{33}^X} \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 \\ 0 & \lambda & -\lambda \\ 0 & 0 & \lambda \end{pmatrix} \begin{pmatrix} \sigma_{11}^X & \sigma_{12}^X & \sigma_{13}^X \\ 0 & \sigma_{22}^X & \sigma_{23}^X \\ 0 & 0 & \sigma_{33}^X \end{pmatrix}
$$

$$
= \begin{pmatrix} 0 & -\lambda \frac{\sigma_{12}^X}{\sigma_{11}^X} & \lambda \frac{\sigma_{12}^X \sigma_{33}^X - \sigma_{22}^X \sigma_{13}^X}{\sigma_{11}^X \sigma_{22}^X} \\ 0 & \lambda & -\lambda \frac{(\sigma_{23}^X)^2}{\sigma_{22}^X \sigma_{33}^X} \\ 0 & 0 & \lambda \end{pmatrix}.
$$

Thus, two restrictions need to be imposed on the upper triangular mean-reversion matrix K_Z^Q:

$$
K_{11}^{Z,Q} = 0, \quad K_{33}^{Z,Q} = K_{22}^Z.
$$

Furthermore, notice that $K_{23}^{Z,Q}$ will always have the opposite sign of $K_{22}^{Z,Q}$ and $K_{33}^{Z,Q}$, but its absolute size can vary independently of the two other parameters.

Next we study the factor loadings in the affine function for the short rate process. In the AFDNS models, $r_t = X_t^1 + X_t^2$, which is equivalent to fixing

$$
\delta_0^X = 0, \quad \delta_1^X = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}.
$$

From the relation $(\delta_1^X)' = (\delta_1^Z)' A^{-1}$, it follows that

$$
(\delta_1^Z)' = (\delta_1^X)' A = \begin{pmatrix} 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} \sigma_{11}^X & \sigma_{12}^X & \sigma_{13}^X \\ 0 & \sigma_{22}^X & \sigma_{23}^X \\ 0 & 0 & \sigma_{33}^X \end{pmatrix} = \begin{pmatrix} \sigma_{11}^X & \sigma_{21}^X + \sigma_{22}^X & \sigma_{13}^X + \sigma_{23}^X \end{pmatrix}.
$$

This shows that there are no restrictions on δ_1^Z. For the constant term, we have

$$
\delta_0^X = \delta_0^Z - (\delta_1^Z)' A^{-1} \eta \iff \delta_0^Z = \delta_0^X = 0.
$$

Thus, we have obtained one additional parameter restriction,

$$
\delta_0^Z = 0.
$$

Finally, for the mean-reversion matrix under the P-measure, we have

$$
K_X^P = A K_Z^P A^{-1} \iff K_Z^P = A^{-1} K_X^P A.
$$

Since K_X^P is a free 3×3 matrix, K_Z^P is also a free 3×3 matrix. Thus, no restrictions are imposed on the P-dynamics in the equivalent canonical representation of this model.

References

Ahn, Dong-Hyun, Robert F. Dittmar, and A. Ronald Gallant, 2002, "Quadratic Term Structure Models: Theory and Evidence," Review of Financial Studies, Vol. 15, No. 1, 243-288.

Ang, Andrew and Monika Piazzesi, 2003, "A No-Arbitrage Vector Autoregression of Term Structure Dynamics with Macroeconomic and Latent Variables," Journal of Monetary Economics, Vol. 50, 745-787.

Ang, Andrew , Monika Piazzesi, and Min Wei, 2006, "What Does the Yield Curve Tell Us About GDP Growth?," Journal of Econometrics, Vol. 131, 359-403.

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

Christensen, Jens H., Francis X. Diebold, Glenn Rudebusch, and Georg Strasser, 2007, "Multivariate Comparisons of Predictive Accuracy," manuscript, University of Pennsylvania.

Cox, John C., Jonathan E. Ingersoll, and Stephen A. Ross, 1985, "A Theory of the Term Structure of Interest Rates," Econometrica, Vol. 53, No. 2, 385-407.

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

Dai, Qiang and Kenneth J. Singleton, 2002, "Expectations Puzzles, Time-Varying Risk Premia, and Affine Models of the Term Structure," Journal of Financial Economics, Vol. 63, No. 3, 415-441.

De Pooter, Michiel, Francesco Ravazzolo, and Dick van Dijk, 2007, "Predicting the Term Structure of Interest Rates: Incorporating Parameter Uncertainty and Macroeconomic Information, " Tinbergen Institute Discussion Papers.

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. and Roberto S. Mariano, 1995, "Comparing Predictive Accuracy," Journal of Business and Economic Statistics, Vol. 13, 253-265.

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., 1996, "Idiosyncratic Variation of Treasury Bill Yields," Journal of Finance, Vol. 51, 527-552.

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

Duffie, Darrell and Rui Kan, 1996, "A Yield-factor Model of Interest Rates," Mathematical Finance, Vol. 6, 379-406.

Duffie, Darrell, Jun Pan, and Kenneth J. Singleton, 2000, "Transform Analysis and Asset Pricing for Affine Jump-Diffusions," Econometrica, Vol. 68, 1343-1376.

Fama, Eugene F. and Robert R. Bliss, 1987, "The Information in Long-Maturity Forward Rates," American Economic Review, Vol. 77, 680-692.

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, 2006, "The U.S. Treasury Yield Curve: 1961 to the Present," Finance and Economics Discussion Series 2006-28, Board of Governors of the Federal Reserve System.

Hördahl, Peter, Oreste Tristani, and David Vestin, 2005, "A Joint Econometric Model of Macroeconomic and Term-Structure Dynamics," Journal of Econometrics, Vol. 131, 406-436.

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.

Krippner, Leo, 2006, "A Theoretically Consistent Version of the Nelson and Siegel Class of Yield Curve Models," Applied Mathematical Finance, Vol. 13, No. 1 (March), 39-59.

Mönch, Emanuel, 2006, "Forecasting the Yield Curve in a Data-Rich Environment: A No-Arbitrage Factor-Augmented VAR Approach," manuscrupt, Humboldt University.

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

Rudebusch, Glenn D., Eric Swanson, and Tao Wu, 2006, "The Bond Yield 'Conundrum' from a Macro-Finance Perspective," Monetary and Economic Studies, Vol. 24 (December), 83-128.

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

Vasiček, Oldrich, 1977, "An Equilibrium Characterization of the Term Structure," Journal of Financial Economics, Vol. 5, 177-188.

Williams, David, 1997, *Probability with Martingales*, Cambridge University Press.