FEDERAL RESERVE BANK OF SAN FRANCISCO

WORKING PAPER SERIES

The Affine Arbitrage-Free Class of Nelson-Siegel Term Structure Models

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

Francis X. Diebold
University of Pennsylvania and NBER

Glenn D. Rudebusch
Federal Reserve Bank of San Francisco

March 2010

Working Paper 2007-20
http://www.frbsf.org/publications/economics/papers/2007/wp07-20bk.pdf

The Affine Arbitrage-Free Class of Nelson-Siegel Term Structure Models†

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

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

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

Abstract

We derive the class of affine arbitrage-free dynamic term structure models that approximate the widely-used Nelson-Siegel yield curve specification. These arbitrage-free Nelson-Siegel (AFNS) models can be expressed as slightly restricted versions of the canonical representation of the three-factor affine arbitrage-free model. Imposing the Nelson-Siegel structure on the canonical model greatly facilitates estimation and can improve predictive performance. In the future, AFNS models appear likely to be a useful workhorse representation for term structure research.

† For helpful comments we thank the editors, referees and seminar/conference participants at the University of Chicago, Copenhagen Business School (especially Anders Bjerre Trolle and Peter Feldhütter), Stanford University (especially Ken Singleton), the Wharton School, Cambridge University, HEC Paris, Humboldt University, the NBER summer Institute, the Warwick Royal Economic Society annual meeting, the Coimbra Workshop on Financial Time Series, the New York inaugural meeting of SoFiE, the Montreal CIRANO/CIREQ Financial Econometrics meeting, and the Brazilian Finance Association annual meeting. For research assistance we thank Georg Strasser, Rong Hai, and Justin Weidner. The views expressed are those of the authors and do not necessarily reflect the views of others at the Federal Reserve Bank of San Francisco.

This version: March 2010.

1 Introduction

Understanding the dynamic evolution of the yield curve is important for many tasks, including pricing financial assets and their derivatives, managing financial risk, allocating portfolios, structuring fiscal debt, conducting monetary policy, and valuing capital goods. To investigate yield-curve dynamics, researchers have produced a vast literature with a wide variety of models. However, those models tend to be either theoretically rigorous but empirically disappointing, or empirically successful but theoretically lacking. In this paper, we introduce a theoretically rigorous yield curve model that simultaneously displays empirical tractability, good fit, and superior forecasting performance.

Because bonds trade in deep and well-organized markets, the theoretical restrictions that eliminate opportunities for riskless arbitrage across maturities and over time hold 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). Those models specify the risk-neutral evolution of the underlying yield-curve factors as well as the dynamics of risk premia. Following Duffie and Kan (1996), the affine versions of those 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 often exhibit poor empirical time series performance, especially when forecasting future yields (Duffee, 2002). In addition, and crucially, the estimation of those models is known to be problematic, in large part because of the existence of numerous likelihood maxima that have essentially identical fit to the data but very different implications for economic behavior. These empirical problems appear to reflect a pervasive 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 fewer troublesome likelihood maxima. However, the additional restrictions on model structure are not well motivated theoretically or statistically, and their arbitrary application and the computational burden of estimation effectively preclude robust model validation and thorough simulation studies of the finite-sample properties of the estimators.

In part to overcome the problems with empirical implementation of the canonical affine AF model, we develop in this paper a new class of affine AF models. This new class is based on the popular yield-curve representation that was introduced by Nelson and Siegel (1987) and extended by Diebold and Li (2006) to a dynamic Nelson-Siegel, or DNS, model. Thus, from one perspective, we take the theoretically rigorous but empirically problematic affine

AF model and make it empirically tractable by incorporating DNS elements.

From an alternative perspective, we take the DNS model, which is empirically successful but theoretically lacking, and make it rigorous by imposing absence of arbitrage. This rigor is important because the Nelson-Siegel model is extremely popular in practice, among both financial market practitioners and central banks (e.g., Svensson, 1995; Bank for International Settlements, 2005; Gürkaynak, Sack, and Wright, 2007; Nyholm, 2008). DNS's popularity stems from several sources, both empirical and theoretical, as discussed in Diebold and Li (2006). Empirically, the DNS model is simple and stable to estimate, and it is quite flexible and fits both the cross section and time series of yields remarkably well, in many countries and periods, and for many grades of bonds. Theoretically, DNS imposes certain economically-desirable properties, such as requiring the discount function to approach zero with maturity, and Diebold and Li (2006) show that it corresponds to a modern three-factor model of time-varying level, slope and curvature. However, despite its good empirical performance and a certain amount of theoretical appeal, DNS fails on an important theoretical dimension: it does not impose the restrictions necessary to eliminate opportunities for riskless arbitrage (e.g., Filipović, 1999, and Diebold, Piazzesi, and Rudebusch, 2005). This motivates us in this paper to introduce the class of AF Nelson-Siegel (AFNS) models, which are affine AF term structure models that maintain the DNS factor-loading structure.

In short, the AFNS models proposed here combine the best of the AF and DNS traditions. Approached from the AF side, they maintain the AF theoretical restrictions of the canonical affine models but can be easily estimated, because the dynamic Nelson-Siegel structure helps identify the latent yield-curve factors and delivers analytical solutions (which we provide) for zero-coupon bond prices. Approached from the DNS side, they maintain the simplicity and empirical tractability of the popular DNS models, while simultaneously enforcing the theoretically desirable property of absence of riskless arbitrage.

After deriving the new class of AFNS models, we examine their in-sample fit and out-of-sample forecast performance relative to standard DNS models. For both the DNS and the AFNS models, we estimate parsimonious and flexible versions (with both independent factors and more richly parameterized correlated factors). We find that the flexible versions of both models are preferred for in-sample fit, but that the parsimonious versions exhibit significantly better out-of-sample forecast performance. As a final comparison, we also show that an AFNS model can outperform the canonical affine AF model in forecasting.

We proceed as follows. First we present the main theoretical results of the paper; in Section 2 we derive the AFNS class of models, and in Section 3 we characterize the precise relationship between the AFNS class and the canonical representation of affine AF models. We next provide an empirical analysis of four leading DNS and AFNS models, incorporating both parsimonious and flexible versions; in Section 4 we examine in-sample fit, and in Section

5 we examine out-of-sample forecasting performance. We conclude in Section 6, and we provide proofs and additional technical details in several appendices.

2 Nelson-Siegel Term Structure Models

Here we review the DNS model and introduce the AFNS class of AF 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 months to maturity, and β_0, β_1, β_2, and λ are parameters.

As noted earlier, this representation is commonly used by central banks and financial market practitioners to fit the cross section of yields. Although such a static representation is useful for some purposes, a dynamic version is required to understand the evolution of the bond market over time. Hence Diebold and Li (2006) suggest allowing the β coefficients to vary over time, in which case, given their Nelson-Siegel loadings, the coefficients may be interpreted as time-varying level, slope and curvature factors. To emphasize this, we re-write the model as

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

Diebold and Li assume an autoregressive structure for the factors, which produces the DNS model, a fully dynamic Nelson-Siegel specification. Indeed, it is a state-space model, with the yield factors as state variables, as emphasized in Diebold, Rudebusch and Aruoba (2006).

Empirically, the DNS model is highly tractable and typically fits well. Theoretically, however, it does not require that the dynamic evolution of yields 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 preclude arbitrage at the bond prices implicit in the resulting Nelson-Siegel yield curve. In the next subsection, we show how to remedy this theoretical weakness.

2.2 The Arbitrage-Free Nelson-Siegel Model

Our derivation of the AFNS model starts from the standard continuous-time affine AF structure of 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 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 drifts and dynamics $\theta^Q : [0, T] \rightarrow \mathbf{R}^n$ and $K^Q : [0, T] \rightarrow \mathbf{R}^{n \times n}$ are bounded, continuous functions.[3] Similarly, the volatility matrix $\Sigma : [0, T] \rightarrow \mathbf{R}^{n \times n}$ is a bounded, continuous function, while $D : M \times [0, T] \rightarrow \mathbf{R}^{n \times n}$ has diagonal structure,

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

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

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

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

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

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

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

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

[2] Note that (3) refers to the risk-neutral ("Q") dynamics.

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

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,

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

where $B(t,T)$ and $C(t,T)$ are the solutions to the 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. The pricing functions imply that zero-coupon yields are

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

Given the pricing functions, for a three-factor affine model with $X_t = (X_t^1, X_t^2, X_t^3)$, the closest match to the Nelson-Siegel yield function is a yield function of the form[4]

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

with ODEs for the $B(t,T)$ functions that have the solutions

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

In this case the factor loadings exactly match Nelson-Siegel, but there is an unavoidable "yield-adjustment term", $-\frac{C(t,T)}{T-t}$, which depends only on the maturity of the bond, not on time. As described in the following proposition, there exists a class of affine AF models that satisfies the above ODEs.

