Forecasting the yield curve with the arbitrage-free dynamic Nelson-Siegel model: Brazilian evidence

Fabricio Tourrucôo[12]
João F. Caldeira[3]
Guilherme V. Moura[4]
André A. P. Santos[5]

Resumo

Avaliamos se a imposição de uma restrição de não-arbitragem no modelo dinâmico de Nelson-Siegel contribui para a obtenção de projeções mais acuradas da estrutura a termo. Para tal objetivo, conduzimos uma aplicação empírica para um amplo painel de contratos futuros de taxas de juro no Brasil e testamos as diferenças nos desempenhos das projeções para especificações alternativas incluindo: passeio aleatório, autoregressões (vetoriais) e o modelo dinâmico de Nelson-Siegel com fatores correlacionados e não-correlacionados. Mostramos empiricamente que o modelo livre-de-arbitragem dinâmico de Nelson-Siegel supera todos os modelos alternativos quando horizontes de projeção mais longos são considerados.

Palavras-chave: Curva de juros, modelo livre-de-arbitragem dinâmico de Nelson-Siegel, modelos de fatores dinâmicos, filtro de Kalman

Abstract

We assess the extent to which the imposition of a no-arbitrage restriction on the dynamic Nelson-Siegel model helps obtaining more accurate forecasts of the term structure. For that purpose, we provide an empirical application based on a large panel of Brazilian interest rate future contracts and test for differences in forecasting performance among alternative benchmark specifications including the random walk, (vector) autoregressions, and the dynamic Nelson-Siegel model with both uncorrelated and correlated factors. We show empirically that the arbitrage-free Nelson-Siegel model is able to outperform all other benchmark models when longer forecasting horizons are taken into account.

Keywords: Yield curve, arbitrage-free dynamic Nelson-Siegel model, dynamic factor models, Kalman filter

JEL: E43, E47, G12, G17

Área 4 - Macroeconomia, Economia Monetária e Finanças.

1 Economia - Universidade Federal do Rio Grande do Sul – UFRGS
2 Corresponding author e-mail: fabricio.tourrucoo@ufrgs.br
3 Economia - Universidade Federal do Rio Grande do Sul - UFRGS
4 Economia – Universidade Federal de Santa Catarina - UFSC
5 Economia – Universidade Federal de Santa Catarina - UFSC

Forecasting the yield curve with the arbitrage-free dynamic Nelson-Siegel model: Brazilian evidence

Abstract

We assess the extent to which the imposition of a no-arbitrage restriction on the dynamic Nelson-Siegel model helps obtaining more accurate forecasts of the term structure. For that purpose, we provide an empirical application based on a large panel of Brazilian interest rate future contracts and test for differences in forecasting performance among alternative benchmark specifications including the random walk, (vector) autoregressions, and the dynamic Nelson-Siegel model with both uncorrelated and correlated factors. We show empirically that the arbitrage-free Nelson-Siegel model is able to outperform all other benchmark models when longer forecasting horizons are taken into account.

Keywords: yield curve, arbitrage-free Nelson-Siegel model, dynamic factor models, Kalman filter

1. Introduction

There has been growing interest in the ability to forecast the behavior of the term structure of interest rates. Such forecasts are of paramount importance for macroeconomists, financial economists and fixed income managers since bond portfolio optimization, pricing of financial assets and their derivatives, as well as risk management, rely heavily on interest rate forecasts. Moreover, these forecasts are widely used by financial institutions, regulators, and institutional investors to develop macroeconomic scenarios. One of the most popular approaches to forecast the yield curve is the the dynamic version of the Nelson & Siegel (1987) model proposed by Diebold & Li (2006) (hereafter DNS). Existing evidence suggests that these specifications are remarkably well suited both to fit the term structure and to forecast its movements (see, for example, Diebold & Rudebusch, 2013, and the references therein).

Despite the large empirical evidence favorable to the DNS approach, one drawback is that it fails on an important theoretical dimension: it does not impose the restrictions necessary for riskless arbitrage opportunities, as shown in Björk & Christensen (1999). This drawback is relevant, since many financial applications that rely on interest rate modeling such as the pricing of interest-rate-linked assets require an arbitrage-free setting. This difficulty motivated Christensen *et al.* (2009) and Christensen *et al.* (2011) to develop an arbitrage-free version of the DNS model (hereafter AFNS), thus overcoming the theoretical weakness of the original model specification.

The AFNS model of Christensen *et al.* (2009) and Christensen *et al.* (2011) has many appealing features. First, it preserves the desirable economic interpretation of the three-factor model of time-varying level, slope and curvature of the original DNS specification. Second, the AFNS ensures lack of arbitrage opportunities with a more parsimonious structure in comparison to general affine arbitrage-free models such as those considered in Duffie & Kan (1996) and Duffee (2002). More specifically, the way the AFNS achieves these desirable properties is by adding a yield-adjustment term to the original DNS specification which contains the necessary restrictions to ensure freedom of arbitrage.

That being said, an immediate question that arises is the following: is the imposition of no-arbitrage helpful for forecasting purposes? This question is indeed very controversial. First, as we shall see in

Section 2.2, the yield-adjustment term of the AFNS model puts no restriction on the physical dynamics of yields. In other words, the imposition of no-arbitrage delivers yield-adjustment term that vary with maturity but are constant over time. This suggests that the inclusion of the yield-adjustment term is unlikely to provide forecasting gains. In particular, Duffee (2011) points out that imposition of no-arbitrage based on cross-section restrictions is irrelevant for forecasting.[1] However, empirical work has found predictive gains from imposing no-arbitrage in a DNS framework. For instance, Gimeno & Marqués (2009) and Christensen *et al.* (2011) found that imposition of no-arbitrage leads to substantial forecast improvements. Diebold & Rudebusch (2013) point out that, despite its time constancy, the yield-adjustment term can act as a bias correction and thus produce forecast improvements.[2] Clearly, additional empirical work is necessary in order to help clarify the gains (or lack of) coming from the imposition of no-arbitrage in the DNS class of models.

In this paper we assess the extent to which the imposition of a no-arbitrage restriction on the DNS setting brings additional forecasting gains by providing an empirical evidence based on a large data set of constant-maturity future contracts of the Brazilian Inter Bank Deposit Future Contract (DI1) which is equivalent to a zero-coupon bond and is highly liquid (293 million contracts worth US$ 15 billion traded in 2010). The market for DI1 contracts is one of the most liquid interest rate markets in the world. Many banks, insurance companies, and investors use DI1 contracts as investment and hedging instruments. The data set considered in the paper contains daily observations of DI1 contracts traded on the Brazilian Mercantile and Futures Exchange (BM&F) with fixed maturities of 3, 6, 9, 12, 15, 18, 21, 24, 27, 30, 33, 36, 42, and 48 months. We use a rolling estimation to produce out-of-sample forecasts for 1-week, 1-month, 3-month, and 6-month ahead based on the DNS and AFNS models with both uncorrelated and correlated factors. Moreover, we consider three alternative classical benchmark specifications: the random walk (RW) model, which is taken as our baseline model, and the unrestricted first-order autoregressive (AR(1)) and vector-autoregressive (VAR(1)) models. We assess forecasting accuracy by means of the root mean squared forecast error (RMSFE) and the trace RMSFE (TRMSFE) considered in Hördahl *et al.* (2006) and de Pooter *et al.* (2010). Finally, we also test for the differences in forecasting performance using the test proposed in Giacomini & White (2006).

Our empirical evidence suggests that when a short forecasting horizon is considered (e.g. 1-week-ahead) the differences in forecasting performance among the candidate models is rather inconclusive since in very few instances the candidate models outperform the baseline RW model. However, when longer forecasting horizons are considered, the AFNS model with uncorrelated factors appears to deliver the most accurate forecasts in the vast majority of the cases. Therefore, the most important message from our empirical test is that the imposition of no-arbitrage is indeed helpful but only for longer (e.g. 3-month- and 6-month-ahead) forecasting horizons. These results corroborate the evidence reported in Christensen *et al.* (2011) for US Treasuries data, as they find that the AFNS model with indendent factors outperform the DNS specifications for the 6-month- and 12-month-ahead forecast horizons.

The rest of the paper is organized as follows. Section 2 describes the DNS and AFNS specifications adopted in this paper. Section 3 discusses the implementation details and the empirical results. Finally, Section 5 concludes.

[1] Coroneo *et al.* (2011) and Nyholm & Vidova-Koleva (2012) also found that the imposition of no-arbitrage adds little to forecasting accuracy.

[2] Other authors such as Ang & Piazzesi (2003), Mönch (2008), Carriero (2011), Favero *et al.* (2012), among others, have also found that imposition of no-arbitrage leads to more accurate term structure forecasts. Moreover, Carriero & Giacomini (2011) show that the imposition of no-arbitrage restriction is important specially when an economic measure of accuracy is taken into account.

