

RESEARCH ARTICLE

Time series of functional data with application to yield curves

Rituparna Sen¹ | Claudia Klüppelberg² 

¹Applied Statistics Division, Indian Statistical Institute, Kolkata, India

²Lehrstuhl für Mathematische Statistik, Technische Universität München, Munich, Germany

Correspondence

Rituparna Sen, Applied Statistics Division, Indian Statistical Institute, Kolkata, India.
Email: rsen@isichennai.res.in

Abstract

We develop time series analysis of functional data observed discretely, treating the whole curve as a random realization from a distribution on functions that evolve over time. The method consists of principal components analysis of functional data and subsequently modeling the principal component scores as vector autoregressive moving average (VARMA) process. We justify the method by showing that an underlying ARMAH structure of the curves leads to a VARMA structure on the principal component scores. We derive asymptotic properties of the estimators, fits, and forecast. For term structures of interest rates, these provide a unified framework for studying the time and maturity components of interest rates under one setup with few parametric assumptions. We apply the method to the yield curves of USA and India. We compare our forecasts to the parametric model that is based on Nelson-Siegel curves. In another application, we study the dependence of long term interest rate on the short term interest rate using functional regression.

KEYWORDS

asymptotics, functional principal component, functional regression, prediction, vector ARMA

1 | INTRODUCTION

Functional data analysis (FDA, see the work of Ramsay and Silverman¹ for an overview) is an extension of multivariate data analysis to functional data, where each observation is a curve, rather than a vector in \mathbb{R}^n . An important feature of FDA is its ability to handle dependencies within each observation, especially smoothness, ordering, and neighborhood. Actual observations can be at discrete and irregular points within the curve. The first step of FDA is to replace these actual observations by a simple functional representation. Spline-based approximation is the most commonly used method. Kernel or wavelet-based approximations are also used. FDA has been successfully applied to real-life problems such as analysis of child size evolution,¹ climatic variation,² handwriting in Chinese,³ medical research,⁴ behavioral sciences,⁵ spectrometry data,⁶ etc.

An important tool of FDA is functional principal component analysis (FPCA, see the works of Castro et al⁷ and Rice and Silverman⁸). Functional processes can be characterized by their mean function and the eigenfunctions of the autocovariance operator. This is a consequence of the Karhunen-Loève representation of the functional process. The components of this representation can be estimated. Individual trajectories are then represented by their functional principal component scores, which are available for subsequent statistical analysis. This often leads to substantial dimension reduction.

Most of the development in FDA has been with independent and identical replications of function valued data. This permits the use of information from multiple curves to identify patterns. However, in certain situations, it is unrealistic to

assume that the functions are independent. We do need some process structure. One idea to follow up here is to work with the replication principle implicit in stationary time series, where the values of the process are functions. This problem of dependent functional observations is gaining popularity only recently. Besse et al² develop an AR(1) model for forecasting climatic variations. Kargin and Onatski⁹ use an AR(1) model for forecasting Eurodollar futures. Hörmann and Kokoszka¹⁰ study weakly dependent functional processes, but they ignore the issue of smoothing. This is common to a lot of work following Bosq,¹¹ eg, the works of Bosq¹² and Aue et al,¹³ where the theory is developed assuming that the functions are observed continuously. In practice, however, we only observe the functions at a dense but discrete subset of the support, often with measurement error. Then, we need to interpolate smoothly to infer about the whole function. This raises new questions about the behavior of the estimators. We develop the theory, where the functions follow a stationary ARMA(p, q) model and are observed discretely with noise. We start with kernel smoothing, followed by dimension reduction using FPCA. Based on the time series of the first few significant principal components, we fit a vector autoregressive (VAR) or vector autoregressive moving average (VARMA) model. We provide techniques for estimation of the model parameters and selection of the optimal model.

The methods developed are applied to the modeling and forecast of yield curves. Hays et al¹⁴ have previously used FDA for yield curve modeling. The method in their work¹⁴ is factor analysis with penalized likelihood for estimation, assuming an AR(p) model for the factors. In this paper, the method is PCA with state space modeling for time series, assuming VARMA(p, q) evolution of the components. In addition, we perform a regression of short and long horizon yields, which is new as is the use of Indian government bond data where most of the deviation from the work of Diebold and Li¹⁵ happens.

This paper is organized as follows. In Section 2, we provide some background on principal components analysis of functional data and term structures of interest rates. We then present our main setup and methodology for time series analysis of functional data and some related results. In Section 3, we propose the estimation techniques. In Section 4, we describe two applications, namely, forecasting and regression. We derive asymptotic properties of the estimators, fits, and forecasts in Section 5. We present the analysis of real data of interest rates in Section 6, and finally, we conclude in Section 7.

2 | METHODOLOGY

Consider a sample of n smooth random trajectories $(f_t(u))_{u \in \mathcal{T}}$ for $t = 1, \dots, n$ generated from a process f . Throughout, we assume that f is an element of the Hilbert space $\mathcal{H} := L^2(\mathcal{T})$ endowed with the inner product $\langle f, g \rangle_{\mathcal{H}} = \int_{\mathcal{T}} f(u)g(u)du$ and the norm $\|f\| = \sqrt{\langle f, f \rangle_{\mathcal{H}}} < \infty$ a.s.. The observed measurements are available on a dense grid of support points u_{ti} on the domain $\mathcal{T} = [a_1, a_2]$ with additive white noise error W_{ti} , which is independent of the underlying process. The measurements are for $t = 1, \dots, n$ and $i = 1, \dots, m$,

$$\tilde{f}_t(u_{ti}) = f_t(u_{ti}) + W_{ti} \quad \text{with} \quad E(W_{ti}) = 0, \text{Var}(W_{ti}) = \sigma^2. \tag{1}$$

2.1 | Principal components analysis of functional data

We represent the smooth functional f in terms of its decomposition into functional principal components, a common approach in FDA. For the domain \mathcal{T} , setting

$$\mathbf{G}_f(u, v) = \text{Cov}(f(u), f(v)), \quad E(f(u)) = \mu_f(u), \quad u, v \in \mathcal{T}, \tag{2}$$

the functional principal components are the eigenfunctions of the autocovariance operator $\mathbf{G}_f : \mathcal{H} \mapsto \mathbb{R}$, a linear operator on the space \mathcal{H} , that is given by

$$\mathbf{G}_f(g)(u) = \int_{\mathcal{T}} \mathbf{G}_f(u, v)g(v)dv.$$

We denote the orthonormal eigenfunctions by ϕ_k , with associated eigenvalues λ_k for $k = 1, 2, \dots$, such that $\lambda_1 \geq \lambda_2 \geq \dots$ and $\sum_k \lambda_k < \infty$. The Karhunen-Loève theorem (see the work of Rice and Silverman⁸) provides a representation of individual random trajectories of the functional f , given by

$$f(u) = \mu_f(u) + \sum_{k=1}^{\infty} \xi_k \phi_k(u), \quad u \in \mathcal{T}, \tag{3}$$

where the ξ_k are uncorrelated random variables that satisfy

$$\xi_k = \int (f(u) - \mu_f(u))\phi_k(u) du, \quad E\xi_k = 0, \quad \text{Var}(\xi_k) = \lambda_k. \quad (4)$$

Under the data generating mechanism in (1), one has with indicator function $I(\cdot)$

$$E(\tilde{f}_t(u)) = \mu_f(u), \quad \text{Cov}(\tilde{f}_t(u), \tilde{f}_t(v)) = \mathbf{G}_f(u, v) + \sigma^2 I(u = v). \quad (5)$$

This implies that the smooth mean function μ_f and the smooth covariance surface \mathbf{G}_f can be consistently estimated from available data by pooling the sample of n trajectories and smoothing the resulting scatterplot. The exception for targeting points on \mathbf{G}_f with $u = v$ in (5) is necessitated by the presence of W . This does not pose a problem since it follows from the smoothness of the surface \mathbf{G}_f that the areas of $\mathbf{G}_f(u, v)$, for which $u = v$, can still be consistently estimated. Well-known procedures exist to infer eigenfunctions and eigenvalues.^{8,16}

Processes f are then approximated by substituting estimates and using a judiciously chosen finite number K of terms in sum (3). This choice can be made using one-curve-leave-out cross-validation,⁸ pseudo-AIC criteria¹⁷ or a scree plot, a tool from multivariate analysis, where one uses estimated eigenvalues to obtain a prespecified fraction of variance explained as a function of K or looks for a change point.