[4] One could of course define "closest" in other ways. Our strategy is to find the affine AF model with factor loadings that match Nelson-Siegel exactly.

Proposition 1:

Suppose that the instantaneous risk-free rate is

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

where 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

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

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

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

and

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

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

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

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

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

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

Finally, zero-coupon bond yields are

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

Proof: See Appendix A.

The existence of an AFNS model, as defined in Proposition 1, is related to the work of Trolle and Schwartz (2009), who show that the dynamics of the forward rate curve in a general m-dimensional Heath-Jarrow-Morton (HJM) model can always be represented by a finite-dimensional Markov process with time-homogeneous volatility structure if each volatility function is given by

$$\sigma_i(t,T) = p_{n,i}(T-t)e^{-\gamma_i(T-t)}, \quad i = 1, \ldots, m,$$

where $p_{n,i}(\tau)$ is an n-order polynomial in τ. Because the forward rates in the DNS model satisfy this requirement, there exists such an AF three-dimensional HJM model. However, the simplicity of the solution in the case of the Nelson-Siegel model presented in Proposition 1 is striking.

Proposition 1 also has several interesting implications. First, the three state variables are Gaussian Ornstein-Uhlenbeck processes with a constant volatility matrix Σ.[5] The instantaneous interest rate is the sum of level and slope factors (X_t^1 and X_t^2), while the curvature factor's (X_t^3) 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 AFNS model under the Q-measure and is silent about the dynamics under the P-measure. Still, the very indirect role of curvature 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.[6] Third, Proposition 1 provides insight into the nature of the parameter λ. Although a few authors (e.g., Koopman, Mallee, and van der Wel, 2010) have considered time-varying λ, it is a constant in the AFNS model and has the interpretation as 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, and crucially, AFNS contains an additional maturity-dependent term $-\frac{C(t,T)}{T-t}$

[5]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 because 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.

[6]With a unit root in the level factor, $-\frac{C(t,T)}{T-t} \to -\infty$ as maturity increases, which implies that with an unbounded horizon T the model is not arbitrage-free. Therefore, as is often done in theoretical discussions, we impose an arbitrary maximum horizon. Alternatively, we could modify the mean-reversion matrix K^Q to include a sufficiently small $\varepsilon > 0$ in the upper left-hand position to obtain an AF model that is indistinguishable from the AFNS model in Proposition 1.

relative to DNS. This "yield-adjustment" term is a key difference between DNS and AFNS, and we now examine it in detail.

2.3 The Yield-Adjustment Term

The only parameters in the system of ODEs for the AFNS $B(t,T)$ 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 rather in the yield-adjustment term $-\frac{C(t,T)}{T-t}$. Hence in the AFNS 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 AFNS 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 is of the 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 $\overline{A} = \sigma_{11}^2+\sigma_{12}^2+\sigma_{13}^2$, $\overline{B} = \sigma_{21}^2+\sigma_{22}^2+\sigma_{23}^2$, $\overline{C} = \sigma_{31}^2+\sigma_{32}^2+\sigma_{33}^2$, $\overline{D} = \sigma_{11}\sigma_{21}+\sigma_{12}\sigma_{22}+\sigma_{13}\sigma_{23}$, $\overline{E} = \sigma_{11}\sigma_{31}+\sigma_{12}\sigma_{32}+\sigma_{13}\sigma_{33}$, and $\overline{F} = \sigma_{21}\sigma_{31}+\sigma_{22}\sigma_{32}+\sigma_{23}\sigma_{33}$.

This result has two implications. First, the fact that AFNS zero-coupon bond yields are given by an analytical formula greatly facilitates empirical implementation of AFNS 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; hence the maximally-flexible AFNS specification that can be identified has triangular volatility matrix[7]

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

Later we will quantify the yield-adjustment term and examine how it affects empirical performance in leading specifications, to which we now turn.

2.4 Four Specific Nelson-Siegel Models

In general, the DNS and AFNS 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, we focus on two versions of the DNS model that have featured prominently in recent studies, examining the effects of imposing absence of arbitrage.

In the *independent-factor DNS model*, the three state variables are independent first-order autoregressions, as in Diebold and Li (2006). The state transition equation is

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

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

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

In the *correlated-factor DNS model*, the state variables follow a first-order vector autoregression, as in Diebold, Rudebusch, and Aruoba (2006). The transition equation is

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

[7] The choice of upper or lower triangular is irrelevant.

where the stochastic shocks $\eta_t(L)$, $\eta_t(S)$, and $\eta_t(C)$ have covariance matrix $Q = qq'$, where

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

In both the independent-factor and correlated-factor DNS models, the measurement equation is

$$\begin{pmatrix} y_t(\tau_1) \\ y_t(\tau_2) \\ \vdots \\ y_t(\tau_N) \end{pmatrix} = \begin{pmatrix} 1 & \frac{1-e^{-\lambda\tau_1}}{\lambda\tau_1} & \frac{1-e^{-\lambda\tau_1}}{\lambda\tau_1} - e^{-\lambda\tau_1} \\ 1 & \frac{1-e^{-\lambda\tau_2}}{\lambda\tau_2} & \frac{1-e^{-\lambda\tau_2}}{\lambda\tau_2} - e^{-\lambda\tau_2} \\ \vdots & \vdots & \vdots \\ 1 & \frac{1-e^{-\lambda\tau_N}}{\lambda\tau_N} & \frac{1-e^{-\lambda\tau_N}}{\lambda\tau_N} - e^{-\lambda\tau_N} \end{pmatrix} \begin{pmatrix} L_t \\ S_t \\ C_t \end{pmatrix} + \begin{pmatrix} \varepsilon_t(\tau_1) \\ \varepsilon_t(\tau_2) \\ \vdots \\ \varepsilon_t(\tau_N) \end{pmatrix}, \tag{8}$$

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

The corresponding AFNS 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. To preserve affine dynamics under the P-measure, we limit our focus to essentially affine risk premium specifications (see Duffee, 2002), in which case Γ_t takes 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,

$$dX_t = K^P[\theta^P - X_t]dt + \Sigma dW_t^P, \tag{9}$$

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. Hence we focus on the two AFNS models that correspond to the two DNS models above.

In the *independent-factor AFNS model*, the three state variables are independent under the P-measure,

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

In the *correlated-factor AFNS model*, the three state variables may interact dynamically and/or their shocks may be correlated,

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

This is the most flexible AFNS model with all parameters identified. In both the independent-factor and correlated-factor AFNS models, the measurement equation is

$$\begin{pmatrix} y_t(\tau_1) \\ y_t(\tau_2) \\ \vdots \\ y_t(\tau_N) \end{pmatrix} = \begin{pmatrix} 1 & \frac{1-e^{-\lambda\tau_1}}{\lambda\tau_1} & \frac{1-e^{-\lambda\tau_1}}{\lambda\tau_1} - e^{-\lambda\tau_1} \\ 1 & \frac{1-e^{-\lambda\tau_2}}{\lambda\tau_2} & \frac{1-e^{-\lambda\tau_2}}{\lambda\tau_2} - e^{-\lambda\tau_2} \\ \vdots & \vdots & \vdots \\ 1 & \frac{1-e^{-\lambda\tau_N}}{\lambda\tau_N} & \frac{1-e^{-\lambda\tau_N}}{\lambda\tau_N} - e^{-\lambda\tau_N} \end{pmatrix} \begin{pmatrix} X_t^1 \\ X_t^2 \\ X_t^3 \end{pmatrix} - \begin{pmatrix} \frac{C(\tau_1)}{\tau^1} \\ \frac{C(\tau_2)}{\tau^2} \\ \vdots \\ \frac{C(\tau_N)}{\tau^N} \end{pmatrix} + \begin{pmatrix} \varepsilon_t(\tau_1) \\ \varepsilon_t(\tau_2) \\ \vdots \\ \varepsilon_t(\tau_N) \end{pmatrix}, \quad (10)$$

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

3 The AFNS Subclass of Canonical Affine AF Models

Before proceeding to an empirical analysis of the various DNS and AFNS models, we first answer a key theoretical question: What, precisely, are the restrictions that the AFNS model imposes on the canonical representation of three-factor affine AF model?[8]

Denoting the state variables by Y_t, the canonical $A_0(3)$ model is

$$\begin{aligned} r_t &= \delta_0^n + (\delta_1^n)' Y_t \\ dY_t &= K_n^P [\theta_n^P - Y_t] dt + \Sigma_n dW_t^P \\ dY_t &= K_n^Q [\theta_n^Q - Y_t] dt + \Sigma_n dW_t^Q, \end{aligned}$$

with $\delta_0^n \in \mathbf{R}$, $\delta_1^n, \theta_n^P, \theta_n^Q \in \mathbf{R}^3$, and $K_n^P, K_n^Q, \Sigma_n \in \mathbf{R}^{3\times 3}$.[9] If the essentially affine risk premium

[8]By this we mean the $A_0(3)$ representation, with three state variables and zero square-root processes, as detailed in Singleton (2006), Chap. 12.

