A Dynamic Functional Factor Model for Yield Curves: Identification, Estimation, and Prediction*

Sven Otto**
University of Cologne

Nazarii Salish***
University Carlos III de Madrid

July 5, 2019

Abstract

The problem of yield curve forecasting from a functional time series perspective is discussed. A functional factor model is considered, in which the factors follow some linear autoregressive process. The model is identified by imposing suitable conditions on the factors and the loading functions. By applying the least squares principle, a functional principal components based estimator is obtained, which is shown to be consistent. The minimum mean squared error forecast from the dynamic functional factor model is considered, and pointwise and simultaneous prediction bands are derived. Finally, the accuracy of the predictions and prediction bands is discussed in an out-of-sample experiment with monthly yield curves of U.S. Treasuries.

Keywords: Factor model, Functional principal components, Prediction band, Yield curve

JEL Classification: C30, C58, E43

*We are thankful to Jörg Breitung and Malte Knüppel for very helpful comments and suggestions. Further, we would like to thank the participants attending the Verein für Socialpolitik Jahrestagung 2018 in Freiburg, the Statistische Woche 2018 in Linz and the $(EC)^2$ Conference 2018 in Rome. A previous version of this paper circulated under the name "Prediction Bands for Yield Curves: A Dynamic Functional Factor Model Approach."

**Corresponding author: University of Cologne, Institute of Econometrics and Statistics, Albertus-Magnus-Platz, 50923 Köln, Germany. Tel.: +49-221-470-6186. Mail: sven.otto@statistik.uni-koeln.de.

***Universidad Carlos III de Madrid, Department of Economics, Calle Madrid 126, 28903 Getafe (Madrid), Spain.

1 Introduction

The number of factors that explain the shape of the term structure of bond yields is typically much smaller than the number of available maturities. According to Diebold et al. (2005), yields of the same bond with different maturities are driven by only a few sources of risk. Factor models thus received much attention in the literature on bond yield modeling, as they allow a high dimensional problem to be reduced to a lower dimensional one.

In the classical factor analysis, both the latent factors and the factor loadings are estimated using principal components, while the model is identified by imposing orthogonality conditions (see, e.g., Anderson 2003). Litterman and Scheinkman (1991), Ang et al. (2006), and Joslin et al. (2014) conducted a factor analysis on the yields of U.S. Treasuries with different maturities and demonstrated that the first three principal components explain more than 98% of the variation in the yields. Bliss (1997) proposed a dynamic version of the classical factor model for bond yields, where the first three principal components follow an autoregressive process. For surveys of classical and more recent results in factor analysis, see the review articles by Stock and Watson (2011) and Breitung and Choi (2013).

Another approach that has gained wide acceptance in practice is based on the Nelson-Siegel term structure model (see Nelson and Siegel 1987), where the underlying yield curve is assumed to follow a certain parametric functional structure. The standard version of this model considers three predefined loading functions, which are denoted as the level, the slope, and the curvature function. The corresponding factors are interpreted as the long-term, short-term, and medium-term factors. Extensions and alternative factor loadings were proposed in Svensson (1995), Gürkaynak et al. (2007), Christensen et al. (2009), and Park et al. (2009). The dynamic Nelson-Siegel (DNS) model, which was introduced by Diebold and Li (2006), considers common factor dynamics in the form of a vector autoregressive process. In a two-step procedure, the factors are first estimated by ordinary least squares, and the dynamic process is then estimated given the fitted factor values from the first step. For a review on the DNS approach, see Diebold and Rudebusch (2013).

The dynamic versions of both the classical factor model and the Nelson-Siegel model are particularly useful for predicting the term structure of bond yields, since a predictor for the bond yields can be obtained from the minimum mean squared error (MSE) forecast of the factors. However, both approaches have their advantages and disadvantages. While the classical factor analysis does not depend on a specific form of the loadings and provides an optimal basis system, it considers yield curves as vector-valued objects and does not take into account the functional nature of the underlying curve. On the other hand, while the Nelson-Siegel model considers smooth loading functions and generates a fully

functional representation of the yield curve, its predefined loadings are not optimal in some sense, as argued, for instance, in Lengwiler and Lenz (2010) and Hays et al. (2012).

In this paper, we follow recent advances in functional data analysis (for reviews, see, e.g., Ramsay and Silverman 2005, Hörmann and Kokoszka 2012, and Hsing and Eubank 2015). We consider a dynamic functional factor model, which is akin to the models proposed in Hays et al. (2012) and Bardsley et al. (2017). In contrast to Hays et al. (2012), where the estimation is achieved by applying an expectation maximization algorithm, we suggest a different estimation procedure and identify the model by imposing orthogonality conditions on the loading functions and suitable assumptions on the factors and the error term. We propose a functional principal component (FPC) estimator, which results from applying the least squares principle. It is shown that the model parameters are estimated consistently. Following Aue et al. (2015), the factors are modeled as an autoregressive process. Moreover, in the fashion of Stock and Watson (2002), forecasts are conducted in a two-step procedure, where the predictive model is estimated given the estimated factors. Analogously to Diebold and Li (2006), a forecast for the entire yield curve is obtained from the estimated functional factor model and the optimal MSE forecast for the factors. Furthermore, we derive both pointwise and simultaneous prediction bands for these forecasts. While the pointwise bands are asymptotically exact, the simultaneous bands are conservative.

The paper is organized as follows: In Section 2, the functional factor model and its identification conditions are defined, while in Section 3, the FPC estimator is derived, and consistency is shown. Predictions for the entire yield curve as well as pointwise and simultaneous prediction bands are provided in Section 4. Then, in Section 5, the predictor is applied to a dataset of monthly U.S. Treasury yields, and Section 6 concludes.

2 The dynamic functional factor model

We consider a time series of yield curves $Y_t(r)$ with time to maturity $r \in [a, b]$ at time $t = 1, \ldots, T$, where a denotes the lowest time to maturity, and b denotes the longest one. The curves are assumed to be already given as square-integrable functions on the domain $[a, b]$. Following Hays et al. (2012), a general dynamic functional factor model with K factors can be formulated as

$$Y_t(r) = \mu(r) + \sum_{l=1}^{K} F_{l,t}\psi_l(r) + \epsilon_t(r), \quad t = 1, \ldots, T, \quad r \in [a, b],$$

where $\mu(r)$ is an intercept function, $F_{l,t}$ denotes the l-th factor at time t, and $\psi_l(r)$ is the l-th loading for time to maturity r. While $\mu(r)$ and $\psi_l(r)$ are deterministic terms, the vector of factors $F_t = (F_{1,t}, \ldots, F_{K,t})'$ is assumed to be a time series that follows some

linear autoregressive process, which stems the attribute "dynamic" in the name of the model, and $\epsilon_t(r)$ is an idiosyncratic error term. In vector notation, the model can be equivalently expressed as

$$Y_t(r) = \mu(r) + \Psi'(r)F_t + \epsilon_t(r), \quad t = 1, \dots, T, \quad r \in [a, b], \tag{1}$$

where the loadings at time to maturity r are given by $\Psi(r) = (\psi_1(r), \dots, \psi_K(r))'$.

While model (1) can be viewed as an extension of the vector-valued dynamic factor model for bond yields by Bliss (1997) to a continuous domain, it also nests the widely used DNS framework, in which the loadings $\Psi(r)$ are predefined. Diebold and Li (2006) specified the Nelson-Siegel model with three fixed loading functions given by

$$\psi_1(r) = 1, \quad \psi_2(r) = \frac{1 - e^{-\lambda r}}{\lambda r}, \quad \psi_3(r) = \frac{1 - e^{-\lambda r}}{\lambda r} - e^{-\lambda r}, \tag{2}$$

where the decay parameter is fixed, and the intercept is zero. The Nelson-Siegel loadings are illustrated in Figure 1, where the first loading $\psi_1(r)$ represents the level of the yield curve at time t, the function $\psi_2(r)$ corresponds to the slope and the third function $\psi_3(r)$ has a hump at mid-term maturities and represents the curvature of the yield curve. A plot of the yield curve data from Jungbacker et al. (2014) is depicted in Figure 2, and Figure 3 presents an example of a fitted Nelson-Siegel yield curve.

In this paper, we neither impose restrictive assumptions on the shape of the loading functions, as in the DNS framework, nor do we treat yield curves as vectors, as in the classical factor analysis. We consider model (1) to be a fully functional model and assume that both the factors F_t and loadings $\Psi(r)$ are unknown. Analogously to the vector-valued factor model, this implies an ambiguity, since $\Psi'(r)F_t = \Psi'(r)QQ^{-1}F_t$ for any non-singular $K \times K$ matrix Q. Thus, K^2 restrictions are necessary to make $\Psi(r)$ and F_t separately identifiable. Assuming that the loadings are pairwise orthogonal imposes $K(K-1)/2$ restrictions, and another $K(K-1)/2$ restrictions follow by assuming that the factors are pairwise uncorrelated. The remaining K restrictions can be obtained by normalizing either the loadings or the factors. Since the consideration of orthonormal loading functions is common in functional data analysis, we normalize the loadings. Another ambiguity follows from the fact that $\mu(r) + \Psi'(r)F_t = \mu(r) + \Psi'(r)F_0 + \Psi'(r)(F_t - F_0)$ for any $F_0 \in \mathbb{R}^K$, and K restrictions are obtained by setting the mean of the factors to zero.

To formalize these restrictions, we define a Hilbert space structure for the underlying function space $H = L^2([a, b])$, which denotes the space of real-valued square integrable functions on the domain $[a, b]$. For $x, y \in H$, the inner product is defined as $\langle x, y \rangle = \int_a^b x(r)y(r)\,\mathrm{d}r$, and the norm is given by $\|x\|^2 = \langle x, x \rangle$. The following identification conditions are imposed:

Assumption 1. *(a) The loading functions are pairwise orthogonal; i.e.,* $\langle \psi_k, \psi_l \rangle = 0$, *for all* $k, l = 1, \ldots, K$ *with* $k \neq l$.

(b) The loading functions are normalized, so that $\|\psi_l\| = 1$ *for all* $l = 1, \ldots, K$.

(c) The factors satisfy $F_t \sim \mathcal{N}(0, \Sigma_F)$ *for all* $t = 1, \ldots, T$, *where* $\Sigma_F = diag(\lambda_1, \ldots, \lambda_K)$ *with distinct entries* $\lambda_1 > \lambda_2 > \ldots > \lambda_K > 0$.

The assumptions are similar to the identification conditions for the classical factor model (see Stock and Watson 2002 and Bai and Ng 2013). The loadings are assumed to be pairwise orthonormal, while the factors are uncorrelated. Distinct variances of the factors serve to identify the individual entries of F_t, which could otherwise be consistently estimated only up to an orthonormal transformation. Note that Assumption 1 is a local identification condition, since $\Psi(r)$ and F_t are identified separately only up to a sign change. Changing the sign of both the loadings and the factors will leave the common component $\Psi'(r)F_t$ unchanged. A global identification condition can be obtained by fixing the sign for either the loadings or the factors.

Assumption 2. *(a) The functions* $\mu(r)$ *and* $\psi_l(r)$, $l = 1, \ldots, K$, *are bounded.*

(b) The errors $\{\epsilon_t\}_{t \in \mathbb{N}}$ *are an i.i.d. sequence of zero mean Gaussian random functions on the domain* $[a, b]$ *with covariance kernel* $c_\epsilon(r, s)$.

For each $r, s \in [a, b]$, the error term is normally distributed with $E[\epsilon_t(r)] = 0$ and $Cov[\epsilon_t(r), \epsilon_t(s)] = c_\epsilon(r, s)$. Note that Hays et al. (2012) proposed normality conditions on the factors and errors similar to those made in Assumptions 1 and 2. In contrast to their estimation procedure, we impose normality only to derive the prediction bands. For the estimation of the model and the forecast of the yield curve itself, normality of the errors and the factors is not necessary.

Moreover, Assumptions 1(c), 2(a), and 2(b) imply that $\{Y_t\}_{t \in \mathbb{N}}$ is a sequence of H-valued random functions with $E\|Y_t\|^2 \leq \|\mu\|^2 + E\|\epsilon_t\|^2 < \infty$. It follows that $Y_t(r)$ is covariance stationary with mean $E[Y_t(r)] = \mu(r) < \infty$, and covariance kernel

$$c_Y(r, s) = Cov[Y_t(r), Y_t(s)] = \sum_{l=1}^{K} \lambda_l \psi_l(r) \psi_l(s) + c_\epsilon(r, s) < \infty$$

for all t. The covariance operator of $Y_t(r)$ is given by

$$C_Y(x)(r) = \int_a^b c_Y(r, s) x(s) \, \mathrm{d}s, \quad x \in H, \ r \in [a, b],$$

the eigenequation is defined as

$$\int_a^b c_Y(r,s)\psi(s)\,\mathrm{d}s = \lambda\psi(s), \quad \lambda \in \mathbb{R},\ \psi \in H,\ r \in [a,b],$$

and $(\lambda, \psi(s))$ is called an eigenvalue and eigenfunction pair of C_Y. The covariance operator of $\epsilon_t(r)$ is analogously defined as

$$C_\epsilon(x)(r) = \int_a^b c_\epsilon(r,s)x(s)\,\mathrm{d}s, \quad x \in H,\ r \in [a,b].$$

Since C_Y and C_ϵ are positive semi-definite and symmetric (see Hörmann and Kokoszka 2012), all eigenvalues are nonnegative, and the eigenfunctions are orthogonal. An orthonormal eigenfunction that corresponds to the l-th largest eigenvalue of C_Y is called the l-th FPC. We next establish a link between the functional factor model (1) and the FPCs.

Assumption 3. *(a)* $\langle \epsilon_t, \psi_l \rangle = 0$ *for all* $l = 1, \ldots, K$.

(b) $E\|\epsilon_t\|^2 < \lambda_K$.

(c) C_ϵ *has* $L - K$ *nonzero eigenvalues, which are denoted as* $\lambda_{K+1} > \lambda_{K+2} > \ldots > \lambda_L$.

Assumption 3 is similar to the conditions made in Forni et al. (2000) for the dynamic vector-valued factor model. The first assumption implies that the idiosyncratic component $\epsilon_t(r)$ is orthogonal to the common component $\Psi'(r)F_t$, which separates H into a factor space and its orthogonal complement, while the second assumption guarantees some limited amount of cross-correlation in the idiosyncratic component. The covariances of $\epsilon_t(r)$ are bounded in the sense that there is no direction in H along which the idiosyncratic component has a variance that exceeds $Var[F_{K,t}] = \lambda_K$. The third assumption implies that the error term is an element of an $(L-K)$-dimensional subspace of H, and the yield curve series consequently takes values in an L-dimensional subspace of H, where $K \leq L < \infty$.

Next, we show that under Assumptions 1–3, the variances of the factors $\lambda_1, \ldots, \lambda_K$ coincide with the K largest eigenvalues of C_Y, while the factor loadings $\psi_1(r), \ldots, \psi_K(r)$ are the first K FPCs. Furthermore, the factors are equal to the projections of the demeaned yield curve onto the corresponding FPC. This relation is summarized as follows:

Lemma 1. *Let Assumptions 1–3 hold true.*

(a) The largest K *eigenvalues of* C_Y *are given by* $\lambda_1, \ldots, \lambda_K$*, while* $\psi_1(r), \ldots, \psi_K(r)$ *are corresponding eigenfunctions.*

(b) The factors satisfy $F_{l,t} = \langle Y_t - \mu, \psi_l \rangle$ *for all* $l = 1, \ldots, K$ *and* $t = 1, \ldots, T$.