The above procedure is also known in numerical analysis under the acronym *proper orthogonal decomposition* and as such it is used to price and hedge financial derivatives on forward curves; see the work of Heppenger¹⁸ for examples from the energy market and further references.

2.2 | Term structure modeling

Term structures of interest rates, also known as yield curve, represent the relationship between spot rates of zero-coupon securities and their term to maturity. This interest rate pattern is used to discount cash flows appropriately. The yield curve is also changing over time. Yield curves are used by fixed income analysts, who analyze bonds and related securities, to understand conditions in financial markets and to seek trading opportunities. Economists use the curves to understand economic conditions. Term structure modeling is a very interesting and active field. There are two popular approaches to term structure modeling. The no-arbitrage tradition focuses on perfectly fitting the term structure at a point in time to ensure that no arbitrage possibilities exist, which is important for pricing derivatives. The equilibrium tradition focuses on modeling the dynamics of the instantaneous rate, typically using affine models, after which yields at other maturities can be derived under various assumptions about the risk premium. Prominent contributions in the no-arbitrage vein include the works of Hull and White¹⁹ and Heath et al,²⁰ and prominent contributions in the affine equilibrium tradition include the works of Vasicek,²¹ Cox et al,²² and Duffie and Kan.²³ Diebold and Li¹⁵ use factor models imposing structure on the factor loadings to distill the entire yield curve, period-by-period, by regression onto a three-dimensional parameter that evolves dynamically. This is the closest one comes to simultaneous treatment of maturity and time evolution of the term structure. We propose an FDA analysis of the yield curve, treating the whole curve over different maturities as a random realization from a distribution on functions. Our proposed nonparametric approach requires no assumptions from the yield curve beyond smoothness and integrability in contrast to currently used approaches, which include parametric components and assumptions. Our analysis provides a unified framework for studying the time and maturity components of interest rates under a setup without too many parametric assumptions. This gives better modeling, data visualization, and understanding of the interest rate process.

2.3 | Time series of functional data

In this section, we show that an ARMA(p, q) structure on the curves implies a VARMA(p, q) structure on the principal component scores. Starting with the setup as described in (1), we assume that the series of functions follows a stationary ARMAH(p, q) model with mean $\mu \in \mathcal{H}$,

$$f_t(\cdot) - \mu_f = \psi_1(f_{t-1}(\cdot) - \mu_f) + \dots + \psi_p(f_{t-p}(\cdot) - \mu_f) + \epsilon_t(\cdot), \quad (6)$$

where $\epsilon_t(\cdot) = \eta_t(\cdot) + \theta_1\eta_{t-1}(\cdot) + \dots + \theta_q\eta_{t-q}(\cdot)$, and $\eta_t(\cdot)$ is \mathcal{H} white noise. ψ_1, \dots, ψ_p and $\theta_1, \dots, \theta_q$ are linear continuous functions. Combining (6) and (3), we have

$$\sum_{k=1}^{\infty} \xi_{kt} \phi_k(\cdot) = \psi_1 \left(\sum_{k=1}^{\infty} \xi_{k,t-1} \phi_k(\cdot) \right) + \dots + \psi_p \left(\sum_{k=1}^{\infty} \xi_{k,t-p} \phi_k(\cdot) \right) + \epsilon_t(\cdot). \quad (7)$$

Using linearity and continuity of ψ_1, \dots, ψ_p , this implies

$$\sum_{k=1}^{\infty} \xi_{kt} \phi_k(\cdot) = \sum_{k=1}^{\infty} \xi_{kt-1} \psi_1(\phi_k(\cdot)) + \dots + \sum_{k=1}^{\infty} \xi_{kt-p} \psi_p(\phi_k(\cdot)) + \epsilon_t(\cdot). \tag{8}$$

Using vector notation, we have

$$\Phi(\cdot) \Xi_t = \psi_1(\Phi(\cdot)) \Xi_{t-1} + \dots + \psi_p(\Phi(\cdot)) \Xi_{t-p} + \epsilon_t(\cdot), \tag{9}$$

where $\Phi = (\phi_1, \phi_2, \dots)$ and $\Xi = (\xi_{1t}, \xi_{2t}, \dots)^T$. Since the eigenfunctions ϕ_k are orthonormal, we can premultiply Equation (10) by Φ^T to get

$$\Xi_t = \Phi^T \psi_1(\Phi(\cdot)) \Xi_{t-1} + \dots + \Phi^T \psi_p(\Phi(\cdot)) \Xi_{t-p} + Z_t. \tag{10}$$

It remains to show that $Z_t = \Phi^T \epsilon(\cdot)$ is an MAH(q) process. This can be proved by verifying that the autocovariances of Z vanish for lags of order greater than q . This is immediate as

$$\text{Cov}(Z_t, Z_s) = \Phi^T \text{Cov}(\epsilon_t, \epsilon_s) \Phi$$

and ϵ is itself an MAH(q) process.

This implies a VARMA(p, q) structure on the vector of principal component scores Ξ_t .

Moreover, since $\Xi_t = \Phi^T f_t$, stationarity of f implies stationarity of Ξ .

3 | ESTIMATION

3.1 | Estimating parameters of the functional process

At the core of the estimation procedure is the principal analysis of random trajectories (PART), applied to the data \tilde{f}_{it} from (1), which is an algorithm to obtain mean and eigenfunctions, as well as FPC scores, from densely sampled functional data, as described in the work of Müller et al.¹⁶ The smoothing steps in this algorithm are implemented with weighted local linear smoothing as in the work of Fan and Gijbels,²⁴ which works well in practice; alternative smoothing methods can also be used. In order to estimate the overall mean function μ_f , we pool all available data into one big scatterplot $\{(u_i, \tilde{f}_{it}), t = 1, \dots, n, i = 1, \dots, m\}$, and then obtain the nonparametric regression of \tilde{f} versus u by local linear smoothing. Formally, one finds the minimizers $\hat{\beta}_0(\tau), \hat{\beta}_1(\tau)$ of

$$\sum_{t=1}^n \sum_{i=1}^m \kappa_1 \left(\frac{u_i - \tau}{b_f} \right) \{ \tilde{f}_{it} - \beta_0(\tau) - \beta_1(\tau)(u_i - \tau) \}^2, \tag{11}$$

where b_f is the smoothing bandwidth, chosen in practice by (generalized) cross-validation, and κ_1 is a kernel function, which is required to be a square integrable and compactly supported symmetric density function, with finite variance and absolutely integrable Fourier transform. Then, one sets $\hat{\mu}_f(\tau) = \hat{\beta}_0(\tau)$ for which one has an explicit representation that is linear in W_i .²⁴

Analogously, surface data are smoothed by fitting local planes by weighted least squares. Specifically, estimation of the covariance surface \mathbf{G}_f is based on the collection of all available pairwise “empirical covariances” $\mathbf{G}_t(u_{i_1}, u_{i_2}) = (\tilde{f}_{it_1} - \hat{\mu}_f(u_{i_1}))(\tilde{f}_{it_2} - \hat{\mu}_f(u_{i_2}))$, assembling these into a two-dimensional scatterplot $\{(u_{i_1}, u_{i_2}), \mathbf{G}_t(u_{i_1}, u_{i_2})\}, t = 1, \dots, n, i_1, i_2 = 1, \dots, m\}$, and fitting a two-dimensional smoother to obtain the nonparametric regression of $\mathbf{G}_t(u_{i_1}, u_{i_2})$ versus (u_{i_1}, u_{i_2}) . Formally, one minimizes

$$\sum_{t=1}^n \sum_{1 \leq i_1 \neq i_2 \leq m} \kappa_2 \left(\frac{u_{i_1} - \tau_1}{h_f}, \frac{u_{i_2} - \tau_2}{h_f} \right) \times \{ \mathbf{G}_t(u_{i_1}, u_{i_2}) - [\beta_0(\tau_1, \tau_2) + \beta_1(\tau_1, \tau_2)(\tau_1 - u_{i_1}) + \beta_2(\tau_1, \tau_2)(\tau_2 - u_{i_2})] \}^2 \tag{12}$$

with respect to $\hat{\beta}_0(\tau_1, \tau_2), \hat{\beta}_1(\tau_1, \tau_2), \hat{\beta}_2(\tau_1, \tau_2)$ and defines $\hat{\mathbf{G}}_f(\tau_1, \tau_2) = \hat{\beta}_0(\tau_1, \tau_2)$. In (12), κ_2 is a kernel function, which is required to be a square integrable and compactly supported radially symmetric bivariate density function, with finite variance and absolutely integrable Fourier transform. The smoothing bandwidth h_f can again be chosen by (generalized) cross-validation.