[9]Note that Y_t denotes the state variables of the canonical representation, which are different from the X_t state variables in the AFNS models, and that subscripts or superscripts of "n" denote coefficients in the canonical representation.

specification $\Gamma_t = \gamma_n^0 + \gamma_n^1 Y_t$ is imposed on the model, the drift terms under the P-measure (K_n^P, θ_n^P) can be chosen independently of the drift terms under the Q-measure (K_n^Q, θ_n^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. Singleton (2006) imposes identifying restrictions under the Q-measure. Specifically, he sets the mean $\theta_n^Q = 0$, the volatility matrix Σ_n equal to the identity matrix, and he sets the mean-reversion matrix K_n^Q equal to a triangular matrix.[10] The canonical representation then has Q-dynamics

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

and P-dynamics

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

The instantaneous risk-free rate is

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

Hence there are 22 free parameters in the canonical representation of the $A_0(3)$ model class.[11]

In the AFNS class, the mean-reversion matrix under the Q-measure is triangular, so it is straightforward to derive the restrictions that must be imposed on the canonical affine representation to obtain the class of AFNS models. The procedure through which the restrictions are identified is based on so-called *affine invariant transformations*. Appendix C describes such transformations and derives the restrictions associated with the AFNS models considered in this paper. The results are summarized in Table 1, which shows that for the correlated-factor AFNS model there are three key parameter restrictions on the canonical affine model. First, $\delta_0^n = 0$, 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 $\kappa_{1,1}^{n,Q} = 0$, 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, $\kappa_{2,2}^{n,Q} = \kappa_{3,3}^{n,Q}$, so the own mean-reversion rates of the second and third factors under the Q-measure must be identical. The independent-factor AFNS model maintains the three parameter restrictions and adds nine others under both the

[10]Without loss of generality, we will take it to be upper triangular in what follows.

[11]Note that, given this canonical representation, there is no loss of generality in fixing the AFNS model mean under the Q-measure at 0 and leaving the mean under the P-measure, θ^P, to be estimated.

AFNS Model	δ_0^n, δ_1^n	κ_n^Q	κ_n^P	θ_n^P	Restrictions
Independent-factor	$\delta_0^n = 0$ $\delta_{1,3}^n = 0$	$\kappa_{1,1}^{n,Q} = \kappa_{1,2}^{n,Q} = \kappa_{1,3}^{n,Q} = 0$ $\kappa_{2,2}^{n,Q} = \kappa_{3,3}^{n,Q}$	κ_n^P is diagonal	no restriction	12
Correlated-factor	$\delta_0^n = 0$	$\kappa_{1,1}^{n,Q} = 0$ $\kappa_{2,2}^{n,Q} = \kappa_{3,3}^{n,Q}$	no restriction	no restriction	3

Table 1: **AFNS Parameter Restrictions on the $A_0(3)$ Canonical Representation.**

P- and Q-measures.[12]

The Nelson-Siegel parameter restrictions on the canonical affine AF model greatly facilitate estimation.[13] 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.

4 Estimation and In-Sample Fit of DNS and AFNS Models

Thus far we have derived the affine AF class of Nelson-Siegel term structure models, and we have explicitly characterized the restrictions that it places on the canonical $A_0(3)$ model. Here we undertake estimation of the AFNS model and illustrate its relative simplicity. We proceed in several steps. First we introduce the state-space/Kalman-filter maximum-likelihood estimation framework that we employ throughout. Second, we estimate and compare independent- and correlated-factor DNS models. Third, we estimate independent- and correlated-factor AFNS models, which we compare to each other and to their DNS counterparts, devoting special attention to the estimated yield adjustment terms. Throughout, our estimates are based on end-of-month U.S. Treasury bill and bond zero-coupon yields from January 1987 to December 2002 at sixteen maturities: 3, 6, 9, 12, 18, 24, 36, 48, 60, 84, 96, 108, 120, 180, 240, and 360 months. The data are constructed using the unsmoothed Fama-Bliss (1987) approach as described in Diebold and Li (2006).

4.1 Estimation Framework

We first display the state-space representations of the DNS and AFNS models. For the DNS models, the state transition equation is

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

[12]For both specifications, there is a further modest restriction described in Appendix C: $\kappa_{2,3}^{n,Q}$ must have the opposite sign of $\kappa_{2,2}^{n,Q}$ and $\kappa_{3,3}^{n,Q}$, but its absolute numerical size can vary freely.

[13]Note that in the AFNS 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^n and K_n^Q.

where $X_t = (L_t, S_t, C_t)$, and the measurement equation is

$$y_t = BX_t + \varepsilon_t. \tag{11}$$

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

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

where $\Delta t = T - t$. We compute conditional moments of discrete observations and obtain the AFNS state transition equation

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

where Δt is the time between observations. The AFNS measurement equation is[14]

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

In both the DNS and AFNS environments, the assumed 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 the matrix H is diagonal, and the matrix Q is diagonal in the independent-factor and non-diagonal in the correlated-factor case. In the AFNS case, moreover, Q has special structure,

$$Q = \int_0^{\Delta t} e^{-K^P s}\Sigma\Sigma' e^{-(K^P)' s} ds.$$

In addition, the transition and measurement errors are assumed orthogonal to the initial state.

Now we consider Kalman filtering, which we use to evaluate the likelihood functions of the DNS and AFNS models. We initialize the filter at the unconditional mean and variance of the state variables under the P-measure.[15] For the DNS models we have $X_0 = \mu$ and $\Sigma_0 = V$, where V solves $V = AVA' + Q$. For the AFNS models we have $X_0 = \theta^P$ and $\Sigma_0 = \int_0^\infty e^{-K^P s}\Sigma\Sigma' e^{-(K^P)' s} ds$, which we calculate using the analytical solutions provided in

[14] Note that the matrix B is identical in the DNS and AFNS models (compare equations (8) and (10)). The only difference is the addition of the vector A containing the yield-adjustment terms in the AFNS models.

[15] We ensure covariance stationarity under the P measure in the DNS case by restricting the eigenvalues of A to be less than 1, and in the AFNS case by restricting the real component of each eigenvalue of K^P to be positive.

Fisher and Gilles (1996).

Denote the information available at time t by $Y_t = (y_1, y_2, \ldots, y_t)$, and denote model parameters by ψ. Consider period $t-1$ and suppose that the state update X_{t-1} and its mean square error matrix Σ_{t-1} have been obtained. The prediction step is

$$X_{t|t-1} = E^P[X_t|Y_{t-1}] = \Phi_t^{X,0}(\psi) + \Phi_t^{X,1}(\psi)X_{t-1},$$

$$\Sigma_{t|t-1} = \Phi_t^{X,1}(\psi)\Sigma_{t-1}\Phi_t^{X,1}(\psi)' + Q_t(\psi),$$

where for the DNS models we have $\Phi_t^{X,0} = (I - A)\mu$, $\Phi_t^{X,1} = A$, and $Q_t = Q$, and for the AFNS models we have $\Phi_t^{X,0} = (I - \exp(-K^P\Delta t))\theta^P$, $\Phi_t^{X,1} = \exp(-K^P\Delta t)$, and $Q_t = \int_0^{\Delta t} e^{-K^P s}\Sigma\Sigma' e^{-(K^P)'s} ds$, where Δt is the time between observations.