The eigenvalues of C_Y are given by $\lambda_1 > \lambda_2 > \ldots > \lambda_L > 0$, and the FPCs are the corresponding orthonormal eigenfunctions $\psi_1(r), \ldots, \psi_L(r)$, which are identified up to a sign change. The projection $\langle Y_t - \mu, \psi_l \rangle$ is also called the l-th FPC score of $Y_t(r)$ and describes the contribution of the l-th FPC to the curve $Y_t(r)$. While Aue et al. (2015) imposed a vector autoregressive structure on the empirical counterpart of the FPC scores, we incorporate this assumption into the model (1) and assume that the factors follow a VAR process. This dynamic structure is exploited in Section 4 to obtain predictions of the yield curve and to construct prediction bands.

Assumption 4. *The factors follow the VAR(p) process* $F_t = A(L)F_{t-1} + \eta_t$ *with* $K \times K$ *lag polynomial matrices* $A(L) = A_1 + A_2 L + \ldots + A_p L^{p-1}$. *The VAR process is stable; i.e., all roots of the characteristic polynomial lie outside the unit circle. The innovation vectors* $\{\eta_t\}_{t \in \mathbb{N}}$ *are i.i.d. with* $\eta_t = (\eta_{1,t}, \ldots, \eta_{K,t})' \sim \mathcal{N}(0, \Sigma_\eta)$, *where* $\Sigma_\eta = diag(\sigma_1^2, \ldots, \sigma_K^2)$. *Furthermore,* $E[\eta_{l,t}\epsilon_{t+j}(r)] = 0$ *for all* $t \in \mathbb{N}$, $j \geq 0$, $l = 1, \ldots, K$, *and* $r \in [a, b]$.

Since the factors have zero mean and are uncorrelated, the VAR model does not include a constant, and the entries of the innovation vector are uncorrelated. Although the contemporaneous correlations of the factors are zero by Assumption 1(c), the cross-correlations of the factors might be nonzero, and the lag polynomial matrices are therefore not necessarily diagonal. However, Diebold and Li (2006) argued that VAR models for bond yield factors tend to produce poor forecasts compared to those from univariate autoregressive models, since a model with a large number of parameters is prone to overfitting. Hyndman and Ullah (2007) also argued that a simple univariate autoregressive model for each FPC score is more adequate in the forecasting context, since contemporaneous correlations of the FPC scores are zero, and the cross-correlations at non-zero lags are thus often quite small. Therefore, we discuss both multivariate and univariate predictive time series models for the factors in Section 4.

Remark 1. Throughout the theoretical part of the paper, we assume that the yield curves are already given in functional form as elements of H. In practice, however, only vectors of bond yields with a finite number N of times to maturity are observed. The problem of transforming the vector observations into functions has been extensively studied in the literature on functional data analysis and is well understood (see, e.g., Ramsay and Silverman 2005 for a review of the available techniques). The data is interpolated or approximated by means of a basis expansion, which is typically performed using a spline basis in the case of non-periodic data. In the empirical part of the paper, we employ B-spline techniques to reconstruct yield curves (see Figure 3). Since only a finite number of basis functions are necessary to interpolate N observations, the reconstructed yield curve observations are in practice elements of a finite-dimensional subspace of H. More details on this topic are provided in Section 5.1.

3 The functional principal components estimator

While in the classical factor analysis the principal components estimator is obtained as the solution to some least squares problem, an analogous least squares problem arises for the functional factor model (1). Let

$$\widehat{\mu}(r) = \frac{1}{T}\sum_{t=1}^{T} Y_t(r), \quad r \in [a,b],$$

denote the empirical mean function of $Y_t(r)$. We consider the demeaned curve $Y_t(r) - \widehat{\mu}(r)$ and minimize the objective function

$$\sum_{t=1}^{T} \left\| Y_t - \widehat{\mu} - \sum_{l=1}^{K} F_{l,t}\psi_l \right\|^2, \tag{3}$$

subject to the orthonormality conditions given by Assumptions 1(a) and 1(b). Taking the loading functions as given, we minimize (3) with respect to the factors $F_{l,t}$ for all $l = 1, \dots, K$ and $t = 1, \dots, T$. Note that addition and multiplication for functional objects are defined pointwise, so that $(Y_t - \widehat{\mu})(r) = Y_T(r) - \widehat{\mu}(r)$ and $(F_{l,t}\psi_l)(r) = F_{l,t} \cdot \psi_l(r)$ for all $r \in [a,b]$.

Lemma 2. *Under Assumptiona 1(a) and 1(b),*

$$\min_{F_{lt}} \sum_{s=1}^{T} \left\| Y_s - \widehat{\mu} - \sum_{k=1}^{K} F_{k,s}\psi_k \right\|^2 \tag{4}$$

is reached, if $F_{l,t} = \langle Y_t - \widehat{\mu}, \psi_l \rangle$ *for any* $l = 1, \dots, K$ *and* $t = 1, \dots, T$.

The objective function is minimized by setting $F_{l,t}$ equal to the projection of the demeaned yield curve at time t onto the l-th loading function. The least squares problem is consequently equivalent to minimizing

$$\sum_{t=1}^{T} \left\| Y_t - \widehat{\mu} - \sum_{l=1}^{K} \langle Y_t - \widehat{\mu}, \psi_l \rangle \psi_l \right\|^2 \tag{5}$$

with respect to $\psi_1(r), \dots, \psi_K(r)$. Hence, we have to find a K-dimensional orthonormal system in H that minimizes (5), which is a well known problem in the literature on functional data analysis and extensively studied in Hörmann and Kokoszka (2012). Let

$$\widehat{c}_Y(r,s) = \frac{1}{T}\sum_{t=1}^{T} \big(Y_t(r) - \widehat{\mu}(r)\big)\big(Y_t(s) - \widehat{\mu}(s)\big), \quad r,s \in [a,b],$$

be the sample covariance operator. The eigenequation is given by

$$\int_a^b \widehat{c}_Y(r,s)\widehat{\psi}(s)\,\mathrm{d}s = \widehat{\lambda}\widehat{\psi}(r), \quad r \in [a,b], \tag{6}$$

where $\widehat{\lambda} \in \mathbb{R}$ is an eigenvalue and $\widehat{\psi}(r)$ a corresponding eigenfunction of the sample covariance operator, which is defined as $\widehat{C}_Y(x)(r) = \int_a^b \widehat{c}_Y(r,s)x(s)\,\mathrm{d}s$, $x \in H$, $r \in [a,b]$. Note that all eigenvalues are nonnegative, since $\widehat{C}_Y$ is positive semi-definite, and the eigenspaces are pairwise orthogonal, since $\widehat{C}_Y$ is symmetric. Furthermore, the underlying function space H is countably infinite-dimensional, which yields the existence of a sequence of eigenpairs $\{(\widehat{\lambda}_j, \widehat{\psi}_j(r))\}_{j\in\mathbb{N}}$ with orthonormal eigenfunctions $\{\widehat{\psi}_j(r)\}_{j\in\mathbb{N}}$ and with eigenvalues $\{\widehat{\lambda}_j\}_{j\in\mathbb{N}}$ that are arranged in decreasing order $\widehat{\lambda}_1 \geq \widehat{\lambda}_2 \geq \ldots \geq 0$. Following Hörmann and Kokoszka (2012), the eigenfunction $\widehat{\psi}_l(r)$ is called the l-th empirical FPC, and the projection $\widehat{F}_{l,t} = \langle Y_t - \widehat{\mu}, \widehat{\psi}_l \rangle$ is called the l-th empirical FPC score of $Y_t(r)$. The vector of empirical FPCs is given by $\widehat{\Psi}(r) = (\widehat{\psi}_1(r), \ldots, \widehat{\psi}_K(r))'$, and the vector of empirical scores is defined as $\widehat{F}_t = (\widehat{F}_{1,t}, \ldots, \widehat{F}_{K,t})'$.

Theorem 1. *Under Assumptions 1(a) and 1(b), the solution to the least squares problem*

$$\min_{\substack{\Psi(r) \\ F_1,\ldots,F_T}} \sum_{t=1}^{T} \left\| Y_t - \widehat{\mu} - \Psi'(r)F_t \right\|^2$$

is reached if $F_t = \widehat{F}_t$ *for all* $t = 1, \ldots, T$ *and* $\Psi(r) = \widehat{\Psi}(r)$ *for all* $r \in [a,b]$.

The l-th empirical FPC score has zero mean, since $T^{-1}\sum_{t=1}^{T} \widehat{F}_{l,t} = \langle \widehat{\mu} - \widehat{\mu}, \widehat{\psi}_l \rangle = 0$, and it is uncorrelated with the k-th empirical score for $k \neq l$, while its sample variance is equal to $\widehat{\lambda}_l$, since

$$\frac{1}{T}\sum_{t=1}^{T} \widehat{F}_{k,t}\widehat{F}_{l,t} = \frac{1}{T}\sum_{t=1}^{T} \int_a^b \int_a^b \widehat{c}_Y(r,s)\widehat{\psi}_l(s)\widehat{\psi}_k(r)\,\mathrm{d}s\,\mathrm{d}r = \int_a^b \hat{\lambda}_l\widehat{\psi}_l(r)\widehat{\psi}_k(r)\,\mathrm{d}r = \widehat{\lambda}_l \cdot 1_{\{k=l\}},$$

for any $k, l \in \mathbb{N}$.

Note that the sequence $\{\widehat{\psi}_j(r)\}_{j\in\mathbb{N}}$ of empirical FPCs cannot be unique. If $\widehat{\psi}_l(r)$ is an orthonormal eigenfunction of $\widehat{C}_Y$ with respect to the eigenvalue $\widehat{\lambda}_j$, then $-\widehat{\psi}_l(r)$ possesses the same orthonormality properties. However, if the first K eigenvalues of $\widehat{C}_Y$ are distinct, with $\widehat{\lambda}_1 > \widehat{\lambda}_2 > \ldots > \widehat{\lambda}_K > \widehat{\lambda}_{K+1} \geq 0$, then the orthonormal eigenvectors $\widehat{\psi}_1(r), \ldots, \widehat{\psi}_K(r)$ are unique up to a sign change, and the common components $\widehat{F}_{1,t}\widehat{\psi}_1(r), \ldots, \widehat{F}_{K,t}\widehat{\psi}_K(r)$ are uniquely determined.

Consistency of the empirical mean function, the empirical FPCs, the empirical FPC scores, and the eigenvalues of the empirical covariance operator are established under fairly general conditions. Hörmann and Kokoszka (2010) established consistency for weakly

dependent data. Salish and Gleim (2019) expanded the consistency results to strongly dependent data. The results in Salish and Gleim (2019) can be used to establish uniform consistency under the functional factor model (1), as described in the following lemma.

Lemma 3. *Let Assumptions 1–4 hold true. Then, as* $T \to \infty$,

(a) $\sup_{r\in[a,b]} \left|\widehat{\mu}(r) - \mu(r)\right| = O_P(T^{-1/2})$,

(b) $\sup_{r,s\in[a,b]} \left|\widehat{c}_Y(r,s) - c_Y(r,s)\right| = O_P(T^{-1/2})$,

(c) $\left|\widehat{\lambda}_l - \lambda_l\right| = O_P(T^{-1/2})$,

(d) $\sup_{r\in[a,b]} \left|s_l\widehat{\psi}_l(r) - \psi_l(r)\right| = O_P(T^{-1/2})$,

(e) $\max_{1\le t\le T} \left|s_l\langle Y_t - \widehat{\mu}, \widehat{\psi}_l\rangle - \langle Y_t - \mu, \psi_l\rangle\right| = O_P(T^{-1/2})$,

for all $l = 1, \ldots, L$, *where* $s_l = sign(\langle\widehat{\psi}_l, \psi_l\rangle)$.

Accordingly, it follows that $\sup_{r\in[a,b]} |\widehat{Y}_t(r) - \mu(r) - \Psi'(r)F_t| = O_P(T^{-1/2})$, as $T \to \infty$, where the fitted yield curve is formulated as

$$\widehat{Y}_t(r) = \widehat{\mu}(r) + \sum_{l=1}^{K} \widehat{F}_{lt}\widehat{\psi}_l(r), \quad r \in [a,b].$$

In practice, equation (6) can be solved by using, for instance, the `fda`-package, which is available for R and MATLAB (see Ramsay et al. 2009). If the yield curves are transformed from vector-valued data, as discussed in Remark 1, then they are elements of a finite-dimensional subspace of H by construction, which implies that, in practice, only a finite number of nonzero eigenvalues are obtained.

4 Predictions and prediction bands

The optimal one-step ahead forecast of the factors with respect to MSE loss is given by

$$F_{T+1|T} = E[F_{T+1}|I_T] = \sum_{i=1}^{p} A_i F_{T-i+1},$$

where $I_T = \{Y_t,\ t \le T\}$. For longer horizons $h \in \mathbb{N}$, the h-step ahead forecast is obtained by the chain rule of forecasting as

$$F_{T+h|T} = E[F_{T+h}|I_T] = \sum_{i=1}^{p} A_i F_{T+h-i|T},$$

where $F_{T+j|T} = F_{T+j}$ for $j \leq 0$. A forecast for the entire yield curve can be formulated with the help of the following theorem:

Theorem 2. *Let $g(I_T) \in H$ be a forecast function for Y_{T+h} given the information up to time T. Then, $E\|Y_{T+h} - g(I_T)\|^2$ is minimized if $g(I_T)(r) = E[Y_{T+h}(r)|I_T]$ for any $r \in [a,b]$.*

The minimum MSE h-step ahead forecast for the yield curve is thus given by

$$Y_{T+h|T}(r) = E[Y_{T+h}(r)|I_T] = \mu(r) + \Psi'(r)E[F_{T+1}|I_T] = \mu(r) + \Psi'(r)F_{T+h|T}, \tag{7}$$

where $r \in [a,b]$, and the forecast error curve is obtained as

$$\begin{aligned} e_{T+h|T}(r) = Y_{T+h}(r) - Y_{T+h|T}(r) &= \Psi'(r)(F_{T+h} - F_{T+h|T}) + \epsilon_{T+h}(r) \\ &= \Psi'(r)\Big(\sum_{i=0}^{h-1} \Phi_i \eta_{T+h-i}\Big) + \epsilon_{T+h}(r), \quad r \in [a,b], \end{aligned}$$

where the matrices Φ_i are defined by the recursion $\Phi_i = \sum_{j=1}^{p} \Phi_{i-j} A_j$ with $\Phi_0 = I_K$ and $\Phi_j = 0$ for $j < 0$. From Assumptions 2 and 4, we conclude that $e_{T+h}(r)$ is a Gaussian random function with $E[e_{T+h}(r)] = 0$ and covariance kernel

$$c_{e,h}(r,s) = Cov[e_{T+h}(r), e_{T+h}(s)] = \Psi'(r)\Big(\sum_{i=0}^{h-1} \Phi_i \Sigma_\eta \Phi_i'\Big)\Psi(s) + c_\epsilon(r,s), \quad r,s \in [a,b].$$

For any fixed $r \in [a,b]$, it follows that

$$\frac{Y_{T+h}(r) - Y_{T+h|T}(r)}{\sqrt{c_{e,h}(r,r)}} \sim \mathcal{N}(0,1). \tag{8}$$

The above formulas are infeasible, since $\mu(r)$, $\Psi(r)$, F_T, $A(L)$, and Σ_η are unknown and must be replaced by consistent estimators. For the intercept, the loadings, and the factors, consistent estimators are provided by Lemma 3. Following the two-step procedures of Stock and Watson (2002), Diebold and Li (2006), and Aue et al. (2015), the estimated factors can be used to estimate the VAR model parameters. If the factors are known, then the least squares estimator for the $(p \times Kp)$-matrix $B = (A_1, \ldots, A_p)$ is equal to $\widetilde{B} = FZ'(ZZ')^{-1}$, where $F = (F_{p+1}, \ldots, F_T)$, $Z = (Z_p, \ldots, Z_{T-1})$, and $Z_t = (F_t', \ldots, F_{t-p+1}')'$ (see, e.g., Lütkepohl 2005, Section 3). We thus consider its feasible counterpart which is given by $\widehat{B} = \widehat{F}\widehat{Z}'(\widehat{Z}\widehat{Z}')^{-1}$, where $\widehat{F} = (\widehat{F}_{p+1}, \ldots, \widehat{F}_T)$, $\widehat{Z} = (\widehat{Z}_p, \ldots, \widehat{Z}_{T-1})$, and $\widehat{Z}_t = (\widehat{F}_t', \ldots, \widehat{F}_{t-p+1}')'$. The residuals are obtained as $\widehat{\eta}_t = \widehat{F}_t - \sum_{i=1}^{p} \widehat{A}_i \widehat{F}_{t-i}$, where $(\widehat{A}_1, \ldots, \widehat{A}_p) = \widehat{B}$, and the covariance matrix of η_t is estimated by $\widehat{\Sigma}_\eta = \mathrm{diag}(\widehat{\sigma}_1^2, \ldots, \widehat{\sigma}_K^2)$, where $\widehat{\sigma}_l^2 = (T-p-1)^{-1} \sum_{t=p+1}^{T} \widehat{\eta}_{lt}^2$.

