Arbitrage-free Nelson–Siegel model for multiple yield curves

Riccardo Brignone[1] · Christoph Gerhart[1] · Eva Lütkebohmert[1]

Abstract
We propose an affine term structure model that allows for tenor-dependence of yield curves and thus for different risk categories in interbank rates, an important feature of post-crisis interest rate markets. The model has a Nelson–Siegel factor loading structure and thus economically well interpretable parameters. We show that the model is tractable in terms of estimation and provides good in-sample fit and out-of-sample forecasting performance. The proposed model is arbitrage-free across maturities and tenors, and thus perfectly suited for risk management and pricing purposes. We apply our framework to the pricing of caplets in order to illustrate its practical applicability and its suitability for stress testing.

1 Introduction

The global financial crisis of 2007/2008 has triggered some structural changes in interest rate markets that invalidated the classical notion of a single yield curve. In particular, in post-crisis interest rate markets, yields vary not only with respect to maturity but depend also on the length of the tenor structure of the interest rate derivative from which they are calculated. Thus, nowadays interest rate markets are characterized by multiple (tenor dependent) term structures that reflect different risk categories. This has important implications for pricing, portfolio allocation, risk management and monetary policy.

In this paper, we construct an affine term structure model that (1) allows for tenor-dependent yield curves, (2) is free of arbitrage across maturities and tenors, (3) is computationally tractable with (4) economically interpretable parameters, and (5) provides superior in-sample fit and out-of-sample forecasting performance within the Nelson–Siegel class of models. The model is suitable also for negative interest rate environments and allows for semi-closed pricing formulas for various interest rate derivatives.

✉ Eva Lütkebohmert
eva.luetkebohmert@finance.uni-freiburg.de

[1] Department of Quantitative Finance, Institute for Economic Research, University of Freiburg, Rempartstr. 16, 79098 Freiburg i. Br., Germany

 Springer

Our paper relates to a large literature on term structure models. The models introduced in [28,30] are heavily used by various central banks to fit daily yield curves (see e.g. [1,9]). They are flexible enough to describe various different shapes of interest rate term structures and the dynamic extension of the Nelson–Siegel model introduced in [6] has been shown to be highly tractable and at the same time provides accurate forecasting results. While the Nelson–Siegel parameters (as well as the Svensson parameters) have convenient economic interpretations as level, slope and curvature factors, the model is not arbitrage-free and hence not suitable for pricing purposes and of limited use for risk management. Affine arbitrage-free term-structure models have been introduced e.g. in [5,8] in the single curve setting. However, as shown in [7] the canonical affine models typically provide poor forecasting performance. This shortcoming is overcome in [2] who introduce an arbitrage-free affine specification of the Nelson–Siegel model in the single curve setting.

The financial crisis of 2007/2008, however, has invalidated the classical notion of a single yield curve. Reference [20] points out the importance of taking different risk categories in interbank rates into account. In the past years, affine short rate models in the multiple curve setting have been introduced in [14,16,19,25–27] among others. Affine Libor models for tenor dependent yields have been developed in [17,18]. Reference [4] provides a very general framework to model multiple yield curves through affine processes which comprises short rate models, Libor market models and Heath–Jarrow–Morton (HJM) models of tenor dependent yield curves. In particular, their approach relies on a general numéraire process and the modelling of multiplicative spreads between tenor dependent interbank rates and Overnight Indexed Swap (OIS) rates as first suggested in [3].

Our approach builds on the idea developed in [2] and combines it with the general affine model introduced in [4] in order to develop an affine arbitrage-free multiple curve version of the Nelson–Siegel model. The proposed model inherits advantages from both the affine models as well as the parsimonious Nelson–Siegel class of models: it can be estimated through standard techniques (Kalman filter) and presents good in- and out-of-sample performances. Since it is free of arbitrage across maturities and tenors and has parameters with a sound economic interpretation, the model is well suited for derivative pricing and tailor-made for scenario generation, stress testing and other risk management purposes. We illustrate this by applying our model to the pricing of caplets and by analysing the impact of level, slope and curvature shifts in the discount curve and the multiplicative log spreads on the caplet price.

The remainder of the paper is structured as follows. In Sect. 2 we specify the financial market setting of our model. Section 3 summarizes the general affine multiple curve model introduced in [4] adapted to our market setting. The specification of the arbitrage-free multiple curve Nelson–Siegel model is then derived in Sect. 4. Estimation and forecasting results are provided in Sect. 5 while Sect. 6 discusses the pricing of interest rate derivatives in the proposed model. Section 7 concludes.

2 Financial market instruments

We consider a financial market in which Xibor rates are quoted for a finite and generic set of tenors $\mathscr{D} = \{\delta_1, \ldots, \delta_m\}$ with $\delta_1 < \cdots < \delta_m$ for some $m \in \mathbb{N}$ and we denote by $L_t^k(t, t + \delta_k)$ the (spot) Xibor rate prevailing at time t for the time interval $[t, t + \delta_k]$. The tenor δ_k is typically equal to 1 week (1W) or several months (1M, 2M, 3M, 6M, or 12M). We assume that forward rate agreements with Xibor rates of different tenors as reference rate are traded in the market.

Definition 1 A *forward rate agreement* (FRA) with tenor δ_k, settlement date T and strike K, is a contract stating that the fixed interest rate K will apply to a certain nominal value N, which we assume to be normalized to one, over the specified future period $[T, T + \delta_k]$. The payoff at maturity $T + \delta_k$ is then given by

$$\delta_k(L_T^k(T, T + \delta_k) - K).$$

The interest rate K which ensures that the FRA for period $[T, T + \delta_k]$ has value zero at time t is referred to as *FRA rate* and will be denoted by $L_t^k(T, T + \delta_k)$ in the sequel.

In addition, we assume that overnight indexed swaps are traded in the market. The EONIA rate is the reference rate for overnight borrowing in the interbank market in the Eurozone and represents the underlying of overnight indexed swaps in the Eurozone.

Definition 2 An *overnight indexed swap* (OIS) is a contract where an investor agrees to pay a predetermined fixed rate K on a notional N at some predetermined equidistant future dates $\{T_0, \ldots, T_n\}$ with tenor $\delta = T_i - T_{i-1}$ to the other party. In return, the investor receives interest rate payments at a floating interest rate on the same notional principal. The floating rate in an OIS is indexed to an overnight rate (such as the EONIA in the Eurozone) and is given by simply compounding the consecutive overnight rates between the dates T_i and $T_i + \delta$. In the following, we again normalize the notional N to one. The fixed rate K such that the OIS contract for the time period $[T_0, T_n]$ has zero value at time $t \leq T_0$ is called the time t *forward swap rate* and will be denoted by $K_t^{\text{OIS}}(T_0, T_n)$.

Applying appropriate bootstrapping techniques (see e.g. [15]) to the quoted OIS forward swap rates, we can extract the term structure $T \mapsto P^d(t, T)$ of OIS zero-coupon bond prices $P^d(t, T)$ at time t and we define the simple compounded OIS spot rate at time t for the time period $[t, t + \delta]$ as

$$L_t^d(t, t + \delta) := \frac{1}{\delta}\left(\frac{1}{P^d(t, t + \delta)} - 1\right).$$

From the bootstrapped OIS zero-coupon bond prices, we can extract the simple compounded forward OIS rate $L_t^d(T, T + \delta)$ at time t for the time period $[T, T + \delta]$ which is defined via

$$1 + \delta L_t^d(T, T + \delta) := \frac{P^d(t, T)}{P^d(t, T + \delta)}.$$

The OIS rates will be considered as risk-free. This is motivated by the wide dissemination of collateral agreements to reduce the counterparty risk associated with OTC-traded derivatives.[1] Collateralisation is based on discounting at an overnight rate and reflects a funding and hedging mechanism. As a consequence, prices of OTC-traded derivatives quoted in the interbank market can be considered free of credit and liquidity risk.

We then have the following definition of a multiple curve financial market (compare [4,13]).

Definition 3 For a fixed time horizon $\mathbb{T} < \infty$ we consider a financial market consisting of the following basic traded assets

1. OIS zero-coupon bonds for all maturities $T \in [0, \mathbb{T}]$, and
2. FRA contracts for all maturities $T \in [0, \mathbb{T}]$ and for all tenors $\delta_1, \ldots, \delta_m$.

---

[1] Collateral agreements for OTC-traded derivatives have been suggested in the credit support annex (CSA) to the International Swaps and Derivatives Association master agreement. Compare [21,22] for reference.

The risk-free term structure can be derived from the OIS bonds. The tenor-dependent (risky) yield curves reflect different levels of credit and liquidity risk in the interbank market and can be extracted from quoted FRA and swap rates of corresponding tenors. Since the fair swap rate S_t^k at time $t \leq T_1$ of a swap with tenor δ_k and cashflow dates $T_1, \ldots, T_n$ can be represented through the relation

$$S_t^k = \frac{\sum_{j=1}^n L_t^k(T_{j-1}, T_{j-1} + \delta_k)}{\sum_{j=1}^n \delta_k P^d(t, T_j)} + K$$

as a function of FRA rates for some fixed strike rate $K \in \mathbb{R}$ and tenor length δ_k, it suffices to assume that FRA contracts for all maturities and tenors are traded in the market.

3 Continuous-time affine model for tenor-dependent term structures

Next, we set up the continuous time affine model for multiple yield curves. Therefore, we first fix the dynamics of the risk-free short rate. Afterwards, we specify the dynamics of the spreads between tenor-dependent (risky) yields and the risk-free yields in order to obtain the risky term structures.

In the first step, we follow the classical approach of [8] in modelling the term structure of risk-free interest rates as an affine function of underlying latent factors. We suppose that prices are discounted at the OIS rate, i.e., the numéraire $B = (B_t)_{0 \leq t \leq \mathbb{T}}$ is given by the OIS bank account

$$B_t = \exp\left(\int_0^t r_s^d \, ds\right),$$

for $t \in [0, \mathbb{T}]$, where r_t^d denotes the overnight short rate at time t. Further, we denote by $\mathbb{Q}$ the associated (spot) martingale measure such that prices of traded assets can be calculated as conditional expectations of B-discounted payoffs under $\mathbb{Q}$. We consider the probability space $(\Omega, \mathscr{F}, \mathbb{Q})$ with filtration $(\mathscr{F}_t)_{t \geq 0}$ generated by a standard Brownian motion $(W_t)_{t \geq 0}$ on $\mathbb{R}^n$ under $\mathbb{Q}$. We model the short-rate r^d as an affine function of a state variable X_t that follows an affine process[2] on some open subset $M \subset \mathbb{R}^n$. More specifically, we assume that X_t satisfies the stochastic differential equation

$$dX_t = K(t)[\theta(t) - X_t]dt + \Sigma(t)D(X_t, t)dW_t$$

under $\mathbb{Q}$, where $\theta : [0, T] \to \mathbb{R}^n$ and $K : [0, T] \to \mathbb{R}^{n \times n}$ are bounded, continuous functions. Moreover, the matrix $\Sigma : [0, T] \to \mathbb{R}^{n \times n}$ is assumed to be a bounded, continuous function and $D : \mathbb{R}^n \times [0, T] \to \mathbb{R}^{n \times n}$ has diagonal structure,

$$D(X_t, t) = \text{diag}\left(\sqrt{\gamma^1(t) + \vartheta_1^1(t)X_t^1 + \cdots + \vartheta_n^1(t)X_t^n}, \ldots, \sqrt{\gamma^n(t) + \vartheta_1^n(t)X_t^1 + \cdots + \vartheta_n^n(t)X_t^n}\right)$$

for $\gamma = (\gamma^1, \ldots, \gamma^n) : [0, T] \to \mathbb{R}^n$ and $\vartheta = (\vartheta_i^j)_{i,j=1,\ldots,n} : [0, T] \to \mathbb{R}^{n \times n}$ bounded and continuous functions. We assume that the *overnight (discount) short rate* $(r_t^d)_{t \geq 0}$ is an affine function of the state variable, i.e.,

$$r_t^d = \rho_0^d(t) + (\rho_1^d)^\top X_t,$$

---

[2] That is, a time-homogeneous Markov process whose characteristic function $\mathbb{E}^{\mathbb{Q}}[e^{\langle u, X_t \rangle}] = e^{\tilde{\phi}(t,u) + \tilde{\psi}(t,u)^\top X_0}$ is exponentially affine in the initial state X_0.

for a bounded and continuous function $\rho_0^d : [0, \mathbb{T}] \to \mathbb{R}$ and a vector $\rho_1^d \in \mathbb{R}^n$.