We note that the diagonal terms $(i_1, i_2), i_1 = i_2$, are missing in the summation over i_1, i_2 in (12). This omission is motivated by the dependence structure of the targets \tilde{f}_{it} . Due to the assumed smoothness of the covariance surface \mathbf{G}_f , the diagonal, on the other hand, is not essential in the surface estimation step and can be omitted from the data that are used to construct the surface, without incurring any asymptotic penalty.

Once mean and covariance functions of the functional process f have been determined, the next step is the estimation of the (eigenvalue/eigenfunction) pairs, which are defined as the solutions of the eigenequations $\int \mathbf{G}_f(s, t)\phi_k(s)ds = \lambda_k\phi_k(t)$, substituting the estimated covariance surface $\hat{\mathbf{G}}_f$ for \mathbf{G}_f . Solutions $(\hat{\lambda}_k, \hat{\phi}_k)$ are obtained by numerical eigenanalysis, based on an initial discretization step, under orthonormality constraints for the eigenfunctions. Positive definiteness of the corresponding covariance surface can be guaranteed by a projection of the initial estimate $\hat{\mathbf{G}}_f$ on a positive definite version $\tilde{\mathbf{G}}_f$, as described in the work of Yao et al.²⁵

In the last step, the PART algorithm yields estimates of the individual FPC scores. Motivated by (4), these are implemented as

$$\hat{\xi}_{tk} = \sum_{i=2}^m (\tilde{f}_{ti} - \hat{\mu}_f(u_{ti}))(u_{ti} - u_{ti-1})\hat{\phi}_k(u_{ti}), \quad t = 1, \dots, n, k = 1, 2, \dots \tag{13}$$

Individual trajectories can then be represented by an empirical version of the Karhunen-Loève expansion (3), for appropriate K ,

$$\hat{f}_t^{(K)}(u) = \hat{\mu}_f(u) + \sum_{k=1}^K \hat{\xi}_{tk}\hat{\phi}_k(u). \tag{14}$$

3.2 | VARMA modeling of the principal component scores

The estimated principal component score vectors $\hat{\xi}_t = (\hat{\xi}_{t1}, \dots, \hat{\xi}_{tK})$ form a vector time series of length n . The infinite dimension of the functional data has been reduced to a finite dimension K . We fit VARMA models of order p, q to the finite dimensional time series of estimated principal component scores $\hat{\xi}_t$.

A VARMA(p, q) process is defined in vector notation as

$$\xi_t = \mu + \psi_1\xi_{t-1} + \dots + \psi_p\xi_{t-p} + \epsilon_t + \theta_1\epsilon_{t-1} + \dots + \theta_q\epsilon_{t-q}, \quad t = p + 1, \dots, n,$$

which can be further simplified by adopting the representation of a lag polynomial

$$\Psi(L)\xi_t = \mu + \Theta(L)\epsilon_t, \quad t = p + 1, \dots, n. \tag{15}$$

Here, μ and ξ_t, ϵ_t for $t = 1, \dots, n$ are vectors of dimension K and $\psi_1, \dots, \psi_p, \theta_1, \dots, \theta_q$ are $K \times K$ matrices.

Note that, in the above model, each ξ_{tk} depends not only on its own history but also on other series' history (cross-dependencies). This gives us several additional tools for analyzing causality as well as feedback effects.

A basic assumption in the above model is that the residual vectors follow a multivariate white noise, ie,

$$E(\epsilon_t) = 0 \tag{16}$$

$$E(\epsilon_t\epsilon_s') = \begin{cases} \Sigma_\epsilon, & \text{if } t = s \\ 0, & \text{if } t \neq s. \end{cases} \tag{17}$$

The coefficient matrices must satisfy certain constraints in order that the VARMA model is stationary. It is required that roots of

$$\det(I - \psi_1z - \dots - \psi_pz^p) = 0 \tag{18}$$

lie outside the unit circle. Here, I is the identity matrix. For more details on VARMA models, see Chapter 11 in the work of Brockwell and Davis.²⁶

Model selection and forecasts can be done conveniently by using the equivalent representation of VARMA using state space models proposed by Akaike.²⁷ For details, consider the work of Aoki and Havenner.²⁸ The main advantage of the state space approach is its capability to find the best model in terms of the Akaike information criterion.

$$\text{AIC} = \log |\det(\tilde{\Sigma}_{\epsilon,p,q})| + 2s/n. \tag{19}$$

Here, s is the number of estimated parameters, n is the sample size, and $\tilde{\Sigma}_{\epsilon,p,q}$ is the estimated covariance matrix obtained as

$$\tilde{\Sigma}_{\epsilon,p,q} = \frac{1}{n} \sum_{t=p+1}^n \hat{\epsilon}_{t,p,q}\hat{\epsilon}_{t,p,q}'$$

where $\hat{\epsilon}_{t,p,q}$ is the estimated error vector for the t th data vector after fitting a VARMA(p, q) model.

If the MA part of the VARMA model has coefficients $\theta_1 = \dots = \theta_q = 0$, the VARMA model reduces to a VAR model. Often, as in our data sets, the MA order q of the optimal VARMA model selected by AIC is zero. In such cases,

it is sufficient to fit a VAR models of order p to the finite dimensional time series of estimated principal component scores $\hat{\xi}_{tk}$. Estimation and model specification of the VAR class is, in general, less complicated than of VARMA models. For the VAR(p) model, the number of estimated parameters s in Equation (19) is equal to $K(1 + pK) + K(K + 1)/2$. $|\hat{\Sigma}_{\epsilon,p}|$ is the estimated covariance matrix.

4 | PRACTICAL APPLICATIONS

4.1 | Forecasting

Our primary aim is forecasting the curve for a future date based on the information available upto a certain point of time. The final VARMA(p, q) model, chosen in Section 3.2, is used to produce model forecasts $\tilde{\xi}_{tk}$ of future principal component scores. Plugging these into Equation (14), we obtain the forecasts $\hat{f}_i(u)$ of the original process f .

$$\hat{f}_i(u) = \hat{\mu}_f(u) + \sum_{k=1}^K \tilde{\xi}_{tk} \hat{\phi}_k(u). \quad (20)$$

Diebold and Li,¹⁵ henceforth referred to as DL, use parametric functions involving variations of Nelson-Siegel exponential components to model the yield curve and then use univariate AR(1) models componentwise to estimate and forecast the factors. This method performs very well for forecasting the yield curve since these parametric functions are specifically designed for this situation. However, the problem of forecasting curves can arise in a lot of other situations. In such cases, the DL method fails completely. We need the set of basis functions to be able to adapt to the data to be of broad and general use. In particular, the basis functions we use are eigenfunctions of the covariance of the dataset. Hence, they can be used in any general setup.

Kargin and Onatski⁹ use predictive factors, similar to simultaneous linear predictions and an alternative to canonical correlations, together with an AR(1) model to predict the term structure of Eurodollar futures. It is not clear if and how the canonical correlation idea can be extended beyond AR(1). In addition, in the empirical application presented in their paper, this method performs worse than the DL method.

4.2 | Correlation and regression involving long-term and short-term interest rates

The correlation between long-term and short-term interest rates is a matter of debate among economists, see, eg, the work of Brown and Schaefer.²⁹ The most powerful and widely accepted theory regarding the relationship between short-term and long-term interest rates is the expectations theory of the term structure. Under the expectations theory, long-term interest rates are described as functions of the weighted averages of expected future short rates plus term premia. According to the theory, therefore, it can be surmised that, a rise (fall) in current short rates will lead to an increase (decrease) in long-term rates. In fact, this is the situation that is usually observed in the real world. However, there are occasional exceptions: For example, in the USA, the Federal Reserve increased the federal funds rate by one percentage point in May 1994, but the interest rates of long maturities fell after that. In the UK, the Bank of England decided to decrease the repo rate by 0.25 percentage points in February 2003, but relatively long-maturity interest rates rose compared with those of the previous day.