2. The Nelson-Siegel class of models

Nelson & Siegel (1987) have shown that the term structure can be surprisingly well fitted at a particular point in time by a linear combination of three smooth functions. The Nelson-Siegel model of the yield curve is given by

$$y(\tau) = \beta_1 + \beta_2 \left(\frac{1 - e^{-\lambda\tau}}{\lambda\tau} \right) + \beta_3 \left(\frac{1 - e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau} \right) + \epsilon_\tau \tag{1}$$

where $y(\tau)$ is the zero-coupon yield with τ months to maturity, and β_1, β_2, and β_3 can be interpreted as the level, slope, and curvature of the yield curve, respectively. The parameter λ determines the exponential decay of the β_2 and β_3 loadings.

2.1. Dynamic Nelson-Siegel model (DNS)

Diebold & Li (2006) show that using the static Nelson-Siegel model as basis for a dynamic factor model generates highly accurate interest rate forecasts. The dynamic Nelson-Siegel model (DNS) is

$$y_t(\tau) = X_{1t} + X_{2t} \left(\frac{1 - e^{-\lambda\tau}}{\lambda\tau} \right) + X_{3t} \left(\frac{1 - e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau} \right) + \epsilon_t(\tau), \tag{2}$$

where X_{1t}, X_{2t}, and X_{3t} can be interpreted as the time-varying level, slope, and curvature factors.

The DNS model can be interpreted as a dynamic factor model and written in state-space form. Consider the $N \times T$ matrix of observable yields $(\mathbf{y}_t)$, $t = 1, \ldots, T$, where $\mathbf{y}_t$ is the observation vector at time t, $\mathbf{y}_t = (y_{it})$, $i = 1, \ldots, N$. The DNS model can be represented in state-space form as

$$\mathbf{y}_t = \Gamma + B\mathbf{X}_t + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, \mathbf{\Sigma}), \quad t = 1, \ldots, T, \tag{3}$$

$$\mathbf{X}_t = \mu + A\mathbf{X}_{t-1} + \eta_t, \quad \eta_t \sim \mathcal{N}(0, \Omega), \quad t = 1, \ldots, T, \tag{4}$$

where B is the $N \times K$ matrix of factor loadings that depends on the decay parameter λ, $\mathbf{X}_t = (X_{1t}, \ldots, X_{Kt})'$ is a K-dimensional vector containing the coefficients, ε_t is the $N \times 1$ vector of disturbances with $\mathbf{\Sigma}$ being its $N \times N$ diagonal covariance matrix. μ is a $K \times 1$ vector of constants, A is the $K \times K$ transition matrix, and Ω is the conditional covariance matrix of disturbance vector η_t, which are independent of the residuals ε_t $\forall t$. Γ is a vector of constants, which is fixed to zero in the DNS specification but will be different from zero in the AFNS specification presented bellow. Equations (3) and (4) characterize a general Gaussian linear state-space model.

Empirically, the DNS model is highly tractable and provides good fits. Theoretically, however, arbitrage opportunities cannot be prevented due to an unrestricted dynamic evolution of the yields. Indeed, as implied by Filipovic (1999), it is impossible to prevent arbitrage at bond prices in the resulting Nelson-Siegel yield curve. Christensen *et al.* (2011) showed how to remedy this theoretical weakness.

2.2. The dynamic arbitrage-free Nelson-Siegel model (AFNS)

Aiming to overcome this theoretical weakness, but hoping to maintain the good properties of the DNS model, Christensen *et al.* (2011) have developed an arbitrage-free affine model in the spirit of Duffie & Kan (1996) which has the same factor loadings as the original DNS model. The resulting class of AFNS models involves considering the filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{Q})$ with filtration $(\mathcal{F}_t) = \{\mathcal{F}_t : t \leq 0\}$ satisfying the usual conditions (see Williams, 1997). $\mathbb{Q}$ denotes the risk-neutral measure and we will denote the real world probability measure by $\mathbb{P}$. The risk-neutral dynamic factors $\mathbf{X}_t$ are assumed to follow a Markov process defined on a set $M \subseteq \mathbb{R}^n$ that solves the stochastic differential equation (SDE)

$$\mathsf{d}\mathbf{X}_t = \mathbf{K}^{\mathbb{Q}}(t) \left[\Theta^{\mathbb{Q}}(t) - \mathbf{X}_t \right] \mathsf{d}t + \mathbf{\Sigma}(t) D(\mathbf{X}_t, t) \mathsf{d}\mathbf{W}_t^{\mathbb{Q}}, \tag{5}$$

where $\mathbf{W}^{\mathbb{Q}}$ is a standard Brownian Motion in $\mathbb{R}^n$ defined according to the filtration $(\mathcal{F}_t)$. Furthermore, as in Duffie & Kan (1996) one assumes that the drifts, $\Theta^{\mathbb{Q}} : [0,T] \to \mathbb{R}^n$, dynamics, $\mathbf{K}^{\mathbb{Q}} : [0,T] \to \mathbb{R}^{n\times n}$, and volatility, $\mathbf{\Sigma} : [0,T] \to \mathbb{R}^{n\times n}$, are continuous functions. Finally, it is assumed that the mapping $D : M \times [0,T] \to \mathbb{R}^{n\times n}$ has a diagonal structure with entries

$$[D]_{ii} = \sqrt{\gamma^i(t) + \delta^i(t)X_{1t} + \ldots + \delta^n(t)X_{nt}} \quad \forall i \in \{1,\ldots,n\}, \tag{6}$$

where $\gamma^i : [0,T] \to \mathbb{R}^n$ and $\delta^i : [0,T] \to \mathbb{R}^{n\times n}$ are continuous functions. In addition, Duffie & Kan (1996) assume the instantaneous risk-neutral rate is an affine function of the state variables

$$r_t = \rho_0(t) + \rho_1(t)'\mathbf{X}_t, \tag{7}$$

with continuous functions $\rho_0 : [0,T] \to \mathbb{R}$ and $\rho_1 : [0,T] \to \mathbb{R}^n$. Under this affine formulation Duffie & Kan (1996) proved that a closed-form analytic expression for zero-coupon bond prices is attained as a linear function of the latent dynamic factors. This implies that the zero-coupon yield at time t for a bond with maturity T is given by:

$$y(t,T) = -\frac{1}{T-t}\log\left(\mathbb{E}^{\mathbb{Q}}\left[\exp\left(-\int_t^T r_u \mathsf{d}u\right)\right]\right) = -\frac{1}{T-t}\left[\Gamma(t,T) + B(t,T)'\mathbf{X}_t\right] \tag{8}$$

where $\Gamma(t,T)$ and $B(t,T)$ satisfy the system of ordinary differential equations (ODEs)

$$\begin{aligned} \tfrac{\mathsf{d}\Gamma(t,T)}{\mathsf{d}t} &= \rho_0 - B(t,T)'\left(\mathbf{K}^{\mathbb{Q}}\right)'\theta^{\mathbb{Q}} - \tfrac{1}{2}\textstyle\sum_{i=1}^n \left[\mathbf{\Sigma}'B(t,T)B(t,T)'\mathbf{\Sigma}\right]_{ii}\gamma^i, \\ \tfrac{\mathsf{d}B(t,T)}{\mathsf{d}t} &= \rho_1 + \left(\mathbf{K}^{\mathbb{Q}}\right)'B(t,T) - \tfrac{1}{2}\textstyle\sum_{i=1}^n \left[\mathbf{\Sigma}'B(t,T)B(t,T)'\mathbf{\Sigma}\right]_{ii}\left(\delta^i\right)', \end{aligned} \tag{9}$$

with boundary conditions $\Gamma(T,T) = 0$ and $B(T,T) = 0$.

Christensen *et al.* (2011) objective was to work under the generic affine latent factor dynamic structure proposed in(8) in order to rule out arbitrage opportunities. On the other hand, they had to impose additional restrictions on the ODE for $B(t,T)$ in (9) to ensure that it has solutions equal to the factor loadings of the DNS model:

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

This results in

$$y(t,T) = X_{1t} + X_{2t}\left[\frac{1-e^{-\lambda(T-t)}}{\lambda\left(T-t\right)}\right] + X_{3t}\left[\frac{1-e^{-\lambda(T-t)}}{\lambda\left(T-t\right)} - e^{-\lambda(T-t)}\right] - \frac{\Gamma(t,T)}{T-t}, \tag{10}$$

where DNS factor loadings are matched and the additional term $-\frac{\Gamma(t,T)}{T-t}$ is a correction to ensure the model is arbitrage-free.