In the time-t update step, $X_{t|t-1}$ is improved by using the additional information contained in Y_t. We have

$$X_t = E[X_t|Y_t] = X_{t|t-1} + \Sigma_{t|t-1}B(\psi)'F_t^{-1}v_t,$$

$$\Sigma_t = \Sigma_{t|t-1} - \Sigma_{t|t-1}B(\psi)'F_t^{-1}B(\psi)\Sigma_{t|t-1},$$

where

$$v_t = y_t - E[y_t|Y_{t-1}] = y_t - A(\psi) - B(\psi)X_{t|t-1},$$

$$F_t = cov(v_t) = B(\psi)\Sigma_{t|t-1}B(\psi)' + H(\psi),$$

$$H(\psi) = diag(\sigma_\varepsilon^2(\tau_1), \ldots, \sigma_\varepsilon^2(\tau_N)).$$

At this point, the Kalman filter has delivered all ingredients needed to evaluate the Gaussian log likelihood, the prediction-error decomposition of which is

$$\log l(y_1, \ldots, y_T; \psi) = \sum_{t=1}^{T} \Big(-\frac{N}{2}\log(2\pi) - \frac{1}{2}\log|F_t| - \frac{1}{2}v_t' F_t^{-1} v_t \Big),$$

where N is the number of observed yields. We numerically maximize the likelihood with respect to ψ using the Nelder-Mead simplex algorithm. Upon convergence, we obtain standard errors from the estimated covariance matrix,

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

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

	A Matrix			Mean
	$A_{\cdot,1}$	$A_{\cdot,2}$	$A_{\cdot,3}$	μ
$A_{1,\cdot}$	0.9827 (0.0128)	0	0	0.0696 (0.0137)
$A_{2,\cdot}$	0	0.9778 (0.0166)	0	-0.0249 (0.0151)
$A_{3,\cdot}$	0	0	0.9189 (0.0284)	-0.0108 (0.0079)

	q Matrix		
	$q_{\cdot,1}$	$q_{\cdot,2}$	$q_{\cdot,3}$
$q_{1,\cdot}$	0.0025 (0.0002)	0	0
$q_{2,\cdot}$	0	0.0033 (0.0002)	0
$q_{3,\cdot}$	0	0	0.0075 (0.0004)

Table 2: **Estimated Independent-Factor DNS Model.** The left panel contains the estimated A matrix and μ vector. The right panel contains the estimated q matrix. Standard errors appear in parentheses. The estimated λ is 0.06040 (0.00100) for maturities measured in months. The maximized log likelihood is 16,332.94.

4.2 DNS Model Estimation

Independent-factor DNS estimates appear in Table 2, and correlated-factor DNS estimates appear in Table 3. In both models the level factor is the most persistent, and the curvature factor is least persistent. In the correlated-factor DNS model, only one off-diagonal element of the estimated A-matrix is statistically significant.[16]

Volatility parameters are most easily compared by converting from Cholesky factors to conditional covariance matrices. For independent-factor DNS we have

$$Q^{DNS}_{indep} = qq' = \begin{pmatrix} 6.17\times 10^{-6} & 0 & 0 \\ 0 & 1.11\times 10^{-5} & 0 \\ 0 & 0 & 5.58\times 10^{-5} \end{pmatrix}, \tag{12}$$

and for correlated-factor DNS we have

$$Q^{DNS}_{corr} = qq' = \begin{pmatrix} 6.03\times 10^{-6} & -5.47\times 10^{-6} & 6.76\times 10^{-6} \\ -5.47\times 10^{-6} & 1.01\times 10^{-5} & -4.73\times 10^{-6} \\ 6.76\times 10^{-6} & -4.73\times 10^{-6} & 5.09\times 10^{-5} \end{pmatrix}. \tag{13}$$

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

The independent- and correlated-factor DNS models are nested, so we can test the independent-

[16]Interestingly, the significant parameter is $A_{S_t,C_{t-1}}$, which is the key non-zero off-diagonal element required in Proposition 1 for the AFNS specification.

	A Matrix			Mean
	$A_{\cdot,1}$	$A_{\cdot,2}$	$A_{\cdot,3}$	μ
$A_{1,\cdot}$	0.9874	0.0050	-0.0097	0.0723
	(0.0165)	(0.0183)	(0.0157)	(0.0145)
$A_{2,\cdot}$	0.0066	0.9332	0.0819	-0.0294
	(0.0228)	(0.0229)	(0.0202)	(0.0159)
$A_{3,\cdot}$	0.0152	0.0401	0.9011	-0.0120
	(0.0526)	(0.0418)	(0.0377)	(0.0126)

	q Matrix		
	$q_{\cdot,1}$	$q_{\cdot,2}$	$q_{\cdot,3}$
$q_{1,\cdot}$	0.0025	0	0
	(0.0001)		
$q_{2,\cdot}$	-0.0022	0.0023	0
	(0.0003)	(0.0001)	
$q_{3,\cdot}$	0.0028	0.0006	0.0066
	(0.0007)	(0.0006)	(0.0004)

Table 3: **Estimated Correlated-Factor DNS Model.** The left panel contains the estimated A matrix and μ vector. The right panel contains the estimated q matrix. Standard errors appear in parentheses. The estimated λ is 0.06248 (0.00109) for maturities measured in months. The maximized log likelihood is 16,415.36.

factor restrictions using a standard likelihood-ratio (LR) test. Under the null hypothesis of independent-factor DNS, $LR = 2[\log L(\theta_{corr}) - \log L(\theta_{indep})] \sim \chi^2(9)$. We obtain $LR = 164.8$, with associated p-value less than .0001, so we would formally reject the restrictions imposed in the independent-factor DNS model. This rejection reflects an elevated negative correlation between the shocks to the level and slope factors 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.

Crucially, however, the extra parameters in the correlated-factor model, although statistically significant, appear economically unimportant. That is, the increased flexibility of the correlated-factor DNS model provides little advantage in fitting observed yields, as documented in Table 4, which reports means and root mean squared errors (RMSEs) for model residuals. The RMSE differences appear negligible (typically less than one half of one basis point), maturity-by-maturity, and no consistent advantage across maturities accrues to the correlated-factor model. Interestingly, both models have difficulty fitting yields beyond the 10-year maturity, which suggests that a maturity-dependent yield adjustment term could improve fit. We now examine the empirical performance of AFNS models, which incorporate precisely such yield adjustments.

4.3 AFNS Model Estimation

Thus far we have examined just one simple model (DNS), comparing fit in the independent- and correlated-factor cases. Now we bring AFNS into the mix, and things get more interesting. In particular, we can compare independent- and correlated-factor cases, with and without imposition of absence of arbitrage. 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-

	DNS Indep-Factor		DNS Corr-Factor		AFNS Indep-Factor		AFNS Corr-Factor	
Maturity	Mean	RMSE	Mean	RMSE	Mean	RMSE	Mean	RMSE
3	-1.64	12.26	-1.84	11.96	-2.85	18.53	-2.49	11.55
6	-0.24	1.09	-0.29	1.34	-1.19	7.12	-0.03	0.64
9	-0.54	7.13	-0.51	6.92	-1.24	3.45	-0.33	6.91
12	4.04	11.19	4.11	10.86	3.58	9.60	3.72	10.14
18	7.22	10.76	7.28	10.42	7.14	10.43	5.53	8.33
24	1.18	5.83	1.19	5.29	1.37	5.94	-1.18	4.37
36	-0.07	1.51	-0.19	2.09	0.30	1.98	-1.10	3.16
48	-0.67	3.92	-0.85	4.03	-0.40	3.72	0.93	4.13
60	-5.33	7.13	-5.51	7.31	-5.27	6.82	-2.01	5.22
84	-1.22	4.25	-1.30	4.25	-1.50	4.29	0.89	3.83
96	1.31	2.10	1.29	2.02	1.03	2.11	1.05	1.83
108	0.03	2.94	0.07	3.11	-0.11	3.02	-3.23	5.26
120	-5.11	8.51	-5.01	8.53	-4.95	8.23	-11.65	14.00
180	24.11	29.44	24.40	29.66	27.87	32.66	3.85	16.53
240	25.61	34.99	26.00	35.33	35.96	42.61	4.32	23.97
360	-29.62	37.61	-29.12	37.18	1.37	22.03	-0.81	23.04

Table 4: **Summary Statistics for in-Sample Model Fit.** Residual means and root mean squared errors for sixteen maturities. Maturities are in months; means and RMSEs are in basis points.

	$K^P_{\cdot,1}$	$K^P_{\cdot,2}$	$K^P_{\cdot,3}$	θ^P
$K^P_{1,\cdot}$	0.0816	0	0	0.0710
	(0.0615)			(0.0129)
$K^P_{2,\cdot}$	0	0.2114	0	-0.0282
		(0.1780)		(0.0173)
$K^P_{3,\cdot}$	0	0	1.2330	-0.0093
			(0.4240)	(0.0061)

(Left panel: K^P Matrix; Mean)

	$\Sigma_{\cdot,1}$	$\Sigma_{\cdot,2}$	$\Sigma_{\cdot,3}$
$\Sigma_{1,\cdot}$	0.0051	0	0
	(0.0001)		
$\Sigma_{2,\cdot}$	0	0.0110	0
		(0.0006)	
$\Sigma_{3,\cdot}$	0	0	0.0264
			(0.0014)

(Right panel: Σ Matrix)

Table 5: **Estimated Independent-Factor AFNS Model.** The left panel contains the estimated K^P matrix and θ^P vector. The right panel contains the estimated Σ matrix. Standard errors appear in parentheses. The estimated λ is 0.5975 (0.0115) for maturities measured in years. The maximized log likelihood is 16,279.92.

sample forecast exercise, which we pursue with the AFNS model in the next section.[17] By comparison, the estimation of the AFNS model is 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 AFNS 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.[18]

The estimated independent-factor AFNS model is reported in Table 5. Although the independent-factor DNS and AFNS models are non-nested, they contain the same number of parameters, so their likelihoods can be compared directly. The lower log likelihood value obtained for the AFNS model (16,280 vs. 16,332) suggests weaker in-sample performance, which appears consistent with the RMSEs in Table 4.