In the second step, in order to introduce tenor-dependent term structures, we follow [4] and model the *multiplicative spot spreads*

$$S^k(t, t) := \frac{1 + \delta_k L_t^k(t, t + \delta_k)}{1 + \delta_k L_t^d(t, t + \delta_k)} = (1 + \delta_k L_t^k(t, t + \delta_k)) P^d(t, t + \delta_k)$$

between Xibor rates and simple compounded OIS spot rates for each tenor k as exponentially affine in X_t. Therefore, we assume that

$$\log S^k(t, t) = \rho_0^k(t) + (\rho_1^k)^\top X_t$$

for a bounded and continuous function $\rho_0^k : [0, \mathbb{T}] \to \mathbb{R}$ and a vector $\rho_1^k \in \mathbb{R}^n$. Note that the spot spreads can be directly calculated from the quoted Xibor and OIS rates. Moreover, we define the *multiplicative forward spreads* $S^k(t, T)$ for $0 \le t \le T \le \mathbb{T}$ and tenor k by

$$S^k(t, T) := \frac{1 + \delta_k L_t^k(T, T + \delta_k)}{1 + \delta_k L_t^d(T, T + \delta_k)} = (1 + \delta_k L_t^k(T, T + \delta_k)) \frac{P^d(t, T + \delta_k)}{P^d(t, T)}. \quad (1)$$

The T-forward measure equivalent to the pricing measure $\mathbb{Q}$ is defined via the Radon-Nikodym derivative

$$\frac{d\mathbb{Q}_T}{d\mathbb{Q}}\bigg|_{\mathscr{F}_t} = \frac{P^d(t, T)}{P^d(0, T) \cdot B_t}$$

with $P^d(t, T) = \mathbb{E}^{\mathbb{Q}}[B_t / B_T | \mathscr{F}_t]$ the time t price of a discount (OIS) bond with maturity T. As shown in [4, Prop. 2.5], the fair FRA rate for tenor δ_k can then be expressed as conditional expectation of the (spot) Xibor rate of tenor δ_k, i.e.,

$$L_t^k(T, T + \delta_k) = \mathbb{E}^{\mathbb{Q}_{T+\delta_k}}[L_T^k(T, T + \delta_k) | \mathscr{F}_t],$$

for all $0 \le t \le T \le \mathbb{T}$. Similarly, the multiplicative (forward) spread satisfies

$$S^k(t, T) = \mathbb{E}^{\mathbb{Q}_T}[S^k(T, T) | \mathscr{F}_t],$$

for all $0 \le t \le T \le \mathbb{T}$. It can be shown that X generates exponentially affine discount bond prices $P^d(t, T)$ and forward spreads $S^k(t, T)$. More explicitly, by applying Prop. 3.16 in [4], the discount bond prices and forward multiplicative spreads for tenor k are given by

$$P^d(t, T) = \mathbb{E}^{\mathbb{Q}}\left[\exp\left(-\int_t^T r_u^d du\right) | \mathscr{F}_t\right] = \mathbb{E}^{\mathbb{Q}}\left[\exp\left(-\int_t^T \rho_0^d(u) du - \int_t^T (\rho_1^d)^\top X_u du\right) | \mathscr{F}_t\right]$$

$$= \exp\left(\mathscr{A}^d(t, T) + \mathscr{B}^d(T - t)^\top X_t\right) \quad (2)$$

$$S^k(t, T) = \mathbb{E}^{\mathbb{Q}_T}\left[S^k(T, T) | \mathscr{F}_t\right] = \mathbb{E}^{\mathbb{Q}}\left[S^k(T, T) \frac{d\mathbb{Q}_T}{d\mathbb{Q}}\bigg|_{\mathscr{F}_t}\right] / \left(\frac{d\mathbb{Q}_T}{d\mathbb{Q}}\bigg|_{\mathscr{F}_t}\right)$$

$$= \frac{B_t}{P^d(t, T)} \mathbb{E}^{\mathbb{Q}}\left[S^k(T, T) \frac{P^d(T, T)}{B_T} | \mathscr{F}_t\right]$$

$$= \frac{1}{P^d(t, T)} \mathbb{E}^{\mathbb{Q}}\left[\exp\left(\rho_0^k(T) + (\rho_1^k)^\top X_T - \int_t^T \rho_0^d(u) du - \int_t^T (\rho_1^d)^\top X_u du\right) | \mathscr{F}_t\right]$$

$$= \exp\left(\mathscr{A}^k(t, T) + \mathscr{B}^k(T - t)^\top X_t\right) \quad (3)$$

for all $0 \le t \le T \le \mathbb{T}$ and all $k = 1, \dots, m$, where

$$\mathscr{A}^d(t, T) = -\int_t^T \rho_0^d(u)du + \phi(T - t, 0, -\rho_1^{\mathrm{d}})$$
$$\mathscr{B}^d(T - t) = \psi(T - t, 0, -\rho_1^d)$$
$$\mathscr{A}^k(t, T) = \rho_0^k(T) + \phi(T - t, \rho_1^k, -\rho_1^d) - \phi(T - t, 0, -\rho_1^d)$$
$$\mathscr{B}^k(T - t) = \psi(T - t, \rho_1^k, -\rho_1^d) - \psi(T - t, 0, -\rho_1^d),$$

with ϕ and ψ denoting the characteristic exponents of the process $Y = (X, \int_0^\cdot X_s ds)$, which are given as solutions to the system of ordinary differential equations (ODEs)

$$
\begin{aligned}
\tfrac{d}{dt}\phi(T - t, u, v) &= -\psi(T - t, u, v)^\top K(t)\theta(t) \\
&\quad - \tfrac{1}{2}\sum_{j=1}^n \left(\Sigma^\top(t)\psi(T - t, u, v)\psi(T - t, u, v)^\top \Sigma(t)\right)_{jj} \gamma^j(t) \\
\tfrac{d}{dt}\psi(T - t, u, v) &= K(t)^\top \psi(T - t, u, v) \\
&\quad - \tfrac{1}{2}\sum_{j=1}^n \left(\Sigma^\top(t)\psi(T - t, u, v)\psi(T - t, u, v)^\top \Sigma(t)\right)_{jj} \vartheta^j(t) - v
\end{aligned}
$$
(4)

with boundary conditions $\phi(0, u, v) = 0$ and $\psi(0, u, v) = u$. We refer to "Appendix A" for the derivation of these ODEs.

4 Arbitrage-free dynamic tenor-dependent Nelson–Siegel model

As a special case of the affine multiple term structure model outlined above, we develop in this section an arbitrage-free tenor-dependent Nelson–Siegel model. For simplicity of notation, we consider a model for only two term structures here, the discount curve and a risky curve of tenor k. Therefore, we consider unobservable state variables $X_t = (X_t^1, \ldots, X_t^6) \in \mathbb{R}^6$ where the first three factors correspond to level, slope and curvature factors of the overnight (discount) curve, and the remaining variables influence the level, slope and curvature of the tenor-dependent term structure for tenor k through their impact on the multiplicative spreads. Following [2] we suppose that the overnight (discount) short rate is of the form

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

i.e., r_t^d is determined by the sum of the level and slope factors for the discount curve and the parameters $\rho_0^d \equiv 0$ and $\rho_1^d = (1, 1, 0, 0, 0, 0)^\top$ in the notation of Sect. 3. Further, we assume that the state variables have the following $\mathbb{Q}$-dynamics

$$dX_t = K[\theta - X_t]dt + \Sigma dW_t,$$

with $\theta = (\theta_1, \ldots, \theta_6)$ and matrix $K \in \mathbb{R}^{6\times6}$ of triangular form

$$
K = \left(
\begin{array}{ccc|ccc}
0 & 0 & 0 & & & \\
0 & \lambda_d & -\lambda_d & & 0 & \\
0 & 0 & \lambda_d & & & \\
\hline
& & & 0 & 0 & 0 \\
& 0 & & 0 & \lambda_k & -\lambda_k \\
& & & 0 & 0 & \lambda_k
\end{array}
\right)
$$

for $\lambda_d, \lambda_k > 0$. In the following we will denote the upper left 3×3 submatrix of K by K^d and the lower right 3×3 submatrix by K^k. Further, suppose that $\Sigma \in \mathbb{R}^{6\times6}$ has triangular

form

$$\Sigma = \begin{pmatrix} \Sigma^d & 0 \\ \hline 0 & \Sigma^k \end{pmatrix} = \begin{pmatrix} \begin{array}{ccc} \sigma_{11} & 0 & 0 \\ \sigma_{21} & \sigma_{22} & 0 \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{array} & \text{\Large 0} \\ \hline \text{\Large 0} & \begin{array}{ccc} \sigma_{44} & 0 & 0 \\ \sigma_{54} & \sigma_{55} & 0 \\ \sigma_{64} & \sigma_{65} & \sigma_{66} \end{array} \end{pmatrix}$$

Hence, the first three factors (X_t^1, X_t^2, X_t^3), which influence the discount curve, are independent of the last three factors (X_t^4, X_t^5, X_t^6), which affect the tenor spreads, and vice versa. In analogy to [2] and by application of Eq. (2), we can express the risk-free zero-coupon bond prices in this model as follows

$$P^d(t, T) = \mathbb{E}_t^{\mathbb{Q}} \left[\exp\left(-\int_t^T r_u^d \, du \right) \right] = \exp\left(\mathscr{A}^d(t, T) + \mathscr{B}^d(T - t)X_t \right),$$

where we have

$$\frac{d}{dt} \mathscr{A}^d(t, T) = \rho_0^d + \frac{d}{dt} \phi(T - t, 0, -\rho_1^d)$$

$$= -\psi(T - t, 0, -\rho_1^d)^\top K\theta - \frac{1}{2} \sum_{j=1}^6 \left(\Sigma^\top \psi(T - t, 0, -\rho_1^d) \psi(T - t, 0, -\rho_1^d)^\top \Sigma \right)_{jj},$$

because $\rho_0^d \equiv 0$ and $\gamma_j \equiv 1$ in this setting, and

$$\frac{d}{dt} \mathscr{B}^d(T - t) = \frac{d}{dt} \psi(T - t, 0, -\rho_1^d) = K^\top \psi(T - t, 0, -\rho_1^d) + \rho_1^d,$$

as $\vartheta^j \equiv 0$ in this setting. The latter equation can be rewritten as

$$\frac{d}{dt} \psi(T - t, 0, -\rho_1^d) = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ \lambda_d \, \psi_2(T - t, 0, -\rho_1^d) \\ -\lambda_d \, \psi_2(T - t, 0, -\rho_1^d) + \lambda_d \, \psi_3(T - t, 0, -\rho_1^d) \\ 0 \\ \lambda_k \, \psi_5(T - t, 0, -\rho_1^d) \\ -\lambda_k \, \psi_5(T - t, 0, -\rho_1^d) + \lambda_k \, \psi_6(T - t, 0, -\rho_1^d) \end{pmatrix}$$