There is no clear consensus among researchers in regard to how long-term rates will react to changes in short-term rates. Romer and Romer³⁰ argue that a contractionary monetary policy should be followed by a fall in long rates because the rate of inflation is expected to decline in the future. They explain that a positive correlation between short and long rates, which is typically observed in the real world, is due to the Federal Reserve's information advantage over the public in forecasting inflation. On the other hand, Campbell³¹ asserts that such a usually observed phenomenon stems from bond-market participants' increasing requirement for excess return on long-term bonds. Ellingsen and Söderström³² show that, if market participants consider an unexpected change in the federal funds rate as the Federal Reserve's reaction to economic shocks, then interest rates of all maturities will move in the same direction. In contrast, if a change in short-term rates is regarded as being caused by an unexpected shift in exogenous parameters, such as the relative weight on output variability, the federal funds rate and the interest rates of sufficiently long maturities will move in opposite directions.

We describe a method to quantitatively compute the relation between short-term and long-term interest rates by extending the functional regression techniques of Müller et al³³ to the time series setting. Let the short-term interest rate, say, up to three months maturity, be denoted by $f_{X_t}(u)$ and the long-term interest rate for maturities above three months be denoted

by $f_{Y_t}(u)$. We are interested in predicting $f_{Y_{t+1}}$ given the entire past $(f_{X_1}, f_{Y_1}, \dots, f_{X_t}, f_{Y_t})$ and $f_{X_{t+1}}$. As before, $t = 1, \dots, n$ denotes the time and u denotes the maturity. We carry out a FPCA as described in Section 2.1 of the two series separately. The Karhunen-Loève expansions of the two series in terms of the principal component scores and eigenfunctions as described in Section 3.1 are given by the following:

$$\begin{aligned} f_{X_t}(u) &= \mu_X(u) + \sum_{k=1}^{K_X} \xi_{tk}^X \phi_k^X(u), \\ f_{Y_t}(u) &= \mu_Y(u) + \sum_{k=1}^{K_Y} \xi_{tk}^Y \phi_k^Y(u). \end{aligned} \quad (21)$$

For each $t = 1, \dots, n$, the vector $(\xi_t^X, \xi_t^Y)^T = (\xi_{t1}^X, \dots, \xi_{tK_X}^X, \xi_{t1}^Y, \dots, \xi_{tK_Y}^Y)^T$ is now modeled as a VARMA process

$$\psi(L) \begin{pmatrix} \xi_t^X \\ \xi_t^Y \end{pmatrix} = \Theta(L) \begin{pmatrix} \epsilon_t^X \\ \epsilon_t^Y \end{pmatrix}, \quad (22)$$

where $(\epsilon_t^X, \epsilon_t^Y)^T$ are independent vectors with mean zero and common covariance matrix Σ . The predictor of ξ_{t+1}^Y given $(\xi_1^X, \xi_1^Y, \dots, \xi_t^X, \xi_t^Y, \xi_{t+1}^X)$ is obtained by substituting in (22) the least squares regression prediction of ϵ_{t+1}^Y on ϵ_{t+1}^X , with regression coefficient β_ϵ . Subsequently, these are used in (21) to get the functional regression,

$$\widehat{\mathbb{E}}(f_{Y_{t+1}} | f_{X_1}, f_{Y_1}, \dots, f_{X_t}, f_{Y_t}, f_{X_{t+1}}) \quad (23)$$

$$= \mu_Y(u) + g(f_{X_1}, f_{Y_1}, \dots, f_{X_t}, f_{Y_t}) \quad (24)$$

$$+ \int (f_{X_{t+1}}(v) - \mu_X(v)) \beta(v, u) dv, \quad (25)$$

where the regression surface

$$\beta(v, u) = \sum_{k,m=1}^{\infty} \beta_\epsilon(k, m) \phi_k^X(v) \phi_m^Y(u) \quad (26)$$

and g is a linear function of the past.

5 | ASYMPTOTICS

We derive some consistency results for eigenfunctions, eigenvalues, FPC scores, and fitted trajectories. All proofs and details regarding the assumptions (M1)-(M6) can be found in the Appendix. In the following, the observation interval $\mathcal{T} = [a_1, a_2] \subset (0, T]$.

Recollecting that we estimate the overall mean trajectory μ_f in (11) with bandwidth b_f , and the covariance surface \mathbf{G}_f (2) in (12) with bandwidth h_f , we obtain for the estimation of these key constituents the following result. All convergence results in the following are for $n \rightarrow \infty$ and $\Delta = \sup |u_i - u_{i-1}| \rightarrow 0$.

Theorem 1. *Assuming (M1)-(M4), we have*

$$\sup_{u \in \mathcal{T}} |\hat{\mu}_f(u) - \mu_f(u)| = O_P \left(\frac{1}{\sqrt{nb_f}} \right), \quad (27)$$

$$\sup_{u,v \in \mathcal{T}} |\widehat{\mathbf{G}}_f(u, v) - \mathbf{G}_f(u, v)| = O_P \left(\frac{1}{\sqrt{nh_f^2}} \right). \quad (28)$$

This result provides justification for the mean and covariance function estimates. Next, let \mathfrak{S}' denote the set of indices of the eigenfunctions ϕ_k corresponding to eigenvalues λ_k of multiplicity one. As a consequence of the following theorem, we obtain consistency for the estimation of eigenvalues $\hat{\lambda}_k$ and eigenfunctions $\hat{\phi}_k$ for $k \in \mathfrak{S}'$, justifying the use of these estimates in the subsequent analysis.

Theorem 2. Assume (M1)-(M4). Then,

$$|\hat{\lambda}_k - \lambda_k| = O_P \left(\frac{1}{\sqrt{nh_f^2}} \right) \tag{29}$$

$$\|\hat{\phi}_k - \phi_k\|_H = O_P \left(\frac{1}{\sqrt{nh_f^2}} \right) \quad k \in \mathfrak{S}' \tag{30}$$

$$\sup_{u \in \mathcal{T}} |\hat{\phi}_k(u) - \phi_k(u)| = O_P \left(\frac{1}{\sqrt{nh_f^2}} \right), \quad k \in \mathfrak{S}'. \tag{31}$$

One is also interested in the consistency of estimated principal component scores ξ_t and estimates $\hat{f}_t^{(K)}(u)$ as in (14) of individual trajectories (3).

Theorem 3. Assuming (M1)-(M6),

$$\sup_{1 \leq k \leq K} |\hat{\xi}_{tk} - \xi_{tk}| \xrightarrow{P} 0 \tag{32}$$

$$\sup_{u \in \mathcal{T}} |\hat{f}_t^{(K)}(u) - f_t(u)| \xrightarrow{P} 0. \tag{33}$$

The following result is regarding the forecast $\hat{f}_t(u)$ as in (20) using the fitted ARMA(p, q) model.

Theorem 4. Assuming (M1)-(M6), if the distribution of the innovations is normal, then the forecasted ξ_t are asymptotically normal with variance Σ_ϵ . The forecasted functions $(\phi_1, \dots, \phi_K)^T(\hat{f}_t - f_t)$ are asymptotically normal with variance Σ_ϵ .

6 | EMPIRICAL EXAMPLES

6.1 | Data

The yield curves of two different economies, USA and India, are studied for comparative purpose. The US Department of Treasury webpage lists the daily yield curve from 1990 until date for certain maturities from 1 month to 30 years. The Indian government bond historical data can be obtained from in.investing.com for each maturity separately from 3 months to 15 years. The specific maturities are listed in Table 1. We separate the data into years because for long time horizons, the stationarity assumption of the time series may not be valid. We present the results for the year 2015 for USA and India. They are representative of the other years. In Figure 1, we present the raw data for the countries. For each weekday of the year, we have data of dimension 11 (for US data) and dimension 17 (for Indian data). We think of it as a time series of functions. It is observed that the US curves are pretty smooth, whereas the Indian data have more fluctuations, both with respect to maturity and in time.

TABLE 1 Forecasting the term structure as in Section 6.3

Country	Trading Days (n)	Maturities (τ)	Bandwidth for Mean	Bandwidth for Covariance	PC (K)	VARMA (p, q)
USA	251	1/12,3/12,6/12,1,2,3,5,7,10,20,30	8.23	4.4	1	p = 3, q = 0
India	250	3/12,6/12,1,2, ..., 15	2.75	0.44	3	p = 1, q = 0

Abbreviations: PC, principal components; VARMA, vector autoregressive moving average.