Unlike the DNS models, which admit arbitrage, the AFNS models of Christensen *et al.* (2011) impose a particular structure on both the risk-neutral and real world latent factor dynamic processes. Christensen *et al.* (2011), following Singleton (2006), argue that not all parameters in the specification (5-7) can be identified so that identifying restrictions must be imposed under the $\mathbb{Q}$ measure. Specifically, the mean $\Theta^{\mathbb{Q}} = 0$, the volatility matrix $\mathbf{\Sigma}$ is triangular,

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

$\rho_0 = 0$, $\rho_1 = (1, 1, 0)'$, $\delta = 0$, $\gamma = (1, \ldots, 1)'$, and the state-variable reversion rate

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

Under those restrictions, the yield adjustment term $-\frac{\Gamma(t,T)}{T-t}$ is of the form

$$\frac{\Gamma(t,T)}{T-t} = \frac{1}{2}\frac{1}{T-t}\sum_{i=1}^{3}\int_t^T \left[\mathbf{\Sigma}' B(s,T) B(s,T)' \mathbf{\Sigma}\right]_{ii} \mathsf{d}s, \tag{11}$$

the instantaneous risk-neutral rate is given by $r_t = X_{1t} + X_{2t}$, and the latent state variables $\mathbf{X}_t = (X_{1t}, X_{2t}, X_{3t})$ are described by the system of SDEs under the risk-neutral measure $\mathbb{Q}$

$$\begin{pmatrix} \mathsf{d}X_{1t} \\ \mathsf{d}X_{2t} \\ \mathsf{d}X_{3t} \end{pmatrix} = -K^{\mathbb{Q}} \begin{pmatrix} X_{1t} \\ X_{2t} \\ X_{3t} \end{pmatrix} \mathsf{d}t + \mathbf{\Sigma} \begin{pmatrix} \mathsf{d}W_{1t}^{\mathbb{Q}} \\ \mathsf{d}W_{2t}^{\mathbb{Q}} \\ \mathsf{d}W_{3t}^{\mathbb{Q}} \end{pmatrix}, \quad \lambda > 0.$$

This completes the specification of the AFNS model under the risk-neutral measure.

2.2.1. Independent and correlated AFNS models

We consider two versions of the DNS model and investigate the effect of the arbitrage-free restriction on their corresponding AFNS versions. The first is the uncorrelated-factor DNS model which is specified in terms of a diagonal transition matrix in (4). The correlated-factor DNS model has a full transition matrix instead. In both DNS models the measurement equation is the same.

The corresponding AFNS models are formulated in continuous time and the relation between the dynamics under measures $\mathbb{Q}$ and $\mathbb{P}$ is given by a change of measure that preserves the affine dynamic structure (see Christensen *et al.* (2011) for full details). This is a very convenient property since it guarantees that identical ($\mathbb{P}$-measure) models can be estimated.

Using the solution already presented in (10), yields must be related to the state variables by

$$\mathbf{y}_t = \Gamma + BX_t + \varepsilon_t. \tag{12}$$

Comparing equations (2) and (10), we notice that the matrix B is identical for DNS and AFNS models. The only difference is the added vector Γ containing the yield-adjustment terms in the AFNS models. Since the AFNS is a continuous model, the dynamics is discretized to allow comparison with the discrete-time DNS models. The estimation is outlined in the next section.

For the uncorrelated-factor AFNS model the dynamics of the state variables under the $\mathbb{P}$-measure is

$$\begin{pmatrix} \mathsf{d}X_{1t} \\ \mathsf{d}X_{2t} \\ \mathsf{d}X_{3t} \end{pmatrix} = \begin{pmatrix} \kappa_{11}^{\mathbb{P}} & 0 & 0 \\ 0 & \kappa_{22}^{\mathbb{P}} & 0 \\ 0 & 0 & \kappa_{33}^{\mathbb{P}} \end{pmatrix} \left[\begin{pmatrix} \theta_1^{\mathbb{P}} \\ \theta_2^{\mathbb{P}} \\ \theta_3^{\mathbb{P}} \end{pmatrix} - \begin{pmatrix} X_{1t} \\ X_{2t} \\ X_{3t} \end{pmatrix} \right] \mathsf{d}t + \begin{pmatrix} \sigma_{11} & 0 & 0 \\ 0 & \sigma_{22} & 0 \\ 0 & 0 & \sigma_{33} \end{pmatrix} \begin{pmatrix} \mathsf{d}W_{1t}^{\mathbb{P}} \\ \mathsf{d}W_{2t}^{\mathbb{P}} \\ \mathsf{d}W_{3t}^{\mathbb{P}} \end{pmatrix}. \tag{13}$$

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

$$\begin{pmatrix} \mathsf{d}X_{1t} \\ \mathsf{d}X_{2t} \\ \mathsf{d}X_{3t} \end{pmatrix} = \begin{pmatrix} \kappa^{\mathbb{P}}_{11} & \kappa^{\mathbb{P}}_{12} & \kappa^{\mathbb{P}}_{13} \\ \kappa^{\mathbb{P}}_{21} & \kappa^{\mathbb{P}}_{22} & \kappa^{\mathbb{P}}_{23} \\ \kappa^{\mathbb{P}}_{31} & \kappa^{\mathbb{P}}_{32} & \kappa^{\mathbb{P}}_{33} \end{pmatrix} \left[\begin{pmatrix} \theta^{\mathbb{P}}_1 \\ \theta^{\mathbb{P}}_2 \\ \theta^{\mathbb{P}}_3 \end{pmatrix} - \begin{pmatrix} X_{1t} \\ X_{2t} \\ X_{3t} \end{pmatrix} \right] \mathsf{d}t + \begin{pmatrix} \sigma_{11} & 0 & 0 \\ \sigma_{21} & \sigma_{22} & 0 \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{pmatrix} \begin{pmatrix} \mathsf{d}W^{\mathbb{P}}_{1t} \\ \mathsf{d}W^{\mathbb{P}}_{2t} \\ \mathsf{d}W^{\mathbb{P}}_{3t} \end{pmatrix}. \quad (14)$$

This is the most flexible version of the AFNS models, where all parameters are identified. From these specifications we see the striking resemblance between DNS and AFNS models.

2.2.2. An exact expression for the covariance of the continuous-time AFNS model

In this paper, the models are estimated using a maximum likelihood estimation method based on the Kalman filter discussed in Section 3. The AFNS model can be formulated in a state-space form as follows. Departing from the continuous-time formulation of the AFNS model, the conditional mean vector and the conditional covariance matrix are

$$\mathbb{E}^{\mathbb{P}}\left[\mathbf{X}_T | \mathcal{F}_t\right] = \left[I - \exp(-\mathbf{K}^{\mathbb{P}}\Delta t)\right]\Theta^{\mathbb{P}} + \exp(-\mathbf{K}^{\mathbb{P}}\Delta t)X_t \quad (15)$$

$$\mathbb{V}^{\mathbb{P}}\left[\mathbf{X}_T | \mathcal{F}_t\right] = \int_0^{\Delta t} \exp(-\mathbf{K}^{\mathbb{P}}s)\boldsymbol{\Sigma}\boldsymbol{\Sigma}' \exp(-(\mathbf{K}^{\mathbb{P}})'s)\mathsf{d}s \quad (16)$$

To estimate the AFNS models proposed by Christensen *et al.* (2011) we must compute the conditional covariance matrix

$$\mathbb{V}^{\mathbb{P}}\left[\mathbf{X}_t | \mathcal{F}_{t-1}\right] = \int_0^{\Delta t} \exp(-\mathbf{K}^{\mathbb{P}}s)\boldsymbol{\Sigma}\boldsymbol{\Sigma}' \exp(-(\mathbf{K}^{\mathbb{P}})'s)\mathsf{d}s \quad (17)$$

of discrete observations. Since the estimation is an intense computational process, we need to provide fast intermediate calculations. One approach is to approximate the integral and the matrix exponential. Another approach uses the diagonalization of $\mathbf{K}^{\mathbb{P}}$ to calculate the integral exactly. Due to the reduced size of the matrix $\mathbf{K}^{\mathbb{P}}$, the latter is significantly faster than the former.

The AFNS state transition is

$$\mathbf{X}_t = \left(I - \exp\left(-\mathbf{K}^{\mathbb{P}}\Delta t\right)\right)\Theta^{\mathbb{P}} + \exp\left(-\mathbf{K}^{\mathbb{P}}\Delta t\right)\mathbf{X}_{t-1} + \eta_t,$$

where Δt is the time between the observations at t and $t-1$, with measurement equation

$$\mathbf{y}_t = \Gamma + BX_t + \varepsilon_t,$$

and error structure

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

where H is diagonal, $Q = \mathbb{V}^P\left[X_t | \mathcal{F}_{t-1}\right]$ and the transition and measurement errors are assumed orthogonal to the initial state.

The computation of Q is straightforward if we use the diagonalization of $\mathbf{K}^{\mathbb{P}}$

$$\mathbf{K}^{\mathbb{P}} = V\Lambda V^{-1}, \quad (18)$$

where V contains the eigenvectors of $\mathbf{K}^{\mathbb{P}}$, and Λ is a diagonal matrix containing the eigenvalues (λ_i) of $\mathbf{K}^{\mathbb{P}}$. We refer the reader to Lang (1987) and Golub & Van Loan (1996) for details on linear algebra theory and on the computational aspects of linear algebra, respectively.