Due to the boundary condition $\psi(0, u, v) = u$ we obtain that the fifth and sixth component of the vector $\psi(T - t, 0, -\rho_1^d)$ are equal to zero and we have

$$\psi_i(T - t, 0, -\rho_1^d) = 0 \quad \text{for} \quad i = 4, 5, 6. \tag{5}$$

Thus, due to the special choice of the matrices K and Σ, we can rewrite the discount bond price as

$$P^d(t, T) = \exp\left(\sum_{i=1}^3 \psi_i^d(t, T) X_t^i + \mathscr{A}^d(t, T) \right),$$

where the functions $\psi^d(t, T) = (\psi_1^d(t, T), \psi_2^d(t, T), \psi_3^d(t, T))$ with $\psi_i^d(t, T) = \psi_i(T - t, 0, -\rho_1^d)$, for $i = 1, 2, 3$, are solutions to the system of Riccati equations

$$\begin{pmatrix} \frac{d}{dt} \psi_1^d(t, T) \\ \frac{d}{dt} \psi_2^d(t, T) \\ \frac{d}{dt} \psi_3^d(t, T) \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 \\ 0 & \lambda_d & 0 \\ 0 & -\lambda_d & \lambda_d \end{pmatrix} \begin{pmatrix} \psi_1^d(t, T) \\ \psi_2^d(t, T) \\ \psi_3^d(t, T) \end{pmatrix}$$

and the function $\mathscr{A}^d$ solves

$$\frac{d\mathscr{A}^d(t, T)}{dt} = -\psi^d(t, T)^\top K^d \theta^d - \frac{1}{2} \sum_{i=1}^{3} ((\Sigma^d)^\top \psi^d(t, T) \psi^d(t, T)^\top \Sigma^d)_{j,j}$$

where $\theta^d = (\theta_1^d, \theta_2^d, \theta_3^d)^\top$. Hence, the discount bond prices are determined solely by the first three latent factors. The boundary conditions are $\psi_1^d(T, T) = \psi_2^d(T, T) = \psi_3^d(T, T) = \mathscr{A}^d(T, T) = 0$. Similarly to [2], we obtain that the solutions to the ODEs are equal to the Nelson–Siegel factor loadings

$$\psi_1^d(t, T) = -(T - t)$$

$$\psi_2^d(t, T) = -\frac{1 - e^{-\lambda_d(T-t)}}{\lambda_d}$$

$$\psi_3^d(t, T) = (T - t)e^{-\lambda_d(T-t)} - \frac{1 - e^{-\lambda_d(T-t)}}{\lambda_d}$$

and $\mathscr{A}^d$ is given by

$$\mathscr{A}^d(t, T) = (K^d\theta^d)_2 \int_t^T \psi_2^d(s, T)ds + (K^d\theta^d)_3 \int_t^T \psi_3^d(s, T)ds$$

$$+ \frac{1}{2} \sum_{j=1}^{3} \int_t^T ((\Sigma^d)^\top \psi^d(s, T) \psi^d(s, T)^\top \Sigma^d)_{j,j} ds.$$

Therefore, the risk-free (discount) yields $y_t^d(\tau)$ at time t for time-to-maturity $\tau = T - t > 0$ are described by a function of the form

$$y_t^d(\tau) = -\frac{1}{T - t} \log P^d(t, T)$$

$$= X_t^1 + \frac{1 - e^{-\lambda_d\tau}}{\lambda_d\tau} X_t^2 + \left[\frac{1 - e^{-\lambda_d\tau}}{\lambda_d\tau} - e^{-\lambda_d\tau}\right] X_t^3 - \frac{\mathscr{A}^d(t, t+\tau)}{\tau} \quad (6)$$

Thus, factor loadings agree with those in the Nelson–Siegel model but there is an additional "yield-adjustment term" $-\mathscr{A}^d(t, T)/(T - t)$ which depends on the maturity of the bond. As has been shown in [2], by taking this additional term into account, the classical dynamic Nelson–Siegel model developed in [6] can be turned into an arbitrage-free term structure model. In other words, the yield-adjustment-term ensures absence of arbitrage across maturities in our model.

Next, we turn to the term structure of interest rates for tenor k. Therefore, we assume that the multiplicative spot spread for tenor k is of the form

$$\log S^k(t, t) = X_t^4 + X_t^5,$$

i.e. $\rho_0^k = 0$ and $\rho_1^k = (0, 0, 0, 1, 1, 0)^\top$ in the notation of Sect. 3. We obtain from Eq. (3) that the tenor-dependent multiplicative forward spreads satisfy

$$S^k(t, T) = \exp\left(\mathscr{A}^k(t, T) + \mathscr{B}^k(T - t)X_t\right)$$

with

$$\frac{d}{dt}\mathscr{A}^k(t,T) = \frac{d}{dt}\phi(T-t,\rho_1^k,-\rho_1^d) - \frac{d}{dt}\phi(T-t,0,-\rho_1^d)$$

$$= -\psi(T-t,\rho_1^k,-\rho_1^d)^\top K\theta - \frac{1}{2}\sum_{j=1}^6\left(\Sigma^\top\psi(T-t,\rho_1^k,-\rho_1^d)\psi(T-t,\rho_1^k,-\rho_1^d)^\top\Sigma\right)_{jj}$$

$$+ \psi(T-t,0,-\rho_1^d)^\top K\theta + \frac{1}{2}\sum_{j=1}^6\left(\Sigma^\top\psi(T-t,0,-\rho_1^d)\psi(T-t,0,-\rho_1^d)^\top\Sigma\right)_{jj}$$

since $\rho_0^k \equiv 0$ and $\gamma_j \equiv 1$ in this setting. Moreover, we have

$$\frac{d}{dt}\mathscr{B}^k(T-t) = \frac{d}{dt}\psi(T-t,\rho_1^k,-\rho_1^d) - \frac{d}{dt}\psi(T-t,0,-\rho_1^d)$$

$$= K^\top\psi(T-t,\rho_1^k,-\rho_1^d) + \rho_1^d - K^\top\psi(T-t,0,-\rho_1^d) - \rho_1^d$$

as $\vartheta^j \equiv 0$. The latter equation implies that

$$\frac{d}{dt}\mathscr{B}^k(T-t) = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ \lambda_k\,\psi_5(T-t,\rho_1^k,-\rho_1^d) \\ -\lambda_k\,\psi_5(T-t,\rho_1^k,-\rho_1^d) + \lambda_k\,\psi_6(T-t,\rho_1^k,-\rho_1^d) \end{pmatrix}$$

where we use the fact that $\psi_i(T-t,0,-\rho_1^d) = 0$ for $i = 4, 5, 6$ (compare Eq. (5)) and the fact that we have $\psi_i(T-t,\rho_1^k,-\rho_1^d) = \psi_i(T-t,0,-\rho_1^d)$ for the first three components $i = 1, 2, 3$, as implied by the boundary condition $\psi(0, u, v) = u$ for $u = \rho_1^k = (0, 0, 0, 1, 1, 0)^\top$. Hence, we can express the multiplicative forward spreads as

$$S^k(t,T) = \exp\left(\sum_{i=1}^3 \psi_i^k(t,T)X_t^{3+i} + \mathscr{A}^k(t,T)\right) \tag{7}$$

where the functions $\psi^k(t,T) = (\psi_1^k(t,T), \psi_2^k(t,T), \psi_3^k(t,T))$ with $\psi_i^k(t,T) = \psi_{3+i}(T-t,\rho_1^k,-\rho_1^d)$ for $i = 1, 2, 3$, are solutions to the system of Riccati equations

$$\begin{pmatrix} \frac{d}{dt}\psi_1^k(t,T) \\ \frac{d}{dt}\psi_2^k(t,T) \\ \frac{d}{dt}\psi_3^k(t,T) \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & \lambda_k & 0 \\ 0 & -\lambda_k & \lambda_k \end{pmatrix} \begin{pmatrix} \psi_1^k(t,T) \\ \psi_2^k(t,T) \\ \psi_3^k(t,T) \end{pmatrix}$$

and the function $\mathscr{A}^k$ solves

$$\frac{d\mathscr{A}^k(t,T)}{dt} = -\psi^k(t,T)^\top K^k\theta^k - \frac{1}{2}\sum_{i=1}^3((\Sigma^k)^\top\psi^k(t,T)\psi^k(t,T)^\top\Sigma^k)_{j,j}$$

where $\theta^k = (\theta_4, \theta_5, \theta_6)^\top$. Hence, the multiplicative forward spreads are solely determined by the last three latent factors. The boundary conditions are given by $\psi^k(T,T) = (1, 1, 0)$ and $\mathscr{A}^k(T,T) = 0$. The solution of this system of equations is given by

$$\psi_1^k(t,T) = 1$$
$$\psi_2^k(t,T) = e^{-\lambda_k(T-t)}$$
$$\psi_3^k(t,T) = (T-t)\lambda_k e^{-\lambda_k(T-t)}$$

and

$$\mathscr{A}^k(t, T) = (K^k\theta^k)_2 \int_t^T \psi_2^k(s, T)ds + (K^k\theta^k)_3 \int_t^T \psi_3^k(s, T)ds$$
$$+ \tfrac{1}{2} \sum_{j=1}^3 \int_t^T ((\Sigma^k)^\top \psi^k(s, T)\psi^k(s, T)^\top \Sigma^k)_{j,j}ds.$$

Thus, the bond prices and multiplicative spreads are both exponentially affine in X_t with coefficients that have a similar structure as the Nelson–Siegel factor loadings. We now consider the risky (tenor-dependent) term structure for tenor k. The one-period continuously compounded yields $y_t^k(t + \delta_k)$ can be calculated from the simple compounded spot Xibor rates as follows

$$y_t^k(t + \delta_k) = \tfrac{1}{\delta_k} \log \left(1 + \delta_k L_t^k(t, t + \delta_k)\right)$$
$$= \tfrac{1}{\delta_k} \log \left(\frac{S^k(t,t)}{P^d(t,t+\delta_k)}\right)$$
$$= \tfrac{1}{\delta_k} \log S^k(t, t)) - \tfrac{1}{\delta_k} \log P^d(t, t + \delta_k) = y_t^d(t + \delta_k) + \tfrac{1}{\delta_k} \log S^k(t, t)),$$

where we used relation (1). The multi-period yields $y_t^k(\tau)$ at time t for tenor k and time-to-maturity $\tau = T - t = N\delta_k$ for some $N \in \mathbb{N}$, are obtained as follows

$$y_t^k(\tau) = -\tfrac{1}{\tau} \log \left(\prod_{n=0}^{N-1} \left(1 + \delta_k L_t^k(t + n\delta_k, t + (n+1)\delta_k)\right)^{-1}\right)$$
$$= -\tfrac{1}{\tau} \log \left(\prod_{n=0}^{N-1} \left(\frac{P^d(t,t+(n+1)\delta_k)}{S^k(t,t+n\delta_k)P^d(t,t+n\delta_k)}\right)\right)$$
$$= y_t^d(\tau) + \tfrac{1}{\tau} \sum_{n=0}^{N-1} \log S^k(t, t + n\delta_k),$$

where we again applied relation (1). Since the discount yields are affine in X_t (compare Eq. (6)) and the multiplicative forward spreads are exponentially affine in X_t (compare Eq. (7)), we obtain that the tenor-dependent yields are also affine in X_t. Moreover, we observe that the tenor-dependent term structure is also of Nelson–Siegel shape due to the special structure of the discount short rate and the multiplicative spot spreads. As the multiplicative spreads are usually greater than one in the post crisis period, the level, slope and curvature factors for the spreads lead to a positive shift of the overnight (discount) curve so that the risky term structures are strictly larger than the discount curve. The terms $\mathscr{A}^k(t, t + n\delta_k)/\tau$ in the expression for the tenor-dependent yields represent "tenor-adjustment terms" which ensure absence of arbitrage across tenors.

5 Numerical results

In this section we illustrate the numerical performances of the proposed model by fitting it to real data. We start by illustrating how the historical time series of yields and tenor dependent multiplicative spreads are built, then we show in-sample and out-of-sample performances in terms of estimation and forecasting. Computations are done using Matlab®(Version R2019b) in Microsoft Windows 10®running on a machine equipped with Intel(R) Core(TM) i7-9750HQ CPU @2.60GHz and 16 GB of RAM.

5.1 Data

Our data set consists of daily bootstrapped discount and three month yield curves for maturities up to 10 years. More specifically, we use European market data provided by Bloomberg for the time period from September 2nd, 2005, until May 22nd, 2018. We bootstrap the discount curves from Overnight Indexed Swap (OIS) rates indexed to the EONIA rate with

maturities ranging from one week up to 10 years. Further, we construct the risky 3 month yield curves from market quotes of deposit rates for 3 month maturity and swap rates with maturities between 6 months and 10 years. All yields have been bootstrapped following the procedure in [15]. In total we consider 3316 observations for yields and multiplicative spreads of 3 months tenor. From the constructed yield curves we only consider the interest rates with maturities $\tau \in \{0.25, 0.5, 1, 3, 5, 7, 10\}$ (expressed in years) in the in-sample estimation and out-of-sample forecasting. These have been directly bootstrapped from market data and hence do not allow for any arbitrage opportunities that might arise due to interpolation methods. Including interest rates of further maturities has only a minor impact on the results.

