Functional Autoregressive Model using Signal Compression

Husneara Rahman* Xin Qi†

Abstract

Modern advancement of technology permits us to accumulate more complicated data than before. Unlike the traditional time series where only a scalar or a vector is observed at each time point, in functional time series, a curve is observed at each time point. Correlation exists among the curves observed at different time points. In this paper, we consider a functional autoregressive (FAR) model with general order which is a generalization of the traditional AR model. To fit the FAR model and obtain the estimate of coefficient functions, we propose a signal compression procedure. To determine the optimal tuning parameters and optimal order of FAR model, we propose a window-shifting cross-validation procedure. We compare our method to recently proposed method on both simulated data and real data, which illustrate the good predictive performance of our method.

Key Words: Functional time series, functiona autoregressive model, signal compression, window-shifting cross-validation.

1. Introduction

Functional data analysis is an important area of modern statistics. It allows to incorporate higher dimension and volume of data than regular procedures. Good references are the books by Ramsay and Dalzell (1991), Ramsay and Silverman (2005). Functional Time Series (FTS), as a new field in functional data analysis, arise in situations where a series of functional data are observed over time. We observe such data in different areas like health, biology, economics, environment etc. Some examples are age specific incidence rates of a disease that is observed over several years, hourly consumption of electricity which is recorded for each day over several months, hourly CO_2 emission of a certain chemical plant recorded each day over the year, etc. Correlation exists among the functional observations at different time points.

Various models have been proposed to describe the dependency of the curves at different time points. Among them, a functional autoregressive (FAR) model is a class of important models which are the extension of the autoregressive (AR) models in the classic time series. An extensive discussion about theoretical background of the FAR models are provided in the monograph of Bosq (2000).

Damon and Guillas (2002) have incorporated exogenous predictors to first order functional autoregressive models. They also have discussed different error measures for functional data, functional kernel model, ARH and ARHX models. Hörmann *et al.* (2010), Hörmann and Kokoszka (2012) have generalized the linear model of Bosq (2000) to incorporate non-linear dependence. They have used a moment based notion of weak dependence. Kokoszka and Reimherr (2013) have proposed a hypothesis based approach to determine the order of Functional Autoregressive Models. Aue *et al.* (2015) have discussed autoregressive process of general order and its estimation using FPCA. An useful discussion of existing works and `R` packages on functional time series are also presented.

*Department of Mathematics and Statistics, Georgia State University, Atlanta, GA 30303, U.S.A.*Email*: hrahman3@gsu.edu

†Department of Mathematics and Statistics, Georgia State University, Atlanta, GA 30303, U.S.A., *Email*: xqi3@gsu.edu

The general functional autoregressive model of order p can be expressed as

$$Y_i(t) = \sum_{j=1}^{p} \int \psi_j(t,s) Y_{i-j}(s) ds + \varepsilon_i(t), \tag{1}$$

where $\{Y_1(t), \cdots, Y_n(t)\}$ are functional time series data, $\varepsilon_i(t)$ are independent noise curves, and $\{\psi_1(t,s), \cdots, \psi_p(t,s)\}$ are coefficient kernel functions.

Different dimension reduction methods, such as FPCA, have been used to estimate FAR models. In this paper, we will apply a recently proposed dimension reduction approach proposed by Luo and Qi (2017) to the FAR models and propose new efficient estimation methods. This dimension reduction has a nice property: among all dimension reduction of the same form, it has the smallest prediction error.

To facilitate detailed understanding about functional time series models, we provide some background information in this section. We present a brief discussion on classic timeseries models along with model identification and estimation in the following subsection.

1.1 Classic time series and models

Time series data arise sequentially over time. Two major aspects of time series data analysis are to determine and estimate the underlying model of the observed series and to forecast future data of our interest from previous values. We refer the reader to the books by Box *et al.* (2015), Montgomery *et al.* (1990), Cryer and Chan (2008) for the theory and applications of the classic time series. The most popular time series models include autoregressive (AR), moving average (MA) and mixture of AR and MA models called ARMA models. Here we provide a brief introduction of these models. A moving average model of order q, $(MA(q))$ can be expressed in standard way as follows,

$$Y_t = \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - \cdots - \theta_q \epsilon_{t-q},$$

where $\{Y_t\}$ are sequence of observed time series, $\{\epsilon_t\}$ are i.i.d. random errors with zero mean and $\{\theta_1, \cdots, \theta_p\}$ are weights on the errors.
An autoregressive model of order p, $AR(p)$ can be expressed as,

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \epsilon_t,$$

where $\{\epsilon_t\}$ are random error terms and are independent of $Y_{t-1}, Y_{t-2}, \cdots Y_{t-p}$ with zero mean. A general type of model can be achieved by assuming a mixture of moving average and autoregressive models, termed as autoregressive moving average model of order p and q, $ARMA(p,q)$. A mathematical expression of $ARMA(p,q)$ model is expressed as,

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - \cdots - \theta_q \epsilon_{t-q}.$$

In practice, we do not know the underlying model and corresponding parameters for particular observed data sets. Thus model specification and order detection plays important role in timeseries data analysis. Box-Jenkin's iterative procedure by Box and Jenkins (1976) is a popular approach in this regard. They have used sample autocorrelation functions, sample partial autocorrelation functions and differencing techniques to identify underlying models. Sample autocorrelation function (sample ACF) is a useful tool to determine the order of $MA(q)$ model. For $MA(q)$ model, it has cut-off property after lag q. For large sample, these estimates follows approximate normal distribution. A test of hypothesis can be carried to test the significance of r_k, the sample autocorration at lag k, by

examining ($r_k \pm$ margin of errors). On the other hand, sample ACF does not show useful property to identify $AR(p)$ models. It gradually dies off and make it difficult to identify the model. sample partial autocorrelation funcrions (sample PACF) are useful in this situation. This function for lag k, r_{kk} is evaluated after removing the effect of interveining variables $Y_{t-1}, \cdots, Y_{t-k+1}$. The sample PACF become zero after p lag for underlying $AR(p)$ model. Sample ACF follows asymptotically normal density with zero mean and variance $1/n$ after lag p for the underlying $AR(p)$ model. Thus test of hypothesis can be carried out using $\pm 2/\sqrt{n}$ as critical values for sample PACF estimates larger than lag p. Sample ACF or PACF functions are effective for $MA(q)$ or $AR(p)$ models. For situations like mixed models, when neither ACF nor PACF cuts-off, Box and Jenkins (1976) have suggested an effective procedure to approximate a mixed model. For mixed ARMA models, graphical approaches like corned method by Beguin *et al.* (1980), smallest cannonical correlation method by Tsay and Tiao (1985), extended autocorrelation function EACF, etc. are used. The EACF procedure has good sampling property for large sample. See Cryer and Chan (2008) for details of the procedure.