Substituting (18) in (17), using

$$\exp(-\mathbf{K}^{\mathbb{P}}s) = V\exp(-\Lambda s)V^{-1},$$

and similarly

$$\exp(-\left(\mathbf{K}^{\mathbb{P}}\right)^{T}s) = \left(V^{-1}\right)^{T}\exp(-\Lambda s)V^{T},$$

we obtain

$$Q = V\left(\int_{0}^{\Delta t}\exp(-\Lambda s)\Omega\exp(-\Lambda s)ds\right)V^{T},$$

where $\Omega = \left(\omega_{ij}\right)_{n\times n} = V^{-1}\mathbf{\Sigma}\mathbf{\Sigma}^{T}\left(V^{-1}\right)^{T}$. Since the exponential of a diagonal matrix with entries $-\lambda_i s$ is a diagonal matrix with entries $e^{-\lambda_i s}$, each term of the matrix under the integral is $(\omega_{ij}e^{-(\lambda_i+\lambda_j)s})_{n\times n}$. Integration yields an expression which only involves matrix multiplications

$$Q = V\left(\frac{\omega_{ij}}{\lambda_i+\lambda_j}\left(1-e^{-(\lambda_i+\lambda_j)\Delta t}\right)\right)_{n\times n}V^{T}. \tag{19}$$

Stationarity of the system under the $\mathbb{P}$-measure is ensured if the real parts of all eigenvalues of $\mathbf{K}^{\mathbb{P}}$ are positive, and this condition is imposed in all estimations. For this reason, we can start the Kalman filter at the unconditional mean, $\mathbf{X}_0 = \Theta^{\mathbb{P}}$, and covariance matrix, $\mathbf{\Sigma}_0$. In particular, the unconditional variance $\mathbf{\Sigma}_0 = \int_0^{\infty}\exp(-\mathbf{K}^{\mathbb{P}}s)\mathbf{\Sigma}\mathbf{\Sigma}'\exp(-(\mathbf{K}^{\mathbb{P}})'s)\mathsf{d}s$ used in the initialization of the filter is easily obtained from the above expression. Assuming $\mathbf{K}^{\mathbb{P}}$ has eigenvalues with positive real parts, the integral converges to

$$\mathbf{\Sigma}_0 = V\left(\frac{\omega_{ij}}{\lambda_i+\lambda_j}\right)_{n\times n}V^{T}. \tag{20}$$

3. Data and estimation framework

In this section we describe the estimation method and present the data set used in the empirical analysis. Specifically, we estimate the three-factor dynamic Nelson-Siegel model of Diebold & Li (2006) and the arbitrage-free Nelson and Siegel model developed by Christensen *et al.* (2011) via Maximum Likelihood. Benchmark models used include the Random Walk forecast (RW), linear autoregressive (AR) and vector autoregressive models (VAR).

3.1. Estimation of the DNS and AFNS models

Given the state space formulation of the dynamic factor model presented in (3) and (4), the Kalman filter can be used to obtain the likelihood function via the prediction error decomposition, as well as filtered estimates of the states and of their covariance matrices. However, the computational burden associated with the Kalman filter recursions depends crucially on the dimension of both the state and observation vectors. Moreover, in yield curve models the dimension of the observation vector $(N\times 1)$ is often much larger than that of the state vector $(K\times 1)$. In these circumstances, Jungbacker & Koopman (2014) have shown that significant computational gains can be achieved by a simple transformation. First, define the $N\times N$ and the $K\times N$ matrices:

$$J = \begin{bmatrix} J^L \\ J^H \end{bmatrix}, \qquad J^L = C\Lambda(\lambda)'\Sigma^{-1},$$

respectively, where C can be any $K \times K$ invertible matrix, and J^H is chosen to guarantee that J is full rank. Selecting $C = (\Lambda(\lambda)'\Sigma^{-1}\Lambda(\lambda))^{-1}$ implies:

$$J(y_t - \Gamma) = \begin{bmatrix} J^L(y_t - \Gamma) \\ J^H(y_t - \Gamma) \end{bmatrix} = \begin{bmatrix} y_t^L \\ y_t^H \end{bmatrix} = \begin{bmatrix} f_t \\ 0 \end{bmatrix} + \begin{bmatrix} A^L\varepsilon_t \\ A^H\varepsilon_t \end{bmatrix}, \begin{bmatrix} A^L\varepsilon_t \\ A^H\varepsilon_t \end{bmatrix} \sim N\left(0, \begin{bmatrix} C & 0 \\ 0 & \Sigma^H \end{bmatrix}\right).$$

The law of motion of the factors in (4) is not affected by the transformation. Note that y_t^H is neither dependent on f_t, nor correlated with y_t^L and, therefore, does not need to be considered for the estimation of the factors. This implies that the Kalman filter only needs to be applied to the low dimensional subvector y_t^L for signal extraction, generating large computational gains when $N >> K$ (see Table 1 of Jungbacker & Koopman, 2014).

Denote $l(y)$ the log-likelihood function of the untransformed model in (3) and (4), where $y = (y_1', \ldots, y_T')'$. Evaluation of $l(y)$ can also take advantage of the transformations presented above. Jungbacker & Koopman (2014) show that the log-likelihood of the untransformed model can be represented as

$$l(y) = c + l(y^L) - \frac{T}{2} \log \frac{|\Sigma|}{|C|} - \frac{1}{2} \sum_{t=1}^{T} e_t'\Sigma^{-1}e_t, \tag{21}$$

where c is a constant independent of both y and the parameters, $l\left(y^L\right)$ is the log-likelihood function of the reduced system, and $e_t = y_t - \Gamma - \Lambda(\lambda)f_t$. Note that computation of matrix J^H is not required at any point, as proved in Lemma 2 of Jungbacker & Koopman (2014).

3.2. Data

The analyzed data set consists of Brazilian Interbank Deposit Futures Contract (DI1)[3]. DI1 market is one of the largest fixed-income markets among emerging economies. We use DI1 daily closing yields of available contracts. Since not all maturities are observed on a daily basis, we interpolate the available data using cubic splines to produce yield curves with fixed maturities. The chosen fixed maturities are: 3, 6, 9, 12, 15, 18, 21, 24, 27, 30, 33, 36, 42, and 48 months.

The DI1 contract with maturity τ is a zero-coupon future contract in which the underlying asset is the DI interest rate accrued on a daily basis, capitalized between trading days t and τ[4] The value of contract is set by its value at maturity, $R\$100,000.00$, discounted according to the accrued interest rate negotiated between the seller and the buyer. In 2010 the DI1 market traded a total of 293 million contracts corresponding to US$ 15 billion. The DI1 contract is very similar to the zero-coupon bond, except for the daily payment of margin adjustments. The data set contains liquid maturities from

[3]BM&FBOVESPA is the entity that offers the DI1 contract and determines the number of maturities with authorized contracts.

[4]The DI rate is the average daily rate of Brazilian Interbank Deposits (borrowing/lending), calculated by the Clearinghouse for Custody and Settlements (CETIP) for all business days. The DI rate, which is published on a daily basis, is expressed in annually compounded terms, based on 252 business days. When buying a DI1 contract at the $DI^\star$ rate at time t and keeping it until maturity τ, the gain or loss is given by:

$$100.000 \left(\frac{\prod_{i=1}^{\zeta(t,\tau)} (1 + y_i)^{\frac{1}{252}}}{(1 + DI^*)^{\frac{\zeta(t,\tau)}{252}}} - 1 \right),$$

where y_i denotes the DI rate, $(i-1)$ days after the trading day. The function $\zeta(t,\tau)$ represents the number of working days between t and τ.

January 2006 to December 2012, with a total of $T = 1488$ daily observations. The data source is the Brazilian Mercantile and Futures Exchange (BM&F).

Table 1 reports descriptive statistics for the Brazilian interest rate yield curve based on the DI1 market. For each time series we report the mean, standard deviation, minimum, maximum and the lag-1 sample autocorrelation. The summary statistics confirm some common stylized facts to yield curve data: the sample average curve is upward sloping and concave, volatility is decreasing with maturity, and autocorrelations are very high.

Table 1: **Descriptive statistics for the term structure of interest rates**

The Table reports summary statistics for DI1 yields over the sample period January 2007 - December 2012 (1488 daily observations). We examine daily data, constructed using cubic splines. Maturity is measured in months. We show, for each maturity, mean, standard deviation, minimum, maximum and and a selection of autocorrelation (Acf, $\widehat{\rho}(1)$, $\widehat{\rho}(5)$, and $\widehat{\rho}(21)$, respectively) and partial autocorrelation (Pacf, $\widehat{\alpha}(2)$ and $\widehat{\alpha}(5)$) coefficients.