5.2 Yield curve fitting

In Sect. 4, we introduced an affine term structure model which is free of arbitrages across tenors and maturities and takes into account correlations between factors. We will refer to this model as correlated arbitrage-free multiple curve Nelson Siegel (AFMCNS) model. We provide a state-space formulation of the AFMCNS model and describe the Kalman filter estimation of the model in "Appendix B". Parameters can then be estimated by maximizing numerically the log-likelihood function in (14), for which we use the built-in Matlab® function fmincon.

To evaluate the model's in- and out-of-sample performances, we consider three sub-models which are closely related to the correlated AFMCNS model for comparison. The first is the independent AFMCNS model, which is equivalent to the correlated one with the obvious exception that the off-diagonal elements in the matrices $K^{\mathbb{P}}$ and Σ are equal to 0. The estimation results for the correlated as well as for the independent version of the AFMCNS model are presented below. Second, by forcing the yield and tenor adjustment terms to be equal to 0, we obtain a dynamic Nelson–Siegel model which is not arbitrage-free but adapted for a multiple-curve setting. We refer to this model as MC-DNS model. It can be written in state-space form as

$$
\begin{aligned}
X_t &= (I - A)\mu + AX_{t-1} + \eta_t \\
y_t &= BX_t + \epsilon_t
\end{aligned}
\tag{8}
$$

where $X_t = \{X_t^1, X_t^2, X_t^3, X_t^4, X_t^5, X_t^6\}$, $y_t = \{y_t^d(\tau_n), \log S_t^k(\tau_n)\}_{n=1}^7$, B is the matrix of the Nelson–Siegel factor loading coefficients, and A is a diagonal matrix in the independent MC-DNS and a full matrix in the correlated MC-DNS. Finally,

$$
\begin{pmatrix} \eta_t \\ \epsilon_t \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} Q & 0 \\ 0 & H \end{pmatrix}\right],
\tag{9}
$$

where H is a diagonal matrix and $Q = qq'$ is a diagonal matrix in the independent MC-DNS and a full matrix in the correlated MC-DNS. The Kalman filter for the MC-DNS model is set up following [2, Section 4.1]. Estimation results for the MC-DNS model are reported in Table 1.

5.2.1 AFMCNS model estimation

Table 2 shows the estimation results for the AFMCNS model. For the one-day conditional mean-reversion matrix $\exp\left(-K^{\mathbb{P}}/252\right)$ the elements on the main diagonal are similar in the independent and correlated case, with factors related to the discount yields more persistent

Table 1 Parameter estimates of A and μ as in (8) and matrix $Q = qq'$ as in (9) of the MC-DNS model

	Independent MC-DNS						Correlated MC-DNS					
	$A_{,1}$	$A_{,2}$	$A_{,3}$	$A_{,4}$	$A_{,5}$	$A_{,6}$	$A_{,1}$	$A_{,2}$	$A_{,3}$	$A_{,4}$	$A_{,5}$	$A_{,6}$
$A_{1,}$	0.9981	0	0	0	0	0	0.9982	−0.0002	0.0006	−0.0001	−0.0001	0.0001
$A_{2,}$	0	1.0003	0	0	0	0	0.0014	1.0004	−0.0004	−4.18E−06	0.0001	0.0001
$A_{3,}$	0	0	1.0028	0	0	0	0.0002	−4.78E−05	1.0014	0.0002	−0.0002	−0.0001
$A_{4,}$	0	0	0	0.8123	0	0	0.0005	−1.70E−05	−0.0007	0.8115	0.0009	2.37E−05
$A_{5,}$	0	0	0	0	0.9874	0	0.0005	3.36E−05	−0.0003	0.0001	0.9850	0.0004
$A_{6,}$	0	0	0	0	0	0.8781	0.0002	8.42E−06	−0.0001	0.0001	0.0001	0.8774
μ	0.0840	0.4908	0.0875	0.0004	0.0002	0.0008	0.0831	0.4945	0.0981	−3.75E−07	0.0011	0.0009
	$q_{,1}$	$q_{,2}$	$q_{,3}$	$q_{,4}$	$q_{,5}$	$q_{,6}$	$q_{,1}$	$q_{,2}$	$q_{,3}$	$q_{,4}$	$q_{,5}$	$q_{,6}$
$q_{1,}$	0.0029	0	0	0	0	0	0.0004	0	0	0	0	0
$q_{2,}$	0	0.0005	0	0	0	0	−0.0003	0.0001	0	0	0	0
$q_{3,}$	0	0	0.0013	0	0	0	−0.0003	−0.0001	0.0010	0	0	0
$q_{4,}$	0	0	0	0.0002	0	0	0	0	0	0.0001	0	0
$q_{5,}$	0	0	0	0	0.0001	0	0	0	0	2.05E−05	0.0001	0
$q_{6,}$	0	0	0	0	0	0.0005	0	0	0	−0.0004	−0.0002	0.0003

$\lambda^d = 0.5348$, $\lambda^k = 0.4820$ (for maturity measured in years), Log-Likelihood $= 299607.057$ for the independent MC-DNS and $\lambda^d = 0.5396$, $\lambda^k = 0.4821$ (for maturity measured in years), Log-Likelihood $= 310922.6088$ for the correlated MC-DNS

Table 2 Parameter estimates of the AFMCNS model with 3 month tenor

Independent AFMCNS

$\exp\left(-K^{\mathbb{P}}/252\right)$

0.9999	0	0	0	0	0
0	0.9994	0	0	0	0
0	0	0.9935	0	0	0
0	0	0	0.6290	0	0
0	0	0	0	0.9898	0
0	0	0	0	0	0.8109

$\theta^{\mathbb{P}}$

−0.0036	−0.0423	−0.0355	0.0001	0.0003	0.0011

Σ

0.0049	0	0	0	0	0
0	0.0067	0	0	0	0
0	0	0.0194	0	0	0
0	0	0	0.0016	0	0
0	0	0	0	0.0017	0
0	0	0	0	0	0.0066

Correlated AFMCNS

$\exp\left(-K^{\mathbb{P}}/252\right)$

0.9999	0.0004	0.0014	−0.0065	−0.0092	0.0084
0.0001	0.9998	0.0001	−0.0072	−0.0117	0.0085
0.0013	0.0027	0.9962	−0.0046	−0.0455	−0.0071
−0.0009	−0.0025	0.0004	0.6340	0.0168	0.0151
0.0002	0.0005	−0.0002	0.0137	0.9901	0.0130
0.0015	−0.0022	−0.0003	−0.0119	−0.0068	0.8193

$\theta^{\mathbb{P}}$

−0.0030	−0.0464	−0.0295	0.0005	−0.0011	0.0006

Σ

0.0100	0	0	0	0	0
−0.0103	0.0026	0	0	0	0
−0.0167	−0.0043	0.0198	0	0	0
0	0	0	0.0008	0	0
0	0	0	0.0003	0.0008	0
0	0	0	−0.0049	−0.0011	0.0009

$\lambda^{d} = 0.3836$, $\lambda^{k} = 0.4700$ (for maturity measured in years), Log-Likelihood $= 306598.0796$ for the independent AFMCNS and $\lambda^{d} = 0.3833$, $\lambda^{k} = 0.4701$ (for maturity measured in years), Log-Likelihood $= 311512.2591$ for the correlated AFMCNS
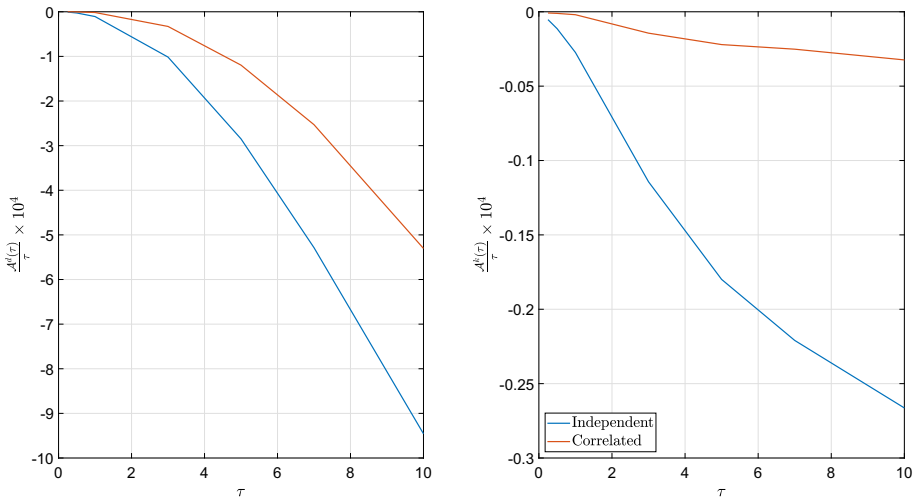
Fig. 1 Yield and tenor adjustments (in basis points) in the independent and correlated AFMCNS models

than those related to the log-spreads. Off-diagonal elements appear non negligible, especially for those elements corresponding to $\{X_t^i\}_{i=4}^6$ for which we find a lot of values higher than 0.01 in absolute value. Concerning the long-run mean $\theta^{\mathbb{P}}$ there is not much difference between the two models. However, the values are shifted when compared to μ (in the MC-DNS) which can be explained by the fact that we are adding the adjustment term. In fact, the yield and tenor adjustment terms are negative for all maturities and decreasing in maturity (compare Fig. 1). Since we subtract the adjustment terms in the measurement equation, the mean level in the AFMCNS model is smaller than in the MC-DNS model. Moreover, the tenor adjustment term is much smaller in size compared to the yield adjustment term (roughly by a factor 0.1) which explains why the mean level of the log spreads is less affected than the mean level of the discount yields.

When converting the volatility matrix Σ into a one-day conditional covariance matrix[3]

$$Q = \int_0^{1/252} \exp\left(-K^{\mathbb{P}}s\right) \Sigma \Sigma' \exp\left(-(K^{\mathbb{P}})'s\right) ds,$$

we obtain:

$$Q = \begin{bmatrix} 9.67\text{E--}08 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1.80\text{E--}07 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1.49\text{E--}06 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1.07\text{E--}08 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1.13\text{E--}08 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1.72\text{E--}07 \end{bmatrix}$$

---

[3] The small magnitude of the components of Q is explained by the fact that we are using daily data. Those numbers are consistent with [2, Eq. 12 and 13] which use monthly data.

in case of independence and

$$
Q = \begin{bmatrix}
4.09\text{E--}07 & -4.26\text{E--}07 & -6.67\text{E--}07 & 0 & 0 & 0 \\
-4.26\text{E--}07 & 4.74\text{E--}07 & 6.68\text{E--}07 & 0 & 0 & 0 \\
-6.67\text{E--}07 & 6.68\text{E--}07 & 2.60\text{E--}06 & 0 & 0 & 0 \\
0 & 0 & 0 & 1.45\text{E--}10 & 2.36\text{E--}10 & -1.96\text{E--}09 \\
0 & 0 & 0 & 2.36\text{E--}10 & 2.62\text{E--}09 & -5.58\text{E--}09 \\
0 & 0 & 0 & -1.96\text{E--}09 & -5.58\text{E--}09 & 3.81\text{E--}08
\end{bmatrix}
$$

in the correlated case. We find that factors related to the discount yields are more volatile than those related to the log-spreads. This is comparable in case of the MC-DNS model.[4] From the covariance estimates, we calculate correlation coefficients between factors

$$
\begin{bmatrix}
1 & -0.9665 & -0.6458 & 0 & 0 & 0 \\
-0.9665 & 1 & 0.6007 & 0 & 0 & 0 \\
-0.6458 & 0.6007 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0.3817 & -0.8350 \\
0 & 0 & 0 & 0.38173 & 1 & -0.5572 \\
0 & 0 & 0 & -0.8350 & -0.5572 & 1
\end{bmatrix}
$$

and find that correlations are not negligible. E.g. there is a very strong negative correlation of -0.9665 (resp. -0.8350) between X_1 and X_2 (resp. between X_4 and X_6). This indicates that for the discount yields level shifts often go along with slope changes while for the log spreads level shifts mostly go along with curvature changes.