Although the two independent-factor models differ statistically, they are quite similar economically, as can be seen in two ways. First, we compare mean reversion matrices, covariance matrices, and mean vectors. To compare the independent-factor AFNS mean-reversion matrix to that of the independent-factor DNS model, we translate the continuous-time matrix in Table 5 into the one-month conditional mean-reversion matrix,

$$\exp\Big(-K^P\frac{1}{12}\Big) = \begin{pmatrix} 0.993 & 0 & 0 \\ 0 & 0.983 & 0 \\ 0 & 0 & 0.902 \end{pmatrix}. \tag{14}$$

[17] For example, Rudebusch, Swanson, and Wu (2006) report difficulty replicating the published estimates of a no-arbitrage model even though they use identical data and estimation programs.

[18] 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).

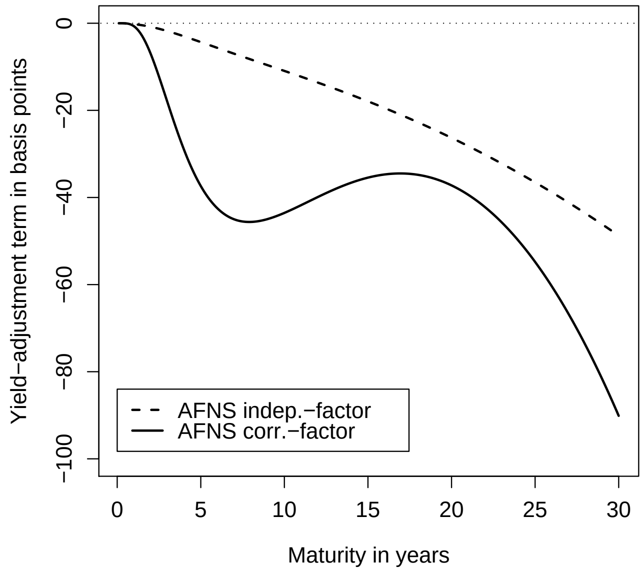

Figure 1: **Yield-Adjustment Terms for AFNS Models.**

Similarly, we convert the volatility matrix into a one-month conditional covariance matrix

$$Q_{indep}^{AFNS} = \int_0^{\frac{1}{12}} e^{-K^P s} \Sigma \Sigma' e^{-(K^P)' s} ds = \begin{pmatrix} 2.15 \times 10^{-6} & 0 & 0 \\ 0 & 9.94 \times 10^{-6} & 0 \\ 0 & 0 & 5.26 \times 10^{-5} \end{pmatrix}. \quad (15)$$

Inspection reveals that the mean reversion matrix and covariance matrix (and also the factor mean vector) are similar across the independent-factor DNS and AFNS models.

Second, the similarity of the independent-factor DNS and AFNS models can be seen by noting that they make identical assumptions about the P-dynamics and therefore differ only by the yield-adjustment term, which is quite rigid in the independent-factor case. In particular, the independent-factor AFNS yield adjustment is

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

and is plotted in Figure 1. It is everywhere negative, monotonically increasing in absolute value, and very smooth. Presumably a more flexible yield-adjustment term is needed to

	K^P Matrix			Mean
	$K^P_{\cdot,1}$	$K^P_{\cdot,2}$	$K^P_{\cdot,3}$	θ^P
$K^P_{1,\cdot}$	5.2740	9.0130	-10.7100	0.0794
	(1.3100)	(1.4200)	(1.4800)	(0.0084)
$K^P_{2,\cdot}$	-0.2848	0.5730	-0.5528	-0.0396
	(1.3200)	(2.3200)	(2.7600)	(0.0200)
$K^P_{3,\cdot}$	-37.3100	-66.7700	80.0900	-0.0279
	(11.0000)	(11.9000)	(12.1000)	(0.0193)

	Σ Matrix		
	$\Sigma_{\cdot,1}$	$\Sigma_{\cdot,2}$	$\Sigma_{\cdot,3}$
$\Sigma_{1,\cdot}$	0.0154	0	0
	(0.0004)		
$\Sigma_{2,\cdot}$	-0.0013	0.0117	0
	(0.0051)	(0.0018)	
$\Sigma_{3,\cdot}$	-0.1641	-0.0590	0.0001
	(0.0069)	(0.0106)	(6.8900)

Table 6: **Estimated Correlated-Factor AFNS Model.** The left panel contains the estimated K^P matrix and θ^P vector. The right panel contains the estimated Σ matrix. Standard errors appear in parentheses. The estimated λ is 0.8244 (0.0122) for maturities measured in years. The maximized log likelihood is 16,494.29.

achieve substantial improvement in fit. The correlated-factor AFNS model, to which we now turn, achieves this.

We begin with two model comparisons that involve correlated-factor AFNS. First consider independent- vs. correlated-factor AFNS. The models are nested, so under the null hypothesis of independent-factor AFNS, $LR = 2[\log L(\theta_{corr}) - \log L(\theta_{indep})] \sim \chi^2(9)$. We obtain $LR =$ 428.7, with associated p-value less than .0001, so independent-factor AFNS is dominated by correlated-factor AFNS. Second, consider correlated-factor DNS vs. correlated-factor AFNS. The models are non-nested but contain equal numbers of parameters, so we compare their log likelihoods directly, with the clear result that correlated-factor DNS is dominated by correlated-factor AFNS.

Combining the model comparison results above with those reported earlier in Section 4.2, correlated-factor AFNS emerges as the clear in-sample favorite among all the various combinations of independent-factor, correlated-factor, DNS and AFNS models. Presumably, this is due to the greater flexibility of the correlated-factor AFNS yield adjustment. We report the estimated correlated-factor AFNS model in Table 6, from which we can infer the estimated yield adjustment. In population, the adjustment is

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

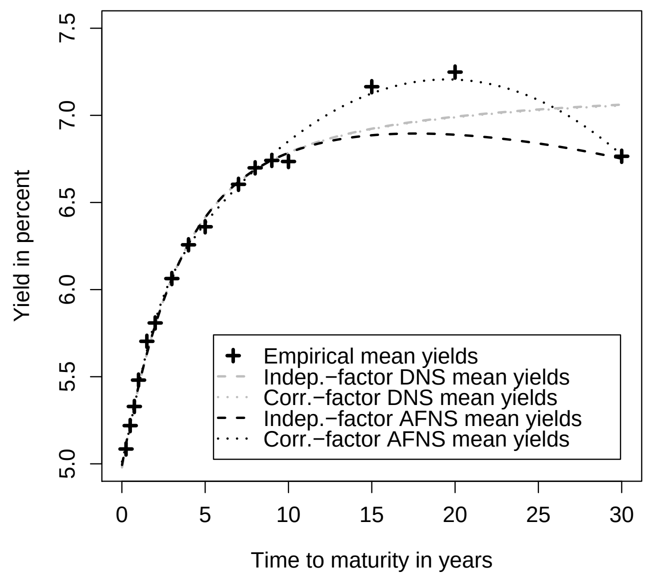

Figure 2: **Mean Yield Curves.** We show the empirical mean yield curve, and the independent- and correlated-factor DNS and AFNS model mean yield curves.

Replacing population parameters with estimates delivers the corresponding estimated yield adjustment, which we plot in Figure 1. It is indeed more flexible, with an interesting hump in the 15- to 20-year maturity range, which improves the fit of those long-term yields in particular, although it also helps with shorter maturities.

Another way to appreciate the role of the yield adjustment term is to compare the mean fitted yield curves from the independent- and correlated-factor AFNS and DNS models to the sample mean yield curve, which is done in Figure 2. All of the models match the mean yield curve well for maturities up to ten years, but their behavior diverges for longer maturities. Note that the DNS model curve is monotonically increasing, while with the yield-adjustment terms, the AFNS models can bend downward and achieve better long-maturity fit.[19]

The enhanced flexibility produced by the correlated-factor AFNS yield adjustment term allows the level factor to become less persistent, as evidenced by the estimated one-month

[19]This result suggests why the DNS model is not arbitrage free. At very long maturities, only the level factor has any appreciable influence on bond yields. To eliminate the arbitrage opportunity from going long on a bond with very long maturity and hedging the risk by shorting a bond with a slightly shorter maturity, eventually the yield curve must slope downwards (an application of Jensen's inequality and an illustration of convexity), which the DNS model cannot support.

conditional mean-reversion matrix

$$\exp\Big(-K^P\frac{1}{12}\Big) = \begin{pmatrix} 0.917 & -0.107 & 0.122 \\ 0.0390 & 0.981 & 0.0112 \\ 0.456 & 0.769 & 0.0667 \end{pmatrix}. \tag{16}$$

Evidently, the level factor becomes less persistent once the flexible correlated-factor AFNS yield adjustment is incorporated, because the level factor is more free to work with slope and curvature to improve fit at shorter maturities, given that the yield adjustment is most helpful at long maturities.