Once we can determine the nature and order of underlying model, we need to estimate the model parameters. Several procedures like method of moments, least-squares estimation, maximum likelihood estimation and unconditional least squares, etc. are used. Here we describe the estimation techniques for classical stationary autoregressive models. Method of moments is implemented by solving sample $Yule - Walker$ equations, which is a well-known technique to estimate autoregression parameters. From the classic $AR(p)$ model, we can obtain the following relation for autocorrelation functions,

$$\rho_k = \phi_1 \rho_{k-1} + \cdots + \phi_k + \cdots + \rho_{p-k};\ k > 0,$$

where ρ_k is the correlation function among values at lag k. We obtain well-known $Yule - Walker$ equations by substituting k with $1, 2, \cdots, p$. These sets of equations form system of linear equations for $\phi_1, \cdots, \phi_p$. Let us assume the following for computational purpose,

$$\boldsymbol{\phi} = \begin{pmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{pmatrix},\ \boldsymbol{\rho} = \begin{pmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_p \end{pmatrix},\ \mathbf{P} = \begin{pmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{p-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{p-2} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \rho_p & \rho_{p-1} & \rho_{p-2} & \cdots & 1 \end{pmatrix},$$

then the linear system becomes, $\boldsymbol{\rho} = \mathbf{P}\boldsymbol{\phi}$. Solving for $\boldsymbol{\phi}$, we can obtain, $\boldsymbol{\phi} = \mathbf{P}^{-1}\boldsymbol{\rho}$. Substituting the sample estimates for the corresponding correlation functions in the relation, we can obtain $Yule - Walker$ estimate of the autocorration parameter of a general stationary $AR(p)$ model. More details can be found in Box *et al.* (2015).

The least-squares and maximum likelihood procedures can also be applied to estimate AR parameters. Using the least squares estimation procedure using conditional sum of squares is straitforward for autoregressive models. Unconditional least squares method requires numerical calculation. Under stationarity and large sample assumption, these procedures provide similar estimates as that of the $Yule - Walker$ equations.

We organize the remaining parts of the paper as follows. In section 2, we discuss existing functional time series models, where we emphasize our discussion on FAR models. In section 3, we describe our proposed procedure for FAR models. Discussions about the computation procedures of the proposed method is presented in section 4. Simulation studies, carried out for the comparison of the performance of proposed method with that of existing ones, are described in section 5. In section 6, we present the results of application of the FAR procedures in Australian fertility data. We conclude with findings and discussion in section 7.

2. Functional time series models

Let, $\{y_i(t)\}$ denote a functional time series (FTS) data obtained over time and indexed by $1 \leq i \leq n$. Hourly consumption of electricity data which is obtained each day, age specific growth rate data which is obtained each year, etc. are examples of functional time series data. In this section, we discuss some functional time series models and estimation procedures.

Functional Principal Componet Analysis (FPCA) is a popular dimension reduction method in functional data analysis. We also observe the use of FPCA in different literatures of functional time series. Hyndman and Ullah (2007), Hyndman and Shang (2009) have considered the following FTS model,

$$y_i(t) = f_i(t) + \sigma_i(t)\epsilon_{it},$$

where ϵ_{it} are independently and identically distributed with zero mean and unit variance. $\sigma_i(t)$ in the model allows heteroscedasticity. The smoothed functions $\{f_i(t)\}$ are decomposed using FPCA or weighted FPCA approach as,

$$f_i(t) = \mu(t) + \sum_{k=1}^{K} \beta_{ik}\phi_k(t) + e_i(t);\ i = 1, 2, \cdots, n,$$

here $\mu(t)$ is the mean function, $\phi_k(t)$ is the k-th principal component function, $\{\beta_{ik}\}$ are the corresponding coefficients, $\{e_i(t)\}$ denote independent and identically distributed random functions with zero mean, $K < n$. Hyndman and Ullah (2007) viewed each $\{\beta_{ik}; i = 1, \cdots, n\}$ as univariate time series and assumed that the K univariate time series are independent. They built model for each of them and predicted the future values of these coefficients. Then they obtained the forecast of the future curves using these predicted coefficient and the PC functions. Their approach was implemented in the `R` package FTSA (see Shang (2013) for details). Hyndman and Shang (2009) have discussed weight based approach for the estimation and forecasting purpose of functional time series, where higher weights were associated with more recent data. They also have proposed bootstrap prediction intervals for the forecasts. Hörmann *et al.* (2010), Hörmann and Kokoszka (2012) have discussed weakly dependent functional data, where they have provided a general framework of temporal dependence for functional observations. Panaretos and Tavakoli (2013) and Hörmann *et al.* (2015) have implemented dynamic functional principal component approach for functional time series. Aue *et al.* (2015) have used functional forecasting procedure using multivariate forecast algorithm for functional principal components. Functional autoregressive models are the extension of the AR models in classic time series. We review the existing FAR models and methods in the following section.

2.1 Existing FAR models and methods

A general FAR model of order p is expressed as the following form,

$$Y_i(t) = \sum_{j=1}^{p} \int \psi_j(t, s) Y_{i-j}(s) ds + \varepsilon_i(t),$$

where $\varepsilon(t)'s$ are uncorrelated errors with mean zero, $Y(t)'s$ are centralized stochastic process.