Lemma 4. *Let $S = diag(s_1, \ldots, s_K)$, where $s_l = sign\langle\widehat{\psi}_l, \psi_l\rangle$. Then, under Assumptions 1–4, it follows that*

(a) $\|S\widehat{A}_i S - A_i\|_M = O_P(T^{-1/2}), \quad$ *for all* $i = 1, \ldots, p,$

(b) $\|\widehat{\Sigma}_\eta - \Sigma_\eta\|_M = O_P(T^{-1/2}),$

as $T \to \infty$, where $\|\cdot\|_M$ denotes some matrix norm.

The feasible h-step ahead factor forecast is defined by the recursion

$$\widehat{F}_{T+h|T} = \sum_{i=1}^{p} \widehat{A}_i \widehat{F}_{T+h-i|T},$$

where $\widehat{F}_{T+j|T} = \widehat{F}_{T+j}$ for $j \leq 0$, and the feasible h-step ahead forecast for the yield curve is obtained as

$$\widehat{Y}_{T+h|T}(r) = \widehat{\mu}(r) + \widehat{\Psi}'(r)\widehat{F}_{T+h|T}. \tag{9}$$

Moreover, the estimated covariance kernel of the forecast error function is given by

$$\widehat{c}_{e,h}(r,s) = \widehat{\Psi}'(r)\Big(\sum_{i=0}^{h-1} \widehat{\Phi}_i \widehat{\Sigma}_\eta \widehat{\Phi}_i'\Big)\widehat{\Psi}(s) + \sum_{l=K+1}^{L} \widehat{\lambda}_l \widehat{\psi}_l(r)\widehat{\psi}_l(s), \quad r, s \in [a,b].$$

Note that the h-step ahead predictor for stationary functional time series by Aue et al. (2015) coincides with (9). The authors assumed a VAR(p) model for the first K empirical FPC scores, and they discussed its relation to the optimal forecast from the functional autoregressive model (FAR) of order p. The consistent estimator by Bosq (2000) for the FAR(p) model is based on a truncated FPC decomposition. Aue et al. (2015) demonstrated that the predictor (9) and the h-step ahead predictor by Bosq (2000) are asymptotically equivalent, as $T \to \infty$, if the truncation parameter for the estimator by Bosq (2000) is set to K.

The following result indicates that (9) is an asymptotically optimal minimum MSE forecast for model (1), as $T \to \infty$.

Lemma 5. *Let Assumptions 1–4 hold true. Then, as $T \to \infty$,*

(a) $\sup_{r\in[a,b]} \big|\widehat{Y}_{T+h|T}(r) - Y_{T+h|T}(r)\big| = O_P(T^{-1/2}),$

(b) $\sup_{r,s\in[a,b]} \big|\widehat{c}_{e,h}(r,s) - c_{e,h}(r,s)\big| = O_P(T^{-1/2}).$

Together with equation (8), we can formulate both pointwise and simultaneous $(1-\alpha)$ prediction bands. The first part of the following theorem provides an interval forecast for

each fixed time to maturity $r \in [a, b]$ and some given significance level α, while the second part presents a conservative simultaneous prediction band for the entire yield curve.

Theorem 3. *Under Assumptions 1–4, it follows that*

$$(a) \lim_{T\to\infty} P\left(\frac{|Y_{T+h}(r) - \widehat{Y}_{T+h|T}(r)|}{\omega(r)} \leq u_{1-\frac{\alpha}{2}}\right) = 1-\alpha, \quad \text{for all } r \in [a,b],$$

$$(b) \lim_{T\to\infty} P\left(\frac{|Y_{T+1}(r) - \widehat{Y}_{T+1|T}(r)|}{\omega(r)} \leq \sqrt{\chi^2_{L,1-\alpha}}, \quad \text{for all } r \in [a,b]\right) \geq 1-\alpha,$$

where u_ν is the ν-quantile of the standard normal distribution, $\chi^2_{L,\nu}$ is the ν-quantile of the chi-squared distribution with L degrees of freedom, and

$$\omega^2(r) = \widehat{\Psi}'(r)\left(\sum_{i=0}^{h-1} \widehat{\Phi}_i \widehat{\Sigma}_\eta \widehat{\Phi}_i'\right)\widehat{\Psi}(r) + \sum_{l=K+1}^{L} \widehat{\lambda}_l \widehat{\psi}_l^2(r).$$

Estimating the functional factor model via FPCs yields a model that asymptotically explains $(\sum_{l=1}^{K}\lambda_l)/(\sum_{l=1}^{\infty}\lambda_l)$ of the variability in Y_t, while $\sum_{l=1}^{\infty}\lambda_l = E\|Y_t - \mu\| < \infty$. The proportion of the variance explained by the l-th FPC is given by $\lambda_l/\sum_{k=1}^{\infty}\lambda_k$. Hence, a scree plot provides a useful selection criterion for the number of relevant factors K. The lag order p can also be identified using an information criterion from the VAR literature. Given the relation of the VAR(p) model for the factors and the FAR(p) model for the curves, as discussed above, the testing procedure for selecting the lag order in the FAR(p) model of Kokoszka and Reimherr (2013) also provides a useful criterion to select p. Furthermore, the functional FPE criterion of Aue et al. (2015) allows for a simultaneous identification of the lag order p and the number of factors K.

5 Application to yields for U.S. Treasuries

To apply the prediction and prediction band methodology, we consider a panel of monthly unsmoothed Fama-Bliss zero-coupon yields of U.S. Treasuries with fixed maturities of 3, 6, 9, 12, 15, 18, 21, 24, 30, 36, 48, 60, 72, 84, 96, 108, and 120 months, from January 1985 until December 2007 (see Figure 2). The dataset is taken from Jungbacker et al. (2014), which is available in the Journal of Applied Econometrics Data Archive.[1] It extends the dataset of Diebold and Li (2006). The time span ranges from the period after the Volcker disinflation until the 2008 financial crisis, which can be treated as a consistent monetary policy regime (see Mönch 2012).

[1] see http://qed.econ.queensu.ca/jae/.

5.1 From discrete data to functional data

The data is given in the form of a panel of bond yields y_{it} at time $t = 1, \ldots, T$ and times to maturity r_i, $i = 1, \ldots, N$, with $r_1 < r_2 < \ldots < r_N$. We convert the panel to a functional time series $Y_t(r)$, $t = 1, \ldots, T$, $r \in [a, b]$, by means of a cubic B-spline expansion, where $a = r_1$ and $b = r_N$. The curve is represented as $Y_t(r) = \sum_{j=1}^{J} c_{tj}\phi_j(r)$, $r \in [a, b]$, where $\{\phi_j\}_{j=1,\ldots,J}$ are basis functions on the domain $[a, b]$, and $c_{t1}, \ldots, c_{tJ}$ are basis coefficients. The B-spline basis is defined with respect to a sequence of knots. At each interior observation point $r_2, \ldots, r_{N-1}$, we set a single interior knot, which yields $N - 2$ basis functions. For cubic B-splines, another four basis functions are necessary to ensure twice continuous differentiability. The number of basis functions is hence $J = N+2$. For the definition of and a comprehensive discussion on the B-spline basis, see de Boor (2001).

Minimizing the sum of squared errors $\sum_{i=1}^{N}(Y_t(r_i) - y_{it})^2$ is not feasible, since $J > N$. Following Ramsay and Silverman (2005), we introduce the roughness penalty $\lambda > 0$ and minimize the criterion

$$\min_{c_{t1},\ldots,c_{tJ}} \sum_{i=1}^{N}(Y_t(r_i) - y_{it})^2 + \lambda \int_a^b \left(D^2 Y_t(r)\right)^2 \mathrm{d}r$$

for each $t = 1, \ldots, T$, where $D^2 Y_t(r)$ is the second derivative of $Y_t(r)$. By choosing a roughness penalty λ that is close to zero, the approximation errors at the observed points $r_1, \ldots, r_N$ are almost zero, such that $Y_t(r)$ interpolates the observed bond yields $y_{1t}, \ldots, y_{Nt}$. For our application, we set $\lambda = 10^{-8}$. A comprehensive overview of the problem of transforming discrete data to functional data is provided in Ramsay and Silverman (2005) and Ramsay et al. (2009). The basis representation not only guarantees a smooth yield curve and but also produces yields at maturities that we do not observe. Furthermore, it handles missing values across maturities in a natural way. Note that the B-Spline functions for a basis of an $(N + 2)$-dimensional subspace of H. Therefore, at most $N + 2$ eigenvalues of the empirical covariance function are nonzero.

5.2 Empirical functional principal components of the yield curve

Since the data consists of 17 times to maturity, the transformed functional objects are elements of a 19-dimensional subspace of H, which implies that $\widehat{C}_Y$ has at most 19 nonzero eigenvalues. Figure 4 depicts a scree plot of the eigenvalues. The estimated explained variance of the l-th FPC is given by $\widehat{\lambda}_l / \sum_{k=1}^{\infty} \widehat{\lambda}_k$, while $\sum_{k=1}^{\infty} \widehat{\lambda}_k = \sum_{k=1}^{19} \widehat{\lambda}_k = 3.4814$. The first empirical FPC explains more than 96% of the variation of the curve.

Figure 5 presents orthonormal eigenfunctions for the empirical covariance operator's six largest eigenvalues, which are identified up to a sign change. In the foreground, the

eigenfunctions belonging to the largest three eigenvectors are illustrated, while the next five eigenfunctions are plotted in the background. The shape of the first three loading functions are similar to the level, shape, and curvature function of the Nelson-Siegel model (see Figure 1). Figure 6 illustrates a plot of $\widehat{\mu}(r) \pm 0.5\widehat{\psi}_l(r)$ for $l = 1, \ldots, 6$, which illustrates the effect of each factor on the mean function across all times to maturity. The series of the first six empirical FPC scores are presented in Figure 7.

5.3 Yield curve prediction

To evaluate and compare the forecasting performances in an out-of-sample experiment, we follow the sequential setting in Diebold and Li (2006) and forecast the yield curve for each month from January 1994 until the end of the sample. Let P denote the set of prediction time points, so that the h-step ahead yield curve forecast for time $t \in P$ is estimated using all observations from the beginning of the sample until the time point $t - h$.

For the predictor from the functional dynamic factor model in equation (9), we consider the fixed numbers of $K = 3$, $K = 4$, and $K = 6$ factors and three different specifications for the factor dynamics, which are given by

$$F_{l,t} = a_l F_{l,t-1} + \eta_{l,t}, \quad l = 1, \ldots, K, \tag{AR1}$$

$$F_{l,t} = a_l F_{l,t-1} + b_l F_{t,t-2} + \eta_{l,t}, \quad l = 1, \ldots, K, \tag{AR2}$$

$$F_t = A F_{t-1} + \eta_t. \tag{VAR1}$$

As reported in many studies, bond yields are close to being nonstationary, which makes it difficult to outperform the simple random walk (RW) no-change forecast (see Duffee 2002, Ang and Piazzesi 2003, Diebold and Li 2006, Mönch 2008, Caldeira and Torrent 2017). This is also reflected by the fact that the first eigenvalue in Figure 4 is quite large compared to the others and that augmented Dickey-Fuller tests are not able to reject the unit root hypothesis for the first and the second empirical FPC score series in Figure 7 at the usual significance levels. We thus consider the RW forecast as the main benchmark.

Another benchmark is the forecast obtained from the DNS model. The factors are estimated by regressing the available yields onto the Nelson-Siegel loadings given by equation (2) for a fixed value of $\lambda = 0.0609$. In a second step, a linear autoregressive model is fitted to the estimated factors from the first step, which, analogously to equation (7), gives rise to a forecast of the yield curve. Diebold and Li (2006) considered VAR(1) and univariate AR(1) factor dynamics and demonstrated that the model has a good forecasting performance for larger forecasting horizons.

The accuracies of the curve forecasts are evaluated by computing root mean square

forecast errors

$$RMSFE(r_i) = \sqrt{\frac{1}{|P|}\sum_{t\in P}\left(\widehat{Y}_{t|t-h}(r_i) - Y_t(r_i)\right)^2}$$

and trace root mean square forecast errors

$$TRMSFE = \sqrt{\frac{1}{N}\frac{1}{|P|}\sum_{i=1}^{N}\sum_{t\in P}\left(\widehat{Y}_{t|t-h}(r_i) - Y_t(r_i)\right)^2},$$

where $r_1, \ldots, r_N$ are the available times to maturity in the data, and $|P|$ denotes the number of prediction time points. The results are presented in Table 1. The forecasts based on the FPC estimator with AR(1) factor dynamics have a lower TRMSFE than the DNS forecasts for all forecasting horizons, and a lower TRMSFE than the RW forecast for the six-month and 12-month ahead forecasts. The Diebold-Maniano test results in Table 2 indicate that the FPC-based forecasts outperform the RW and DNS forecasts significantly in the case of bonds with a shorter time to maturity.