The one-month conditional covariance matrix is

$$Q_{corr}^{AFNS} = \int_0^{\frac{1}{12}} e^{-K^P s}\Sigma\Sigma' e^{-(K^P)' s} ds = \begin{pmatrix} 7.42\times10^{-6} & -6.11\times10^{-6} & -7.62\times10^{-6} \\ -6.11\times10^{-6} & 1.07\times10^{-5} & 5.89\times10^{-7} \\ -7.62\times10^{-6} & 5.89\times10^{-7} & 1.87\times10^{-4} \end{pmatrix}. \tag{17}$$

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 shocks 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. On the other hand, the estimated covariances of the shocks in the DNS models are likely to be unbiased as they are varied to provide the best fit for the P-dynamics without any implications for the cross-sectional fit of the model.

5 Out-of-Sample Predictive Performance

Here we investigate whether the in-sample superiority of the correlated-factor AFNS model carries over to out-of-sample forecast accuracy. We first describe the recursive estimation and prediction procedure employed. Second, we compare performance of the four uncorrelated/correlated factor DNS/AFNS models, exactly as in the in-sample analysis of Section 4 except that we work out-of-sample as opposed to in-sample. Third, we compare the out-of-sample predictive performance of AFNS to that of the canonical $A_0(3)$ model.

5.1 Construction of Out-of-Sample Forecasts

We construct six- and twelve-month-ahead forecasts from the four DNS and AFNS models for yields at various maturities. We estimate and forecast using an expanding sample. The first estimation sample is January 1987 to December 1996; then January 1987 to January 1997, and so on. The largest estimation sample for the one-month-ahead forecasts ends in November 2002 (72 forecasts in all). For the six- and twelve-month horizons, the largest samples end in June 2002 and December 2001 (67 and 61 forecasts), respectively.

Under quadratic loss the optimal forecast is simply the relevant conditional expectation. The optimal DNS forecast for a maturity-τ yield made at time t for time $t+h$ is therefore

$$y_{t+h,t}^{DNS}(\tau) \equiv E_t^P[y_{t+h}(\tau)] = E_t^P[L_{t+h}] + E_t^P[S_{t+h}]\Big(\frac{1-e^{-\lambda\tau}}{\lambda\tau}\Big) + E_t^P[C_{t+h}]\Big(\frac{1-e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau}\Big). \tag{18}$$

But from the first-order transition dynamics we have immediately

$$E_t^P[X_{t+h}] = \big(\sum_{i=0}^{h-1} A^i\big)(I-A)\mu + A^h X_t, \tag{19}$$

where $X_t = (L_t, S_t, C_t)$. The straightforward forecasting of the state vector (19) translates into straightforward forecasting of the yield vector via (18).

Similarly, the optimal AFNS forecast for a maturity-τ yield made at time t for time $t+h$ is

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

where

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

and $X_t = (X_t^1, X_t^2, X_t^3)$.[20]

5.2 Evaluation of Out-of-Sample Forecasts

Predictive accuracy has been a key metric to evaluate the adequacy of yield-curve models; recent analyses include Ang and Piazzesi (2003), Hördahl, Tristani, and Vestin (2005), De Pooter, Ravazzolo, and van Dijk (2007), Chua et al. (2008), Mönch (2008), and Zantedeschi, Damien and Polson (2009). Define the h-step-ahead forecast error for maturity τ as $\hat{e}_{t+h,t}(\tau) = y_{t+h}(\tau) - \hat{y}_{t+h,t}(\tau)$. Then the forecast performances of the four models (DNS/AFNS, independent/correlated) are compared using the root mean squared forecast error (RMSFE) for $\tau = 3, 12, 36, 60, 120, 360$, and $h = 6, 12$ (in months). These RMSFEs are shown in Table 7. For each of the 12 combinations of yield maturity and forecast horizon, the most accurate model's RMSFE is boxed. The results are striking. In 11 of the 12 combinations, the most accurate model is the independent-factor AFNS model. In particu-

[20] Making the formulae operational of course requires replacing population system parameters with estimates. We denote the operational forecasts by $\hat{y}_{t+h,t}^{DNS}(\tau)$ and $\hat{y}_{t+h,t}^{AFNS}(\tau)$.

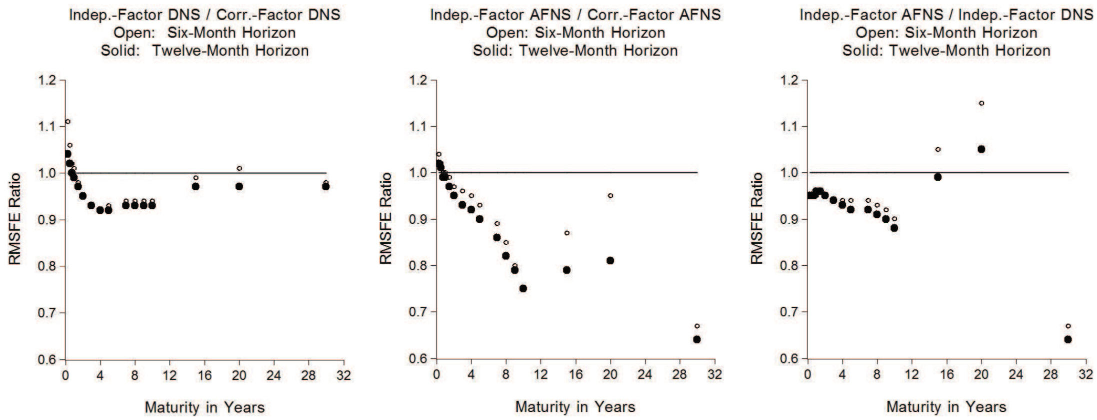

Figure 3: **Out-of-Sample Root Mean Squared Forecast Error Ratios.**

lar, the in-sample advantage of the correlated-factor AFNS model disappears out of sample. Evidently, the correlated-factor AFNS model is prone to in-sample overfitting due to its rich P-dynamics.[21]

In examining forecast performance, we are interested in two broad questions. First, how does the forecast performance of the correlated-factor models compare to that of the independent-factor models, and second, how does the imposition of AF structure affect forecast performance. Figure 3 suggests the answers, showing ratios of RMSFEs for various combinations of model, maturity and forecast horizon. The first question is addressed in the left and middle panels, which show the ratios of the independent-factor and correlated-factor DNS models and the independent-factor and correlated-factor AFNS models, respectively. The ratios are almost uniformly below one, which supports the parsimonious models.

The second question is addressed in the right panel, which shows RMSFE ratios of the independent-factor AFNS and DNS models. The evidence is somewhat mixed—due largely to anomalous behavior at the twenty-year maturity—but overall the AF version dominates. Therefore, out-of-sample forecast performance appears largely improved by imposing freedom from arbitrage, especially at the longer 12-month forecast horizon.

5.3 Comparison to Duffee (2002)

An important remaining issue is the forecasting performance of AFNS relative to the canonical AF $A_0(3)$ model. In this sub-section we address that issue, and in so doing we provide insight

[21]The single case in which the independent-factor AFNS model is not the most accurate pertains to the 3-month yield. This advantage likely reflects idiosyncratic fluctuations in short-term Treasury bill yields from institutional factors unrelated to yields on longer-maturity Treasuries, as described by Duffee (1996). The more flexible models appear to have a slight advantage in fitting such idiosyncratic movements.