Bosq (2000) has provided theoretical details for Autoregressive process in Hilbert and Banach spaces. Let $\{X_n, n \in \mathbb{Z}\}$ be a stationary random sequence in seperable Hilbart

space $H = L^2([0,1])$ with mean μ. Then it is said to be an autoregressive Hilbertian process of order P associated with $(\mu, \epsilon, \Psi_1, \cdots, \Psi_p)$ if it is such that,

$$X_n - \mu = \Psi_1(X_{n-1} - \mu) + \cdots + \Psi_p(X_{n-p} - \mu) + \epsilon_n, \quad n \in \mathbb{Z}, \tag{2}$$

where $\epsilon = (\epsilon_n, n \in (Z))$ is a strong sense *H-white noise*, $\mu \in H$, and $\Psi_1, \cdots, \Psi_p \in \mathcal{L}$, the space of bounded linear operators in H and $\Psi_p \neq 0$.

Let, $Y = (Y_n, n \in \mathbb{Z})$, where $Y_n = (X_n, \cdots, X_{n-p+1})$; $\mu^{'} = (\mu, \cdots, \mu) \in H^p$, $\epsilon^{'} = (\epsilon_n^{'}, n \in \mathbb{Z})$, $\epsilon_n^{'} = (\epsilon_n, 0, \cdots, 0) \in H^p, n \in \mathbb{Z}$ and the operator on H^p be defined as,

$$\Psi^{'} = \begin{pmatrix} \Psi_1 & \Psi_2 & \cdots & \Psi_{p-1} & \Psi_p \\ I & 0 & \cdots & 0 & 0 \\ 0 & I & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & I & 0 \end{pmatrix}; I \text{ is the identity operator.}$$

It is shown in Bosq (2000) that, If X is an autoregressive process of order p in H^1 associated with $(\mu, \epsilon, \Psi_1, \cdots, \Psi_p)$, then Y is an autoregressive process of order one in H^p associated with $(\mu^{'}, \epsilon^{'}, \Psi^{'})$. Thus autoregressive process of order one plays the central role in functional linear autoregressive process.

There exists a unique stationary solution of the process in equation (2), under the following assumption,

$(c_0^{'})$: $\quad ||\Psi^{'j0}||_{\mathcal{L}_p} < 1$ for some $j_0 \geq 1$,

with the scalar product in separable H^p is defined as,

$\langle (x_1, \cdots, x_p), (y_1, \cdots, y_p) \rangle_p := \sum_{j=1}^{p} \langle x_j, y_j \rangle; \{x_i; i = 1, \cdots, p\}, \{y_j; j = 1, \cdots, p\} \in H$.

Since the assumption $(c_0^{'})$ is difficult to verify and is not a necessary condition, Bosq (2000) has introduced another condition that is directly related with linear operators $\Psi_1, \cdots, \Psi_p$. Let us define,

$$Q(z) = z^p I - z^{p-1}\Psi_1 - \cdots - z\Psi_{p-1} - \Psi_p, \; z \in \mathbb{C},$$

where $Q(z)$ is a bounded linear operator over complex extension H' of H. Now the alternative condition is stated as,

$(c_1^{'})$: $\quad Q(z)$ not invertible $\implies |z| < 1$.

If condition $(c_1^{'})$ holds then $(c_0^{'})$ holds and thus a unique stationary solution of equation (2) exists. Bosq (2000) also has stated that if the following condition,

(c_2) : $\quad \sum_{j=1}^{p} \|\Psi_j\|_{\mathcal{L}} < 1$,

holds then condition $(c_1^{'})$ holds, which implies the existence of unique stationary solution of the functional autoregressive process in equation (2).

2.1.1 Estimation of FAR Models:

Functional autoregressive model of order one, FAR(1) plays the basic role in the study of functional autoregressive porcess. Bosq (2000), Hörmann and Kokoszka (2012), Horváth and Kokoszka (2012) and other authors have studied the FAR(1) model extensively. Here we discuss the existing estimation porcedure for FAR models specially FAR(1) process. Let us consider FAR(1) model from model (2),

$$X_n = \Psi(X_{n-1}) + \epsilon_n, \quad n \in \mathbb{Z},$$

where $\{X_n \in L^2([0,1])\}$, have zero mean. The lag-1 autocovariance operator be defined as:

$$C_1(x) = E[\langle X_n, x\rangle X_{n+1}], \quad x \in H.$$

C_1^T be adjoint operator so that,

$$C_1^T = E[\langle X_n, x\rangle X_{n-1}] = E[\langle \Psi(X_{n-1}, x\rangle X_{n-1}] = C(\Psi^T)$$

$$C_1 = \Psi C$$

An estimate like Yule-Walker estimation for the scalar auto regressive (AR) model can be obtained using finite sample version of the relation $\Psi = C_1 C^{-1}$. Since C is positive definite Hilbert-Schmidt operator, we have the following decomposition:

$$C(x) = \sum_{j=1}^{\infty} \lambda_j \langle x, v_j\rangle v_j,$$

where $\{v_j\}$ are orthonormal eigen functions of C with the relation $C(v_j) = \lambda_j v_j$. The relation $C^{-1}(x) = \sum_{j=1}^{\infty} \lambda_j^{-1}\langle x, v_j\rangle v_j$ makes C^{-1} unbounded as some λ_j converge to zero as $j \longrightarrow \infty$. To solve this problem, only first p important estimated eigen functions are taken.

$$\hat{\Psi} = \hat{C}_1 \sum_{j=1}^{p} \hat{\lambda}_j^{-1}\langle x, \hat{v}_j\rangle \hat{v}_j = \frac{1}{N-1}\sum_{k=1}^{N-1}\Big\langle X_k, \sum_{j=1}^{p}\hat{\lambda}_j^{-1}\langle x, \hat{v}_j\rangle \hat{v}_j\Big\rangle X_{k+1}$$

$$= \frac{1}{N-1}\sum_{k=1}^{N-1}\sum_{j=1}^{p}\hat{\lambda}_j^{-1}\langle x, \hat{v}_j\rangle\langle X_k \hat{v}_j\rangle X_{k+1},$$

is the finite sample estimate for lag 1 coefficient.