To evaluate the prediction bands, we consider the same out-of-sample setting as that for the curve forecasts. The coverage rates of the pointwise and simultaneous bands from Theorem 3 are listed in Tables 3 and 4, where we set $L = 10$ for the simultaneous bands (see Figure 4). Figure 8 presents four exemplary forecasts, pointwise prediction bands, and simultaneous prediction bands. The accuracies of the interval forecasts from the pointwise prediction bands are evaluated using the test by Christoffersen (1998), which compares the nominal coverage of the interval forecast with the true coverage. The results are provided in Table 3, and the null hypothesis that the expected coverage coincides with the nominal coverage is not rejected for the medium and longer maturity bonds. While the true coverage rates for the pointwise prediction bands do not deviate significantly from the nominal coverage for many times to maturity, the simultaneous bands are conservative and have a 100% coverage at a nominal coverage level of 85%, so that there is no yield curve in our sample that exceeds the simultaneous prediction bands.

6 Conclusion

We have introduced an identification strategy for a functional factor model by imposing orthogonality conditions for the loading functions that are elements of the Hilbert space of square integrable functions. The conditions are similar to those from the vector-valued factor models by Stock and Watson (2002) and Bai (2003).

Using results from functional data analysis, an FPC estimator is derived, and consistency results are presented. The minimum MSE h-step ahead forecast coincides with the

predictor proposed in Aue et al. (2015) and thus provides a model-based justification of the prediction strategy in Aue et al. (2015). Furthermore, pointwise and simultaneous prediction bands are derived from the forecast error curve distribution. In an out-of-sample experiment with yield curve data, the forecasting procedure provides a higher accuracy when compared to the conventional DNS model.

Since the number of factors K and the lag order p are assumed to be known or selected by heuristic arguments, further research is required to answer the question of how these numbers can be identified.

References

Anderson, T. W. (2003). *An introduction to multivariate statistical analysis, 3rd Edition*. John Wiley & Sons.

Ang, A. and 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.

Ang, A., Piazzesi, M., and Wei, M. (2006). What does the yield curve tell us about gdp growth? *Journal of econometrics*, 131(1-2):359–403.

Aue, A., Norinho, D. D., and Hörmann, S. (2015). On the prediction of stationary functional time series. *Journal of the American Statistical Association*, 110(509):378–392.

Bai, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica*, 71(1):135–171.

Bai, J. and Ng, S. (2013). Principal components estimation and identification of static factors. *Journal of Econometrics*, 176(1):18–29.

Bardsley, P., Horváth, L., Kokoszka, P., and Young, G. (2017). Change point tests in functional factor models with application to yield curves. *The Econometrics Journal*, 20(1):86–117.

Bliss, R. R. (1997). Movements in the term structure of interest rates. *Economic Review-Federal Reserve Bank of Atlanta*, 82(4):16.

Bosq, D. (2000). *Linear processes in function spaces: Theory and applications*, volume 149. Springer Science & Business Media.

Breitung, J. and Choi, I. (2013). Factor models. *Handbook Of Research Methods And Applications In Empirical Macroeconomics*.

Caldeira, J. and Torrent, H. (2017). Forecasting the us term structure of interest rates using nonparametric functional data analysis. *Journal of Forecasting*, 36(1):56–73.

Christensen, J. H., Diebold, F. X., and Rudebusch, G. D. (2009). An arbitrage-free generalized nelson–siegel term structure model. *The Econometrics Journal*, 12(3).

Christoffersen, P. F. (1998). Evaluating interval forecasts. *International economic review*, pages 841–862.

de Boor, C. (2001). *A Practical Guide to Splines, revised edition*. Applied Mathematical Sciences. Springer.

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

Diebold, F. X., Piazzesi, M., and Rudebusch, G. D. (2005). Modeling bond yields in finance and macroeconomics. *American Economic Review*, 95(2):415–420.

Diebold, F. X. and Rudebusch, G. D. (2013). *Yield Curve Modeling and Forecasting: The Dynamic Nelson-Siegel Approach*. Princeton University Press.

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

Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2000). The generalized dynamic-factor model: Identification and estimation. *Review of Economics and statistics*, 82(4):540–554.

Gürkaynak, R. S., Sack, B., and Wright, J. H. (2007). The us treasury yield curve: 1961 to the present. *Journal of monetary Economics*, 54(8):2291–2304.

Hays, S., Shen, H., Huang, J. Z., et al. (2012). Functional dynamic factor models with application to yield curve forecasting. *The Annals of Applied Statistics*, 6(3):870–894.

Hörmann, S. and Kokoszka, P. (2010). Weakly dependent functional data. *The Annals of Statistics*, 38(3):1845–1884.

Hörmann, S. and Kokoszka, P. (2012). Functional time series. *Handbook of Statistics: Time Series Analysis: Methods and Applications*, 30:157.

Horváth, L. and Kokoszka, P. (2012). *Inference for functional data with applications*, volume 200. Springer Science & Business Media.

Hsing, T. and Eubank, R. (2015). *Theoretical foundations of functional data analysis, with an introduction to linear operators*. John Wiley & Sons.

Hyndman, R. J. and Ullah, M. S. (2007). Robust forecasting of mortality and fertility rates: a functional data approach. *Computational Statistics & Data Analysis*, 51(10):4942–4956.

Joslin, S., Priebsch, M., and Singleton, K. J. (2014). Risk premiums in dynamic term structure models with unspanned macro risks. *The Journal of Finance*, 69(3):1197–1233.

Jungbacker, B., Koopman, S. J., and Van der Wel, M. (2014). Smooth dynamic factor analysis with application to the us term structure of interest rates. *Journal of Applied Econometrics*, 29(1):65–90.

Kokoszka, P. and Reimherr, M. (2013). Determining the order of the functional autoregressive model. *Journal of Time Series Analysis*, 34(1):116–129.

Lengwiler, Y. and Lenz, C. (2010). Intelligible factors for the yield curve. *Journal of Econometrics*, 157(2):481–491.

Litterman, R. and Scheinkman, J. (1991). Common factors affecting bond returns. *Journal of fixed income*, 1(1):54–61.

Lütkepohl, H. (2005). *New introduction to multiple time series analysis*. Springer Science & Business Media.

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

Mönch, E. (2012). Term structure surprises: the predictive content of curvature, level, and slope. *Journal of Applied Econometrics*, 27(4):574–602.

Nelson, C. R. and Siegel, A. F. (1987). Parsimonious modeling of yield curves. *The Journal of Business*, 60(4):473–489.

Park, B. U., Mammen, E., Härdle, W., and Borak, S. (2009). Time series modelling with semiparametric factor dynamics. *Journal of the American Statistical Association*, 104(485):284–298.

Ramsay, J., Hooker, G., and Graves, S. (2009). *Functional data analysis with R and MATLAB*. Springer Science & Business Media.

Ramsay, J. and Silverman, B. (2005). *Functional data analysis*. Springer Science & Business Media.

Salish, N. and Gleim, A. (2019). A moment-based notion of time dependence for functional time series. *Journal of Econometrics*.

Stock, J. H. and Watson, M. (2011). Dynamic factor models. *Oxford Handbook on Economic Forecasting*.

Stock, J. H. and Watson, M. W. (2002). Forecasting using principal components from a large number of predictors. *Journal of the American statistical association*, 97(460):1167–1179.

Svensson, L. E. (1995). Estimating forward interest rates with the extended nelson & siegel method. *Sveriges Riksbank Quarterly Review*, 3(1):13–26.

Yao, F., Müller, H.-G., and Wang, J.-L. (2005). Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association*, 100(470):577–590.

Appendix: Proofs

We first show the following auxiliary result:

Lemma A.1. *Let $\{Y_t\}_{t\in\mathbb{N}}$ be a functional time series that satisfies (1) and Assumptions 1, 2, 3, and 4. Then, the demeaned series $X_t(r) = Y_t(r) - \mu(r)$, $r \in [a,b]$, admits the Karhunen-Loève representation, $X_t(r) = \sum_{l=1}^{\infty} \theta_{l,t}\psi_l(r)$, and, for its FPC scores $\theta_{l,t}$, some constants $R < \infty$ and $\beta \in (1,\infty)$ exist, such that*

$$\left|E\left[\theta_{l_1,t}\theta_{l_2,t-h}\right]\right| \leq R\, h^{-\beta}\sqrt{\lambda_{l_1}\lambda_{l_2}}, \tag{A.1}$$

$$\sum_{\tau_1,\tau_2,\tau_3=-\infty}^{\infty} \left|\kappa_{l_1,l_2,l_3,l_4}(\tau_1,\tau_2,\tau_3)\right| \leq R\sqrt{\lambda_{l_1}\lambda_{l_2}\lambda_{l_3}\lambda_{l_4}}, \tag{A.2}$$

for all $t, h, l_1, l_2, l_3, l_4, \in \mathbb{N}$, where $\kappa_{l_1,l_2,l_3,l_4}(\tau_1,\tau_2,\tau_3)$ denotes the 4-th order cumulant of $(\theta_{l_1,t}, \theta_{l_2,t+\tau_1}, \theta_{l_3,t+\tau_2}, \theta_{l_4,t+\tau_3})$, and $\lambda_l = E[\theta_{l,t}]$.

Proof. Since $E\|X_t\|^2 < \infty$, the demeaned series admits the Karhunen-Loève representation; that is,

$$X_t(r) = \sum_{l=1}^{\infty} \theta_{l,t}\psi_l(r),$$

where $\{\psi_l\}_{l\in\mathbb{N}}$ represents the corresponding sequences of eigenfunctions, and $\theta_{l,t} = \langle X_t, \psi_l\rangle$ denotes the l-th FPC score of $X_t(r)$. From (1), we have that

$$X_t(r) = \sum_{l=1}^{K} F_{l,t}\psi_l(r) + \sum_{l=1}^{\infty} \varepsilon_{l,t}\varphi_l(r),$$

where the second term on the right-hand side is the Karhunen-Loève representation of the error term $\epsilon_t(r)$. Following the identification results of Theorem 1, it follows that

$$\theta_{l,t} = \begin{cases} F_{l,t} & \text{for } 1 \leq l \leq K, \\ \varepsilon_{l-K,t} & \text{for } l > K. \end{cases} \tag{A.3}$$

Finally, since $F_{l,t}$ satisfies Assumption 4, it can be written as a linear process based on the innovation sequence $\{\eta_t\}_{t\in\mathbb{N}}$, and $\varepsilon_{k,t}$ are i.i.d. and Gaussian by Assumption 2 and the properties of the Karhunen-Loève representation. Then, following the discussion in Salish and Gleim (2019) in Section 2.2, the scores $\theta_{l,t}$ satisfy restrictions (A.1) and (A.2). □

Proof of Lemma 1

Since the covariance kernel of $\epsilon_t(r)$ is bounded and $\langle \epsilon_t, \psi_l\rangle = 0$ from Assumption 3, it follows that $\int_a^b c_\epsilon(r,s)\psi_l(s)\,\mathrm{d}s = E[\epsilon_t(r)\langle\epsilon_t,\psi_l\rangle] = 0$ for all $l = 1,\ldots,K$ and $t = 1,\ldots,T$,

and

$$\int_a^b c_Y(r,s)\psi_l(s)\,\mathrm{d}s = \int_a^b \Big(\sum_{k=1}^{K}\lambda_k\psi_k(r)\psi_k(s)\Big)\psi_l(s)\,\mathrm{d}s + \int_a^b c_\epsilon(r,s)\psi_l(s)\,\mathrm{d}s = \lambda_l\psi_l(r)$$

for all $r \in [a,b]$. As a consequence, λ_l is an eigenvalue of C_Y, and $\psi_l(r)$ is a corresponding eigenfunction. Let $\{\xi_j\}_{j\in\mathbb{N}}$ be the sequence of eigenvalues of C_ϵ, and let $\{v_j(r)\}_{j\in\mathbb{N}}$ be a sequence of corresponding orthonormal eigenfunctions. Then, $\{\xi_j, v_j(r)\}_{j\in\mathbb{N}}$ are also eigenpairs of C_Y. Furthermore, $\{v_j\}_{j\in\mathbb{N}}$ forms a basis of H, and, by Parseval's identity, $E\|\epsilon_t\|^2 = \sum_{j=1}^{\infty} E\langle\epsilon_t,\psi_j\rangle^2 = \sum_{j=1}^{\infty}\langle C_\epsilon(\psi_j),\psi_j\rangle = \sum_{j=1}^{\infty}\xi_j$, and $E\|\epsilon_t\|^2 < \lambda_K < \ldots < \lambda_1$ by Assumption 2(c). As a result, $\lambda_1,\ldots,\lambda_K$ are the K largest eigenvalues of C_Y. The second result follows from

$$\langle Y_t - \mu, \psi_l\rangle = \sum_{k=1}^{K} F_{k,t}\langle\psi_k,\psi_l\rangle + \langle\epsilon_t,\psi_l\rangle = F_{l,t},$$

since the eigenfunctions are orthonormal, and $\langle\epsilon_t,\psi_l\rangle = 0$.

Proof of Lemma 2

From Assumptions 1(a) and 1(b), it follows that $\sum_{t=1}^{T}\|\sum_{l=1}^{K}F_{lt}\psi_l\|^2 = \sum_{t=1}^{T}\sum_{l=1}^{K}F_{lt}^2$, which yields

$$\sum_{t=1}^{T}\Big\|Y_t - \widehat{\mu} - \sum_{l=1}^{K}F_{lt}\psi_l\Big\|^2 = \sum_{t=1}^{T}\Big(\|Y_t-\widehat{\mu}\|^2 - 2\sum_{l=1}^{K}F_{lt}\langle Y_t-\widehat{\mu},\psi_l\rangle + \sum_{l=1}^{K}F_{lt}^2\Big).$$

The minimization problem (4) is then equivalent to minimizing $F_{lt}^2 - 2F_{lt}\langle Y_t-\widehat{\mu},\psi_l\rangle$ for all $l = 1,\ldots,K$ and $t = 1,\ldots,T$, and the unique minimum is attained if $F_{lt} = \langle Y_t - \widehat{\mu},\psi_l\rangle$.

Proof of Theorem 1

Let $X_t(r) = Y_t(r) - \widehat{\mu}(r)$. The goal is to find an orthonormal basis $\psi_1(r),\ldots,\psi_K(r)$, such that $\sum_{t=1}^{T}\|X_t - \sum_{l=1}^{K}\langle X_t,\psi_l\rangle\psi_l\|^2$ is minimized. We follow Hörmann and Kokoszka (2012) and Horváth and Kokoszka (2012), where this problem is treated in detail. From Assumptions 1(a) and 1(b), it follows that $\sum_{t=1}^{T}\|Y_t-\mu-\sum_{l=1}^{K}\langle Y_t-\mu,\psi_l\rangle\psi_l\|^2 = \sum_{t=1}^{T}\|X_t\|^2 - \sum_{t=1}^{T}\sum_{l=1}^{K}\langle X_t,\psi_l\rangle^2$. Note that $\sum_{t=1}^{T}\sum_{l=1}^{K}\langle X_t,\psi_l\rangle^2 = T\sum_{l=1}^{K}\langle\widehat{C}_Y(\psi_l),\psi_l\rangle$, and the spectral decomposition yields $\langle\widehat{C}_Y(\psi_l),\psi_l\rangle = \sum_{j=0}^{\infty}\widehat{\lambda}_j\langle\psi_l,\widehat{\psi}_j\rangle^2$. Then, following Theorem 3.2 in Horváth and Kokoszka (2012), the maximum of $\sum_{l=1}^{K}\sum_{j=0}^{\infty}\widehat{\lambda}_j\langle\psi_l,\widehat{\psi}_j\rangle^2$ is attained if $\psi_l(r) = \widehat{\psi}_l(r)$ for all $l = 1,\ldots,K$. Finally, the assertion follows with Lemma 2.