When comparing the log-likelihood values reported in the caption of Table 2, it is evident that the correlated version of the AFMCNS model performs better in-sample than the independent specification. This is not only due to the higher number of parameters, as largely confirmed by the likelihood ratio (LR) test for which we obtain a very high LR of 10755.21, but this is also confirmed by the results in Table 3, where root mean squared errors (RMSEs) are consistently smaller in the correlated case than in the independent factor case. RMSEs and mean errors (ME) displayed in Table 3 are computed from the differences between the historical yields and model estimated yields. Figure 2 shows a plot of these time series. This figure also appreciates the ability of the proposed model to replicate the time series of observed yields. Further, comparing the log-likelihood ratios of the MC-DNS and the AFMCNS models (reported in the captions of Tables 1 and 2) confirms superiority of the AFMCNS model. This is also confirmed by the smaller RMSEs in Table 3.

5.3 Forecasting

We evaluate the forecasting performance of our proposed model by comparing six- and twelve-months ahead forecasts of discount and tenor dependent yields with its non-arbitrage free counterpart. For convenience of exposition, we denote $y_t := \{y_t^d(\tau_n), y_t^k(\tau_n)\}_{n=1}^{7}$. Forecasting yields in the AFMCNS model consists of two steps:

(1) Estimate model parameters over the sample period ending at time T as in Sect. 5.2.
(2) Compute optimal h months ahead forecasts as

$$
y_{T+h}^{AF} = \begin{pmatrix}
E_T^{\mathbb{P}}[X_{T+h}^1] + \frac{1-e^{-\lambda_d \tau}}{\lambda_d} E_T^{\mathbb{P}}[X_{T+h}^2] + \left(\frac{1-e^{-\lambda_d \tau}}{\lambda_d} - e^{-\lambda_d \tau} \right) E[X_{T+h}^3] \\
E[X_{T+h}^4] + e^{-\lambda_k \tau} E_T^{\mathbb{P}}[X_{T+h}^5] + \tau \lambda_k e^{-\lambda_k \tau} E_T^{\mathbb{P}}[X_{T+h}^6]
\end{pmatrix} - \begin{pmatrix}
\frac{\mathscr{A}^d(\tau)}{\tau} \\
-\mathscr{A}^k(\tau)
\end{pmatrix}
\tag{10}
$$

---

[4] Results are available upon request.

Table 3 Summary statistics for in-sample model fit

	τ	DNS-Indep		DNS-Corr		AFNS-Indep		AFNS-Corr	
		ME	RMSE	ME	RMSE	ME	RMSE	ME	RMSE
y^d	0.25	−2.72	*5.96*	0.00	**0.16**	2.78	5.62	0.63	1.58
	0.5	−1.75	2.83	−0.05	**0.81**	7.43	*9.68*	−0.71	3.86
	1	2.81	4.89	−0.76	**5.56**	−4.53	*7.23*	−4.23	7.02
	3	−3.17	5.56	−1.78	3.67	−2.23	*5.59*	0.59	**0.37**
	5	3.45	*6.11*	0.83	1.82	−2.34	3.41	−1.32	**1.19**
	7	5.44	*8.24*	−0.72	**4.69**	2.75	5.22	−0.88	5.34
	10	−4.65	*7.13*	−4.22	6.90	−3.03	5.40	−1.49	**3.53**
	Average	−0.08	5.82	−0.96	3.37	0.12	*6.02*	−1.06	**3.27**
y^k	0.25	−1.16	*5.55*	0.00	**0.33**	1.16	5.45	0.36	2.57
	0.5	−0.49	2.43	0.00	**0.01**	3.42	*7.95*	−0.09	3.10
	1	3.36	*5.66*	−0.25	5.50	−3.68	5.33	−3.02	**4.35**
	3	−1.93	3.06	−1.13	2.16	−1.04	*5.40*	0.02	**0.49**
	5	1.04	*5.33*	0.22	1.57	−0.48	2.50	0.00	**0.01**
	7	3.33	*8.09*	−0.32	**3.70**	3.45	5.86	0.26	4.87
	10	−2.95	*4.83*	−2.77	4.31	−1.95	3.32	−0.91	**2.51**
	Average	0.17	4.99	−0.61	**2.51**	0.13	*5.11*	−0.48	2.56
	Total average		5.41		2.94		*5.57*		**2.91**

Mean errors (ME) and root mean squared errors (RMSEs) are reported in basis points

Maturities are in years

Values in bold indicate the best performance, values in italic denote the worst performance

where $E_T^{\mathbb{P}}[X_{T+h}] = (I - \exp(-K^{\mathbb{P}}h))\theta^{\mathbb{P}} + \exp(-K^{\mathbb{P}}h)X_T$. We are interested in computing 6 and 12 months ahead forecasts which are typically better achieved using monthly data instead of daily data. Therefore, we split our data set with daily observations in 21 datasets with monthly observations by taking one data point each month. More specifically, we construct historical time series of yields taking every 21st observation. We do so starting from each day of the first month of the dataset. In this way, we get 21 historical time series with monthly observations. Then, on each sub-dataset, we estimate the model on a 6 years (72 months) rolling window and forecast yields using Eq. (10). This produces 73 forecasts for each sub-dataset and in total $73 \times 21 = 1533$ six and twelve months ahead forecasts. In Table 4 we report the root mean squared forecasting errors (RMSFE) for the proposed arbitrage free multiple curve Nelson Siegel model (AFMCNS) and for the multiple curve dynamic Nelson Siegel (MCDNS) model for benchmark comparison. Note that we only consider here the independent versions of those models. In fact, in unreported tests, we obtained superior out of sample performances of those models with respect to their correlated counterparts. This finding is consistent with [2, Table 7]. Results in Table 4 are striking and show that the AFMCNS is better in forecasting than the MCDNS throughout all the various yields and maturities with a RMSFE around 40-60% smaller than the benchmark model for short maturities and 12-40% for higher maturities.
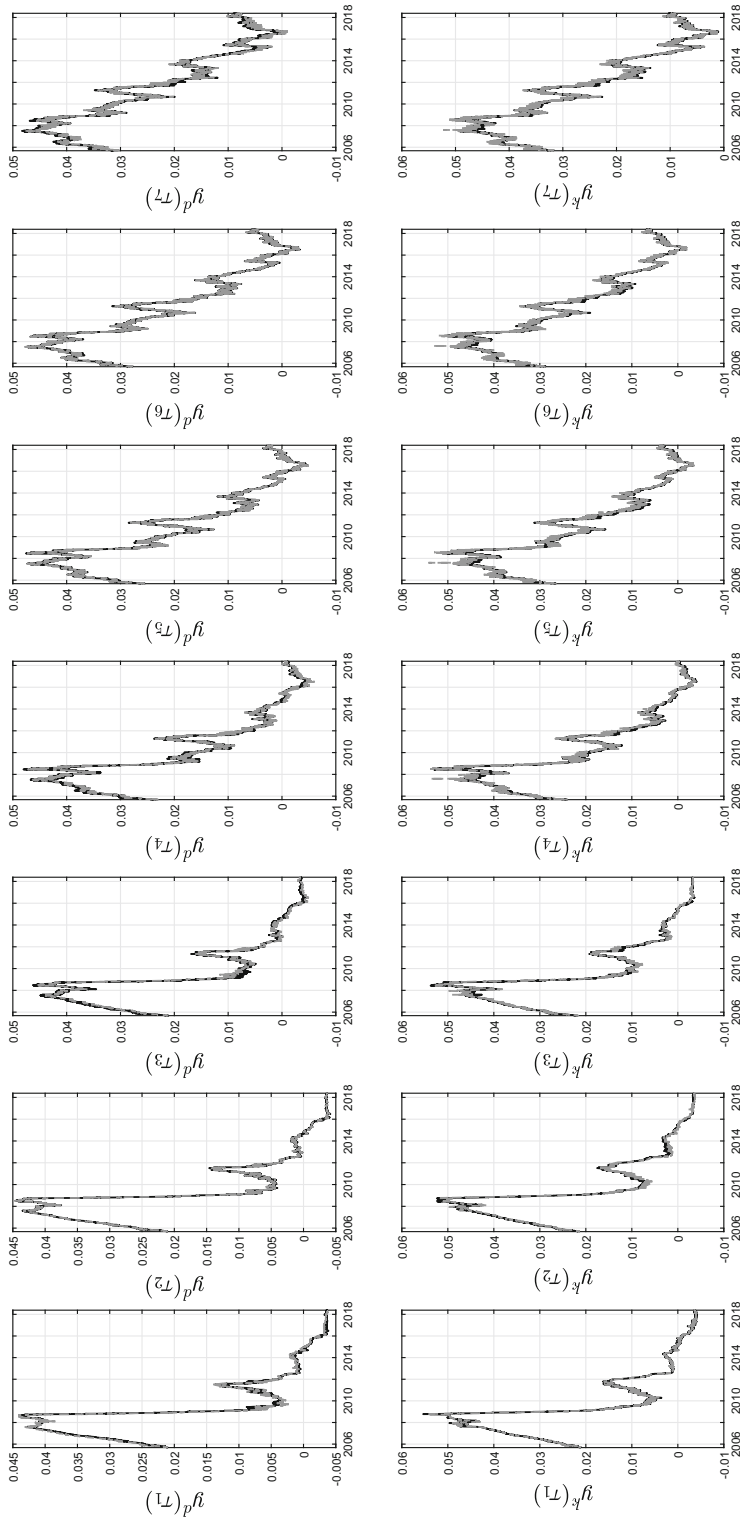
Fig. 2 Historical (black line) yields and yields estimated with the correlated AFMCNS model (grey line) for 7 different maturities ($\tau = \{0.25, 0.5, 1, 3, 5, 7, 10\}$, expressed in years) and tenor k equal to 3 months

Table 4 Out-of-sample forecasting results

	6 months		12 months	
	MCDNS	AFMCNS	MCDNS	AFMCNS
$y_d(\tau_1)$	45.2447	**25.5872**	66.1119	**35.6187**
$y_d(\tau_2)$	43.3077	**25.4037**	65.0016	**35.9040**
$y_d(\tau_3)$	44.0420	**26.7215**	66.7115	**38.1650**
$y_d(\tau_4)$	56.9866	**36.3356**	84.7662	**53.4556**
$y_d(\tau_5)$	64.5509	**43.2467**	96.0734	**65.3436**
$y_d(\tau_6)$	63.6161	**47.1426**	96.6505	**71.0293**
$y_d(\tau_7)$	57.9311	**50.0400**	90.1028	**73.5512**
$y_k(\tau_1)$	30.3770	**11.9630**	61.5200	**30.0458**
$y_k(\tau_2)$	34.7462	**16.3992**	65.9578	**36.0147**
$y_k(\tau_3)$	41.3085	**23.1866**	72.6095	**42.6233**
$y_k(\tau_4)$	59.1535	**36.7901**	93.0645	**58.9721**
$y_k(\tau_5)$	67.3193	**44.4750**	103.1823	**69.9925**
$y_k(\tau_6)$	65.3142	**48.2985**	101.8745	**74.7712**
$y_k(\tau_7)$	57.5005	**50.5924**	92.7184	**76.1396**

Root mean squared forecasting errors (in basis points) for yields and log-spreads for 7 different maturities $\tau = \{0.25, 0.5, 1, 3, 5, 7, 10\}$ (expressed in years) of the multiple curve dynamic Nelson–Siegel model (MCDNS) and the arbitrage-free multiple curve Nelson–Siegel model (AFMCNS)