2.1.2 Order detection of FAR process

There have been extensive literature on order detection of classical auto regressive models. In case of functional autoregressive models, very few discussions about the methodology of order determination have been found. Bosq (2000) have discussed two possible approaches. In Section 5.5 of Bosq (2000), a partial theoretical solution for a special case using **Shibata-Mourid** statistics has been described. An empirical method of order detection was discussed for the general case. We provide a brief overview of the procedure. Based on predictive ability of the model, order p of a $ARH(p)$ model can be determined for a number of choices of order p from $\{1, 2, \cdots, p_{max}\}$. The estimate $\hat{p}$ of order p using n observations is,

$$\hat{p}_n = \underset{1 \le p \le p_{max}}{\arg\min}\ \hat{\sigma}_n^2(p),$$

where $\hat{\sigma}_n^2(p)$ is the estimate of prediction error using n ovservations and $ARH(p)$ model. Details of the procedure can be found in Section 9.2 of Bosq (2000).

Kokoszka and Reimherr (2013) have introduced hypothesis testing procedure to determine order of functional autoregressive models. They described model assumptions and representation for the FAR(p) process. The kernel matrix was estimate using least-squared estimation procedure. Sequence of hypothesis testing was proposed to estimate the order of FAR process with a chi-square approximation to the test statistic.

3. FAR using Signal compression approach

In this section, we describe our proposed procedure for functional autoregressive process. We present the discussion by introducting the signal compression approach proposed by Luo and Qi (2017) in section 3.1. In section 3.2, we describe the proposed procedure for FAR models followed by corresponding sample estimates in section 3.3.

3.1 Signal compression approach

Signal compression approach for funcion-on-function linear model by Luo and Qi (2017) is an efficient dimension reduction framework. A general function-on-function linear regression model can be expressed as follows,

$$Y = \mu + \sum_{q=1}^{Q} \Psi_q(X_q) + \varepsilon,$$

$$\text{or, } Y(t) - \mu(t) = \sum_{q=1}^{Q} \int_{a_q}^{b_q} \psi_q(t, s_q) X_q(s_q) ds_q + \varepsilon(t), \quad c \leq t \leq d, \tag{3}$$

where $\{X_1, \cdots, X_Q\}$ are Q predictor functions with zero mean, $Y(t); c \leq t \leq d$ is the observed response function, $\varepsilon(t)'s$ are independent noise functions with mean zero, these are independent of all the Q predictor functions and $\{\psi_q \in \mathcal{L}^2; 1 \leq q \leq Q\}$ are coefficient kernel functions. The signal part of model (1) is, $F(t) = \sum_{j=1}^{p} \int \psi_j(t, s) X_q(s) ds$. Thus the model can be re-expressed as,

$$Y(t) = F(t) + \varepsilon(t),$$

where, without loss of generality, we assume that $Y(t)'s$ are centralized stochastic process. Let, $\mathbf{\Lambda}(t, t')$ is the cross-covariance function of $(F(t), F(t'))$ and $\mathbf{\Sigma_{qp}}(s_q, s_p')$ is the cross-covariance funtion of $(X_q(s_q), X_p(s_p'))$. Let, $\widetilde{\boldsymbol{\phi}}_{\boldsymbol{k}}, k \geq 1$ be scaled eigen functions of $\mathbf{\Lambda}(t, t')$ with σ_k^2 as corresponding eigen value. The Karhunen-Loeve (KL) expansion of $F(t)$ is $F(t) = \sum_{k=1}^{\infty} \widetilde{Z}_k \widetilde{\phi}_k(t)$, where $\widetilde{Z}_k$ are uncorrelated random variables with mean zero and variance σ_k^2. The truncated KL expansion have minimum mean integrated squared error. This leads to the best finite dimensional approximation of $F(t)$ as $\sum_{k=1}^{K} Z_k \phi_k(t) = \sum_{k=1}^{K} \widetilde{Z}_k \widetilde{\phi}_k(t)$, where $Z_k = \widetilde{Z}_k / \sigma_k$ and $\phi_k(t) = \sigma_k \widetilde{\phi}_k(t)$. This procedure greatly reduce the number of parameters to be estimated, improve the computational efficiency and prediction accuracy. Let $\hat{\psi}_q^K$ be estimates of karnel functions using the K dimensional approximation of F and $y_{pred}^{(K)}(t)$ be the prediction based on this. Then, for any positive integer K, $y_{pred}^{(K)}(t)$ has the smallest mean integrated square prediction error among all predicted functions of the form $\tilde{y}_{pred} = \sum_{q=1}^{Q} \int_{a_q}^{b_q} \tilde{\psi}_q^K(t, s_q) x_{new,q}(s_q) ds_q$ with $\tilde{\psi}_q^K(t, s) = \sum_{i=1}^{K} \sum_{j=1}^{K} a_{ij}^{(q)} \xi_i(t) \zeta_{jq}(s)$, where $\xi_i(t), \zeta_{jq}(s)$ are arbitrary square integrable functions and $a_{ij}^{(q)}$ is an arbitrary number.

That is for any positive integer K, $(\sum_{k=1}^{K}\phi_k(t)\varphi_{k1}(s_1),\cdots,\sum_{k=1}^{K}\phi_k(t)\varphi_{kQ}(s_Q))$, as an estimate of $(\psi_1(t,s),\cdots,\psi_Q(t,s))$ of model 3, has the smallest prediction error among all estimates of the form

$$(\sum_{i=1}^{K}\sum_{j=1}^{K}a_{ij}^{(1)}\xi_i(t)\zeta_{j1}(s_1),\sum_{i=1}^{K}\sum_{j=1}^{K}a_{ij}^{(2)}\xi_i(t)\zeta_{j2}(s_2),\cdots,\sum_{i=1}^{K}\sum_{j=1}^{K}a_{ij}^{(Q)}\xi_i(t)\zeta_{jQ}(s_Q));$$
$$1\le i,j\le K,\quad\text{and}\quad 1\le q\le Q.$$