								Acf			Pacf	
Maturity	τ	Mean	Std Dev	Min	Max	Skew	Kurt	$\widehat{\rho}(1)$	$\widehat{\rho}(5)$	$\widehat{\rho}(21)$	$\widehat{\alpha}(2)$	$\widehat{\alpha}(5)$
Month	1	10.66	1.75	6.97	14.13	-0.337	2.219	0.998	0.987	0.936	-0.016	-0.012
	3	10.66	1.79	7.02	14.52	-0.286	2.296	0.998	0.989	0.940	-0.017	-0.025
	6	10.73	1.86	6.91	15.32	-0.201	2.472	0.998	0.989	0.939	-0.012	-0.023
	9	10.85	1.90	6.86	16.04	-0.132	2.650	0.998	0.988	0.934	0.004	-0.010
	12	11.00	1.92	6.87	16.40	-0.091	2.784	0.997	0.987	0.929	0.019	0.000
	15	11.16	1.92	6.90	16.91	-0.053	2.907	0.997	0.985	0.923	0.024	0.003
	18	11.31	1.88	7.00	17.12	-0.046	3.002	0.997	0.984	0.917	0.020	0.002
	21	11.42	1.85	7.14	17.26	-0.042	3.118	0.997	0.983	0.912	0.023	-0.002
	24	11.52	1.81	7.32	17.44	-0.032	3.243	0.996	0.982	0.907	0.020	-0.001
	27	11.60	1.77	7.49	17.62	-0.004	3.375	0.996	0.981	0.901	0.015	-0.002
	30	11.66	1.74	7.65	17.78	0.028	3.504	0.996	0.980	0.895	0.010	-0.006
	36	11.75	1.66	7.91	17.83	0.080	3.731	0.995	0.978	0.887	0.015	-0.013
	42	11.83	1.59	8.13	17.93	0.104	3.937	0.995	0.976	0.880	0.017	-0.006
	48	11.89	1.56	8.29	18.00	0.181	4.156	0.995	0.975	0.873	0.018	-0.011
	Level	12.41	1.42	9.34	18.99	0.673	5.185	0.992	0.957	0.800	-0.064	-0.014
	Slope	-1.92	1.95	-6.37	2.84	-0.267	2.814	0.995	0.974	0.866	-0.039	-0.036
	Curvature	-1.00	3.94	-9.48	8.60	0.120	2.192	0.997	0.978	0.903	-0.160	-0.011

Figure 1 displays a three-dimensional plot of the data set and illustrates how yield levels and spreads vary substantially throughout the sample. The plot also suggests the presence of an underlying factor structure. Although the yield series vary substantially over time for each maturity, a strong common pattern in the 14 series is apparent for most of them: the yield curve is an upward-sloping function of maturity. For example, the last year of the sample is characterized by rising interest rates, especially for the shorter maturities, which respond faster to the contractionary monetary policy implemented by the Brazilian Central Bank in the first half of 2010. It is clear from Figure 1 that not only the term structure level fluctuates over time but also its slope and curvature. The curve takes on various forms ranging from nearly flat to (inverted) S-type shapes.

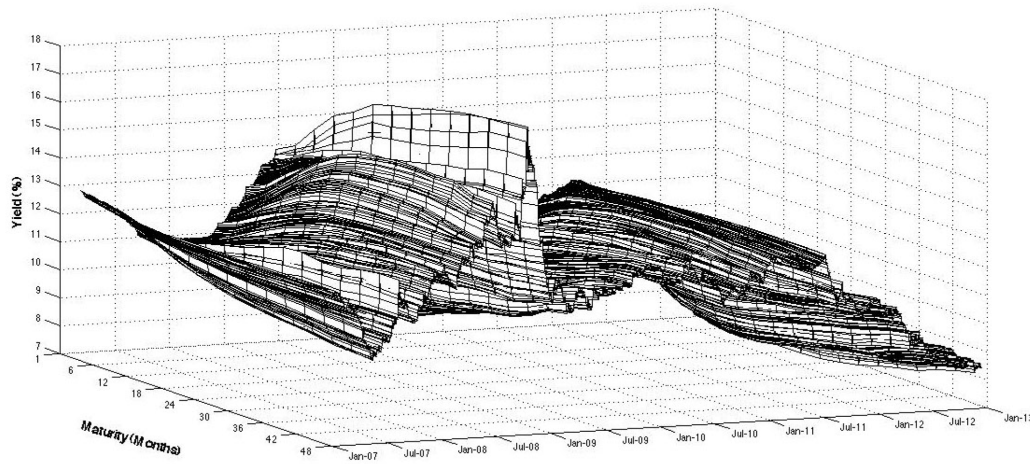
4. Empirical analysis

In this section we investigate whether the in-sample superiority of the flexible correlated-factor model carries over to out-of-sample forecast accuracy.

Estimates for the vector of factor means, μ, and the λ parameter are very similar for both the uncorrelated and correlated DNS models, as can be seen on Table 2 and on Table 3. The mean of

Figure 1: **Evolution of the yield curve**

Evolution of the term structure of interest rates (based on DI1 contracts) for the 2006:01-2012:12 period. The sample consists of the daily yields for maturities of 1, 3, 4, 6, 9, 12, 15, 18, 24, 27, 30, 36, 42, and 48 months.



the level, slope, and curvature factors are 12.56%, −2.21%, and −0.84% in the uncorrelated model, respectively, and 12.03%, −3.44%, and 0.80% for the correlated model, respectively. All figures are in accordance with observable empirical yield curves. The transition matrices reveal that the factors are highly persistent, being well above 0.9 for both estimated DNS models. The correlated transition matrix shows only very mild correlation.

Table 2: **Estimates for the uncorrelated DNS model**

Estimated uncorrelated DNS model. All numbers are in decimals. State transition matrix A is enforced to be diagonal. Corresponding error matrix Q is enforced to be diagonal and positive semi-definite (using Cholesky factorization $Q = qq'$). Vector μ contains the level, slope, and curvature factor means. The estimated λ is 1.4377, where time measured in years.

	A matrix			Means		Q matrix		
	$A_{.,1}$	$A_{.,2}$	$A_{.,3}$	μ		$q_{.,1}$	$q_{.,2}$	$q_{.,3}$
$A_{1,.}$	0.9873	0	0	0.1256	$q_{1,.}$	4.61×10^{-6}	0	0
$A_{2,.}$	0	0.9496	0	−0.0221	$q_{2,.}$	0	6.00×10^{-6}	0
$A_{3,.}$	0	0	0.9809	−0.0084	$q_{3,.}$	0	0	5.30×10^{-5}

Variances of the state variables are similar across the uncorrelated and correlated DNS models, except for the variance of shocks to the curvature parameter. The covariance estimates obtained in the correlated-factor DNS model translate into a correlation of −0.795 for innovations to the level and slope factor, a correlation of 0.330 for innovations to the level and curvature factor, and a correlation of −0.234 for innovations to the slope and curvature factor. These correlations allow interesting geometrical interpretations: when the yield curve level increases, it gets steeper, and more concave. The slope effect is higher than the curvature effect, and the correlation between slope and curvature reinforces slightly the concavity effect.

Table 3: **Estimates for the correlated DNS model**

Estimated correlated DNS model. All numbers are in decimals. Corresponding error matrix Q are enforced to be positive semi-definite (using Cholesky factorization $Q = qq'$). Vector μ contains the level, slope, and curvature factor means. The estimated λ is 1.3272, where time is measured in years.

	A matrix			Means		Q matrix		
	$A_{.,1}$	$A_{.,2}$	$A_{.,3}$	μ		$q_{.,1}$	$q_{.,2}$	$q_{.,3}$
$A_{1,.}$	0.9898	0.0318	−0.0063	0.1203	$q_{1,.}$	4.84×10^{-6}	0	0
$A_{2,.}$	0.0394	0.9945	0.0392	−0.0344	$q_{2,.}$	-4.62×10^{-6}	6.97×10^{-6}	0
$A_{3,.}$	−0.0032	−0.0610	0.9462	0.0080	$q_{3,.}$	6.38×10^{-6}	-5.43×10^{-6}	7.71×10^{-5}

Estimates of the level factor mean are noticeably higher in AFNS models when compared to both DNS formulations, while slope and curvature are of similar magnitude, as can be seen on Table 4 and on Table 5. λ is slightly higher for correlated DNS and AFNS models. The higher λ makes the slope and curvature factor loadings decay to zero faster.

Table 4: **Estimates for the uncorrelated AFNS model**

Estimated uncorrelated AFNS model. Estimated state transition matrix $\mathbf{K}$ is enforced to be diagonal. Corresponding error matrix Σ is enforced to be diagonal and positive semi-definite (using Cholesky factorization). θ is the vector of factor means (level, slope, and curvature). The associated estimated λ is 1.4039, where time is measured in years.

	$\mathbf{K}^{\mathbb{P}}$ matrix			Mean		Σ matrix		
	$\kappa^{\mathbb{P}}_{.,1}$	$\kappa^{\mathbb{P}}_{.,2}$	$\kappa^{\mathbb{P}}_{.,3}$	$\theta^{\mathbb{P}}$		$\Sigma_{.,1}$	$\Sigma_{.,2}$	$\Sigma_{.,3}$
$\kappa^{\mathbb{P}}_{1,.}$	3.1873	0	0	0.1294	$\Sigma_{1,.}$	0.0334	0	0
$\kappa^{\mathbb{P}}_{2,.}$	0	1.1755	0	−0.0233	$\Sigma_{2,.}$	0	0.0315	0
$\kappa^{\mathbb{P}}_{3,.}$	0	0	1.1027	−0.0258	$\Sigma_{3,.}$	0	0	0.0572

While λ and θ are directly comparable to the decay parameter and vector of factor means in the DNS model, the state transition and covariance matrices are modeled continuously and need to be converted to be comparable. The continuous state transition matrix ($\mathbf{K}$) and volatility matrix ($\mathbf{\Sigma}$) are converted into one-day (the time between observations) conditional matrices. Given the mean-reversion rate matrix $\mathbf{K}$, its daily counterpart can be computed as $\exp\left(-\mathbf{K}^{\mathbb{P}} \frac{1}{252}\right)$. For the volatility matrix $\mathbf{\Sigma}$, its discrete version is computed as

$$Q^{\text{AFNS}} = \int_0^{\frac{1}{252}} \exp(-\mathbf{K}^{\mathbb{P}} s) \mathbf{\Sigma} \mathbf{\Sigma}' \exp(-(\mathbf{K}^{\mathbb{P}})' s) \mathsf{d} s,$$

which is based on equation (17) with $\Delta t = \frac{1}{252}$. Results are presented on Table 6.