Model	Forecast Horizon in Months h=6	h=12
3-Month Yield		
DNS$_{indep}$	96.87	173.39
DNS$_{corr}$	$\boxed{87.43}$	166.91
AFNS$_{indep}$	91.63	164.70
AFNS$_{corr}$	88.49	$\boxed{161.94}$
1-Year Yield		
DNS$_{indep}$	103.25	170.85
DNS$_{corr}$	102.71	173.14
AFNS$_{indep}$	$\boxed{98.49}$	$\boxed{163.46}$
AFNS$_{corr}$	98.63	165.50
3-Year Yield		
DNS$_{indep}$	92.22	135.24
DNS$_{corr}$	99.55	145.82
AFNS$_{indep}$	$\boxed{86.99}$	$\boxed{126.95}$
AFNS$_{corr}$	90.64	135.79
5-Year Yield		
DNS$_{indep}$	87.87	122.09
DNS$_{corr}$	94.95	132.40
AFNS$_{indep}$	$\boxed{82.41}$	$\boxed{112.85}$
AFNS$_{corr}$	88.15	124.87
10-Year Yield		
DNS$_{indep}$	74.71	105.02
DNS$_{corr}$	79.48	112.37
AFNS$_{indep}$	$\boxed{67.48}$	$\boxed{92.39}$
AFNS$_{corr}$	90.21	123.89
30-Year Yield		
DNS$_{indep}$	71.35	96.90
DNS$_{corr}$	72.71	99.68
AFNS$_{indep}$	$\boxed{48.06}$	$\boxed{61.97}$
AFNS$_{corr}$	71.38	96.75

Table 7: **Out-of-Sample Root Mean Squared Forecast Errors, Four Models.** For each maturity and horizon, the smallest RMSFE is boxed. Units are basis points.

	K^P Matrix			Mean
	$K^P_{\cdot,1}$	$K^P_{\cdot,2}$	$K^P_{\cdot,3}$	θ^P
$K^P_{1,\cdot}$	0.0299 (0.0249)	0	0	0.0609 (0.0224)
$K^P_{2,\cdot}$	0	0.7436 (0.1550)	0	-0.0162 (0.0054)
$K^P_{3,\cdot}$	0	0	2.5250 (0.3540)	-0.0043 (0.0026)

	Σ Matrix		
	$\Sigma_{\cdot,1}$	$\Sigma_{\cdot,2}$	$\Sigma_{\cdot,3}$
$\Sigma_{1,\cdot}$	0.0069 (0.0002)	0	0
$\Sigma_{2,\cdot}$	0	0.0208 (0.0004)	0
$\Sigma_{3,\cdot}$	0	0	0.0363 (0.0009)

Table 8: **Estimated Independent-Factor AFNS Model, Duffee (2002) Data Set.** The left panel contains the estimated K^P matrix and θ^P vector. The right panel contains the estimated Σ matrix. Standard errors appear in parentheses. The estimated λ is 0.8131 (0.0183) for maturities measured in years. The maximized log likelihood is 14,948.79.

into the benefits of imposing the Nelson-Siegel restrictions.

We hasten to add that, quite apart from any effects on forecasting performance, imposition of the Nelson-Siegel restrictions delivers clear benefits simply in achieving estimation tractability. The simple estimation of AFNS contrasts starkly with the "challenging" estimation of the maximally-flexible $A_0(3)$ model, whose recalcitrance is well-known. Our earlier-implemented expanding-sample AFNS estimation, for example, is infeasible for the maximally-flexible $A_0(3)$ model. Hence, instead of estimating a somewhat arbitrary $A_0(3)$ model for our data set, we take an existing optimized empirical $A_0(3)$ model from the literature, specifically Duffee (2002), and we compare it to an AFNS model estimated on the same data.

Duffee (2002) examines the predictive performance of the $A_0(3)$ model class, estimating both the maximally-flexible version (given an essentially affine risk premium structure) and a more parsimonious "preferred" specification on a single sample from January 1952 to December 1994.[22] Fixing the parameters at estimated values, Duffee sequentially updates the state variables and produces three-, six- and twelve-month-ahead yield forecasts.[23]

We extend Duffee's forecast comparison to include the independent-factor AFNS model, estimated using three-month, six-month, one-year, two-year, five-year, and ten-year yields from January 1952 to December 1994, as reported in Table 8.[24] Fixing parameters at estimated values, we sequentially update the state variables using the Kalman filter. Based on the updated state variables, we produce six- and twelve-month-ahead yield forecasts as above.

RMSFEs appear in Table 9 for the two models examined by Duffee (2002) (random walk and $A_0(3)$) plus the independent-factor AFNS model, for the six-month, two-year and ten-

[22] The data used are available at http://econ.jhu.edu/People/Duffee/index.htm.

[23] The estimation method used by Duffee (2002) differs from ours in that he avoids filtering by assuming that the six-month, two-year, and ten-year yields are observed without error. Duffee therefore evaluates out-of-sample forecast performance only at those maturities.

[24] There are 21 parameters estimated in Duffee's preferred $A_0(3)$ model and 16 parameters estimated in our AFNS model, including the six measurement error standard deviations.

	Forecast Horizon in Months	
Maturity/Model	h=6	h=12
<u>6-Month Yield</u>		
Random Walk	40.0	48.4
Preferred $A_0(3)$	36.5	42.1
AFNS$_{indep}$	$\boxed{34.0}$	$\boxed{41.3}$
<u>2-Year Yield</u>		
Random Walk	65.2	76.2
Preferred $A_0(3)$	56.6	60.0
AFNS$_{indep}$	$\boxed{54.3}$	$\boxed{59.0}$
<u>10-Year Yield</u>		
Random Walk	66.9	81.5
Preferred $A_0(3)$	63.6	73.8
AFNS$_{indep}$	$\boxed{60.7}$	$\boxed{71.8}$

Table 9: **Out-of-Sample Root Mean Squared Forecast Errors, Three Models.** We show RMSFEs for the random walk model, the preferred $A_0(3)$ model as selected and estimated by Duffee (2002, Table 8), and the independent-factor AFNS estimated using Duffee's data set. For each maturity and horizon, the smallest RMSFE is boxed. Units are basis points.

year yield maturities examined by Duffee. RMSFEs for each forecasting model are based on 42 six-month-ahead forecasts from January 1995 to June 1998, and 36 twelve-month-ahead forecasts from January 1995 to December 1997. For each maturity/horizon combination, the independent-factor AFNS forecasts are the most accurate, consistently outperforming both the random walk and Duffee's preferred $A_0(3)$ model. This superior out-of-sample forecast performance indicates the AFNS class is a leading and, not least, well-identified member of the general $A_0(3)$ class of models.

6 Concluding Remarks

Asset pricing, portfolio allocation, and risk management are 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.

We have bridged the divide, proposing Nelson-Siegel-inspired models that enforce absence of arbitrage. We analyzed our models theoretically, relating them to the canonical Dai-Singleton (2000) representation of three-factor arbitrage-free affine models. We also analyzed our models empirically, both in terms of in-sample fit and out-of-sample prediction. As regards in-sample fit, we showed that the Nelson-Siegel parameter restrictions greatly facilitate estimation, enabling one to escape the challenging $A_0(3)$ estimation environment in favor of the simple and robust AFNS environment, and that the data strongly favor the correlated-factor AFNS specification.

As regards out-of-sample prediction, we showed that the tables are turned: the more parsimonious independent-factor models fare better. The results also suggest that gains may be achieved by imposing absence of arbitrage, particularly for moderate to long yield maturities and forecast horizons, although the evidence is much less conclusive than for in-sample fit. All told, the independent-factor AFNS model fares well in out-of-sample prediction, consistently outperforming, for example, the canonical $A_0(3)$.

Going forward, this new AFNS structure appears likely to be a useful representation for term structure research, as its embedded three-factor structure (level, slope, curvature) maintains fidelity to key aspects of term-structure data that have been recognized at least since

Litterman and Scheinkman (1991), while simultaneously imposing absence of arbitrage. On the theoretical side, it has recently been significantly enriched to include nonlinear regime-switching dynamics by Zantedeschi, Damien and Polson (2009). On the applied side, it has recently been extended in Christensen, Lopez, and Rudebusch (2008) to provide a joint empirical model of nominal and real yield curves and in Christensen, Lopez, and Rudebusch (2009) to model the interbank lending market.

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

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

Because

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

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

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

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.