Here $\xi_1(t),\cdots,\xi_K(t)$, $\zeta_{11}(s),\cdots,\zeta_{KK}(s)$ are arbitrary square integrable functions and $a_{11}^{(1)},\cdots,a_{KK}^{(Q)}$ are arbitrary numbers. $\{\boldsymbol{\varphi}_k:k\ge 1\}$ are charactarized as the solutions to a generalized eigenvalue problem. Assuming the multiplicity of each positive eigenvalue σ_k^2 of $F(t)$ is one, $\varphi_{\mathbf{k}}=(\varphi_{k1}\cdots\varphi_{kQ})$ is the solution to

$$\boldsymbol{\varphi}=\max_{\varphi_1\cdots\varphi_Q}\sum_{q=1}^{Q}\sum_{p=1}^{Q}\int_{a_q}^{b_q}\int_{ap}^{b_p}\varphi_q(s_q)\mathbf{B_{qp}}(s_q,s_p')\varphi_p(s_p')ds_pds_p';\tag{4}$$

$$\text{subject to}\quad\sum_{q=1}^{Q}\sum_{p=1}^{Q}\int_{a_q}^{b_q}\int_{a_p}^{b_p}\varphi_q(s_q)\boldsymbol{\Sigma_{qp}}(s_q,s_p')\varphi_p(s_p')ds_qds_p'=1,$$

$$\text{and}\quad\sum_{q=1}^{Q}\sum_{p=1}^{Q}\int_{a_q}^{b_q}\int_{a_p}^{b_p}\varphi_q(s_q)\boldsymbol{\Sigma_{qp}}(s_q,s_p')\varphi_{lp}(s_p')ds_qds_p'=0\quad\text{for all}\quad 1\le l\le k-1.$$

Where the maximum value is σ_k^2.

Now suppose that we have n independent observations $\{(y_i(t),x_{i1}(t),x_{i2}(t)\cdots,x_{iQ}(t)),1\le i\le n\}$ satisfying

$$\begin{aligned}y_i(t)&=\mu(t)+\sum_{q=1}^{Q}\int_{a_q}^{b_q}\psi_q(t,s_q)x_{iq}(s_q)ds_q+\varepsilon_i(t)\\&=\mu(t)+\sum_{k:\sigma_k>0}z_{ik}\phi_k(t)+\varepsilon_i(t);\quad 1\le i\le n,\end{aligned}\tag{5}$$

where $z_{ik}=\sum_{q=1}^{Q}\int_{a_q}^{b_q}x_{iq}(s_q)\varphi_{kq}(s_q)ds_q$.
Let us define,

$$\widehat{\mathbf{B}}_{qp}(s_q,s_p')=\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}[x_{iq}(s_q)-\bar{x}_q(s_q)]\left[\int_c^d[y_i(r)-\bar{y}(r)][y_j(r)-\bar{y}(r)]dr\right][x_{jp}(s_p')-\bar{x}_p(s_p')],$$

$$\widehat{\boldsymbol{\Sigma}}_{qp}(s_q,s_p')=\frac{1}{n}\sum_{i=1}^{n}[x_{iq}(s_q)-\bar{x}_q(s_q)][x_{ip}(s_p')-\bar{x}_p(s_p')],\quad 1\le p,\quad q\le Q.\tag{6}$$

Then the estimates $\widehat{\boldsymbol{\varphi}}_k=(\widehat{\varphi}_{k1}(s_1),\widehat{\varphi}_{k2}(s_2),\cdots,\widehat{\varphi}_{kQ}(s_Q))$ of $\boldsymbol{\varphi}_k$, $1\le k\le K$, are obtained by solving

$$\max_{(\varphi_1,\cdots,\varphi_Q)}g(\varphi_1,\cdots,\varphi_Q),\quad\text{subject to}\quad\sum_{q=1}^{Q}\sum_{p=1}^{Q}\int_{a_q}^{b_q}\int_{a_p}^{b_p}\varphi_q(s_q)\widehat{\boldsymbol{\Sigma}}_{qp}(s_q,s_p')\varphi_p(s_p')ds_qds_p'=1,$$

$$\text{and}\quad\sum_{q=1}^{Q}\sum_{p=1}^{Q}\int_{a_q}^{b_q}\int_{a_p}^{b_p}\varphi_q(s_q)\widehat{\boldsymbol{\Sigma}}(s_q,s_p')\widehat{\varphi}_{lp}(s_p')ds_qds_p'=0,\quad\text{for}\quad 1\le l\le k-1,\tag{7}$$

where the objective function $g(\varphi_1, \varphi_2, \cdots, \varphi_Q)$ is given by

$$\frac{\sum_{q=1}^{Q}\sum_{p=1}^{Q}\int_{a_q}^{b_q}\int_{a_p}^{b_p}\varphi_q(s_q)\widehat{\mathbf{B}}_{qp}(s_q,s_p')\varphi_p(s_p')ds_q ds_p'}{\sum_{q=1}^{Q}\sum_{p=1}^{Q}\int_{a_q}^{b_q}\int_{a_p}^{b_p}\varphi_q(s_q)\widehat{\mathbf{\Sigma}}_{qp}(s_q,s_p')\varphi_p(s_p')ds_q ds_p' + \sum_{q=1}^{Q}\lambda\left[\int_{a_q}^{b_q}\varphi_q(s_q)^2 ds_q + \tau\int_{a_q}^{b_q}\varphi_q''(s_q)^2 ds_q\right]}.$$