Values in bold indicate best predictive performance

6 Derivative pricing

As an application, we illustrate the pricing of derivative instruments under the proposed AFMCNS model. More specifically, we consider the price at time t of a caplet with notional N, reset date T, and settlement date $T + \delta_k$. Its payoff at the settlement date is given by $N\delta_k(L(T, T + \delta_k) - K)^+$. Following [4] the time t price of the caplet can be derived in semi-closed form. Therefore, we first define

$$Y_t \equiv \log\left(\frac{S^k(t, t)}{P^d(t, T + \delta_k)}\right).$$

Then Eqs. (2) and (3) imply that

$$
\begin{aligned}
Y_t &= \mathscr{A}^k(t, t) + \mathscr{B}^k(0)^\top X_t - \mathscr{A}^d(t, T + \delta_k) - \mathscr{B}^d(T + \delta_k - t)^\top X_t \\
&= \rho_0^k(t) + \phi(0, \rho_1^k, -\rho_1^d) - \phi(0, 0, -\rho_1^d) + \left(\psi(0, \rho_1^k, -\rho_1^d) - \psi(0, 0, -\rho_1^d)\right)^\top X_t \\
&\quad + \int_t^{T+\delta_k} \rho_0^d(u)du - \phi(T + \delta_k - t, 0, -\rho_1^d) - \psi(T + \delta_k - t, 0, -\rho_1^d)^\top X_t \\
&= \rho_0^k(t) + \int_t^{T+\delta_k} \rho_0^d(u)du - \phi(T + \delta_k - t, 0, -\rho_1^d) - \psi(T + \delta_k - t, 0, -\rho_1^d)^\top X_t + (\rho_1^k)^\top X_t \\
&= -\phi(T + \delta_k - t, 0, -\rho_1^d) - \psi(T + \delta_k - t, 0, -\rho_1^d)^\top X_t + (\rho_1^k)^\top X_t
\end{aligned}
$$

where we used that $\rho_0^d = \rho_0^k \equiv 0$ in our Nelson–Siegel setting as well as $\phi(0, u, v) = 0$ and $\psi(0, u, v) = u$. Then, the modified moment generating function of Y_T can be calculated

(using Eq. (2) and $\rho_0^d = \rho_0^k \equiv 0$) as

$$
\begin{aligned}
\varphi_{Y_T}(u) &\equiv \mathbb{E}^{\mathbb{Q}}\left[\frac{P^d(T, T + \delta_k)}{B_T} e^{iuY_T}\,\Big|\,\mathscr{F}_t\right] \\
&= \mathbb{E}^{\mathbb{Q}}\left[e^{iu\left(-\phi(\delta_k, 0, -\rho_1^d) - \psi(\delta_k, 0, -\rho_1^d)^\top X_T + (\rho_1^k)^\top X_T\right)} \cdot e^{\phi(\delta_k, 0, -\rho_1^d) + \psi(\delta_k, 0, -\rho_1^d)^\top X_T}\right. \\
&\qquad \left. \cdot e^{-\int_0^T (\rho_1^d)^\top X_u)du}\,\Big|\,\mathscr{F}_t\right] \\
&= e^{(1-iu)\phi(\delta_k, 0, -\rho_1^d)} \cdot \mathbb{E}^{\mathbb{Q}}\left[e^{(1-iu)\psi(\delta_k, 0, -\rho_1^d)^\top X_T + iu(\rho_1^k)^\top X_T} \cdot e^{-\int_t^T (\rho_1^d)^\top X_u du}\right. \\
&\qquad \left. \cdot e^{-\int_0^t (\rho_1^d)^\top X_u du}\,\Big|\,\mathscr{F}_t\right] \\
&= e^{(1-iu)\phi(\delta_k, 0, -\rho_1^d)} \cdot e^{\phi(T-t, (1-iu)\psi(\delta_k, 0, -\rho_1^d) + iu\rho_1^k, -\rho_1^d)} \cdot \\
&\qquad e^{\psi(T-t, (1-iu)\psi(\delta_k, 0, -\rho_1^d) + iu\rho_1^k, -\rho_1^d)^\top X_t} \cdot e^{-\int_0^t (\rho_1^d)^\top X_u du}
\end{aligned}
$$

with $\phi(T - t, u, v)$ and $\psi(T - t, u, v)$ given by the solution to the system of ODEs in Eq. (4) adapted to our Nelson–Siegel setting, i.e.

$$
\begin{aligned}
\frac{d}{dt}\phi(T - t, u, v) &= -\psi(T - t, u, v)^\top K\theta^{\mathbb{Q}} - \frac{1}{2}\sum_{j=1}^6 \left(\Sigma^\top \psi(T - t, u, v)\psi(T - t, u, v)^\top \Sigma\right)_{jj} \\
\frac{d}{dt}\psi(T - t, u, v) &= K^\top \psi(T - t, u, v) - v
\end{aligned}
$$

with boundary conditions $\phi(0, u, v) = 0$ and $\psi(0, u, v) = u$. This ODE system can be solved analytically and the solution for $\psi(T - t, u, v)$ is given by[5]

$$
\psi(T - t, u, v) = \begin{bmatrix}
v_1(T - t) + u_1 \\
\frac{e^{\lambda_d(t-T)}(\lambda_d u_2 - v_2) + v_2}{\lambda_d} \\
\frac{-e^{\lambda_d(t-T)}\left(\lambda_d^2 u_2(t-T) - \lambda_d(v_2(t-T) + u_3) + v_2 + v_3\right) + v_2 + v_3}{\lambda_d} \\
v_4(T - t) + u_4 \\
\frac{e^{\lambda_k(t-T)}(\lambda_k u_5 - v_5) + v_5}{\lambda_k} \\
\frac{-e^{\lambda_k(t-T)}\left(\lambda_k^2 u_5(t-T) - \lambda_k(v_5(t-T) + u_6) + v_5 + v_6\right) + v_5 + v_6}{\lambda_k}
\end{bmatrix}
$$

where v_i and u_i are i^{th} entries of the vectors u and v.

The price of the caplet can then be expressed as

$$
\begin{aligned}
\Pi(t, T, T + \delta_k, K, N) &= NB_t\delta_k\,\mathbb{E}^{\mathbb{Q}}\left[\frac{1}{B_{T+\delta_k}}(L_T(T, T + \delta_k) - K)^+\,\Big|\,\mathscr{F}_t\right] \\
&= N\,\mathbb{E}^{\mathbb{Q}}\left[\frac{B_t}{B_T}\left(S^k(T, T) - (1 + \delta_k K)P^d(T, T + \delta_k)\right)^+\,\Big|\,\mathscr{F}_t\right] \\
&= N\,\mathbb{E}^{\mathbb{Q}}\left[\frac{B_t}{B_T}P^d(T, T + \delta_k)\left(e^{Y_T} - (1 + \delta_k K)\right)^+\,\Big|\,\mathscr{F}_t\right].
\end{aligned}
$$

Using the above modified moment generating function, we thus obtain (compare [4], Proposition 4.2)

$$
\begin{aligned}
&\Pi(t, T, T + \delta_k, K, N) \\
&= NB_t\left(\frac{1}{2}\varphi_{Y_T}(-i) + \frac{1}{\pi}\int_0^\infty \text{Re}\left(\exp\left(-iu\log(1 + \delta_k K)\right)\frac{\varphi_{Y_T}(u - i)}{-u(u - i)}\right)du\right) \quad (11)
\end{aligned}
$$

---

[5] $\phi(T - t, u, v)$ has been computed with Mathematica®, code snippets are available upon requests.

where

$$\varphi_{Y_T}(u) = \exp\left((1 - iu)\phi(\delta_k, 0, -\rho_1^d) + \phi(T - t, (1 - iu)\psi(\delta_k, 0, -\rho_1^d) + iu\rho_1^k, -\rho_1^d)\right) \times$$

$$\exp\left(\psi\left(T - t, (1 - iu)\psi(\delta_k, 0, -\rho_1^d) + iu\rho_1^k, -\rho_1^d\right)^\top X_t - \left(\rho_1^d\right)^\top \int_0^t X_s ds\right)$$

This illustrates the applicability of our model to the pricing of caplets, which also holds in a negative interest rate environment. To further confirm the practical relevance of the model for derivative pricing we propose a simple calibration exercise. We build a surface of caplet prices on 15 Sep 2016 using discount bond values and cap implied volatilities, following the procedure outlined in [10]. We end up with market caplet prices for $n_T = 15$ maturities $T = \{2.5, 3, 3.5, 4, 4.5, 5, 5.5, 6, 6.5, 7, 7.5, 8, 8.5, 9, 9.5\}$ years and $n_K = 5$ strikes $K = \{-0.005, -0.0013, 0.0025, 0.01, 0.02\}$. Then, we calibrate the proposed model by solving numerically the following minimization problem:

$$\min_\Theta \frac{1}{n_T n_K} \sum_{k=1}^{n_k} \sum_{t=1}^{n_T} |\Pi(K_k, T_t)^{mkt} - \Pi(0, T_t, T_t + \delta_k, K_k, 1, \Theta)|$$

where Θ is the model parameters vector, Π^{mkt} is the market caplet prices surface and $\Pi(\Theta)$ is the model caplet prices surface (where we have put in evidence the dependence on the model parameters) computed as in (11). Let us consider for this numerical illustration the independent version of the AFMCNS model. We obtain the following estimates for the model parameters: $\lambda^d = 0.3540$, $\lambda^k = 0.4680$, $X_0 = \{0.0121, -0.0237, -0.0212, 0.0003, -0.0004, 0.0004\}$ and

$$\Sigma = \begin{bmatrix} 0.0016 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.0201 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.0156 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.0026 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.0021 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.0031 \end{bmatrix}.$$

Figure 3 shows the final calibration output. We find that the model is able to replicate correctly observed caplet prices with a mean absolute error (MAE) around 1.10E-04 (and squared pricing error of order 10^{-7}).

The pricing of swaptions in our framework can in principle be done by adapting the results in [4, Section 4.2] to the proposed AFMCNS model along the same lines to what has been done in the case of caplets. It should be pointed out, however, that the pricing problem is more delicate since the semi-closed formulas in the mentioned paper rely on an approximation of the exercise boundary by an event which is defined in terms of an affine function of the driving process. Finally, for what concerns derivative instruments with more involved payoff structure, we remark that, since X_t evolves according to a multi-dimensional Gaussian Ornstein-Uhlenbeck process, which can be simulated efficiently using for example the Euler scheme or, alternatively, the state transition Eq. (13), interest rate derivatives can be efficiently priced under the proposed model via Monte Carlo simulation. Other analytical or semi-closed pricing formulas for interest rate derivatives in the multiple curve setting are derived e.g. in [12] using continuous-state branching processes with immigration as driving processes or in [11] using time-inhomogeneous Lévy processes to model forward swap rates.