Appendix B: The AFNS Yield-Adjustment Term

In the AFNS models the yield-adjustment term is in general

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

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:

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

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

$$\begin{aligned} I_3 &= \frac{\overline{C}}{2}\frac{1}{T-t}\int_t^T B^3(s,T)ds = \frac{\overline{C}}{2}\frac{1}{T-t}\int_t^T \Big\{(T-s)e^{-\lambda(T-s)} - \frac{1-e^{-\lambda(T-s)}}{\lambda}\Big\}^2 ds \\ &= \overline{C}\Big[\frac{1}{2\lambda^2} + \frac{1}{\lambda^2}e^{-\lambda(T-t)} - \frac{1}{4\lambda}(T-t)e^{-2\lambda(T-t)} - \frac{3}{4\lambda^2}e^{-2\lambda(T-t)} - \frac{2}{\lambda^3}\frac{1-e^{-\lambda(T-t)}}{T-t} + \frac{5}{8\lambda^3}\frac{1-e^{-2\lambda(T-t)}}{T-t}\Big]. \end{aligned}$$

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

$$\begin{aligned} I_5 &= \overline{E}\frac{1}{T-t}\int_t^T B^1(s,T)B^3(s,T)ds = \overline{E}\frac{1}{T-t}\int_t^T \Big[-(T-s)\Big]\Big[(T-s)e^{-\lambda(T-s)} - \frac{1-e^{-\lambda(T-s)}}{\lambda}\Big]ds \\ &= \overline{E}\Big[\frac{3}{\lambda^2}e^{-\lambda(T-t)} + \frac{1}{2\lambda}(T-t) + \frac{1}{\lambda}(T-t)e^{-\lambda(T-t)} - \frac{3}{\lambda^3}\frac{1-e^{-\lambda(T-t)}}{T-t}\Big]. \end{aligned}$$

$$\begin{aligned} I_6 &= \overline{F}\frac{1}{T-t}\int_t^T B^2(s,T)B^3(s,T)ds = \overline{F}\frac{1}{T-t}\int_t^T \Big[-\frac{1-e^{-\lambda(T-s)}}{\lambda}\Big]\Big[(T-s)e^{-\lambda(T-s)} - \frac{1-e^{-\lambda(T-s)}}{\lambda}\Big]ds \\ &= \overline{F}\Big[\frac{1}{\lambda^2} + \frac{1}{\lambda^2}e^{-\lambda(T-t)} - \frac{1}{2\lambda^2}e^{-2\lambda(T-t)} - \frac{3}{\lambda^3}\frac{1-e^{-\lambda(T-t)}}{T-t} + \frac{3}{4\lambda^3}\frac{1-e^{-2\lambda(T-t)}}{T-t}\Big]. \end{aligned}$$

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

Appendix C: Restrictions Imposed in the AFNS Model

Derivation of the AFNS restrictions imposed on the canonical representation of the $A_0(3)$ class of affine models starts with 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 and 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

$$\begin{aligned} dX_t &= AdY_t = [AK_Y^Q\theta_Y^Q - AK_Y^Q Y_t]dt + A\Sigma_Y dW_t^Q = AK_Y^Q A^{-1}[A\theta_Y^Q - AY_t - \eta + \eta]dt + A\Sigma_Y dW_t^Q \\ &= AK_Y^Q A^{-1}[A\theta_Y^Q + \eta - X_t]dt + A\Sigma_Y dW_t^Q = K_X^Q[\theta_X^Q - X_t]dt + \Sigma_X dW_t^Q. \end{aligned}$$

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

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

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

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

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

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; hence, $\mathcal{T}_X$ is called an affine invariant transformation.

In the canonical representation of the subset of $A_0(3)$ affine term structure models considered here, the Q-dynamics are

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

and the P-dynamics are

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

Finally, the instantaneous risk-free rate is

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

There are 22 parameters in this maximally flexible canonical representation of the $A_3(0)$ class of models, and here we present the parameter restrictions needed to arrive at the affine AFNS models.

(1) The AFNS model with independent factors

The independent-factor AFNS model has P-dynamics

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

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, 12 parameter restrictions need to be imposed on the canonical $A_0(3)$ model.

It is easy to verify that the affine invariant transformation $\mathcal{T}_A(Y_t) = AY_t + \eta$ 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 AFNS model. For the mean-reversion matrices, the relationship between the two representations is

$$\begin{aligned} K_X^P = AK_Y^P A^{-1} \quad &\Longleftrightarrow \quad K_Y^P = A^{-1} K_X^P A \\ K_X^Q = AK_Y^Q A^{-1} \quad &\Longleftrightarrow \quad K_Y^Q = A^{-1} K_X^Q A. \end{aligned}$$

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

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

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

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

Furthermore, notice that $K_{23}^{Y,Q}$ will always have the opposite sign of $K_{22}^{Y,Q}$ and $K_{33}^{Y,Q}$, but its absolute size can vary independently of these two parameters. Because K_X^P, A, and A^{-1} are all diagonal matrices, K_Y^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 AFNS 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 $(\delta_1^X)' = (\delta_1^Y)' A^{-1}$ it follows that

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

For the constant term it holds that

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

Thus, we have obtained two additional parameter restrictions

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

(2) The AFNS model with correlated factors

In the correlated-factor AFNS model, the P-dynamics are

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

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, three parameter restrictions are needed.

It is easy to verify that the affine invariant transformation $\mathcal{T}_A(Y_t) = AY_t + \eta$ 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 AFNS model. For the mean-reversion matrices, the relationships between the two representations are

$$\begin{aligned} K_X^P = AK_Y^P A^{-1} \quad &\Longleftrightarrow \quad K_Y^P = A^{-1}K_X^P A \\ K_X^Q = AK_Y^Q A^{-1} \quad &\Longleftrightarrow \quad K_Y^Q = A^{-1}K_X^Q A. \end{aligned}$$

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

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

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

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

Furthermore, notice that $K_{23}^{Y,Q}$ will always have the opposite sign of $K_{22}^{Y,Q}$ and $K_{33}^{Y,Q}$, 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 AFNS 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 $(\delta_1^X)' = (\delta_1^Y)' A^{-1}$, it follows that

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

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

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

Thus, we have obtained one additional parameter restriction,

$$\delta_0^Y = 0.$$

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

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

Because K_X^P is a free 3×3 matrix, K_Y^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, 15, 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, 50, 745-787.

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

Chua, Choong Tze, Dean Foster, Krishna Ramasnwamy, and Robert Stine, 2008, "A Dynamic Model for the Forward Curve," Review of Financial Studies, 21, 265-310.

Christensen, Jens H. E., Jose A. Lopez, and Glenn D. Rudebusch, 2008, "Inflation Expectations and Risk Premiums in an Arbitrage-Free Model of Nominal and Real Bond Yields," Working Paper #2008-34, Federal Reserve Bank of San Francisco.

Christensen, Jens H. E., Jose A. Lopez, and Glenn D. Rudebusch, 2009, "Do Central Bank Liquidity Facilities Affect Interbank Lending Rates?," Working Paper #2009-13, Federal Reserve Bank of San Francisco.

Cox, John C., Jonathan E. Ingersoll, and Stephen A. Ross, 1985, "A Theory of the Term Structure of Interest Rates," Econometrica, 53, 385-407.

Dai, Qiang and Kenneth J. Singleton, 2000, "Specification Analysis of Affine Term Structure Models," Journal of Finance, 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, 63, 415-441.

De Pooter, Michiel, Francesco Ravazzolo, and Dick van Dijk, 2007, "Predicting the Term Structure of Interest Rates: Incorporating Parameter Uncertainty, Model Uncertainty and Macroeconomic Information," available at SSRN: http://ssrn.com/abstract=967914.

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

Diebold, Francis X., Monika Piazzesi, and Glenn D. Rudebusch, 2005, "Modeling Bond Yields in Finance and Macroeconomics," American Economic Review, 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, 131, 309-338.

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

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

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

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

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

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

Fisher, Mark and Christian Gilles, 1996, "Term Premia in Exponential-Affine Models of the Term Structure," Manuscript, Board of Governors of the Federal Reserve System.

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

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

Koopman, Siem J., Max I. P. Mallee, and Michel van der Wel, 2010, "Analyzing the Term Structure of Interest Rates using the Dynamic Nelson-Siegel Model with Time-Varying Parameters," Journal of Business and Economic Statistics, in press.

Krippner, Leo, 2006, "A Theoretically Consistent Version of the Nelson and Siegel Class of Yield Curve Models," Applied Mathematical Finance, 13, 39-59.

Litterman, Robert, and Jose Scheinkman, 1991, "Common Factors Affecting Bond Returns," Journal of Fixed Income, 1, 77-85.

Mönch, Emanuel, 2008, "Forecasting the Yield Curve in a Data-Rich Environment: A No-Arbitrage Factor-Augmented VAR Approach," Journal of Econometrics, 146, 26-43.

Nelson, Charles R. and Andrew F. Siegel, 1987, "Parsimonious Modeling of Yield Curves," Journal of Business, 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, 24, 83-128.

Singleton, Kenneth J., 2006, *Empirical Dynamic Asset Pricing*. Princeton: Princeton University Press.

Svensson, Lars E. O., 1995, "Estimating Forward Interest Rates with the Extended Nelson-Siegel Method," Quarterly Review, 3, Sveriges Riksbank, 13-26.

Trolle, Anders B. and Eduardo S. Schwartz, 2009, "A General Stochastic Volatility Model for the Pricing of Interest Rate Derivatives," Review of Financial Studies, 22, 2007-2057.

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

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

Zantedeschi, D., Damien, P. and Polson, N.G. (2009), "Predictive Macro-Finance: A Sequential Term Structure Modeling Approach Based on Regime Switches," Manuscript, University of Chicago and University of Texas.