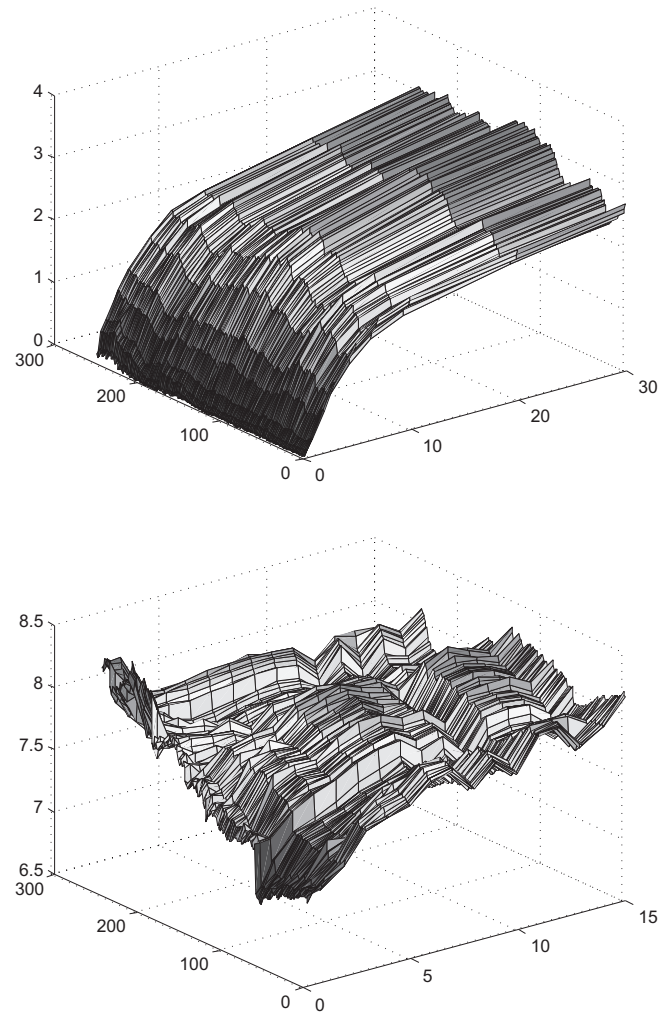


FIGURE 1 Raw data of interest rates. Top: USA, bottom: India; x-axis: day, y-axis: maturity in months, z-axis: interest rate

6.2 | Programs

The initial fitting of functional data to obtain mean, covariance, and principal components is done by employing the PACE package for FDA written in MATLAB. We use the Gaussian kernel. The package is available at

<http://www.stat.ucdavis.edu/PACE/>.

The VAR model fitting and diagnostics are done using the econometrics Toolbox in MATLAB.

The VARMA and related state space model computations are done using the dynamic systems estimation (dse) package in R available at

<http://cran.r-project.org/web/packages/dse/index.html>.

It should be noted that, in all the actual data applications, the models chosen by AIC criterion had the MA degree zero.

6.3 | Forecast

The smoothing bandwidth choices, number of significant principal components, and degree of VARMA models are presented in Table 1. These are obtained by cross-validation. The mean functions, covariance surfaces, and significant eigenfunctions for both countries are presented in Figures 2 to 4. Although the mean functions are similar for the two economies, the covariance surface, and consequently, the eigenfunctions are very different. For the US data, one eigenfunction is enough to explain 86.4% of the variation. For the Indian data, the first three eigenfunctions explain 57.4%, 37%, and 3.6% of the variation, respectively.

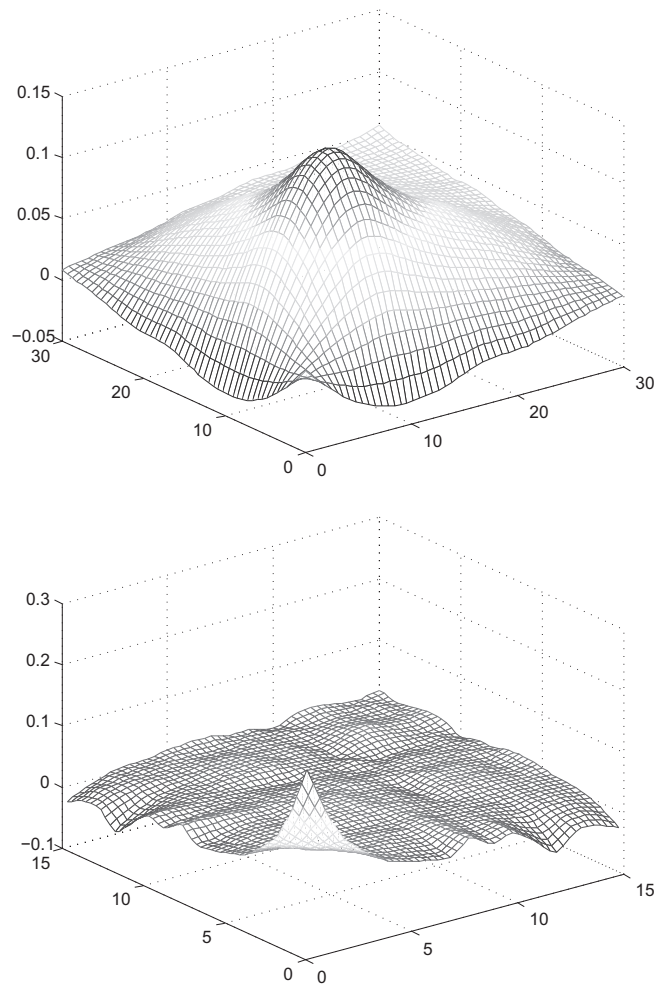


FIGURE 2 Estimated smooth covariance surface of interest rates. Top: USA, bottom: India; x-axis: maturity in months, y-axis: covariance between interest rates

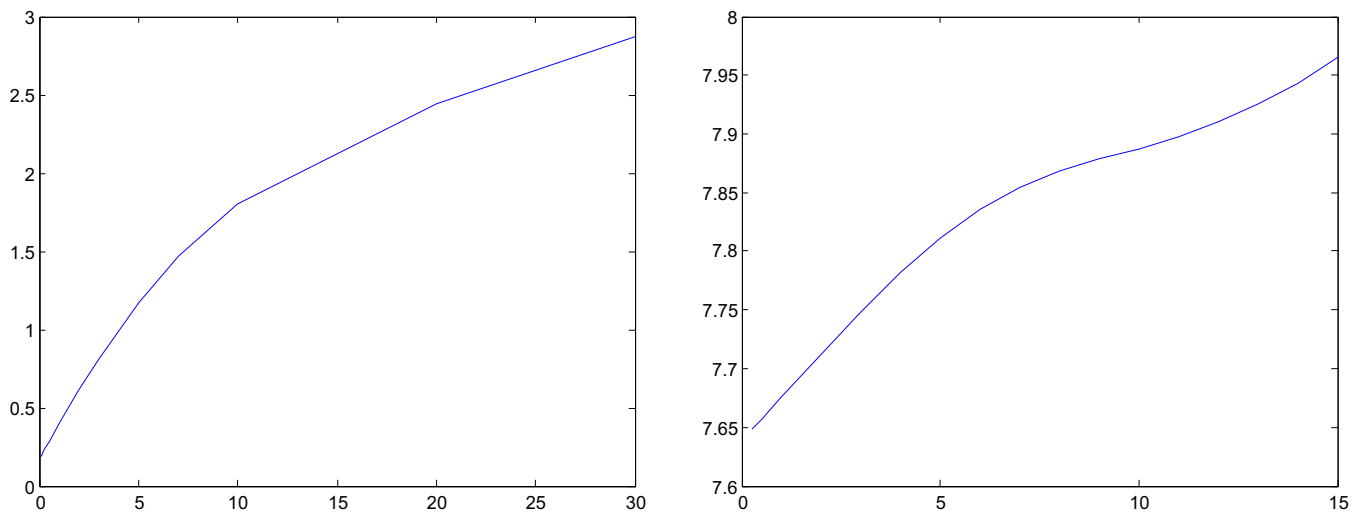


FIGURE 3 Estimated smooth mean function of interest rates. Left: USA, right: India; x-axis: maturity in months, y-axis: mean interest rate [Colour figure can be viewed at wileyonlinelibrary.com]