Proof of Lemma 3

Since $E\|Y_t(r) - \mu(r)\|^2 < \infty$, the demeaned series $X_t(r) = Y_t(r) - \mu(r)$ admits the Karhunen-Loève representation,

$$X_t(r) = \sum_{l=1}^{\infty} \theta_{l,t}\psi_l(r),$$

where $\{\lambda_l\}_{l\in\mathbb{N}}$ is the sequence of eigenvalues of the covariance operator of $X_t(r)$ in decreasing order, and $\{\psi_l(r)\}_{l\in\mathbb{N}}$ is an orthonormal sequence of corresponding eigenfunctions. The FPC scores $\theta_{l,t} = \int_a^b X_t(r)\psi_l(r)\,\mathrm{d}r$ satisfy $E[\theta_{l,t}\theta_{m,t}] = \lambda_l \cdot 1_{\{l=m\}}$. Furthermore, since $\int_a^b (\psi_l(r))^2\,\mathrm{d}r = 1$, Lebesgue's criterion for Riemann integrability implies that, for any $l \geq 1$, the eigenfunction $\psi_l(r)$ is bounded. That is, a constant $M < \infty$ exists, such that

$$\sup_{r\in[a,b]} |\psi_\ell(r)| \leq M \quad \text{for all } l \in \mathbb{N}. \tag{A.4}$$

Proof of (a): From equation (A.4) we have

$$\begin{aligned}
&\Big(\sup_{r\in[a,b]} \big|\widehat{\mu}(r) - \mu(r)\big|\Big)^2 = \sup_{r\in[a,b]} \big(\widehat{\mu}(r) - \mu(r)\big)^2 = \sup_{r\in[a,b]} \left(\frac{1}{T}\sum_{t=1}^{T} X_t(r)\right)^2 \\
&= \sup_{r\in[a,b]} \frac{1}{T^2} \sum_{t_1,t_2=1}^{T} \sum_{l_1,l_2=1}^{\infty} \theta_{l_1,t_1}\theta_{l_2,t_2}\psi_{l_1}(r)\psi_{l_2}(r) \leq \frac{M^2}{T^2} \sum_{t_1,t_2=1}^{T} \sum_{l_1,l_2=1}^{\infty} \theta_{l_1,t_1}\theta_{l_2,t_2} \\
&= \frac{M^2}{T^2}\left(\sum_{t=1}^{T}\sum_{l_1,l_2=1}^{\infty} \theta_{l_1,t}\theta_{l_2,t} + 2\sum_{h=1}^{T-1}\sum_{t=h+1}^{T}\sum_{l_1,l_2=1}^{\infty} \theta_{l_1,t}\theta_{l_2,t-h}\right).
\end{aligned}$$

With the identification (A.3), it follows that $E[\sum_{l_1,l_2=1}^{\infty} \theta_{l_1,t}\theta_{l_2,t}] = \sum_{l=1}^{\infty} \lambda_l$, and

$$E\Big[\sum_{l_1,l_2=1}^{\infty} \theta_{l_1,t}\theta_{l_2,t-h}\Big] = \sum_{l_1,l_2=1}^{K} E[\theta_{l_1,t}\theta_{l_2,t-h}] \leq \frac{R}{h}\sum_{l_1,l_2=1}^{K} \sqrt{\lambda_{l_1}\lambda_{l_2}},$$

where the last inequality follows from Lemma A.1. Then,

$$E\left[\Big(\sup_{r\in[a,b]} \big|\widehat{\mu}(r) - \mu(r)\big|\Big)^2\right] \leq \frac{M^2}{T}\sum_{l=1}^{\infty}\lambda_l + \frac{2M^2R}{T^2}\sum_{h=1}^{T-1}\frac{T-h}{h}\sum_{l_1,l_2=1}^{K}\sqrt{\lambda_{l_1}\lambda_{l_2}} = O(T^{-1})$$

Chebyshev's inequality thus concludes the first part of the lemma.

Proof of (b):

From (a), it follows that $\sup_{r,s\in[a,b]} |\widehat{c}_Y(r,s) - \widehat{c}_X(r,s)| = O_P(T^{-1/2})$, where $\widehat{c}_X(r,s) = T^{-1}\sum_{l=1}^{T} X_t(r)X_t(s)$. It hence suffices to show consistency for $\widehat{c}_X(r,s)$. From equation

(A.4), it follows that

$$\sup_{r,s\in[a,b]}\left|\widehat{c}_X(r,s)-c_X(r,s)\right| = \sup_{r,s\in[a,b]}\left|\frac{1}{T}\sum_{t=1}^{T}\left(X_t(r)X_t(s)-E[X_t(r)X_t(s)]\right)\right|$$
$$= \sup_{r,s\in[a,b]}\left|\frac{1}{T}\sum_{l_1,l_2=1}^{\infty}\sum_{t=1}^{T}\left(\theta_{l_1,t}\theta_{l_2,t}-E[\theta_{l_1,t}\theta_{l_2,t}]\right)\psi_{l_1}(r)\psi_{l_2}(s)\right|$$
$$\leq \frac{M^2}{T}\sum_{l_1,l_2=1}^{\infty}\left|\sum_{t=1}^{T}\theta_{l_1,t}\theta_{l_2,t}-E[\theta_{l_1,t}\theta_{l_2,t}]\right|.$$

Then, by Jensen's inequality,

$$E\Big[\Big|\sup_{r,s\in[a,b]}\left|\widehat{c}_X(r,s)-c_X(r,s)\right|\Big|\Big] \leq \frac{M^2}{T}\sum_{l_1,l_2=1}^{\infty}E\Big[\Big|\sum_{t=1}^{T}\theta_{l_1,t}\theta_{l_2,t}-E[\theta_{l_1,t}\theta_{l_2,t}]\Big|\Big]$$
$$\leq \frac{M^2}{T}\sum_{l_1,l_2=1}^{\infty}\sqrt{E\Big[\Big(\sum_{t=1}^{T}\theta_{l_1,t}\theta_{l_2,t}-E[\theta_{l_1,t}\theta_{l_2,t}]\Big)^2\Big]}$$
$$= \frac{M^2}{T}\sum_{l_1,l_2=1}^{\infty}\sqrt{\sum_{t_1,t_2=1}^{T}E\big[\theta_{l_1,t_1}\theta_{l_2,t_1}\theta_{l_1,t_2}\theta_{l_2,t_2}\big]-E\big[\theta_{l_1,t_1}\theta_{l_2,t_1}\big]E\big[\theta_{l_1,t_2}\theta_{l_2,t_2}\big]}.$$

The 4-th order cumulants have the following property:

$$\kappa_{(x_1,x_2,x_3,x_4)} = E[x_1x_2x_3x_4]-E[x_1x_2]E[x_3x_4]-E[x_1x_3]E[x_2x_4]-E[x_1x_4]E[x_2x_3],$$

which yields

$$\sum_{t_1,t_2=1}^{T}E\big[\theta_{l_1,t_1}\theta_{l_2,t_1}\theta_{l_1,t_2}\theta_{l_2,t_2}\big]-E\big[\theta_{l_1,t_1}\theta_{l_2,t_1}\big]E\big[\theta_{l_1,t_2}\theta_{l_2,t_2}\big]$$
$$= \sum_{t_1,t_2=1}^{T}\kappa_{l_1,l_2,l_1,l_2}(0,|t_1-t_2|,|t_1-t_2|)+E\big[\theta_{l_1,t_1}\theta_{l_1,t_2}\big]E\big[\theta_{l_2,t_1}\theta_{l_2,t_2}\big]+E\big[\theta_{l_1,t_1}\theta_{l_2,t_2}\big]E\big[\theta_{l_2,t_1}\theta_{l_1,t_2}\big].$$

From Lemma A.1, it follows that

$$\sum_{t_1,t_2=1}^{T}\kappa_{l_1,l_2,l_1,l_2}(0,|t_1-t_2|,|t_1-t_2|)\leq R\lambda_{l_1}\lambda_{l_2},$$

and

$$\begin{aligned}
&\sum_{t_1,t_2=1}^{T} E\big[\theta_{l_1,t_1}\theta_{l_1,t_2}\big]E\big[\theta_{l_2,t_1}\theta_{l_2,t_2}\big] + E\big[\theta_{l_1,t_1}\theta_{l_2,t_2}\big]E\big[\theta_{l_2,t_1}\theta_{l_1,t_2}\big] \\
&= \sum_{t=1}^{T} \big(E\big[\theta_{l_1,t}^2\big]E\big[\theta_{l_2,t}^2\big] + E\big[\theta_{l_1,t}\theta_{l_2,t}\big]^2\big) \\
&\quad + 2\sum_{h=1}^{T-1}\sum_{t=h+1}^{T} \big(E\big[\theta_{l_1,t}\theta_{l_1,t+h}\big]E\big[\theta_{l_2,t}\theta_{l_2,t+h}\big] + E\big[\theta_{l_1,t}\theta_{l_2,t+h}\big]E\big[\theta_{l_2,t}\theta_{l_1,t+h}\big]\big) \\
&\leq T\lambda_{l_1}\lambda_{l_2}\Big(2 + \frac{4R}{T}\sum_{h=1}^{T-1}\frac{T-h}{h^\beta}\Big).
\end{aligned}$$

Hence, a constant $C < \infty$ exists, such that

$$E\Big[\Big|\sup_{r,s\in[a,b]}\big|\widehat{c}_X(r,s) - c_X(r,s)\big|\Big|\Big] \leq \frac{C}{\sqrt{T}}\sum_{l_1,l_2=1}^{\infty}\sqrt{\lambda_{l_1}\lambda_{l_2}} = \frac{C}{\sqrt{T}}\sum_{l_1,l_2=1}^{L}\sqrt{\lambda_{l_1}\lambda_{l_2}} = O(T^{-1/2}),$$

since $Y_t(r)$ takes values in an L-dimensional subspace of H, and the assertion follows with Markov's inequality.

Proof of (c): The result follows from Lemma A.1 and Corollary 2 in Salish and Gleim (2019).

Proof of (d): Lemma A.1 and Corollary 2 in Salish and Gleim (2019) imply that $\|s_l\widehat{\psi}_l - \psi_l\| = O_P(T^{-1/2})$. Then, from the Cauchy-Schwarz inequality, it follows that

$$\begin{aligned}
&\sup_{r\in[a,b]}\Big|\widehat{\lambda}_l s_l\widehat{\psi}_l(r) - \lambda_l\psi_l(r)\Big| = \sup_{r\in[a,b]}\Big|\int_a^b \widehat{c}_Y(r,s)s_l\widehat{\psi}_l(s)\,\mathrm{d}s - \int_a^b c_Y(r,s)\psi_l(s)\,\mathrm{d}s\Big| \\
&= \sup_{r\in[a,b]}\Big|\int_a^b \big(\widehat{c}_Y(r,s) - c_Y(r,s)\big)s_l\widehat{\psi}_l(s)\,\mathrm{d}s + \int_a^b c_Y(r,s)\big(s_l\widehat{\psi}_l(s) - \psi_l(s)\big)\,\mathrm{d}s\Big| \\
&\leq \|\psi_l\|\Big(\sup_{r\in[a,b]}\sqrt{\int_a^b\big(\widehat{c}_Y(r,s) - c_Y(r,s)\big)^2\,\mathrm{d}s}\Big) + \big\|s_l\widehat{\psi}_l - \psi_l\big\|\Big(\sup_{r\in[a,b]}\sqrt{\int_a^b c_Y^2(r,s)\,\mathrm{d}s}\Big) \\
&\leq \sup_{r,s\in[a,b]}\big|\widehat{c}_Y(r,s) - c_Y(r,s)\big| + O_P(T^{-1/2}) = O_P(T^{-1/2}),
\end{aligned}$$

which follows from (b). Finally, with (c), and Slutsky's theorem, we obtain

$$\big|s_l\widehat{\psi}_l(r) - \psi_l(r)\big| = \frac{1}{\lambda_l}\Big|\widehat{\lambda}_l s_l\widehat{\psi}_l(r) - \lambda_l\psi_l(r)\Big| + O_P(T^{-1/2}) = O_P(T^{-1/2}).$$

Proof of (e): The triangle inequality and the Cauchy-Schwarz inequality yield

$$\begin{aligned}
&\max_{1\le t\le T} \left|s_l\langle Y_t-\widehat{\mu},\widehat{\psi}_l\rangle - \langle Y_t-\mu,\psi_l\rangle\right| \\
&= \max_{1\le t\le T} \left|\langle Y_t-\widehat{\mu}, s_l\widehat{\psi}_l\rangle - \langle Y_t-\widehat{\mu},\psi_l\rangle + \langle Y_t-\widehat{\mu},\psi_l\rangle - \langle Y_t-\mu,\psi_l\rangle\right| \\
&\le \max_{1\le t\le T} \left(\left|\langle Y_t-\widehat{\mu}, s_l\widehat{\psi}_l-\psi_l\rangle\right| + \left|\langle \mu-\widehat{\mu},\psi_l\rangle\right|\right) \\
&\le \|s_l\widehat{\psi}_l-\psi_l\|\big(\max_{1\le t\le T}\|Y_t-\widehat{\mu}\|\big) + \|\widehat{\mu}-\mu\| = O_P(T^{-1/2}),
\end{aligned}$$

which follows from (a) and (d).

Proof of Theorem 2

Note that

$$\begin{aligned}
&E\|Y_{T+h}-g(I_T)\|^2 = E\|Y_{T+h}-E[Y_{T+h}|I_T]+E[Y_{T+h}|I_T]-g(I_T)\|^2 \\
&= E\|Y_{T+h}-E[Y_{T+h}|I_T]\|^2 - 2E[\langle Y_{T+h}-E[Y_{T+h}|I_T], E[Y_{T+h}|I_T]-g(I_T)\rangle] \\
&\quad + E\|E[Y_{T+h}|I_T]-g(I_T)\|^2.
\end{aligned}$$

While $E\|Y_{T+h}-E[Y_{T+h}|I_T]\|^2$ does not depend on $g(I_T)$, the second term satisfies

$$\begin{aligned}
&E\big[\big\langle Y_{T+h}-E[Y_{T+h}|I_T],\ E[Y_{T+h}|I_T]-g(I_T)\big\rangle\big] \\
&= E\big[E\big[\big\langle Y_{T+h}-E[Y_{T+h}|I_T],\ E[Y_{T+h}|I_T]-g(I_T)\big\rangle\big|I_T\big]\big] \\
&= E\big[\big\langle E\big[Y_{T+h}-E[Y_{T+h}|I_T]\big|I_T\big],\ E[Y_{T+h}|I_T]-g(I_T)\big\rangle\big] \\
&= E\big[\big\langle 0,\ E[Y_{T+h}|I_T]-g(I_T)\big\rangle\big] = 0
\end{aligned}$$

because of the law of iterated expectation. Finally, $E\|E[Y_{T+h}|I_T]-g(I_T)\|^2$ takes the smallest value when $g(I_T)=E[Y_{T+h}|I_T]$.