The same conversion procedure is used in the correlated AFNS model to ensure comparison (see Table 7). Results are similar to the ones reported for the correlated-factor DNS model. The curvature factor becomes slightly more persistent in the AFNS model, while the level becomes slightly less persistent.

The yield adjustment term is the key difference in the measurement update between the AFNS and

Table 5: **Estimates for the correlated AFNS model**

Estimated correlated AFNS model. Estimated state transition matrix $\mathbf{K}$ is fully flexible. Corresponding error matrix Σ is enforced to be lower-triangular and positive semi-definite (using Cholesky factorization). θ is the vector of factor means (level, slope, and curvature). The associated estimated λ is 1.3419, where time is measured in years.

	$\mathbf{K}^{\mathbb{P}}$ matrix			Mean		Σ matrix		
	$\kappa^{\mathbb{P}}_{.,1}$	$\kappa^{\mathbb{P}}_{.,2}$	$\kappa^{\mathbb{P}}_{.,3}$	$\theta^{\mathbb{P}}$		$\Sigma_{.,1}$	$\Sigma_{.,2}$	$\Sigma_{.,3}$
$\kappa^{\mathbb{P}}_{1,.}$	3.1408	−0.0604	0.0254	0.1418	$\Sigma_{1,.}$	0.0401	0	0
$\kappa^{\mathbb{P}}_{2,.}$	−2.6261	1.2257	−1.4691	−0.0344	$\Sigma_{2,.}$	−0.0406	0.0220	0
$\kappa^{\mathbb{P}}_{3,.}$	−0.0412	1.8640	1.9921	−0.0192	$\Sigma_{3,.}$	−0.0355	−0.0926	0.0684

Table 6: **One-day parameters for uncorrelated AFNS model**

$\mathbf{K}$ and Σ converted into daily estimates to ensure comparison with A and Q.

One-day mean reversion matrix			One-day volatility matrix		
0.9874	0	0	4.38×10^{-6}	0	0
0	0.9953	0	0	3.93×10^{-6}	0
0	0	0.9956	0	0	5.29×10^{-5}

DNS models. The adjustment terms for both AFNS models are shown in Figure 2. The yield adjustment term increases the flexibility of the curve fitting, especially at longer maturities. Furthermore the off-diagonal elements of matrix $\mathbf{\Sigma}$ are heavily influenced by the shape of the adjustment terms.

The shape of the almost linear adjustment term is not surprising for the uncorrelated AFNS model since all off-diagonals in $\mathbf{\Sigma}$ are forced to be zero, meaning there are no negative covariances to offset the decreasing function of maturity. However, the more flexible correlated AFNS model can have covariances that change the shape of the adjustment term. Nevertheless, due to the low off-diagonal for the correlated covariance matrix, the effect on the shape of the adjustment term is small. In fact, the very low yield adjustment term in both AFNS cases (highlighted by the considerably lower variances than the DNS models) implies that very little adjustment is needed to make the DNS model arbitrage-free.

To further investigate in-sample model performance, residual means for each maturity together with root-mean-square-errors (RMSE) are presented on Table 8. Based on Table 8 alone, it is very hard to rank models, but it does indicate that the simpler DNS models outperform the AFNS counterpart at least in-sample. The lowest RMSEs of 12 out of 13 maturities belong to the DNS class, out of which, eight belong to the uncorrelated DNS model. The absence of arbitrage restrictions do not seem to improve in-sample fit. Given the relatively small yield adjustment terms presented in Figure 2, the lack of improvement in fit is not very surprising.

4.1. Out-of-sample analysis

The forecasting test is performed in pseudo-real time, i.e., we never use information which is not available at the time the forecast is made. We use a rolling estimation window of 500 daily observations

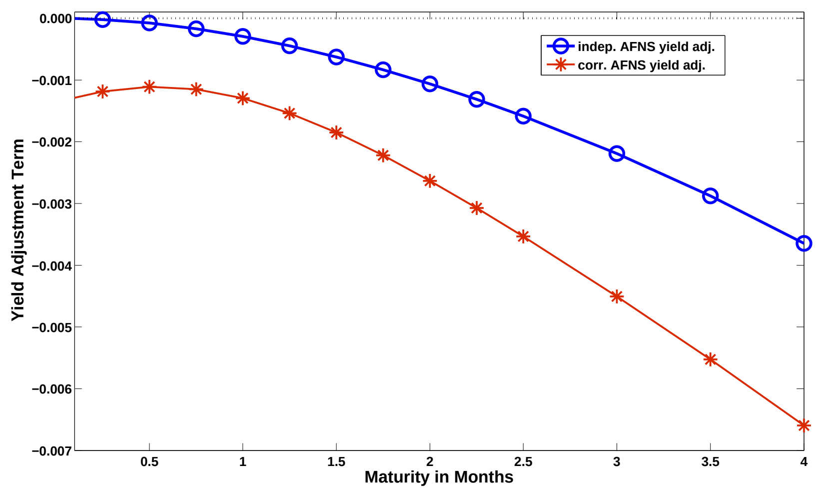
Table 7: **One-day parameters for correlated AFNS model**

K and Σ converted into daily estimates to ensure comparison with A and Q.

One-day mean reversion matrix			One-day volatility matrix		
0.9876	0.0002	-0.0001	4.38×10^{-6}	-6.39×10^{-6}	-5.57×10^{-6}
0.0103	0.9951	0.0058	-6.39×10^{-6}	8.34×10^{-6}	-2.24×10^{-6}
-0.0557	-0.0224	0.9921	-5.57×10^{-6}	-2.24×10^{-6}	5.72×10^{-5}

Figure 2: **Yield adjustment term for the AFNS models**

Yield adjustment term, $\frac{-A(\tau)}{\tau}$, in basis points. Maturity is measured in years. The solution is provided in Christensen *et al.* (2011).



(2 years). We produce forecasts for 1-week, 1-month, 3-months, and 6-months ahead. The choice of a rolling scheme is suggested by two reasons: first, it is a natural way to avoid problems of instability, see e.g. Pesaran *et al.* (2011). Second, having a fixed number of observations used to compute the forecasts, the resulting time series of forecast errors can be tested for accuracy by the Giacomini & White (2006) test. It is worth noting that we use iterated forecasts instead of direct forecasts for the multi-period ahead predictions. Marcellino *et al.* (2005) argue that iterated forecasts are more efficient when the model is correctly specified.

In order to evaluate out-of-sample forecasts, we compute popular error metrics. Given a sample of $\mathcal{M}$ out-of-sample forecasts for an h-period-ahead forecast horizon, we compute the root mean squared forecast error (RMSFE) for maturity τ_i and for model m as follows:

$$\text{RMSFE}_m(\tau_i) = \sqrt{\frac{1}{\mathcal{M}} \sum_{t=1}^{\mathcal{M}} \left(\hat{y}_{t+h|t,m}(\tau_i) - y_{t+h}(\tau_i)\right)^2} \tag{22}$$

Table 8: **Descriptive statistics of in-sample fit**

The means and the root mean squared errors for 13 different maturities. All numbers are measured in basis points.

		DNS				AFNS			
		indep-factor		dep-factor		indep-factor		dep-factor	
Maturity	τ	Mean	RMSE	Mean	RMSE	Mean	RMSE	Mean	RMSE
	3	-7.05	19.54	-4.06	18.00	-0.02	0.03	-14.64	24.28
	6	-0.01	0.01	1.07	1.07	8.38	14.77	-0.03	0.02
	9	3.18	6.71	3.56	6.74	11.01	17.28	6.52	9.50
	12	3.06	6.65	3.33	6.70	9.35	14.01	7.21	10.63
	15	1.11	4.39	1.53	4.54	5.49	8.83	4.68	7.97
Months	18	-0.29	2.31	0.36	2.32	2.19	3.94	2.07	4.09
	21	-0.94	2.41	-0.06	2.19	-0.16	0.91	0.04	0.07
	24	-0.92	3.07	0.15	2.95	-1.49	2.79	-1.3	3.13
	27	-0.56	3.22	0.68	3.24	-2.05	3.75	-1.98	4.53
	30	0.66	3.39	2.02	3.85	-1.30	3.57	-1.43	4.49
	36	1.94	3.39	3.42	4.35	0.42	2.37	-0.21	3.55
	42	0.43	2.54	1.88	3.04	1.13	2.74	0.00	0.00
	48	-2.37	5.40	-1.09	5.20	2.22	5.36	0.65	4.38

where $y_{t+h}(\tau_i)$ is the yield for the maturity τ_i observed at time $t+h$, and $\hat{y}_{t+h|t,m}(\tau_i)$ is the corresponding forecasting made at time t.