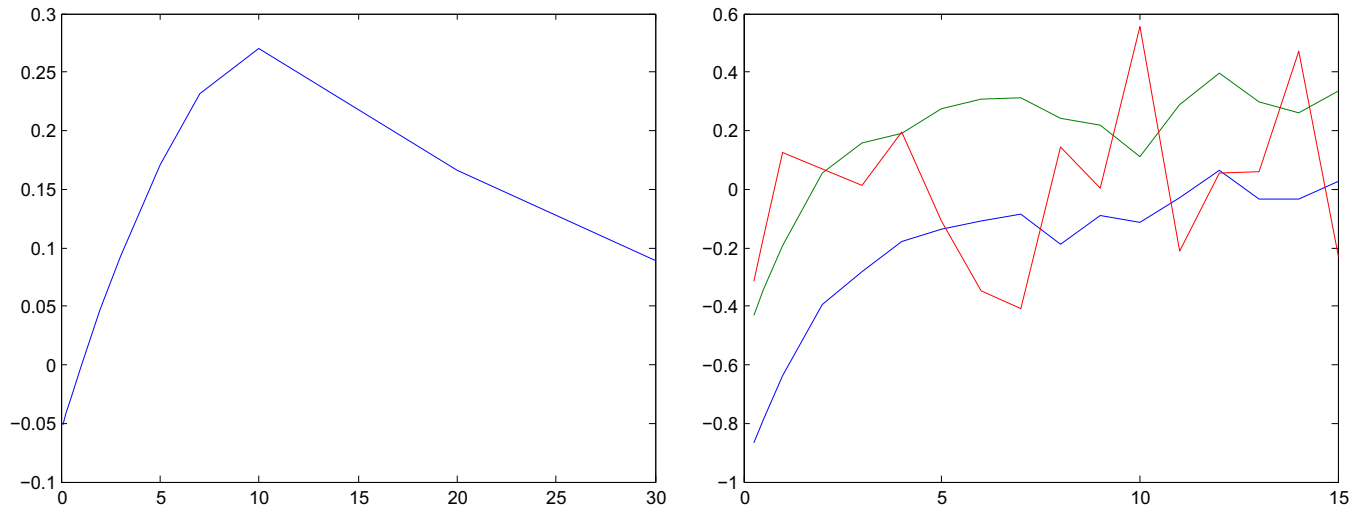


FIGURE 4 Estimated significant eigenfunctions of interest rates. Left: USA, solid line(86.4). Right: India, solid line(57.4), dashed line(37), dotted line(3.6). Numbers in brackets denote percentage of variance explained. x-axis: maturity in months, y-axis: estimated eigenfunctions [Colour figure can be viewed at wileyonlinelibrary.com]

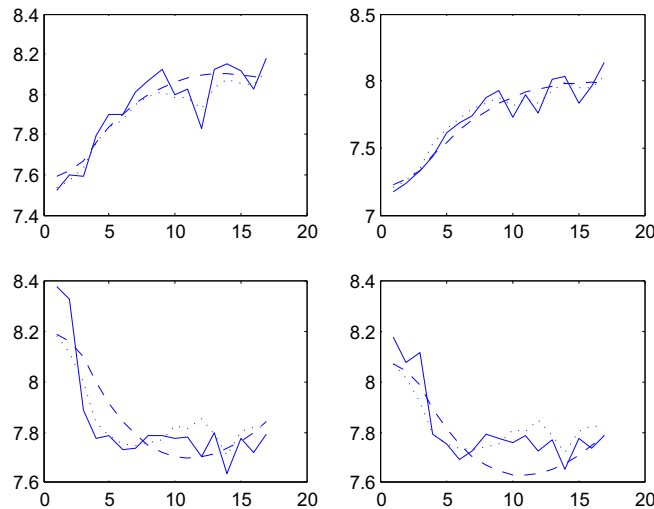


FIGURE 5 For 4 random days, the solid curve is the observed yield curve for the Indian data, the dotted line is forecast using functional data analysis (FDA) and dashed line is forecast using the work of Diebold and Li¹⁵ [Colour figure can be viewed at wileyonlinelibrary.com]

Four random days for the Indian data, we present the observed curve and the forecasts using the FDA method and DL method in Figure 5. It is seen that the FDA method forecasts match the observed curves better. We summarize the distance between the observed and predicted curves for each day in Figure 6. Distance is defined as

$$\int \left(f_t(u) - \hat{f}_t^{(K)}(u) \right)^2 dt.$$

It is seen that for the US data, the estimation based on FDA performs comparably with the DL method, which was specifically designed for the purpose of forecasting the term structure of interest rates. For the Indian data, the FDA method performs better as seen from the forecast error, which is smaller in most cases for the FDA method. The parametric model of Nelson and Siegel is not a good fit for the Indian yield curve.

6.4 | Regression

For regression, the smoothing bandwidth choices and number of significant principal components for different maturity horizons, that is, short and long, for the two years are presented in Table 2. In that table, we also present the degree

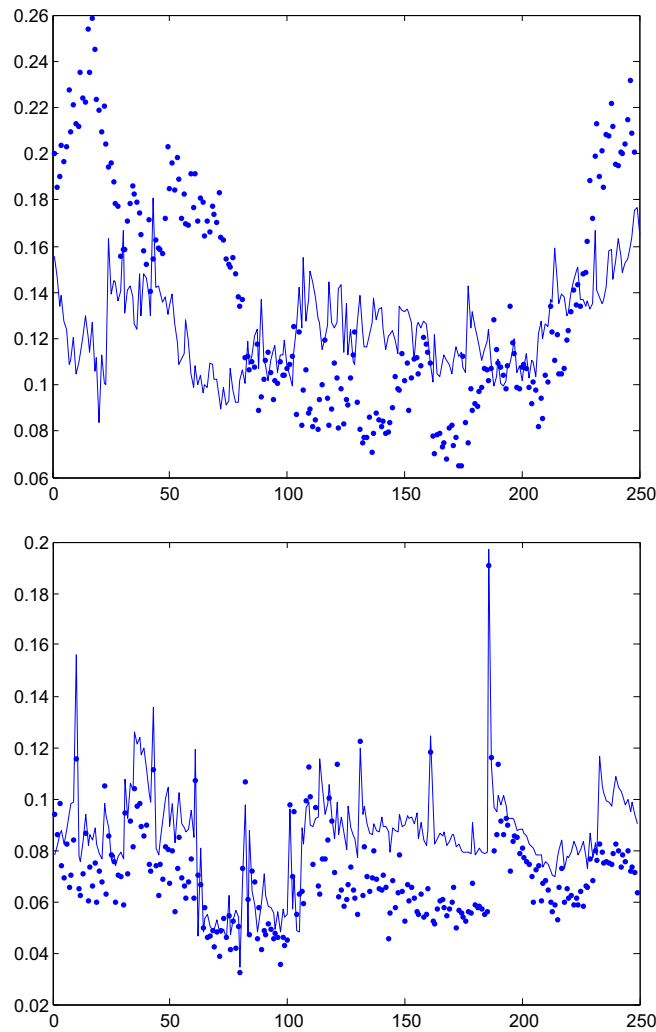


FIGURE 6 Root-mean-square error (RMSE) between observed functions and predicted functions plotted over days. The dotted line is RMSE for predictions using principal components. The solid line is RMSE for predictions using the method of Diebold and Li.¹⁵ Top: USA, the order of autoregressive moving average (ARMA) is chosen to be (3,0). Bottom: India, the order of ARMA is chosen to be (1,0) [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 2 Regression of the long term on the short-term interest rates as in Section 6.4

Maturity	Bandwidth for Mean	Bandwidth for Covariance	PC	VARMA	R^2
Short	0.25	0.25	1	$p = 1; q = 0$	0.17
Long	7.7	4.4	1		

Abbreviations: PC, principal components; VARMA, vector autoregressive moving average.

of VARMA models and regression R^2 . The estimated mean curve and significant principal components are presented in Figure 7. The regression coefficient is 7.85. Since this is positive, it indicates an increasing relationship between the short-term and long-term rates. However, the relation is far from linear as one needs to account for the eigenfunctions in Figure 7 through Equation (26).

In Figure 8 we present the distance between the observed and predicted functions of long term interest rates. Prediction based on time series and regression on the short term interest rates far outperforms the prediction based on only time series of long term rates.