Proof of Lemma 4

From Proposition 3.1 in Lütkepohl (2005), it follows that $\|\widetilde{B}-B\|_M = O_P(T^{-1/2})$. Furthermore, Lemma 3(e) implies that $\|S\widehat{F}-F\|_M = O_P(T^{-1/2})$ and $\|(S\otimes I_p)\widehat{Z}-Z\|_M = O_P(T^{-1/2})$, where $\otimes$ denotes the Kronecker product. The continuous mapping theorem yields $\|\widehat{Z}'(\widehat{Z}\widehat{Z}')^{-1}(S\otimes I_p)-Z'(ZZ')^{-1}\|_M = O_P(T^{-1/2})$. Then,

$$\begin{aligned}
\|S\widehat{B}(S\otimes I_p)-\widetilde{B}\|_M &\le \|S\widehat{F}-F\|_M\cdot\|\widehat{Z}'(\widehat{Z}\widehat{Z}')^{-1}(S\otimes I_p)\|_M \\
&\quad + \|F\|_M\cdot\|\widehat{Z}'(\widehat{Z}\widehat{Z}')^{-1}(S\otimes I_p)-Z'(ZZ')^{-1}\|_M = O_P(T^{-1/2}).
\end{aligned}$$

Therefore, by the triangle inequality, $\|S\widehat{B}(S \otimes I_p) - B\|_M = O_P(T^{-1/2})$, and $\|S\widehat{A}_i S - A_i\|_M = O_P(T^{-1/2})$ for all $i = 1, \ldots, p$. The second result follows from Proposition 3.2 in Lütkepohl (2005) and the fact that $\widehat{\eta}_{lt}^2 = (s_l \widehat{\eta}_{lt})^2$ for all t and l.

Proof of Lemma 5

For the first result, let S be defined as in Lemma 4. Then,

$$\begin{aligned}
&\sup_{r\in[a,b]} \left|\widehat{\Psi}'(r)\widehat{F}_{T+h|T} - \Psi'(r)F_{T+h|T}\right| \\
&= \sup_{r\in[a,b]} \left| \sum_{i=1}^{p} (S\widehat{\Psi}(r))'(S\widehat{A}_i S)(S\widehat{F}_{T+h-i|T}) - \Psi'(r)A_i F_{T+h-i|T} \right| = O_P(T^{-1/2}),
\end{aligned}$$

which follows from Lemmas 3(d) and 3(e), and 4. Finally, by the triangle inequality,

$$\begin{aligned}
&\sup_{r\in[a,b]} \left|\widehat{Y}_{T+h|T}(r) - Y_{T+h|T}(r)\right| \\
&\leq \sup_{r\in[a,b]} \left|\widehat{\mu}(r) - \mu(r)\right| + \sup_{r\in[a,b]} \left|\widehat{\Psi}'(r)\widehat{F}_{T+h|T} - \Psi'(r)F_{T+h|T}\right| = O_P(T^{-1/2})
\end{aligned}$$

because of Lemma 3(a). For the second result, note that

$$c_{e,h}(r,s) = \Psi'(r)\Big(\sum_{i=0}^{h-1} \Phi_i \Sigma_\eta \Phi_i'\Big)\Psi(s) + \sum_{l=K+1}^{L} \lambda_l \psi_l(r)\psi_l(s), \quad r,s \in [a,b],$$

since $c_\epsilon(r,s) = \sum_{l=K+1}^{L} \lambda_l \psi_l(r)\psi_l(s)$ by Mercer's theorem. Then, analogously to the proof for the first result,

$$\begin{aligned}
&\sup_{r,s\in[a,b]} \left|\widehat{c}_{e,h}(r,s) - c_{e,h}(r,s)\right| \\
&\leq \sum_{i=0}^{h} \sup_{r,s\in[a,b]} \left|(S\widehat{\Psi}(r))'S\widehat{\Phi}_i S\widehat{\Sigma}_\eta (S\widehat{\Phi}_i S)'S\widehat{\Psi}(s) - \Psi'(r)\Phi_i\Sigma_\eta\Phi_i\Psi(s)\right| \\
&\quad + \sum_{l=K+1}^{L} \sup_{r,s\in[a,b]} \left|\widehat{\lambda}_l s_l \widehat{\psi}_l(r) s_l \widehat{\psi}_l(s) - \lambda_l \psi_l(r)\psi_l(s)\right| = O_P(T^{-1/2}),
\end{aligned}$$

by Lemmas 3 and 4.

Proof of Theorem 3

From Equation (8), it follows that for any fixed $r \in [a, b]$,

$$P\left(\frac{|Y_{T+h}(r) - Y_{T+h|T}(r)|}{\sqrt{\Psi'(r)(\sum_{i=0}^{h-1} \Phi_i \Sigma_\eta \Phi_i')\Psi(s) + \sum_{l=K+1}^{L} \lambda_l \psi_l(r)\psi_l(s)}} \leq u_{1-\frac{\alpha}{2}}\right) = 1 - \alpha.$$

Then, (a) follows by Lemma 5 and Slutsky's theorem. For (b), let $\theta_{l,T} = \langle Y_t - \mu, \psi_l \rangle$ for $l > K$, and consider the $(L \times 1)$-vector $\delta = ((\sum_{i=0}^{h-1} \Phi_i \eta_{T+h-i})', \theta_{K+1,T+h}, \ldots, \theta_{L,T+h})'$. Furthermore, let $V(r) = (\psi_1(r), \ldots, \psi_L(r))'$, Then, $e_{T+h|T}(r) = V'(r)\delta$, where $\delta \sim \mathcal{N}(0, \Sigma_\delta)$ with

$$\Sigma_\delta = \begin{pmatrix} \sum_{i=0}^{h-1} \Phi_i \Sigma_\eta \Phi_i' & & 0 \\ & \lambda_{K+1} & \\ 0 & & \ddots \\ & & & \lambda_L \end{pmatrix},$$

and $\delta' \Sigma_\delta^{-1} \delta \sim \chi^2(K)$. Therefore,

$$1 - \alpha = P\big(\delta' \Sigma_\delta^{-1} \delta \leq \chi^2_{L,1-\alpha}\big) = P\big((\Sigma_\delta^{-1/2}\delta)'(\Sigma_\delta^{-1/2}\delta) \leq \chi^2_{L,1-\alpha}\big).$$

A result from linear algebra states that for any fixed vector $x \in \mathbb{R}^L$ and any constant $c > 0$, $x'x \leq c^2$ if and only if $|a'x| \leq c\sqrt{a'a}$ for all $a \in \mathbb{R}^L$ (see Lemma A.4 in Yao et al. 2005). Then,

$$P\big((\Sigma_\delta^{-1/2}\delta)'(\Sigma_\delta^{-1/2}\delta) \leq \chi^2_{L,1-\alpha}\big) = P\Big(\big|a'(\Sigma_\delta^{-1/2}\delta)\big| \leq \sqrt{\chi^2_{L,1-\alpha} a'a},\ \forall a \in \mathbb{R}^k\Big).$$

Let $\mathcal{E} = \{a \in \mathbb{R}^L :\ a = \Sigma_\delta^{1/2} V(r),\ r \in [a, b]\}$, which is a subset of $\mathbb{R}^L$. Therefore,

$$P\Big(\big|a'(\Sigma_\delta^{-1/2}\delta)\big| \leq \sqrt{\chi^2_{L,1-\alpha} a'a},\ \forall a \in \mathbb{R}^k\Big) \leq P\Big(\big|a'(\Sigma_\delta^{-1/2}\delta)\big| \leq \sqrt{\chi^2_{L,1-\alpha} a'a},\ \forall a \in \mathcal{E}\Big).$$

Furthermore,

$$P\Big(\big|a'(\Sigma_\delta^{-1/2}\delta)\big| \leq \sqrt{\chi^2_{L,1-\alpha} a'a},\ \forall a \in \mathcal{E}\Big) = P\left(\frac{|V'(r)\delta|}{\sqrt{V'(r)\Sigma_\delta V(r)}} \leq \sqrt{\chi^2_{L,1-\alpha}},\ \forall r \in [a, b]\right).$$

As a consequence,

$$P\left(\frac{|e_{T+h|T}(r)|}{\sqrt{\sum_{i=0}^{h-1}\Psi'(r)\Phi_i\Sigma_\eta\Phi_i'\Psi(r)+\sum_{l=K+1}^{L}\lambda_l\psi_l(r)}}\le\sqrt{\chi^2_{L,1-\alpha}}\quad\forall r\in[a,b]\right)$$
$$=P\left(\frac{\left|V'(r)\delta\right|}{\sqrt{V'(r)\Sigma_\delta V(r)}}\le\sqrt{\chi^2_{L,1-\alpha}},\ \forall r\in[a,b]\right)\ge 1-\alpha.$$

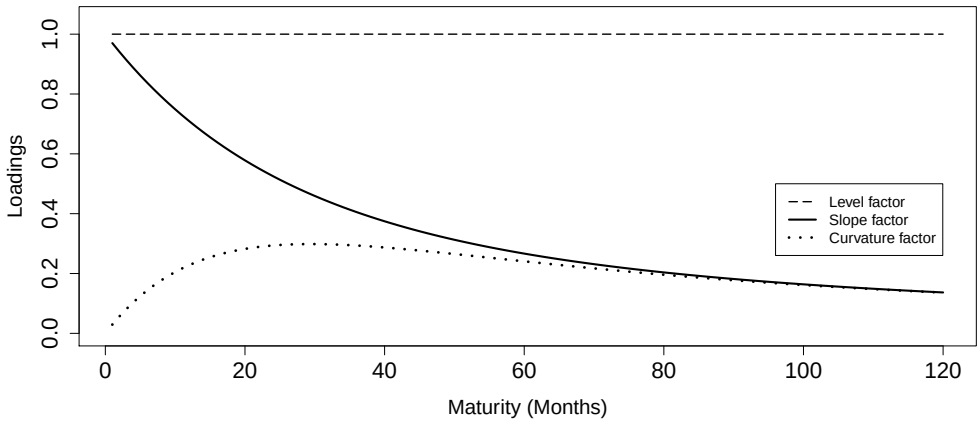
Finally, by Lemmas 3–5 and Slutsky's theorem,

$$\frac{|Y_{T+h}(r)-\widehat{Y}_{T+h|T}(r)|}{\sqrt{\widehat{\Psi}'(r)(\sum_{i=0}^{h-1}\widehat{\Phi}_i\widehat{\Sigma}_\eta\widehat{\Phi}_i')\widehat{\Psi}(r)+\sum_{l=K+1}^{L}\widehat{\lambda}_l\widehat{\psi}_l(r)}}$$
$$=\frac{|e_{T+h|T}(r)|}{\sqrt{\Psi'(r)(\sum_{i=0}^{h-1}\Phi_i\Sigma_\eta\Phi_i')\Psi(r)+\sum_{l=K+1}^{L}\lambda_l\psi_l(r)}}+o_P(1).$$

Then, the assertion follows.

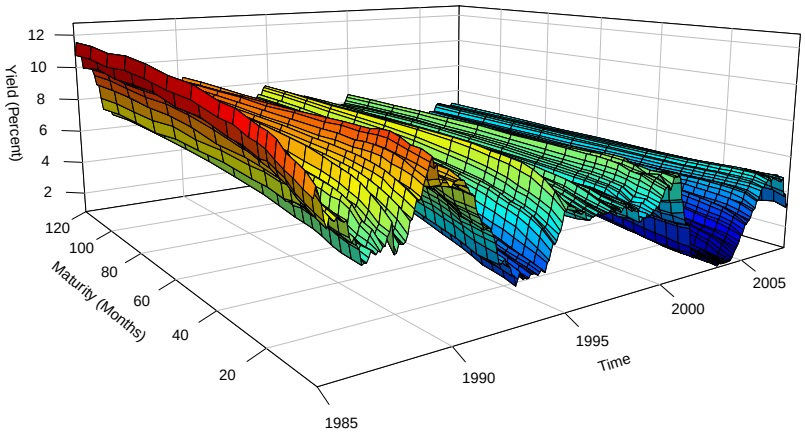
Figures and Tables

Figure 1: Nelson-Siegel loadings



Note: The figure presents a plot of the loading functions considered in Diebold and Li (2006). The decay parameter in equation (2) is set to $\lambda = 0.0609$, which maximizes the curvature factor at a maturity of $r = 30$ months.

Figure 2: Yields of U.S. Treasuries



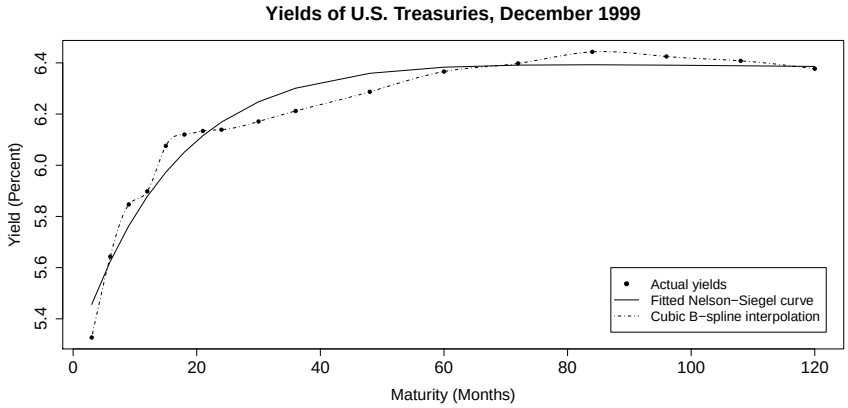
Note: The figure depicts a plot of the monthly yield curves of U.S. Treasuries from January 1985 until December 2007. The dataset is taken from Jungbacker et al. (2014) and is an extension of the dataset considered in Diebold and Li (2006).