As the proposed model has a Nelson–Siegel factor loading structure, the latent variables have convenient economic interpretations as level, slope and curvature factors. Due to this
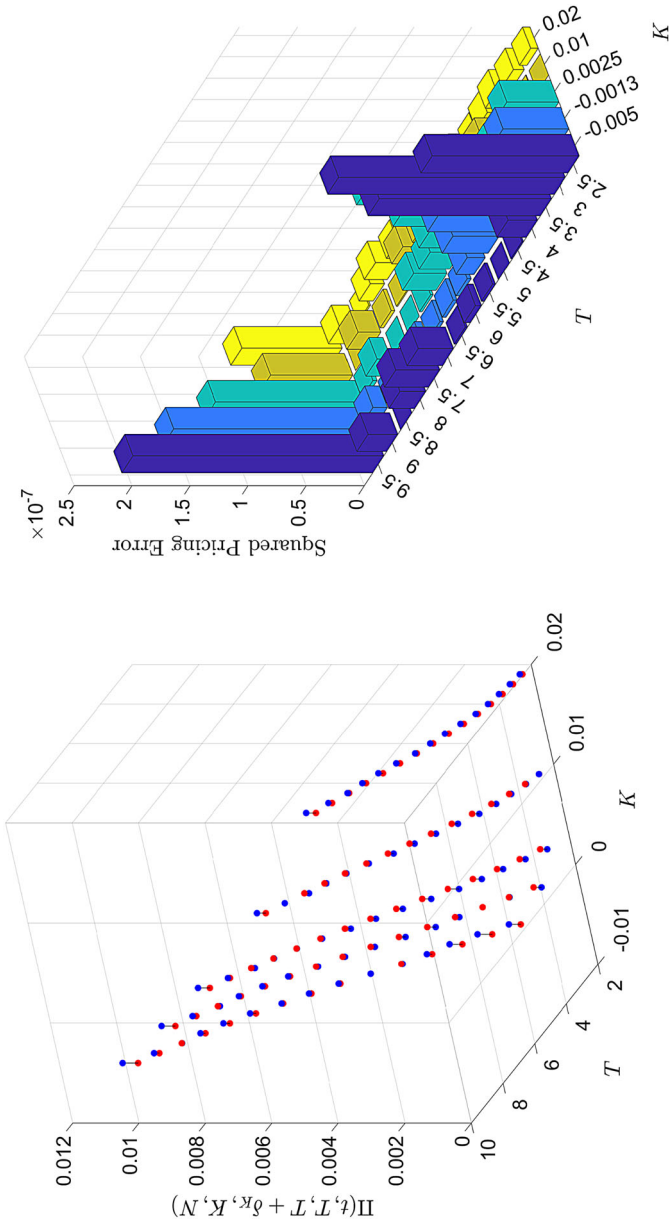
Fig. 3 Model prices against market prices as of 15 September 2016. On the left panel, market prices are represented by blue circles while model prices by red circles. On the right panel, price squared errors are reported. Further notes: $\delta_K = 6/12$, $N = 1$
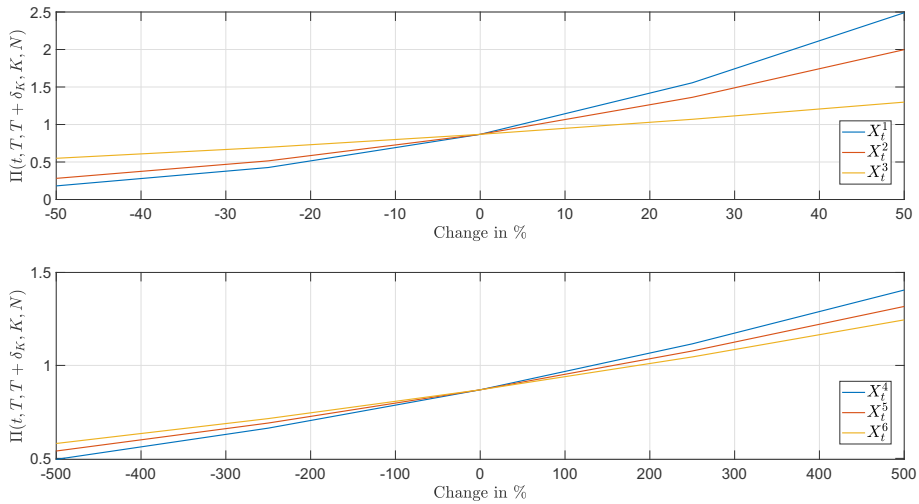
Fig. 4 Price of the Caplets computed from (11) for varying $\{X_t^i\}_{i=1}^6$. Model parameters as in Table 2 (independent case) except $\theta^{\mathbb{Q}}$ which we set equal to 0 in agreement with [2], with $X_t = \{0.0233; -0.0271; -0.0277; 0.0002; -0.0003; 0.0005\}$. Contract parameters: $N = 1000$, $t = 0$, $T = 1$, $\delta_k = 3/12$, $K = 0.0025$

feature and the fact that the model is free of arbitrage, our approach is tailor-made for risk management purposes. In the sequel, we illustrate such an application by studying the price of the caplet for varying initial state variables $(X_t^1, \ldots, X_t^6)$. Results reported in Fig. 4 show that the caplet price is most severely affected by shifts in the level X_t^1 of the discount curve, followed by slope changes X_t^2, and finally by curvature shifts X_t^3. Corresponding changes in X_t^4, X_t^5, and X_t^6 representing shifts in level, slope and curvature of the multiplicative log spreads have a smaller but non-negligible impact on the caplet price. This is particularly interesting when considering the fact that during the global financial crisis, spreads between interbank rates and overnight (discount) rates increased from less than 10bps to levels up to 250bps at the peak of the crisis. In this way, our proposed approach has important implications for risk management as economically meaningful stress scenarios can be easily simulated and due to absence of arbitrage can be used for calculating portfolio values under adverse market situations.

7 Conclusions

In this paper, we proposed an arbitrage-free affine term structure model for multiple yield curves that has a Nelson–Siegel factor loading structure. Our numerical results document superior in-sample and out-of-sample performance of our approach within the Nelson–Siegel class of models. Due to the sound economic interpretation of the latent variables of our model and the absence of arbitrage, the setting is very well suited for risk management purposes. In particular, it allows to study the sensitivity of a portfolio of interest rate related products with respect to level, slope and curvature shocks to the risk-free yield curve and/or to the tenor spreads. We illustrated this by applying the proposed model to the pricing of caplets. Since the valuation of various interest rate derivatives relies on forward looking interest rates, such

as LIBOR, and hence on potential spread adjustments to overnight rates, we believe that our results remain relevant also beyond a discontinuation of LIBOR after 2021.

Data availability Data is available through Bloomberg. Code snippets are available upon request.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

A Derivation of characteristic exponents of *Y*

To derive explicit expressions for the characteristic exponents ϕ and ψ of the process $Y = (X, \int_0^{\cdot} X_s ds)$, we first calculate the characteristic exponents of the process X starting in X_0. Therefore, suppose that

$$\mathbb{E}\left[e^{uX_t}\right] = e^{\tilde{\phi}(t,u) + \tilde{\psi}(t,u)^\top X_0} \tag{12}$$

for functions $\tilde{\phi}(t, u)$ and $\tilde{\psi}(t, u)$. Define the process

$$M_t := f(t, X_t) = \exp(\tilde{\phi}(T - t, u) + \tilde{\psi}(T - t, u)^\top X_t)$$

with $\tilde{\phi}(0, u) = 0$ and $\tilde{\psi}(0, u) = u$. Then

$$M_T = \exp(\tilde{\phi}(0, u) + \tilde{\psi}(0, u)^\top X_T) = e^{\langle u, X_T \rangle}$$

and

$$\mathbb{E}[e^{\langle u, X_T \rangle}] = \mathbb{E}[M_T] = M_0$$

if M is a martingale. In that case, we then obtain

$$\mathbb{E}[e^{\langle u, X_T \rangle}] = e^{\tilde{\phi}(T,u) + \tilde{\psi}(T,u)^\top X_0}$$

and (12) indeed gives the correct characteristic function. Thus, we need to prove that $f(t, X_t)$ is a martingale. Denoting the time derivative by $'$ and applying Itô's formula to $f(t, X_t)$, we obtain

$$\frac{df(t,X_t)}{f(t,X_t)} = -\left(\tilde{\phi}'(T - t, u) + \tilde{\psi}'(T - t, u)^\top X_t\right) dt + \tilde{\psi}(T - t, u)^\top dX_t$$
$$+ \tfrac{1}{2} \sum_{j=1}^n \left(\Sigma^\top(t)\tilde{\psi}(T - t, u)\tilde{\psi}^\top(T - t, u)\Sigma(t)\right)_{jj} (\vartheta^j(t)^\top X_t + \gamma^j(t))dt.$$

Hence, $f(t, X_t)$ is a local martingale if

$$\tilde{\phi}'(T - t, u) + \tilde{\psi}'(T - t, u)^\top X_t = \tilde{\psi}(T - t, u)^\top K(t) [\theta(t) - X_t]$$

$$+ \frac{1}{2} \sum_{j=1}^n \left(\Sigma^\top(t) \tilde{\psi}(T - t, u) \tilde{\psi}^\top(T - t, u) \Sigma(t) \right)_{jj} (\vartheta^j(t)^\top X_t + \gamma^j(t)),$$

i.e. if we have

$$\frac{d}{dt} \tilde{\phi}(T - t, u) = -\tilde{\psi}(T - t, u)^\top K(t)\theta(t) - \frac{1}{2} \sum_{j=1}^n \left(\Sigma(t)^\top \tilde{\psi}(T-t, u) \tilde{\psi}(T-t, u)^\top \Sigma(t) \right)_{jj} \gamma^j(t)$$

$$\frac{d}{dt} \tilde{\psi}(T-t, u) = K(t)^\top \tilde{\psi}(T-t, u) - \frac{1}{2} \sum_{j=1}^n \left(\Sigma(t)^\top \tilde{\psi}(T-t, u) \tilde{\psi}(T-t, u)^\top \Sigma(t) \right)_{jj} \vartheta^j(t)^\top$$

with boundary conditions $\tilde{\phi}(0, u) = 0$ and $\tilde{\psi}(0, u) = u$. We define the functional characteristics $(\tilde{F}, \tilde{R})$ of X via

$$\frac{d}{dt} \tilde{\phi}(t, u) = \tilde{F}(\tilde{\psi}(t, u)), \quad \tilde{\phi}(0, u) = 0,$$

$$\frac{d}{dt} \tilde{\psi}(t, u) = \tilde{R}(\tilde{\psi}(t, u)), \quad \tilde{\psi}(0, u) = u.$$

We can then apply Theorem 4.10 in [23] to infer the functional characteristics (F, R) of Y as

$$F(t, u, v) = \tilde{F}(t, u) \quad \text{and} \quad R(t, u, v) = \begin{pmatrix} \tilde{R}(u) + v \\ 0 \end{pmatrix}.$$

Thus, the characteristic exponents ϕ and ψ of the process Y can be written as

$$\frac{d}{dt} \phi(T - t, u, v) = -\psi(T - t, u, v)^\top K(t)\theta(t)$$

$$- \frac{1}{2} \sum_{j=1}^n \left(\Sigma(t)^\top \psi(T - t, u, v) \psi(T - t, u, v)^\top \Sigma(t) \right)_{jj} \gamma^j(t)$$

$$\frac{d}{dt} \psi(T - t, u, v) = K(t)^\top \psi(T - t, u, v)$$

$$- \frac{1}{2} \sum_{j=1}^n \left(\Sigma(t)^\top \psi(T - t, u, v) \psi(T - t, u, v)^\top \Sigma(t) \right)_{jj} \vartheta^j(t)^\top - v$$

with boundary conditions $\phi(0, u, v) = 0$ and $\psi(0, u, v) = u$ (see also [24], Sec. 3).

B State-space formulation and estimation framework

We fit the parameters of the model under the real-world measure $\mathbb{P}$. The relationship of the process dynamics between the real-world and the risk-neutral measure are characterised by Girsanov's Theorem. In particular we have

$$dW_t = dW^\mathbb{P} - b_t dt,$$

where the risk premium $b_t(\omega)$ is a predictable process of affine form

$$b_t = b_0 + BX_t$$

with $b_0 \in \mathbb{R}^6$, $B \in \mathbb{R}^{6 \times 6}$. This specification ensures the affine structure of the state variables X under $\mathbb{P}$ (see [7]). Hence, we are able to choose any vector $\theta^{\mathbb{P}}$ and matrix $K^{\mathbb{P}}$ under $\mathbb{P}$ and still ensure the required $\mathbb{Q}$-dynamics of the model.

Following [2], the state transition equation is given by

$$X_t = (I - \exp(-K^{\mathbb{P}}\Delta t))\theta^{\mathbb{P}} + \exp(-K^{\mathbb{P}}\Delta t)X_{t-1} + \eta_t \tag{13}$$

with $X_t = (X_t^1, \dots, X_t^6)^{\top}$ and the measurement equation is

$$
\underbrace{\begin{pmatrix} y_t^d(\tau_1) \\ y_t^d(\tau_2) \\ \vdots \\ y_t^d(\tau_N) \\ \log S^k(t, t+\tau_1) \\ \log S^k(t, t+\tau_2) \\ \vdots \\ \log S^k(t, t+\tau_N) \end{pmatrix}}_{\equiv y_t}
=
\underbrace{\begin{pmatrix}
1 & \frac{1-e^{-\lambda_d \tau_1}}{\lambda_d \tau_1} & \frac{1-e^{-\lambda_d \tau_1}}{\lambda_d \tau_1} - e^{-\lambda_d \tau_1} & 0 & 0 & 0 \\
1 & \frac{1-e^{-\lambda_d \tau_2}}{\lambda_d \tau_2} & \frac{1-e^{-\lambda_d \tau_2}}{\lambda_d \tau_2} - e^{-\lambda_d \tau_2} & 0 & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
1 & \frac{1-e^{-\lambda_d \tau_N}}{\lambda_d \tau_N} & \frac{1-e^{-\lambda_d \tau_N}}{\lambda_d \tau_N} - e^{-\lambda_d \tau_N} & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & e^{-\lambda_k \tau_1} & \tau_1 \lambda_k\, e^{-\lambda_k \tau_1} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
0 & 0 & 0 & 1 & e^{-\lambda_k \tau_N} & \tau_N \lambda_k e^{-\lambda_k \tau_N}
\end{pmatrix}}_{\equiv B}
\begin{pmatrix} X_t^1 \\ X_t^2 \\ X_t^3 \\ X_t^4 \\ X_t^5 \\ X_t^6 \end{pmatrix}
$$

$$
+ \underbrace{\begin{pmatrix} -\frac{\mathscr{A}^d(\tau_1)}{\tau_1} \\ -\frac{\mathscr{A}^d(\tau_2)}{\tau_2} \\ \vdots \\ -\frac{\mathscr{A}^d(\tau_N)}{\tau_N} \\ \mathscr{A}^k(\tau_1) \\ \mathscr{A}^k(\tau_2) \\ \vdots \\ \mathscr{A}^k(\tau_N) \end{pmatrix}}_{\equiv A}
+ \begin{pmatrix} \epsilon_t^d(\tau_1) \\ \epsilon_t^d(\tau_2) \\ \vdots \\ \epsilon_t^d(\tau_N) \\ \epsilon_t^k(\tau_1) \\ \epsilon_t^k(\tau_2) \\ \vdots \\ \epsilon_t^k(\tau_N) \end{pmatrix},
$$

where the transition and measurement errors are assumed to be orthogonal to the initial state and