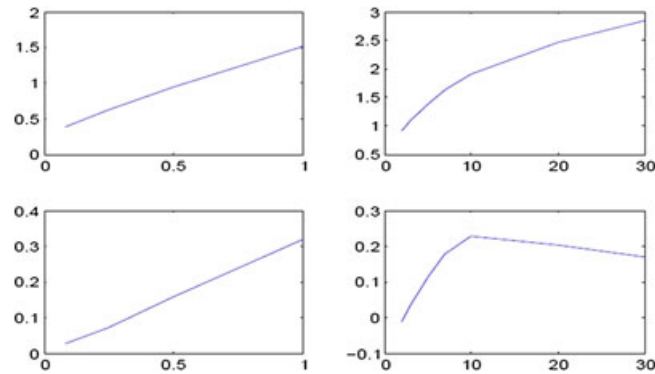


FIGURE 7 For the US data, the mean curves (top) and significant eigenfunctions (bottom). Left: short term (up to 1 year), right: long term (2-30 years). The percentage of variance explained are 95.6 and 89.3 for the short and long term, respectively [Colour figure can be viewed at wileyonlinelibrary.com]

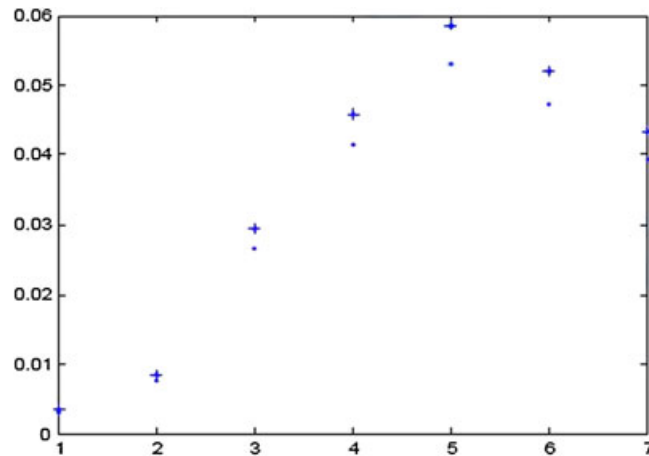


FIGURE 8 Distance between observed functions and predicted functions of long-term interest rates for US data. Root-mean-square error (RMSE) plotted over maturity. The plus marks denote errors using only the time series of long-term maturity rates. The dots are errors using time series and regression on short-term maturity rates [Colour figure can be viewed at wileyonlinelibrary.com]

7 | FUTURE DIRECTIONS

We have developed the method for studying the time series of functional data. It should be straightforward to extend this method to incorporate seasonality. Further work needs to be done for extensions to nonstationary time series, eg, change point, regime switching, heavy-tailed innovations, etc. One can use the methods of Aneiros-Pérez and Vieu³⁴ and Damon and Guillas³⁵ to incorporate the effects of covariates. Another recent paper based on ARH(p) model is the work of Kowal et al,³⁶ which uses hierarchical Gaussian process model that is especially suited for sparsely sampled curves with nonnegligible measurement error. It will be interesting to try this method for ARMAH(p,q), particularly since ARH(p) with additive white noise is ARMAH(p,p) as shown in Proposition 1 in the work of Kowal et al.³⁶ On the applications side, one can use similar techniques to model and forecast the yield curves of corporate bonds, which have been studied using hierarchical Bayesian models in the work of Cruz-Marcelo et al.³⁷

ORCID

Claudia Klüppelberg  <https://orcid.org/0000-0002-0189-2384>

REFERENCES

1. Ramsay JO, Silverman BW. *Functional Data Analysis*. New York, NY: Springer Science+Business Media; 2005. *Springer Series in Statistics*.
2. Besse PC, Cardot H, Stephenson DB. Autoregressive forecasting of some functional climatic variations. *Scand J Stat*. 2000;27(4):673-687.
3. Ramsay JO. Functional components of variation in handwriting. *J Am Stat Assoc*. 2000;95(449):9-15.

4. Ratcliffe SJ, Heller GZ, Leader LR. Functional data analysis with application to periodically stimulated foetal heart rate data. II: functional logistic regression. *Statist Med.* 2002;21(8):1115-1127.
5. Rossi N, Wang X, Ramsay JO. Nonparametric item response function estimates with the EM algorithm. *J Educ Behav Stat.* 2002;27(3):291-317.
6. Reiss PT, Ogden RT. Functional principal component regression and functional partial least squares. *J Am Stat Assoc.* 2007;102(479):984-996.
7. Castro PE, Lawton WH, Sylvestre EA. Principal modes of variation for processes with continuous sample curves. *Technometrics.* 1986;28(4):329-337.
8. Rice JA, Silverman BW. Estimating the mean and covariance structure nonparametrically when the data are curves. *J R Stat Soc Ser B Methodol.* 1991;53(1):233-243.
9. Kargin V, Onatski A. Curve forecasting by functional autoregression. *J Multivar Anal.* 2008;99(10):2508-2526.
10. Hörmann S, Kokoszka P. Weakly dependent functional data. *Ann Stat.* 2010;38(3):1845-1884.
11. Bosq D. *Linear Processes in Function Spaces: Theory and Applications.* Vol. 149. New York, NY: Springer Science+Business Media; 2012.
12. Bosq D. Computing the best linear predictor in a Hilbert space. Applications to general ARMAH processes. *J Multivar Anal.* 2014;124:436-450.
13. Aue A, Norinho DD, Hörmann S. On the prediction of stationary functional time series. *J Am Stat Assoc.* 2015;110(509):378-392.
14. Hays S, Shen H, Huang JZ. Functional dynamic factor models with application to yield curve forecasting. *Ann Appl Stat.* 2012;6(3):870-894.
15. Diebold FX, Li C. Forecasting the term structure of government bond yields. *J Econom.* 2006;130(2):337-364.
16. Müller HG, Stadtmüller U, Yao F. Functional variance processes. *J Am Stat Assoc.* 2006;101(475):1007-1018.
17. Yao F, Müller HG, Wang JL. Functional data analysis for sparse longitudinal data. *J Am Stat Assoc.* 2005;100(470):577-590.
18. Heppenger P. Option pricing in Hilbert space-valued jump-diffusion models using partial integro-differential equation. *SIAM J Financial Math.* 2010;1(1):454-489.
19. Hull J, White A. Pricing interest-rate-derivative securities. *Rev Financial Stud.* 1990;3(4):573-592.
20. Heath D, Jarrow R, Morton A. Bond pricing and the term structure of interest rates: a new methodology for contingent claims valuation. *Econometrica.* 1992;60(1):77-105.
21. Vasicek O. An equilibrium characterization of the term structure. *J Financial Econ.* 1977;5(2):177-188.
22. Cox JC, Ingersoll JE Jr, Ross SA. A theory of the term structure of interest rates. *Econometrica.* 1985;53(2):385-407.
23. Duffie D, Kan R. A yield-factor model of interest rates. *Math Financ.* 1996;6(4):379-406.
24. Fan J, Gijbels I. *Local Polynomial Modelling and Its Applications.* London, UK: Chapman and Hall; 1996.
25. Yao F, Müller HG, Clifford AJ, et al. Shrinkage estimation for functional principal component scores with application to the population kinetics of plasma folate. *Biometrics.* 2003;59(3):676-685.
26. Brockwell PJ, Davis RA. *Time Series: Theory and Methods.* New York, NY: Springer Science+Business Media; 2009.
27. Akaike H. An information criterion (AIC). *Math Sci.* 1976;14(153):5-9.
28. Aoki M, Havenner A. State space modeling of multiple time series. *Econ Rev.* 1991;10(1):1-59.
29. Brown RH, Schaefer SM. The term structure of real interest rates and the Cox, Ingersoll, and Ross model. *J Financial Econ.* 1994;35(1):3-42.
30. Romer CD, Romer DH. Federal reserve information and the behavior of interest rates. *Am Econ Rev.* 2000;90(3):429-457.
31. Campbell JY. Some lessons from the yield curve. *J Econ Perspect.* 1995;9(3):129-152.
32. Ellingsen T, Söderström U. Monetary policy and market interest rates. *Am Econ Rev.* 2001;91(5):1594-1607.
33. Müller H-G, Sen R, Stadtmüller U. Functional data analysis for volatility. *J Econom.* 2011;165(2):233-245.
34. Aneiros-Pérez G, Vieu P. Nonparametric time series prediction: a semi-functional partial linear modeling. *J Multivar Anal.* 2008;99(5):834-857.
35. Damon J, Guillas S. The inclusion of exogenous variables in functional autoregressive ozone forecasting. *Environmetrics.* 2002;13(7):759-774.
36. Kowal DR, Matteson DS, Ruppert D. Functional autoregression for sparsely sampled data. *J Bus Econ Stat.* 2019;37(1):97-109.
37. Cruz-Marcelo A, Ensor KB, Rosner GL. Estimating the term structure with a semiparametric Bayesian hierarchical model: an application to corporate bonds. *J Am Stat Assoc.* 2011;106(494):387-395.
38. Kolmogorov AN, Rozanov YA. On strong mixing conditions for stationary Gaussian processes. *Theory Probab Its Appl.* 1960;5(2):204-208.
39. Masry E, Fan J. Local polynomial estimation of regression functions for mixing processes. *Scand J Stat.* 1997;24(2):165-179.
40. Yao F, Lee T. Penalized spline models for functional principal component analysis. *J R Stat Soc Ser B Stat Methodol.* 2006;68(1):3-25.
41. Hung K, Alt FB. The approximation of the one-step ahead forecast error covariance for vector ARMA models. *Int J Forecast.* 1994;10(1):59-64.