Let $\widehat{z}_{ik} = \sum_{q=1}^{Q}\int_{a_q}^{b_q}(x_{iq}(s_q) - \bar{x}_q(s_q))\widehat{\varphi}_{kq}(s_q)ds_q$, $1 \le k \le K$ and $1 \le i \le n$. The estimates of $\mu(t), \phi_1(t), \cdots, \phi_k(t)$ are obtained by solving linear function-on-scalar regression model 5. In the model, the vector of sample response functions is $\mathbf{y}(t) = (y_1(t), \cdots, y_n(t))^{\mathrm{T}}$ with the predictor vectors $\mathbf{1}_n, \widehat{\mathbf{z}}_1, \cdots, \widehat{\mathbf{z}}_K$, where $\mathbf{1}_n$ is an n-dimensional vector with all elements equal to one and $\widehat{\mathbf{z}}_k = (\widehat{z}_{1k}, \cdots, \widehat{z}_{nk})^{\mathrm{T}}; 1 \le k \le K$. Using penalized least squares method as in Chapter 13 of Ramsay and Silverman (2005), estimates can be obtained by solving,

$$\min_{\substack{\mu(t),\\ \phi_1(t),\cdots,\phi_K(t)}}\left[\frac{1}{n}\sum_{i=1}^{n}\int_c^d\left\{y_i(t) - \mu(t) - \sum_{k=1}^{K}\widehat{z}_{ik}\phi_k(t)\right\}^2 dt + \eta\int_c^d\left\{\mu''(t)^2 + \sum_{k=1}^{K}\phi_k''(t)^2\right\}dt\right], \tag{8}$$

The minimum is taken over all possible functions $\mu(t)$ and $\phi_i(t)$, for $1 \le i \le K$, with square-integrable second derivatives in $[c, d]$ and tuning parameter η. Let $\widehat{\mu}_0(t), \widehat{\phi}_1(t), \cdots \widehat{\phi}_K(t)$ be the estimates obtained by solving equation 8. These estimates can be obtained seperately because of orthogonality property of the predictor vectors. Thus, $\widehat{\mu}_0(t)$ and $\widehat{\phi}_k(t)$ are the solutions to

$$\min_{\mu(t)}\left[\int_c^d |\mu(t) - \bar{y}(t)|^2 dt + \eta\int_c^d |\mu''(t)|^2 dt\right], \tag{9}$$

$$\min_{v_k(t)}\left[\int_c^d |v_k(t) - \widehat{\phi}_k^0(t)|^2 dt + \eta\int_c^d |v_k''(t)|^2 dt\right], \quad 1 \le k \le K, \tag{10}$$

respectively, where $\widehat{\phi}_k^0(t) = \frac{1}{n}\widehat{\mathbf{z}}_k^{\mathrm{T}}\mathbf{y}(t)$ is the least squares estimate of $\phi_k(t)$ without smoothness penalty. Finally, for each $1 \le q \le Q$, $\widehat{\psi}_q^K(t, s_q) = \widehat{\phi}_1(t)\widehat{\varphi}_{1q}(s_q) + \widehat{\phi}_2(t)\widehat{\varphi}_{2q}(s_q) + \cdots + \widehat{\phi}_K(t)\widehat{\varphi}_{Kq}(s_q)$ and $\widehat{\mathbf{F}}^K(t) = \sum_{q=1}^{Q}\int_a^b [\mathbf{x_q}(s) - \bar{x}_q(s_q)\mathbf{1}_n]\,\widehat{\psi}_q^K(t, s_q)ds_q$ which is the estimate of $\mathbf{F}(t) = \sum_{q=1}^{Q}\int_a^b \mathbf{x_q}(s_q)\psi_q(t, s_q)ds_q$, the vector of n sample signal functions, where $\mathbf{x_q}(s_q) = (x_{q1}(s), \cdots, x_{qn}(s))^{\mathrm{T}}$ and $\bar{x}_q(s_q) = \frac{1}{n}\sum_{i=1}^{n} x_{qi}(s_q)$.

$\{\boldsymbol{\varphi}_k : 1 \le k \le K\}$ are estimated sequentially, and $\phi_k(t)$, $1 \le k \le K$ are estimated after $\{\boldsymbol{\varphi}_k : 1 \le k \le K\}$ have been estimated. $\phi_1(t), \cdots, \phi_K(t)$ can be estimated separately. These greatly reduce the number of parameters to be estimated in each step. This procedure also improve the computational efficiency and the prediction accuracy. We use this procedure for functional autoregressive models. The following sections illustrates the procedure for FAR(1) model.

3.2 SigComp Procedure for FAR models

We consider FAR(1) procedure:

$$Y_n = \Psi(Y_{n-1}) + \epsilon_n, \tag{11}$$

here $(Y_n; n \in N)$ are stochastic processes observed over $t \in [0, 1]$, $\Psi \in \mathcal{L}$, the space of bounded and continuous linear operators on H with $\Psi(y)(t) = \int_0^1 \psi(t, s)y(s)ds$. The error term $(\epsilon_n; n \in N)$ are random noise processes with zero mean, $E(\|\epsilon_n\|^2) < \infty$ and are independent of Y_{n-1}. Without loss of generality, we assume that $E(Y_n) = 0\ \forall\ n \in$

$N, t \in [0, 1]$.
We can re-express the model in equation (11) as follows:

$$\begin{aligned} Y_n(t) &= \Psi(Y_{n-1})(t) + \epsilon_n(t) \\ &= \int_0^1 \psi(t,s) Y_{n-1}(s) ds + \epsilon_n(t), \end{aligned}$$

with signal part of the model as $\int_0^1 \psi(t,s) Y_{n-1}(s) ds$. The Kernel operator $\psi(t,s) \in L^2\{[0,1] \times [0,1]\}$ satisfies $\int\int \psi^2(t,s)\, dt\, ds < 1$. we have the response functions as $Y = Y_2, \cdots, Y_n$ and the predictors as $X = Y_1, \cdots, Y_{n-1}$. In order to improve prediction accuracy, we want to get the best finite dimentional approximation to the signal function.