Following Hördahl *et al.* (2006) and de Pooter *et al.* (2010), we also summarize the forecasting performance of each model by computing the trace root mean squared forecast error (TRMSFE). For each forecast horizon, we compute the trace of the covariance matrix of the forecast errors across all N maturities. Hence, lower TRMSFE indicate more accurate forecasts. The TRMSFE can be computed as

$$\mathrm{TRMSFE}_m(\tau_i) = \sqrt{\frac{1}{N}\frac{1}{\mathcal{M}}\sum_{i=1}^{N}\sum_{t=1}^{\mathcal{M}}\left(\hat{y}_{t+h|t}(\tau_i) - y_{t+h}(\tau_i)\right)^2} \tag{23}$$

Finally, in order to assess the statistical significance of these forecasting differences, we use the test proposed by Giacomini & White (2006). The Giacomini & White (2006) (GW) test is a conditional forecasting ability test constructed under the assumption that forecasts are generated using a moving data window. This is a test of equal forecasting accuracy and as such can handle forecasts based on both nested and non-nested models, regardless from the estimation procedures used in the derivation of the forecasts. The test is based on the loss differential $d_{m,t} = (e_{rw,t})^2 - (e_{m,t})^2$, where e_{mt} is the forecast error of model m at time t. We assume that the loss function is quadratic but it can be replaced by other loss functions depending on the forecast goal. The null hypothesis of equal forecasting accuracy can be written as

$$\mathbb{H}_0 : E\left[d_{m,t+h}|\delta_{m,t}\right] = 0, \tag{24}$$

where $\delta_{m,t}$ is a $p \times 1$ vector of test functions or instruments and h is the forecast horizon. If a constant is used as instrument, the test can be interpreted as an unconditional test of equal forecasting accuracy. The GW test statistic $GW_{m,t}$ can be computed as the Wald statistic:

$$GW_{m,n} = n\left(n^{-1}\sum_{t=\omega+1}^{n-h}\delta_{m,t}d_{m,t+h}\right)'\hat{\Omega}_n^{-1}\left(n^{-1}\sum_{t=\omega+1}^{n-h}\delta_{m,t}d_{m,t+h}\right)\xrightarrow{d}\chi^2_{\dim(\delta)} \tag{25}$$

where $\hat{\Omega}_n$ is a consistent HAC estimator for the asymptotic variance of $\delta_{m,t}d_{m,t+h}$, and $n = (T-\omega)$ the number of out-of-sample observations. Under the null hypothesis given in (24), the test statistic $GW_{i,t}$ is asymptotically distributed as χ^2_p. The asterisks on Tables 9 and 10 denote significant differences in out-of-sample model performance relative to RW at the 5, and 10 percent levels.

Tables 9 and 10 report the RMSFE and TRMSFE for each of the forecasting models and for each of the forecasting horizons. The first column in each panel of each Table reports the value of RMSFE and TRMSFE (expressed in basis points) for the random walk model (RW), while all other columns report statistics relative to the RW. The following model abbreviations are used on Tables 9 and 10: AR(1) for the first-order univariate autoregressive model, VAR(1) for the first-order vector autoregressive model, DNS_{AR} for the dynamic Nelson-Siegel model with a (V)AR specification for the factors, and AFNS for the dynamic Arbitrage-Free Nelson Siegel model with a (V)AR specification for the factors. First, the RW confirms to be a very competitive benchmark in forecasting the term structure of bond yields, especially at short horizons.

For each forecasting horizon considered, we highlight the most accurate model, in terms of RMSFE, for each of the 13 maturities analyzed. Bold values indicate relative differences below one, which means that a particular model outperforms the random walk. Stars denote rejection of the null of equal forecasting ability according to the GW test at 5% and 1% level.

Comparing correlated to uncorrelated factor models, we notice that the most parsimonious specification performs better for AFNS but not for DNS. For the latter, VAR parameterization of the factor dynamics achieves better forecasting results. As pointed out, the DNS model outperformed the AFNS model in-sample in terms of RMSE. When analyzing out-of-sample performance, we observe that the DNS model is superior to the other competitors in almost all maturities when short-horizon forecasts are considered (1-week and 1-month). These results are in line with Duffee (2002), who conclude that affine term structure models can have very poor performance out-of-sample.

However, for longer forecasting horizons (3 to 6-months ahead), the AFNS model outperforms all competitors for all maturities. Nevertheless, DNS still outperforms RW, showing its good forecasting capabilities. What is striking here is that the arbitrage-free restrictions seem to improve a lot the long run forecasts, since most of the differences in RMSFE are statistically significant. Moreover, the uncorrelated AFNS version is the best performing model for almost all maturities, except for short maturities, when it is beaten by the correlated AFNS model.

For the 1-week ahead forecasts, the best performing model overall in terms of T-RMSFE is the RW model. However, for 1-month ahead forecasts the correlated factors DNS model performs best, while for the longer forecast horizons of 3- and 6-months ahead the best performing model in terms of T-RMSFE was the AFNS, evidencing the forecast benefits of imposing the arbitrage-free restriction.

Table 9: **Out-of-sample Yield Forecasts : [T] RMSFEs**

Note: This table summarizes Relative Mean Squared Forecast Errors and Trace Root Mean Squared Forecast Error (TRMPSE) relative to the Random Walk obtained by using each of the competing models, for the horizons 1-week, 1-, 3-, and 6-month-ahead. The first column in the table reports the value of (T) RMSFE (expressed in basis points) for the Random Walk model (RW), while all other columns report statistics relative to the RW. The following model abbreviations are used in the table: RW stands for the Random Walk, (V)AR for the first-order (Vector) Autoregressive Model, DNS for the one-step dynamic Nelson-Siegel model with a (V)AR specification for the factors, AFNS refers to the one-step Arbitrage-Free Nelson Siegel model with a (V)AR specification for the factors. Numbers smaller than one (shown in bold) indicate that models outperform the random walk, whereas numbers larger than one indicate underperformance. The gray box indicate outperformance in the maturity. The stars on the right of the cell entries signal the level at which the Giacomini & White (2006) test rejects the null of equal forecasting accuracy (* and **mean respectively rejection at 1% and 5% level).

Maturity	1-week-ahead forecasts							1-month-ahead forecasts						
	RW	AR(1)	VAR(1)	DNS_{AR}	DNS_{VAR}	$AFNS_{AR}$	$AFNS_{VAR}$	RW	AR(1)	VAR(1)	DNS_{AR}	DNS_{VAR}	$AFNS_{AR}$	$AFNS_{VAR}$
3	0.120	**0.986**	**0.764***	**0.990****	2.211	1.792	1.494	0.409	**0.994**	**0.573***	**0.995****	1.113	1.192	**0.576***
6	0.141	**0.994**	**0.928**	1.384	**0.961**	1.149	**0.858***	0.430	**0.997**	**0.749****	1.008	**0.960****	1.162	**0.693***
9	0.154	**0.999**	1.015	1.430	1.120	1.222	1.001	0.442	1.001	**0.858**	1.017	**0.955**	1.164	**0.845**
12	0.164	1.003	1.062	1.269	1.152	1.194	1.033	0.445	1.006	**0.963**	1.021	**0.964**	1.166	**0.956**
15	0.174	1.006	1.105	1.104	1.090	1.126	1.031	0.455	1.012	1.040	1.011	**0.958**	1.146	1.021
18	0.181	1.008	1.130	1.023	1.045	1.083	1.033	0.459	1.017	1.101	1.006	**0.954**	1.130	1.064
21	0.183	1.011	1.152	1.005	1.045	1.081	1.052	0.458	1.023	1.139	1.002	**0.953**	1.128	1.101
24	0.184	1.013	1.162	**0.993**	1.067	1.079	1.032	0.453	1.029	1.162	**0.990**	**0.948**	1.128	1.104
27	0.187	1.013	1.159	**0.988**	1.083	1.074	1.030	0.453	1.031	1.158	**0.977****	**0.942**	1.120	1.100
30	0.188	1.014	1.160	**0.993**	1.060	1.080	1.053	0.452	1.036	1.159	**0.980***	**0.944**	1.119	1.109
36	0.188	1.016	1.162	1.004	1.035	1.085	1.072	0.437	1.045	1.198	1.005	**0.969**	1.134	1.129
42	0.189	1.018	1.169	1.010	1.040	1.053	1.039	0.420	1.056	1.230	1.009	**0.978**	1.125	1.107
48	0.192	1.019	1.166	1.044	1.058	1.042	1.054	0.417	1.061	1.230	1.011	**0.983**	1.092	1.077
T-RMSFE	0.170	1.010	1.116	1.086	1.127	1.132	1.052	0.440	1.024	1.065	1.002	**0.969**	1.139	1.011

Table 10: **Out-of-sample Yield Forecasts : [T] RMSFEs**

Note: This table summarizes Relative Mean Squared Forecast Errors and Trace Root Mean Squared Forecast Error (TRMPSE) relative to the Random Walk obtained by using each of the competing models, for the horizons 1-week, 1-, 3-, and 6-month-ahead. The first column in the table reports the value of (T) RMSFE (expressed in basis points) for the Random Walk model (RW), while all other columns report statistics relative to the RW. The following model abbreviations are used in the table: RW stands for the Random Walk, (V)AR for the first-order (Vector) Autoregressive Model, DNS for the one-step dynamic Nelson-Siegel model with a (V)AR specification for the factors, AFNS refers to the one-step Arbitrage-Free Nelson Siegel model with a (V)AR specification for the factors. Numbers smaller than one (shown in bold) indicate that models outperform the random walk, whereas numbers larger than one indicate underperformance. The gray box indicate outperformance in the maturity. The stars on the right of the cell entries signal the level at which the Giacomini & White (2006) test rejects the null of equal forecasting accuracy (* and **mean respectively rejection at 1% and 5% level).