How to cite this article: Sen R, Klüppelberg C. Time series of functional data with application to yield curves. *Appl Stochastic Models Bus Ind.* 2019;35:1028–1043. <https://doi.org/10.1002/asmb.2443>

APPENDIX

We begin by listing some assumptions, on moments of \tilde{f}_{it} as in (1) and smoothing bandwidths b_f and h_f as used in (11) and (12). Throughout, we consider $\Delta \rightarrow 0$ and $n \rightarrow \infty$.

- (M1) $\sup_i E[\tilde{f}_{it}]^4 < \infty$.
 (M2) $b_f \rightarrow 0$, $nb_f^4 \rightarrow \infty$, $\limsup_n^{1/2} b_f^6 < \infty$,
 $h_f \rightarrow 0$, $nh_f^6 \rightarrow \infty$, $\limsup_n^{1/2} h_f^8 < \infty$,
 $\limsup_n^{1/2} b_f \Delta < \infty$, $\limsup_n^{1/2} h_f^2 \Delta < \infty$.
 (M3) The kernel functions κ_1 and κ_2 are compactly supported kernel functions of order (0,2) and (0,0,2), respectively, as defined in eq. (25) in the work of Yao et al.¹⁷ The Fourier transforms of $\kappa_1(u)$ and $\kappa_2(u, v)$, namely, $\zeta_1(t) = \int e^{-iut} \kappa_1(u) du$ and $\zeta_2(t, s) = \int e^{-(iut+ivs)} \kappa_2(u, v) dudv$ are absolutely integrable. That is, $\int |\zeta_1(t)| dt < \infty$ and $\int \int |\zeta_2(t, s)| dt ds < \infty$.

Let \mathfrak{F}_t^k be the σ -algebra of events generated by the random functions $\{f_s(u), u \in \mathcal{T}, s \leq t \leq k\}$, and let $L_2(\mathfrak{F}_t^k)$ denote the collection of all second-order random variables, which are \mathfrak{F}_t^k measurable. The stationary process $f_t(\cdot)$ is called ρ -mixing³⁸ if

$$\sup_{U \in L_2(\mathfrak{F}_{-\infty}^0), V \in L_2(\mathfrak{F}_k^\infty)} \frac{\text{Cov}(U, V)}{\text{Var}^{1/2}(U)\text{Var}^{1/2}(V)} = \rho(k) \rightarrow 0 \quad \text{as } \rho \rightarrow \infty.$$

- (M4) The process f_t is ρ -mixing with $\sum \rho(l) < \infty$.

Proof of Theorem 1. The proof borrows arguments from the proofs of Lemmas 1, 2, and Theorem 1 in the work of Yao et al.¹⁷ Assumptions (M1)-(M3) ensure that proper versions of the abovementioned theorem apply here. The difference of this paper from the setup in the work of Yao et al.¹⁷ is that the functions are no longer independent but come from an ARMA process. Assumption (M4) allows us to suitably modify the last step of Lemma 1 to bound the variance term under this dependence structure. Compare to Masry and Fan³⁹ on local polynomial estimation of regression functions for mixing processes. \square

Proof of Theorem 2. The proof is analogous to that of Theorem 2 in the work of Yao et al.¹⁷ \square

Further assumptions are needed for the remaining results. For each $j \geq 0$, define $\delta_j^f = \frac{1}{2} \min\{|\lambda_l - \lambda_j| : l \neq j\}$, and $\Lambda_{\delta_j^f} = \{z \in C : |z - \lambda_j| = \delta_j^f\}$, where C are the complex numbers. Furthermore, define $A_{\delta_j^f} = \sup\{\|\mathbf{R}_f(z)\|_F : z \in \Lambda_{\delta_j^f}\}$, where $\mathbf{R}_f(z) = (\mathbf{G}_f - zI)^{-1}$ is the resolvent of operator \mathbf{G}_f and $\|\cdot\|_F$ is an operator norm, defined on the separable Hilbert space F generated by the Hilbert-Schmidt operators on H , endowed with the inner product $\langle T_1, T_2 \rangle_F = \sum_j \langle T_1 u_j, T_2 u_j \rangle_H$ and the norm $\|T\|_F^2 = \langle T, T \rangle_F$, where $T_1, T_2, T \in F$, and $\{u_j : j \geq 1\}$ is any complete orthonormal system in H . Then, we assume

- (M5) $K = K(n) \rightarrow \infty$. and $\sum_{j=1}^K (\delta_j^f A_{\delta_j^f} \sup_{t \in [0, T]} |\phi_j(t)|) / (\sqrt{nh_f^2} - A_{\delta_j^f}) \rightarrow 0$.
 (M6) $\sum_{j=1}^K \sup_{t \in [0, T]} |\phi_j(t)| = o(\min\{\sqrt{nb_f}, \sqrt{\Delta^{-1}}\})$, and
 $\sum_{j=1}^K \sup_{t \in [0, T]} |\phi_j(t)| \sup_{t \in [0, T]} |\phi_j'(t)| = o(\Delta^{-1})$.

Assumptions (M5) and (M6) describe how the number of included eigenfunctions K increases when n tends to infinity. The quantities δ_j reflect the decay of the eigenvalues of the covariance operators, whereas $A_{\delta_j^f}$ depend on the local properties of the covariance operator G around the eigenvalues λ_j . In practice, the eigenvalues usually decrease rapidly to zero, the number of included eigenfunctions K is much less than n ; ie, $n \ll K$, which suggests that Assumptions (M5) and (M6) can be easily fulfilled for such processes.

Proof of Theorem 3. The proof is immediate from Theorem 1 in the work of Yao and Lee.⁴⁰ \square

Proof of Theorem 4. Replacing f_t by $\hat{f}_t^{(K)}$ in (6) and following through the argument of (7)-(10), we have

$$\hat{\xi}_t = \mu_\xi + \psi_{1\xi} \hat{\xi}_{t-1} + \dots + \psi_{p\xi} \hat{\xi}_{t-p} + \epsilon_\xi + (\phi_1 \dots \phi_K)^T R_t. \tag{A1}$$

$$\begin{aligned} R_t(u) &= \hat{f}_t^{(K)}(u) - \mu - \psi_1 \left(\hat{f}_{t-1}^{(K)}(u) - \mu \right) - \dots - \psi_p \left(\hat{f}_{t-p}^{(K)}(u) - \mu \right) - \epsilon_t(u). \\ &= \hat{f}_t^{(K)}(u) - f_t(u) - \psi_1 \left(\hat{f}_{t-1}^{(K)}(u) - f_{t-1}(u) \right) - \dots - \psi_p \left(\hat{f}_{t-p}^{(K)}(u) - f_{t-p}(u) \right). \end{aligned}$$

Combining this with (32)

$$\sup_{t \in [0, T]} |R_t(u)| \leq p \left(1 + \|\psi_1\|^2 + \dots + \|\psi_p\|^2 \right) \max_{t-p \leq i \leq t} \sup_{u \in [0, T]} \left| \hat{f}_i^{(K)}(u) - f_i(u) \right| \xrightarrow{P} 0,$$

where $\|\psi_i\| = \sup_f (\psi_i(f)(u)/f(u))$ is the maximum eigenvalue of the operator ψ_i . For a vector ARMA(p, q) model with normally distributed innovations, the MSE of the forecast error is $(1 + K(p + q)/n)\Sigma_\epsilon$ as shown and given in the work of Hung and Alt.⁴¹ Following the same argument and combining with (A1) using Slutsky's theorem gives the result. □