Let us denote $\Lambda(t,t') = E[F(t)F(t')] = \int_0^1 \int_0^1 \psi(t,s)\Sigma(s,s')\psi(t',s') ds ds'$ as the covariance function of $F(t)$ with $\Sigma(u,u') = E[X(u)X(v)]$ as the covariance function of $X(s)$. For continuous $\Lambda(t,t')$, the KL expansion of $F(t)$ is given by, $F(t) = \sum_{k=1}^{\infty} \tilde{Z}_k \tilde{\phi}_k(t)$, where $\tilde{\phi}_k(t)$ are unit norm eigen functions of $\Sigma(t,t')$ with eigen values $\sigma_1^2 \geq \sigma_2^2 \geq \cdots \geq 0$ and $\{\tilde{Z}_k = \int_0^1 F(t)\tilde{\phi}_k(t)dt; k \geq 1\}$ be sequence of uncorrelated random variables with zero mean and $E(\tilde{Z}_k^2) = \sigma_k^2$. This KL expansion has the minimum mean integrated square error. To get the unit variance random variables $\{Z_k\}$, we consider the scaled versions as $Z_k = \tilde{Z}_k/\sigma_k$ and $\phi_k(t) = \sigma_k \tilde{\phi}_k(t)$ for $k \geq 1$ such that $\sigma_k > 0$. Then according to Theorem 3 of Luo and Qi (2017), $\sum_{k=1}^K Z_k \phi_k(t) = \sum_{k=1}^K \tilde{Z}_k \tilde{\phi}_k(t)$ is the best $K-$dimensional approximation to $F(t)$ for any $K > 0$. The transformed function-on-scalar form of model 11 becomes as follows:

$$Y(t) = \sum_{k:\sigma_k > 0} Z_k \phi_k(t) + \epsilon(t), \tag{12}$$

where $\{Z_k\}$ be uncorrelated scalar predictors and $\{\phi_k\}$ be coefficient functions. We want to make prediction on $Y(t)$ based on these scalar $\{Z_k\}$.

If we define $\xi_k(s)$ as $\xi_k(s) = \frac{1}{\sigma_k}\int_0^1 \tilde{\phi}_k(t)\psi(t,s)dt = \frac{1}{\sigma_k^2}\int_0^1 \phi_k(t)\psi(t,s)dt$, for integers $k > 0$ such that $\sigma_k^2 > 0$, then Z_k can be defined as,

$$Z_k = \sigma_k^{-1}\tilde{Z}_k = \sigma_k^{-1}\int_0^1 [F(t)]\tilde{\phi}_k(t)dt = \sigma_k^{-1}\int_0^1 \Big[\int_0^1 \psi(s,t)X(s)ds\Big]\tilde{\phi}_k(t)dt = \int_0^1 X(s)\xi_k(s)ds.$$

In practice, in order to avoid too small values of σ_k and for some other computation issues, we consider only first few reasonable number of eigen functions and thus can estimate only few $\{Z_k \; ; k = 1, \cdots\}$. These gives the finite dimensional approximation of $F(t)$ as $\sum_{k=1}^K Z_k\phi_k(t) = \int_0^1 \psi^K(s,t)X(s)ds$, with $\psi^K(t,s) = \sum_{k=1}^K \phi_k(t)\xi_k(s)$. Based on Theorem 3 of Luo and Qi (2017), $\psi^K(t,s)$ has the minimum prediction error over a large number of families and the expression $\int_0^1 \psi^K(s,t)X(s)ds$ is the best K dimensional expression of $F(t)$ for finite $K > 0$.

In order to get an estimate of the model, we first get estimate of $\xi_k(s)$, then the estimate of $\phi_k(t)$. Assuming that the multiplicity of each $(\sigma_k^2 > 0)$ is one, $\xi_k(t)$ can be obtained as the solution of the generalized eigen value problem. This solution may not be unique but will have the minimum mean integrated squared prediction error property. Now we discuss the sample version of the procedure in the following section.

3.3 Sample estimates

After theoretical demonstration in the previous section, we relate the stated SigComp procedure with observed sample curves in this section. Let us assume that there are n observed curves $\{y_i(t)\}$, which can be expressed as $n-1$ pairs of sample curves as $\{(y_i(t),\, y_{i-1}(s));\, 2 \leq$

$i \leq n\}$. For the underlying $AR(1)$ model, observations for dependend or response function $Y(t)$ are $\{y_2(t), y_3(t), \cdots, y_n(t);\ 0 \leq t \leq 1\}$ and observations from explanatory or covariate functions $X(s)$ are $\{y_1(s), y_2(s), \cdots, y_{n-1}(s);\ 0 \leq s \leq 1\}$. Without losss of generality, we assume that the obsevations are centered from the mean. The observed $AR(1)$ model can be expressed as,

$$y_i(t) = \sum_{k:\sigma_k>0} z_{ik}\phi_k(t) + \epsilon_i(t);\ 2 \leq i \leq n,$$

where z_{ik} is the *i*-th observation for Z_k. Let $\hat{\mathbf{\Sigma}}(s,s') = \frac{1}{n-1}\sum_{i=2}^{n} y_{i-1}(s)\, y_{i-1}(s')$ be the sample covariance function of the covariate functions. Since we can not observe the signals $F(t)$, the estimates $\hat{\mathbf{B}}$ for $\mathbf{B}$ are obtained as,

$$\hat{\mathbf{B}}(s,s') = \frac{1}{(n-1)^2}\sum_{i=2}^{n}\sum_{j=2}^{n} y_{i-1}(s)\big[\int_0^1 y_i(r)\, y_j(r)\, dr\big] y_{j-1}(s').$$

It has been shown in Luo and Qi (2017) that,

$$\sum_{i-2}^{n} y_{i-1}(s)y_i(r)/(n-1) = \sum_{i-2}^{n} y_{i-1}(s)f_i(r)/(n-1) + \sum_{i=2}^{n} y_{i-1}(s)\epsilon_i(r)/(n-1)$$

converges to $E[X(s)F(r)]$ as the error terms are independent of the covariates and $f_i(t) = \int_0^1 \psi(s,t)y_{i-1}(s)ds$ is the sample signal function. Now, the sample estimates $\{\hat{\xi}_k;\ k = 1, 2, \cdots\}$ of $\{\xi_k;\ k = 1, 2, \cdots\}$ can be obtained as follows:

$$\begin{aligned}\hat{\xi}_k = &\max_{\psi} \frac{\int_0^1\int_0^1 \xi(s)\hat{\mathbf{B}}(s,s')\xi(s')dsds'}{\int_0^1\int_0^1 \xi(s)\hat{\mathbf{\Sigma}}(s,s')\xi(s')dsds' + \lambda\big[\int_0^1 \xi(s)^2ds + \int_0^1 \xi''(s)^2ds\big]} \\ &\text{subject to} \int_0^1\int_0^1 \xi(s)\hat{\mathbf{\Sigma}}(s,s')\xi(s')dsds' = 1, \\ &\text{and} \int_0^1\int_0^1 \xi(s)\hat{\mathbf{\Sigma}}(s,s')\xi_l(s')dsds' = 0; \forall\ 1 \leq l \leq k-1.\end{aligned} \tag{13}$$