Table 1: Root mean square forecast errors

	RW	Nelson-Siegel		Functional principal components estimator								
K		3	3	3	4	6	3	4	6	3	4	6
dynamics	–	AR1	VAR1	AR1			AR2			VAR1		
1-month-ahead forecast												
3m	0.215	0.216	0.194	0.206	0.202	0.202	0.205	0.199	0.199	0.187	0.185	0.190
6m	0.213	0.229	0.206	0.205	0.209	0.210	0.197	0.201	0.201	0.199	0.194	0.196
9m	0.228	0.250	0.236	0.231	0.227	0.227	0.221	0.217	0.217	0.233	0.216	0.218
12m	0.246	0.264	0.249	0.250	0.244	0.243	0.242	0.236	0.236	0.255	0.240	0.242
15m	0.258	0.277	0.261	0.267	0.263	0.259	0.259	0.255	0.252	0.271	0.259	0.258
18m	0.266	0.280	0.269	0.275	0.271	0.269	0.267	0.264	0.262	0.279	0.268	0.268
21m	0.273	0.286	0.279	0.285	0.282	0.280	0.278	0.275	0.273	0.288	0.278	0.280
24m	0.279	0.297	0.293	0.292	0.288	0.288	0.285	0.282	0.282	0.296	0.286	0.288
30m	0.288	0.297	0.298	0.294	0.294	0.294	0.289	0.288	0.288	0.298	0.292	0.295
36m	0.293	0.297	0.300	0.294	0.294	0.295	0.290	0.290	0.290	0.298	0.295	0.297
48m	0.299	0.299	0.304	0.295	0.298	0.299	0.293	0.295	0.296	0.300	0.298	0.301
60m	0.289	0.298	0.303	0.289	0.291	0.291	0.287	0.289	0.288	0.295	0.294	0.296
72m	0.284	0.290	0.294	0.282	0.285	0.285	0.281	0.283	0.283	0.288	0.288	0.290
84m	0.276	0.282	0.286	0.278	0.279	0.277	0.277	0.278	0.276	0.285	0.283	0.282
96m	0.270	0.272	0.275	0.274	0.272	0.272	0.274	0.273	0.272	0.280	0.278	0.279
108m	0.262	0.266	0.268	0.267	0.264	0.265	0.268	0.265	0.265	0.273	0.267	0.270
120m	0.263	0.271	0.273	0.270	0.262	0.263	0.271	0.263	0.264	0.274	0.264	0.268
TRMSFE	0.266	0.276	0.272	0.269	0.268	0.267	0.265	0.264	0.263	0.273	0.266	0.268
6-month-ahead forecast												
3m	0.814	0.820	0.723	0.725	0.734	0.734	0.744	0.752	0.753	0.724	0.667	0.669
6m	0.829	0.867	0.806	0.794	0.798	0.798	0.804	0.808	0.808	0.820	0.745	0.750
9m	0.837	0.879	0.853	0.829	0.826	0.826	0.834	0.832	0.832	0.871	0.785	0.792
12m	0.852	0.889	0.874	0.856	0.851	0.851	0.863	0.859	0.859	0.904	0.823	0.832
15m	0.864	0.895	0.887	0.874	0.870	0.870	0.884	0.881	0.880	0.924	0.851	0.861
18m	0.859	0.888	0.897	0.873	0.870	0.870	0.883	0.880	0.880	0.926	0.861	0.872
21m	0.860	0.884	0.906	0.873	0.871	0.870	0.883	0.881	0.880	0.928	0.869	0.881
24m	0.861	0.885	0.920	0.872	0.869	0.869	0.881	0.879	0.878	0.928	0.874	0.887
30m	0.849	0.858	0.902	0.847	0.846	0.846	0.858	0.858	0.858	0.906	0.863	0.877
36m	0.835	0.838	0.889	0.825	0.825	0.826	0.839	0.839	0.839	0.886	0.852	0.867
48m	0.799	0.801	0.859	0.790	0.792	0.793	0.805	0.808	0.808	0.852	0.831	0.847
60m	0.776	0.785	0.848	0.768	0.770	0.770	0.784	0.786	0.786	0.834	0.820	0.837
72m	0.739	0.749	0.809	0.734	0.736	0.736	0.752	0.754	0.754	0.799	0.790	0.809
84m	0.716	0.717	0.779	0.712	0.712	0.712	0.731	0.732	0.731	0.776	0.770	0.791
96m	0.694	0.688	0.749	0.694	0.694	0.694	0.718	0.717	0.717	0.759	0.754	0.776
108m	0.674	0.671	0.732	0.683	0.681	0.681	0.705	0.703	0.704	0.744	0.737	0.760
120m	0.659	0.672	0.732	0.675	0.671	0.671	0.698	0.695	0.695	0.732	0.723	0.749
TRMSFE	0.798	0.815	0.836	0.793	0.792	0.792	0.807	0.806	0.806	0.845	0.803	0.817
12-month-ahead forecast												
3m	1.392	1.348	1.355	1.272	1.279	1.279	1.294	1.301	1.301	1.369	1.187	1.192
6m	1.418	1.382	1.437	1.344	1.346	1.347	1.350	1.352	1.352	1.462	1.265	1.273
9m	1.402	1.372	1.463	1.359	1.358	1.358	1.357	1.355	1.355	1.490	1.282	1.291
12m	1.408	1.370	1.469	1.378	1.375	1.375	1.376	1.373	1.373	1.509	1.308	1.317
15m	1.400	1.357	1.461	1.379	1.377	1.377	1.383	1.381	1.380	1.509	1.322	1.332
18m	1.374	1.330	1.446	1.354	1.352	1.352	1.359	1.357	1.357	1.484	1.310	1.321
21m	1.348	1.305	1.432	1.328	1.326	1.326	1.334	1.333	1.332	1.458	1.297	1.307
24m	1.326	1.289	1.424	1.304	1.302	1.302	1.310	1.308	1.307	1.432	1.282	1.293
30m	1.280	1.237	1.373	1.248	1.248	1.248	1.260	1.260	1.260	1.375	1.247	1.259
36m	1.233	1.193	1.328	1.196	1.196	1.196	1.213	1.213	1.214	1.320	1.213	1.226
48m	1.146	1.119	1.249	1.115	1.117	1.116	1.142	1.143	1.143	1.235	1.163	1.179
60m	1.092	1.081	1.207	1.061	1.062	1.063	1.093	1.094	1.095	1.184	1.133	1.150
72m	1.026	1.030	1.148	1.011	1.013	1.013	1.050	1.052	1.052	1.132	1.097	1.117
84m	0.983	0.986	1.101	0.972	0.973	0.972	1.014	1.014	1.014	1.091	1.067	1.091
96m	0.946	0.947	1.060	0.946	0.945	0.945	0.994	0.993	0.993	1.064	1.047	1.073
108m	0.913	0.924	1.037	0.928	0.927	0.927	0.975	0.974	0.975	1.044	1.029	1.057
120m	0.895	0.929	1.040	0.918	0.916	0.916	0.965	0.963	0.964	1.030	1.015	1.046
TRMSFE	1.226	1.200	1.306	1.195	1.195	1.195	1.214	1.214	1.214	1.317	1.197	1.211

Note: Root mean square forecast errors (RMSFEs) and trace root mean square forecast errors (TRMSFEs) are reported. Furthermore, RW is the random walk forecast, and the Nelson-Siegel forecasts are based on AR(1) and VAR(1) factor dynamics.

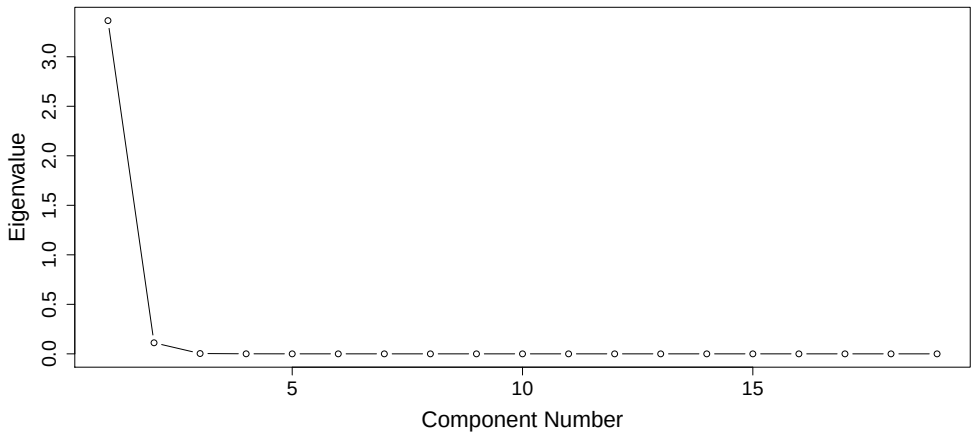
Figure 3: Fitted Nelson-Siegel curve and cubic B-spline representation



Note: The figure illustrates the yield curve from the dataset of Jungbacker et al. (2014) for December 1999. While the solid line presents the fitted Nelson-Siegel curve, the dash-dotted line is a functional representation of the yields from the B-spline basis expansion (see Remark 1 and Section 5.1).

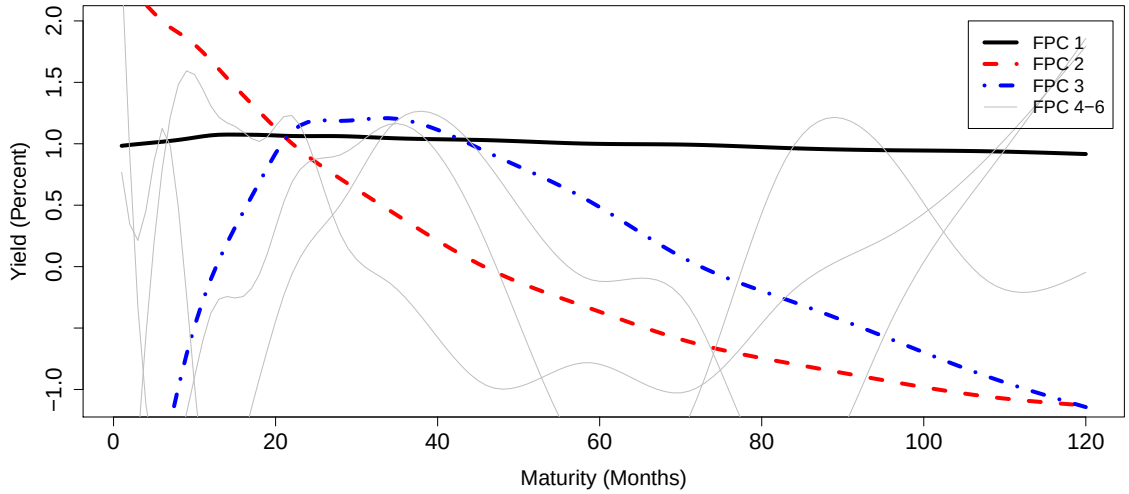
Figure 4: Explained variance of the factors and scree plot

FPC	1	2	3	4	5	6	7	8	9	10
eigenvalue	3.3649	0.1119	0.0033	0.0006	0.0002	0.0001	0.0001	0.0001	0.0001	0.0000
expl.var.	96.65%	3.21%	0.10%	0.02%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
cumulative	96.65%	99.87%	99.96%	99.98%	99.99%	99.99%	99.99%	99.99%	99.99%	100.00%



Note: The table at the top presents the proportion of the explained variance by the l-th FPC score and the cumulative proportions, which are given by $\widehat{\lambda}_l / \sum_{k=1}^{\infty} \widehat{\lambda}_k$ and $(\sum_{j=1}^{l} \widehat{\lambda}_j) / \sum_{k=1}^{\infty} \widehat{\lambda}_k$. The figure at the bottom presents a scree plot of the eigenvalues of the empirical covariance operator of the yield curves.

Figure 5: Empirical functional principal components



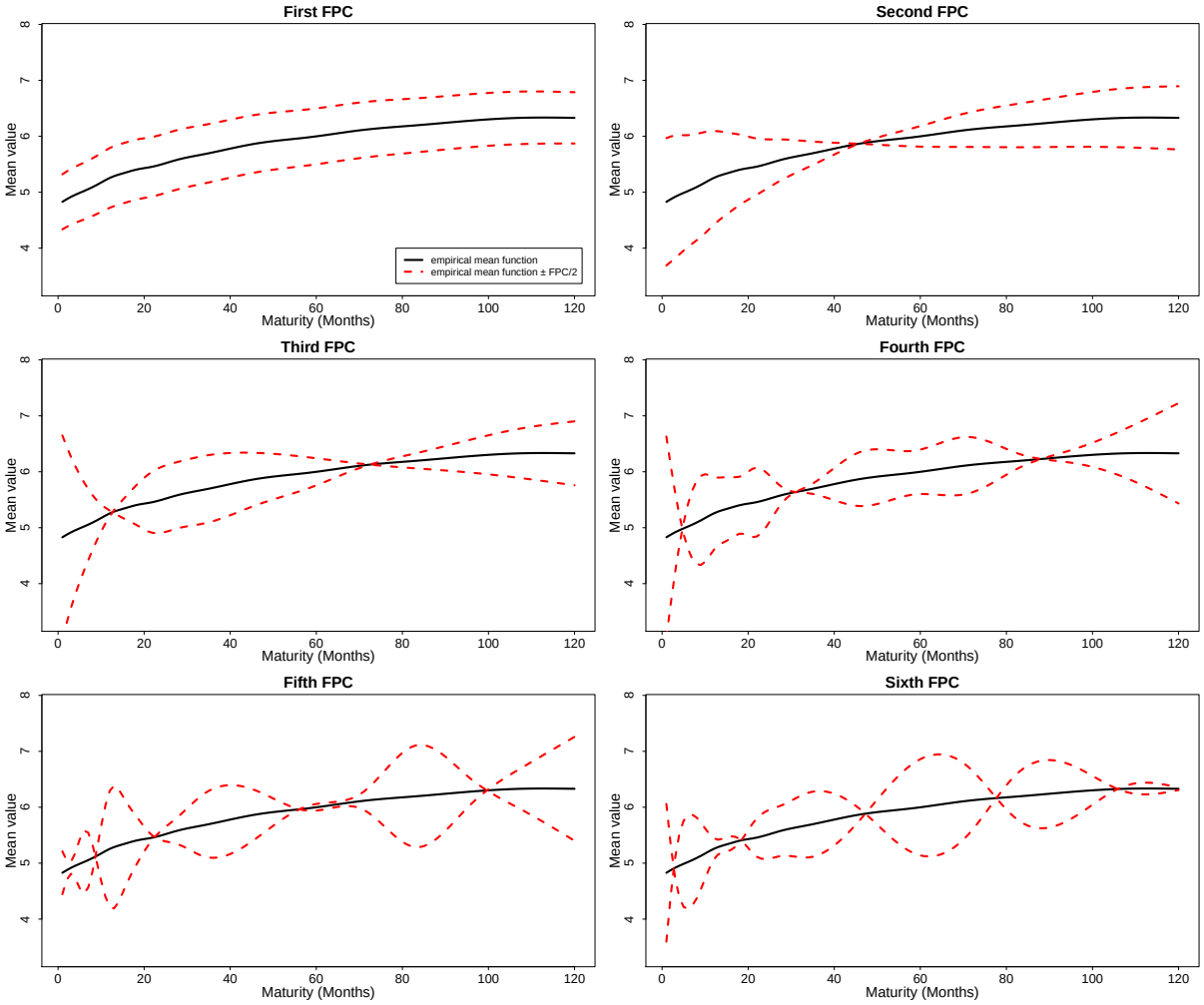
Note: The plot presents the first six empirical functional principal components (FPCs) of the empirical covariance operator of the yield curves from the full dataset of Figure 2.