Maturity	3-month-ahead forecasts							6-month-ahead forecasts						
	RW	AR(1)	VAR(1)	DNS_{AR}	DNS_{VAR}	AFNS_{AR}	AFNS_{VAR}	RW	AR(1)	VAR(1)	DNS_{AR}	DNS_{VAR}	AFNS_{AR}	AFNS_{VAR}
3	1.134	1.077	**0.727***	**0.998***	1.020	**0.970***	**0.546***	2.023	1.252	1.117*	**0.999***	1.113	**0.758***	**0.583***
6	1.192	1.085	**0.950****	**0.980**	**0.981***	**0.910**	**0.705***	2.092	1.325	1.380	**0.984**	**0.960****	**0.710***	**0.697***
9	1.234	1.078	1.119	**0.981**	**0.967***	**0.881**	**0.806***	2.122	1.289	1.609	**0.987**	**0.955**	**0.699***	**0.772***
12	1.248	1.071	1.254	**0.991**	**0.967***	**0.873**	**0.875***	2.110	1.233	1.798	**0.993**	**0.964**	**0.703***	**0.825***
15	1.252	1.067	1.358	**0.996**	**0.966***	**0.869**	**0.919***	2.074	1.196	1.969	**0.994**	**0.958**	**0.709***	**0.859***
18	1.236	1.066	1.437	1.003	**0.969***	**0.871**	**0.953***	2.004	1.163	2.114	**0.999**	**0.954**	**0.721**	**0.892***
21	1.221	1.071	1.495	1.001	**0.966***	**0.867**	**0.975***	1.930	1.144	2.252	1.000	**0.953**	**0.733**	**0.921****
24	1.200	1.085	1.540	**0.997**	**0.962***	**0.867**	**0.992***	1.856	1.154	2.377	**0.999**	**0.948**	**0.744**	**0.947****
27	1.196	1.105	1.566	**0.985****	**0.952***	**0.859****	**0.996***	1.806	1.182	2.485	**0.992****	**0.942**	**0.749****	**0.936***
30	1.190	1.124	1.597	**0.985****	**0.952***	**0.857****	1.006	1.760	1.206	2.587	**0.990***	**0.944**	**0.755***	**0.928***
36	1.154	1.144	1.669	1.001	**0.969***	**0.862**	1.013	1.641	1.233	2.797	**0.999**	**0.969**	**0.774**	**0.919****
42	1.106	1.176	1.709	1.007	**0.976****	**0.859**	**0.981****	1.534	1.284	2.949	1.005	**0.978**	**0.785**	**0.905****
48	1.084	1.205	1.724	1.003	**0.973***	**0.839**	**0.932***	1.477	1.334	3.054	1.001	**0.983**	**0.778**	**0.876****
T-RMSFE	1.190	1.102	1.418	**0.994**	**0.970**	**0.876**	**0.910**	1.890	1.230	2.151	**0.995**	**0.972**	**0.735**	**0.846**

5. Concluding remarks

The dynamic version of the Nelson-Siegel model has been shown in the literature to be remarkably well suited both to fit and to forecast the term structure of interest rates. More recently, Christensen *et al.* (2011) have developed an arbitrage-free version of this model in order to bring theoretical rigor to a empirically successful model. In this paper, we have analyzed the forecasting power of no-arbitrage restrictions on the dynamic Nelson-Siegel model in order to determine the empirical relevance of this new theoretical restriction.

Different specification for the factor dynamics are allowed in both versions of the DNS model. Their forecasts are compared on an yield to yield basis using the root mean squared forecast error, and their overall performance is measured via the trace root mean squared error. Results indicate that, for longer horizons (3- and 6-months ahead),the AFNS outperforms not only the DNS model, but also all other benchmarks considered. However, arbitrage-free restrictions do not seem to improve short horizon forecasts (1-weak and 1-month ahead).

References

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

Björk, Tomas, & Christensen, Bent Jesper. 1999. Interest rate dynamics and consistent forward rate curves. *Mathematical Finance*, **9**(4), 323–348.

Carriero, Andrea. 2011. Forecasting the yield curve using priors from no-arbitrage affine term structure models. *International Economic Review*, **52**(2), 425–459.

Carriero, Andrea, & Giacomini, Raffaella. 2011. How useful are no-arbitrage restrictions for forecasting the term structure of interest rates? *Journal of Econometrics*, **164**(1), 21–34.

Christensen, J. H. E., Diebold, F. X., & Rudebusch, G. D. 2009. An arbitrage-free generalized Nelson-Siegel term structure model. *Econometrics Journal*, **12**, 33–64.

Christensen, Jens H.E., Diebold, Francis X., & Rudebusch, Glenn D. 2011. The affine arbitrage-free class of Nelson-Siegel term structure models. *Journal of Econometrics*, **164**(1), 4–20.

Coroneo, Laura, Nyholm, Ken, & Vidova-Koleva, Rositsa. 2011. How arbitrage-free is the Nelson-Siegel model? *Journal of Empirical Finance*, **18**(3), 393–407.

de Pooter, Michiel, Ravazzolo, Francesco, & van Dijk, Dick. 2010 (Mar.). *Term structure forecasting using macro factors and forecast combination*. Working Paper 2010/01. Norges Bank.

Diebold, F., & Li, C. 2006. Forecasting the term structure of government bond yields. *Journal of Econometrics*, **130**(2), 337–364.

Diebold, Francis X., & Rudebusch, Glen D. 2013. *The Dynamic Nelson-Siegel Approach to Yield Curve Modeling and Forecasting*. Princeton University Press.

Duffee, G. R. 2002. Term premia and interest rate forecasts in affine models. *Journal of Finance*, **57**(1), 405–443.

Duffee, Gregory R. 2011. Forecasting with the term structure: The role of no-arbitrage restrictions. *Working paper, Johns Hopkins University, Department of Economics.*

Duffie, Darrell, & Kan, Rui. 1996. A Yield-Factor Model Of Interest Rates. *Mathematical Finance*, **6**(4), 379–406.

Favero, Carlo A., Niu, Linlin, & Sala, Luca. 2012. Term Structure Forecasting: No-Arbitrage Restrictions versus Large Information Set. *Journal of Forecasting*, **31**(2), 124–156.

Filipovic, Damir. 1999. A Note on the Nelson–Siegel Family. *Mathematical Finance*, **9**(4), 349–359.

Giacomini, Raffaella, & White, Halbert. 2006. Tests of Conditional Predictive Ability. *Econometrica*, **74**(6), 1545–1578.

Gimeno, Ricardo, & Marqués, J Manuel. 2009. Extraction of financial market expectations about inflation and interest rates from a liquid market. *Working Paper 0906, Banco de España.*

Golub, Gene H., & Van Loan, Charles F. 1996. *Matrix Computations*. 3rd edn. Johns Hopkins University Press.

Hördahl, P., Tristani, O., & Vestin, D. 2006. A joint econometric model of macroeconomic and term-structure dynamics. *Journal of Econometrics*, **131**(1-2), 405–444.

Jungbacker, Borus, & Koopman, Siem Jan. 2014. Likelihood-based dynamic factor analysis for measurement and forecasting. *The Econometrics Journal*, *forthcoming.*

Lang, Serge. 1987. *Linear Algebra*. 3rd edn. Springer.

Marcellino, M., Stock, J., & Watson, M. 2005. A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series. *Journal of Econometrics*, **135**(1-2), 499–526.

Mönch, Emmanuel. 2008. Forecasting the yield curve in a data-rich environment: a no-arbitrage factor-augmented VAR approach. *Journal of Econometrics*, **146**(1), 26–43.

Nelson, Charles R., & Siegel, Andrew F. 1987. Parsimonious modeling of yield curves. *The Journal of Business*, **60**(4), 473–489.

Nyholm, Ken, & Vidova-Koleva, Rositsa. 2012. Nelson-Siegel, Affine and Quadratic Yield Curve Specifications: Which One is Better at Forecasting? *Journal of Forecasting*, **31**(6), 540–564.

Pesaran, M. Hashem, Pick, Andreas, & Timmermann, Allan. 2011. Variable selection, estimation and inference for multi-period forecasting problems. *Journal of Econometrics*, **164**(1), 173–187.

Singleton, K.J. 2006. *Empirical Dynamic Asset Pricing: Model Specification and Econometric Assessment*. Princeton University Press.

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