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

with diagonal matrix H and

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

Following [29], we set the mean level $\theta = (\theta_1, \dots, \theta_6)$ under $\mathbb{Q}$ equal to zero. Therefore we obtain

$$\frac{\mathscr{A}^d(t, T)}{T-t} = \frac{1}{2(T-t)} \sum_{j=1}^{3} \int_t^T ((\Sigma^d)^{\top} \psi^d(s, T) \psi^d(s, T)^{\top} \Sigma^d)_{j,j} ds$$

$$= d_1 \frac{(T-t)^2}{6} + d_2 \left[\frac{1}{2\lambda_d^2} - \frac{1 - e^{-\lambda_d(T-t)}}{\lambda_d^3(T-t)} + \frac{1 - e^{-2\lambda_d(T-t)}}{4\lambda_d^3(T-t)} \right]$$

$$+ d_3 \left[\frac{1}{2\lambda_d^2} + \frac{e^{-\lambda_d(T-t)}}{\lambda_d^2} - \frac{(T-t)e^{-2\lambda_d(T-t)}}{4\lambda_d} - \frac{3e^{-2\lambda_d(T-t)}}{4\lambda_d^2} \right.$$

$$\left. - \frac{2(1 - e^{-\lambda_d(T-t)})}{\lambda_d^3(T-t)} + \frac{5(1 - e^{-\lambda_d(T-t)})}{8\lambda_d^3(T-t)} \right]$$

$$+ d_4 \left[\frac{T-t}{2\lambda_d} + \frac{e^{-\lambda_d(T-t)}}{\lambda_d^2} - \frac{1 - e^{-\lambda_d(T-t)}}{\lambda_d^3(T-t)} \right]$$

$$+ d_5 \left[\frac{3e^{-\lambda_d(T-t)}}{\lambda_d^2} + \frac{T-t}{2\lambda_d} + \frac{(T-t)e^{-\lambda_d(T-t)}}{\lambda_d} - \frac{3(1 - e^{-\lambda(T-t)})}{\lambda_d^3(T-t)} \right]$$

$$+ d_6 \left[\frac{1}{\lambda_d^2} + \frac{e^{-\lambda_d(T-t)}}{\lambda_d^2} - \frac{e^{-2\lambda_d(T-t)}}{2\lambda_d^2} - \frac{3(1 - e^{-\lambda_d(T-t)})}{\lambda_d^3(T-t)} + \frac{3(1 - e^{-2\lambda_d(T-t)})}{4\lambda_d^3(T-t)} \right]$$

and

$$\frac{\mathscr{A}^k(t,T)}{T-t} = \frac{1}{2(T-t)} \sum_{j=1}^{3} \int_t^T ((\Sigma^k)^\top \psi^k(s,T)\psi^k(s,T)^\top \Sigma^k)_{j,j} ds$$

$$= c_1 \frac{T-t}{2} + c_2 \frac{1 - e^{-\lambda_k(T-t)}}{4\lambda_k}$$

$$+ c_3 \left[\frac{1 - e^{-2\lambda_k(T-t)}}{8\lambda_k} - \frac{\lambda_k(T-t)^2 e^{-2\lambda_k(T-t)} + (T-t)e^{-2\lambda_k(T-t)}}{4} \right]$$

$$+ c_4 \frac{1 - e^{-\lambda_k(T-t)}}{\lambda_k} + c_5 \left[\frac{1 - e^{-\lambda_k(T-t)}}{\lambda_k} - (T-t)e^{-\lambda_k(T-t)} \right]$$

$$+ c_6 \left[\frac{1 - e^{-2\lambda_k(T-t)}}{4\lambda_k} - \frac{(T-t)e^{-2\lambda_k(T-t)}}{2} \right]$$

where the coefficients are given by $d_1 = \sigma_{11}^2$, $d_2 = \sigma_{21}^2 + \sigma_{22}^2$, $d_3 = \sigma_{31}^2 + \sigma_{32}^2 + \sigma_{33}^2$, $d_4 = \sigma_{11}\sigma_{21}$, $d_5 = \sigma_{11}\sigma_{31}$, $d_6 = \sigma_{21}\sigma_{31} + \sigma_{22}\sigma_{32}$ and $c_1 = \sigma_{44}^2$, $c_2 = \sigma_{54}^2 + \sigma_{55}^2$, $c_3 = \sigma_{64}^2 + \sigma_{65}^2 + \sigma_{66}^2$, $c_4 = \sigma_{44}\sigma_{54}$, $c_5 = \sigma_{44}\sigma_{64}$, $c_6 = \sigma_{54}\sigma_{64} + \sigma_{55}\sigma_{65}$.

The Kalman filter can be now set up in analogy to [2]. Therefore, the filter is initialised at the unconditional mean and variance of the state variables under $\mathbb{P}$: $X_0 = \theta^{\mathbb{P}}$ and $\Sigma_0 = \int_0^\infty e^{-K^{\mathbb{P}}s} \Sigma\Sigma' e^{-(K^{\mathbb{P}})'s} ds$. The prediction step is

$$X_{t|t-1} = \Phi_t + \Psi_t X_{t-1}$$
$$\Sigma_{t|t-1} = \Psi_t \Sigma_{t-1} \Psi_t' + Q$$

where $\Phi_t = (I - \exp(-K^{\mathbb{P}}\Delta t))\theta^{\mathbb{P}}$, $\Psi_t = \exp(-K^{\mathbb{P}}\Delta t)$ where Δt is the time between observations, which we set equal to $1/252$ (respectively, $1/12$) wherever we deal with daily (monthly) data. X_t is updated at time t via

$$X_t = X_{t|t-1} + \Sigma_{t|t-1} B' F_t^{-1} v_t,$$
$$\Sigma_t = \Sigma_{t|t-1} - \Sigma_{t|t-1} B' F_t^{-1} B \Sigma_{t|t-1},$$

where

$$v_t = y_t - A - BX_{t|t-1}$$
$$F_t = B\Sigma_{t|t-1}B' + H,$$
$$H = \text{diag}\left(\sigma_{\epsilon,d}^2(\tau_1), \ldots, \sigma_{\epsilon,d}^2(\tau_N), \sigma_{\epsilon,k}^2(\tau_1), \ldots, \sigma_{\epsilon,k}^2(\tau_N)\right)$$

The Gaussian log-likelihood is then computed according to:

$$\log l(y_1, \ldots, y_T) = \sum_{t=1}^{T} -\frac{N}{2}\log(2\pi) - \frac{\log\det(F_t)}{2} - \frac{1}{2}v_t' F_t^{-1} v_t \qquad (14)$$

where N is the number of observed yields and $\det(\cdot)$ denotes the determinant of a square matrix.

References

1. Bank for International Settlements: Zero-coupon yield curves: technical documentation. BIS papers no. 25 (2005)
2. Christensen, J.H.E., Diebold, F.X., Rudebusch, G.D.: The affine arbitrage-free class of Nelson–Siegel term structure models. J. Econom. **164**, 4–20 (2011)
3. Cuchiero, C., Fontana, C., Gnoatto, A.: A general HJM framework for multiple yield curve modelling. Finance Stoch. **20**, 267–320 (2016)
4. Cuchiero, C., Fontana, C., Gnoatto, A.: Affine multiple yield curve models. Math. Finance **29**, 568–611 (2019)
5. Dai, Q., Singleton, K.J.: Specification analysis of affine term structure models. J. Finance **55**, 1943–1978 (2000)
6. Diebold, F.X., Li, C.: Forecasting the term structure of government bond yields. J. Econom. **130**, 337–364 (2006)
7. Duffee, G.: Term premia and interest rate forecasts in affine models. J. Finance **57**, 405–443 (2002)
8. Duffie, D., Kan, R.: A yield-factor model of interest rates. Math. Finance **6**, 379–406 (1996)
9. ECB, Technical notes on Eurozone yield calibration. European Central Bank Technical Documentation, vol. 29, pp. 568–611 (2015)
10. Eberlein, E., Gerhart, C.: A multiple-curve Lévy forward rate model in a two-price economy. Quant. Finance **18**(4), 537–561 (2018)
11. Eberlein, E., Gerhart, C., Lütkebohmert, E.: A multiple curve Lévy swap market model. Appl. Math. Finance **27**(5), 396–421 (2020)
12. Fontana, C., Gnoatto, A., Szulda, G.: Multiple yield curve modelling with CBI processes. Math. Financ. Econ. **15**, 579–610 (2021)
13. Fontana, C., Grbac, Z., Gümbel, S., Schmidt, T.: Term structure modeling for multiple curves with stochastic discontinuities. Finance Stoch. **24**, 465–511 (2020)
14. Filipović, D., Trolle, A.B.: The term structure of interbank risk. J. Financ. Econ. **109**, 707–733 (2013)
15. Gerhart, C., Lütkebohmert, E.: Empirical analysis and forecasting of multiple yield curves. Insur. Math. Econ. **95**, 59–78 (2020)
16. Grasselli, M., Miglietta, G.: A flexible spot multiple-curve model. Quant. Finance **6**, 1465–1477 (2016)
17. Grbac, Z.: Credit risk in Lévy LIBOR modeling: rating based approach. Ph.D. thesis, vol. 109, pp. 191–226. University of Freiburg (2010)
18. Grbac, Z., Papapantoleon, A., Schoenmakers, J., Skovmand, D.: Affine LIBOR models with multiple curves: theory, examples and calibration. SIAM J. Financ. Math. **6**, 984–1025 (2015)
19. Grbac, Z., Meneghello, L., Runggaldier, W.: Derivative pricing for a multi-curve extension of the Gaussian, exponentially quadratic short rate model. Innov. Deriv. Mark. **109**, 191–226 (2016)
20. Henrard, M.: The irony in the derivatives discounting. Wilmott Mag. **1**, 92–98 (2007)
21. International Swaps and Derivatives Association, ISDA Margin Survey 2012. Available at https://www.isda.org/category/research/surveys (2012)
22. International Swaps and Derivatives Association, ISDA Margin Survey 2014. Available at https://www.isda.org/category/research/surveys (2014)

23. Keller-Ressel, M.: Affine processes—theory and applications in finances. Ph.D. thesis, Vienna University of Technology (2009)
24. Keller-Ressel, M., Mayerhofer, E.: Exponential moments of affine processes. Ann. Appl. Probab. **25**, 714–752 (2015)
25. Kenyon, C.: Post-shock short-rate pricing. Risk Mag. **1**, 83–87 (2010)
26. Kijima, M., Tanaka, K., Wong, T.: A multi-quality model of interest rates. Quant. Finance **2**, 133–145 (2009)
27. Morino, L., Runggaldier, W.: On multicurve models for the term structure. In Dieci, R., He, X.-Z., Hommes, C. (eds) Nonlinear economic dynamics and financial modelling: essays in honour of Carl Chiarella. Springer International, pp. 275–290 (2014)
28. Nelson, C.R., Siegel, A.F.: Parsimonious modelling of yield curves. J. Bus. **60**, 473–489 (1987)
29. Singleton, K.J.: Empirical Dynamic Asset Pricing. Princeton University Press, Princeton (2006)
30. Svensson, L.E.: Estimating and interpreting forward interest rates: Sweden 1992–1994. Centre for Economic Policy Research: Discussion paper, vol. 60, pp. 473–489 (1994)