Table 2: Diebold-Mariano test statistics

	Functional principal components estimator								
K	3	4	6	3	4	6	3	4	6
dynamics	AR1			AR2			VAR1		
Test against the RW forecast									
1-month-ahead forecast									
3m	-0.96	-2.19*	-2.31*	-0.96	-2.11*	-2.25*	-2.80*	-3.22*	-2.60*
12m	0.77	-0.63	-0.87	-0.55	-1.50	-1.66*	1.25	-0.82	-0.47
60m	-0.02	0.72	0.65	-0.38	0.02	-0.09	1.31	0.82	1.12
6-month-ahead forecast									
3m	-4.83*	-4.48*	-4.46*	-2.64*	-2.35*	-2.32*	-3.29*	-3.92*	-3.77*
12m	0.26	-0.08	-0.08	0.45	0.27	0.27	2.27	-0.79	-0.54
60m	-0.66	-0.50	-0.50	0.38	0.45	0.45	2.37	1.41	1.92
12-month-ahead forecast									
3m	-4.21*	-4.04*	-4.05*	-2.23*	-2.09*	-2.10*	-0.62	-3.49*	-3.45*
12m	-1.21	-1.33	-1.33	-0.69	-0.75	-0.75	3.51	-1.81*	-1.69*
60m	-1.40	-1.34	-1.34	0.04	0.07	0.08	2.58	0.83	1.20
Test against the DNS forecast with AR(1) factor dynamics									
1-month-ahead forecast									
3m	-2.18*	-2.71*	-2.35*	-1.70*	-2.70*	-2.53*	-3.70*	-4.03*	-3.32*
12m	-2.98*	-4.06*	-4.12*	-3.13*	-3.84*	-3.91*	-0.97	-2.12*	-1.80*
60m	-1.98*	-1.31	-1.44	-2.19*	-1.63	-1.80*	-0.58	-1.14	-0.35
6-month-ahead forecast									
3m	-6.60*	-5.92*	-5.90*	-4.28*	-3.77*	-3.72*	-3.25*	-4.01*	-3.87*
12m	-2.25*	-2.60*	-2.61*	-2.28*	-2.72*	-2.73*	0.49	-1.59	-1.36
60m	-1.16	-1.03	-1.03	-0.10	0.05	0.06	2.37	1.29	1.92
12-month-ahead forecast									
3m	-3.14*	-2.82*	-2.83*	-2.13*	-1.86*	-1.87*	0.46	-2.64*	-2.56*
12m	0.27	0.17	0.17	0.37	0.20	0.21	3.08	-1.07	-0.93
60m	-0.81	-0.76	-0.75	0.84	0.91	0.94	3.56	1.25	1.70

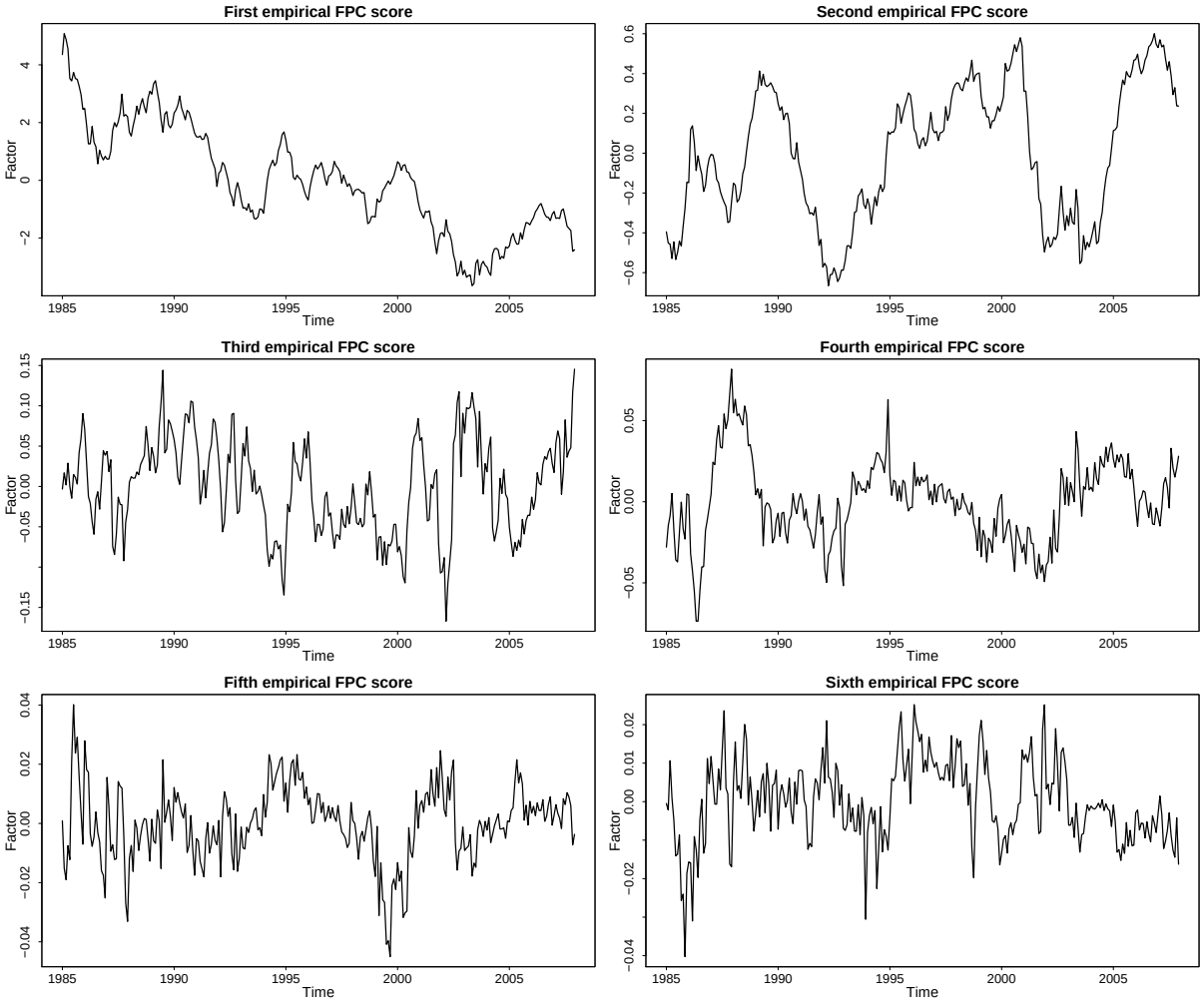
Note: The benchmark statistics for the Diebold-Mariano tests are the random-walk (RW) forecast and the dynamic Nelson-Siegel (DNS) forecast with AR(1) factor dynamics. A negative value indicates the superiority of the functional principal components (FPCs) based forecast, while the asterisks denote significance at the 5% level.

Figure 6: Effects of the first eight functional principal components



Note: The figure presents the plot of $\widehat{\mu}(r) \pm 0.5 \cdot \widehat{\psi}_l(r)$ for $l = 1, \ldots, 6$, where the empirical functional principal components (FPCs) are identified up to a sign change.

Figure 7: Empirical functional principal component scores



Note: The figure presents the plot of the empirical functional principal component (FPC) scores $\widehat{F}_{l,t}$ for $l = 1, \ldots, 6$, which are estimated using the full dataset of Figure 2.

Table 3: Coverage rates and average widths of one-month-ahead interval forecasts

	Functional principal components estimator with $K = 3$ components								
nominal coverage	85			90			95		
dynamics	AR(1)	AR(2)	VAR(1)	AR(1)	AR(2)	VAR(1)	AR1	AR2	VAR1
3m	97.02	96.43	95.24	97.02	97.02	96.43	**97.62**	**97.62**	**97.02**
	(1.09)	(1.06)	(0.73)	(1.25)	(1.21)	(0.84)	(1.48)	(1.44)	(1.00)
6m	96.43	97.02	95.24	97.02	97.02	96.43	98.21	98.21	**97.62**
	(1.05)	(1.02)	(0.77)	(1.20)	(1.16)	(0.88)	(1.43)	(1.39)	(1.05)
9m	96.43	96.43	**89.88**	97.62	97.62	96.43	98.21	98.21	98.21
	(1.03)	(1.00)	(0.82)	(1.18)	(1.15)	(0.94)	(1.41)	(1.36)	(1.12)
12m	94.64	95.24	**89.88**	97.02	97.02	**92.86**	98.21	**97.62**	**97.02**
	(1.03)	(1.00)	(0.87)	(1.18)	(1.14)	(1.00)	(1.41)	(1.36)	(1.19)
15m	92.26	92.86	**89.29**	95.24	95.83	**91.67**	**97.62**	**97.62**	**97.02**
	(1.03)	(1.00)	(0.91)	(1.18)	(1.14)	(1.04)	(1.41)	(1.36)	(1.24)
18m	92.26	91.67	90.48	95.24	94.64	**91.67**	98.21	**97.62**	**95.83**
	(1.02)	(0.99)	(0.93)	(1.17)	(1.13)	(1.06)	(1.39)	(1.35)	(1.27)
21m	90.48	91.07	**89.88**	**92.86**	**92.26**	**91.67**	**97.62**	**97.02**	**96.43**
	(1.01)	(0.97)	(0.95)	(1.15)	(1.11)	(1.08)	(1.37)	(1.32)	(1.29)
24m	**89.88**	90.48	**89.88**	**92.86**	**91.07**	**92.26**	**97.02**	**95.24**	**96.43**
	(0.99)	(0.96)	(0.97)	(1.13)	(1.10)	(1.10)	(1.35)	(1.31)	(1.31)
30m	**89.29**	90.48	**89.29**	**92.26**	**92.26**	**91.67**	**96.43**	**94.64**	**96.43**
	(0.97)	(0.94)	(0.97)	(1.11)	(1.08)	(1.11)	(1.33)	(1.28)	(1.32)
36m	**87.50**	**89.29**	**87.50**	**92.26**	**91.07**	**92.26**	**94.64**	**93.45**	**97.02**
	(0.95)	(0.91)	(0.97)	(1.08)	(1.05)	(1.11)	(1.29)	(1.25)	(1.32)
48m	**86.31**	**87.50**	**86.90**	**91.67**	**89.88**	**91.67**	**93.45**	**92.86**	**94.64**
	(0.92)	(0.89)	(0.97)	(1.05)	(1.01)	(1.10)	(1.25)	(1.21)	(1.32)
60m	**86.90**	**86.31**	**88.69**	**91.07**	**90.48**	**92.26**	**94.05**	**93.45**	**94.64**
	(0.89)	(0.86)	(0.94)	(1.01)	(0.98)	(1.08)	(1.21)	(1.17)	(1.28)
72m	**86.90**	**87.50**	**88.10**	**92.26**	**90.48**	**92.26**	**95.24**	**94.05**	**95.83**
	(0.89)	(0.86)	(0.94)	(1.01)	(0.98)	(1.07)	(1.21)	(1.17)	(1.28)
84m	**87.50**	**86.90**	**88.69**	**91.07**	**89.88**	**92.26**	**95.83**	**93.45**	**96.43**
	(0.87)	(0.84)	(0.92)	(0.99)	(0.96)	(1.05)	(1.18)	(1.14)	(1.25)
96m	**87.50**	**87.50**	**89.29**	**92.86**	**90.48**	**92.86**	**95.24**	**94.05**	**95.83**
	(0.86)	(0.83)	(0.90)	(0.98)	(0.95)	(1.03)	(1.17)	(1.13)	(1.23)
108m	**88.69**	**88.10**	**89.29**	**92.26**	**91.67**	**93.45**	**96.43**	**96.43**	**97.62**
	(0.86)	(0.84)	(0.91)	(0.99)	(0.95)	(1.04)	(1.18)	(1.14)	(1.24)
120m	90.48	**87.50**	**89.29**	**92.86**	**92.26**	**93.45**	**97.62**	**96.43**	**97.62**
	(0.86)	(0.83)	(0.90)	(0.98)	(0.95)	(1.02)	(1.17)	(1.13)	(1.22)

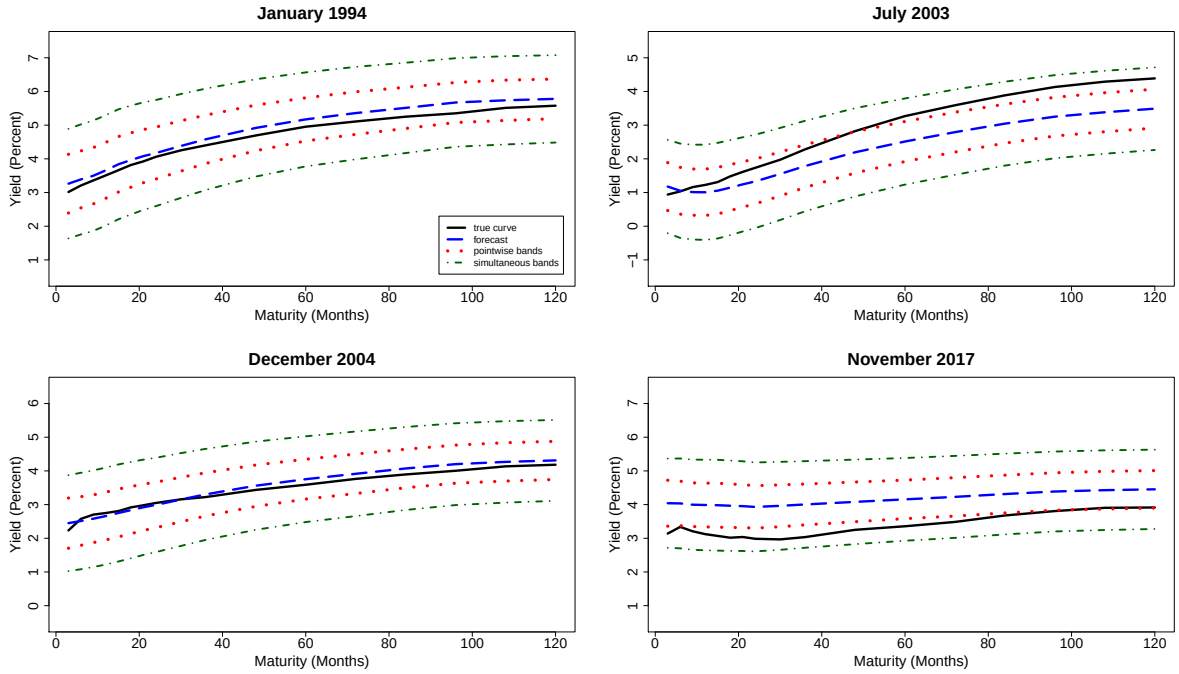
Note: The out-of-sample performance of the pointwise $(1-\alpha)$-prediction bands from Theorem 3 is presented for the levels $\alpha = 15\%$, $\alpha = 10\%$, and $\alpha = 5\%$. The true coverage rates are presented in percentage points, while the average widths of the interval forecasts are indicated in the brackets . Bold numbers indicate that the coverage test by Christoffersen (1998) does not reject the hypothesis that the expected coverage coincides with the nominal coverage at the 5% level.

Table 4: Coverage rates of one-month-ahead simultaneous prediction bands

	Functional principal components estimator with $K = 3$ components								
nominal coverage	85			90			95		
dynamics	AR(1)	AR(2)	VAR(1)	AR(1)	AR(2)	VAR(1)	AR(1)	AR(2)	VAR(1)
true coverage	100	100	100	100	100	100	100	100	100
average width	(2.43)	(2.35)	(2.44)	(2.55)	(2.46)	(2.56)	(2.73)	(2.64)	(2.74)

Note: The out-of-sample performance of the simultaneous $(1-\alpha)$-prediction bands from Theorem 3 is presented for the levels $\alpha = 15\%$, $\alpha = 10\%$, and $\alpha = 5\%$, and $L = 10$. The true coverage rates are presented in percentage points, while the average widths of the interval forecasts are indicated in the brackets.

Figure 8: One-month ahead curve predictions and prediction bands ($\alpha = 5\%$)



Note: Four exemplary one-step ahead forecasts, pointwise prediction bands and simultaneous prediction bands ($\alpha = 5\%$) are illustrated for the model with $K = 3$ factors and AR(1) factor dynamics. While the predictions for January 1994 and December 2014 are usual cases, the ones for July 2003 and November 2017 are exceptional, since the true curve crosses the pointwise prediction band.