After determining reasonable $K > 0$, we estimate the K functions $\hat{\xi}_1(s), \cdots, \hat{\xi}_K(s)$ and corresponding predictors $z_{i1}, \cdots, z_{iK}$ where $z_{ik} = \int_0^1 y_{i-1}(s)\hat{\xi}_k(s)ds$, for $2 \leq i \leq n$ and $1 \leq k \leq K$. Now the coefficient functions $\{\xi_k(t); 1 \leq k \leq K\}$ of the function-on-scalar regression model (12) is estimated using the observed sample response functions $\mathbf{y}(t) = (y_2(t), \cdots, y_n(t))$ and K predictors as $\{\hat{\mathbf{z}}_k = (\hat{z}_{2k}, \cdots, \hat{z}_{nk})^T;\ k = 1, 2, \cdots K\}$.

The co-efficients functions can be estimated using penalized least square method stated in Ramsay and Silverman (2005),

$$\min_{\phi_1(t),\cdots,\phi_K(t)} \big[\frac{1}{n}\sum_{i=2}^{n}\int_0^1 \big\{y_i(t) - \sum_{k=1}^{K}\hat{z}_{ik}\phi_k(t)\big\}^2 dt + \eta\int_0^1 \big\{\sum_{k=1}^{K}\phi_k''(t)^2\big\}dt\big], \tag{14}$$

where η is the tuning parameter. The estimates $\{\hat{\phi}_k(t);\ 1 \leq k \leq K\}$ can be found separately since $\{\hat{\mathbf{z}}_k\}$ are orthogonal. This greatly reduce the computational cost. Moreover, it can be shown that, $\{\hat{\phi}_k\}$ are the solutions of,

$$\min_{v_k(t)} \big[\int_0^1 |v_k(t) - \hat{\phi}_k^0|^2 dt + \eta\int_0^1 |v_k''(t)|^2 dt\big],\ 1 \leq k \leq K, \tag{15}$$

where $\hat{\phi}_k^0 = \frac{1}{n}\hat{\mathbf{z}}_k^T\mathbf{y}(t)$ is the simple unconstrained least squares estimate of $\phi(t)$. Using the solutions, estimate of $\psi^K(t,s)$ can be obtained as, $\hat{\psi}^K(t,s) = \sum_{k=1}^{K}\hat{\phi}_k(t)\hat{\xi}_k(s)$ and

the estimate of the signal function $\mathbf{F}(t) = \int_0^1 \mathbf{x}(s)\psi(t,s)ds$ can be obtained as $\hat{\mathbf{F}}^K(t) = \int_0^1 \mathbf{x}(s)\hat{\psi}^K(t,s)ds$, where $\mathbf{x}(s) = (y_1(s), \cdots, y_{n-1}(s))^T$. In this section, we have described the procedure for functional autoregressive model of order one. Higher order models can be estimated in similar procedure.

4. Computational procedures

In this section, the computational procedures of solving the optimization problems, choosing number of components and tuning parameters, the cross validation procedure will be discussed. Due to high dimesionality of the data and complexity of constraints optimization problems, we need to incorporate some numerical techniques to obtain the estimates. For the current problem of our interest, we assume that the n functional observations $\{y_i(t)\}$ are densely and regularly observed over the interval $[0, 1]$. Each functional observations of $\{y_i(t);\ i = 1, \cdots, n\}$ are observed at a common set of l discrete observation points $a = t_1 = 0 < t_2 < \cdots < t_l < 1 = b$. Here the choice of $a = 0, b = 1$ can be assumed without loss of generality. Now, for any continuous function $g(t);\ 0 \le t \le 1$, we can approximate the integral by the weighted summation $\int_a^b g(t)dt = \sum_{k=1}^l \delta_k g(t_k)$, with the weights $\{\delta_k\}$. These weights can be chosen in different procedures. For equally spaced observation points, we use equal weights for all k, $\delta_k = (b-a)/l;\ 1 \le k \le l$. If the observation points are unequally spaced, then we can choose the weights using trapezoidal formula, where $\delta_1 = (t_2 - t_1)/2$, $\delta_l = (t_l - t_{l-1})/2$, and $\delta_k = (t_{k+1} - t_{k-1})/2$ for $1 < k < l$.

4.1 Solving optimization problem

Let $\{b_1, b_2, \cdots, b_m\}$ be m basis functions in $L^2[0,1]$. We need to choose a large enough number for m. The optimization problem (13) can be solved in the space spanned by these basis functions. Then any $\xi(s)$ can be expressed as $\xi(s) = \sum_{j=1}^m c_j b_j(s) = \mathbf{c}^T\mathbf{b}$, where $\mathbf{b} = (b_1, \cdots, b_m)^T$ and $\mathbf{c} = (c_1, \cdots, c_m)^T$. Using this basis function expansion, the numerator of the objective function of (13) can be expressed as follows:

$$\int_0^1\int_0^1 \xi(s)\hat{\mathbf{B}}(s,s')\xi(s')dsds' = \mathbf{c}^T\Big[\int_0^1\int_0^1 \mathbf{b}(s)\hat{\mathbf{B}}(s,s')\mathbf{b}(s')^T dsds'\Big]\mathbf{c}.$$

Using the numerical approximation of integrals stated above, we have,

$$\begin{aligned}
&\int_0^1\int_0^1 \mathbf{b}(s)\hat{\mathbf{B}}(s,s')\mathbf{b}(s')^T dsds' \\
&= \int_0^1\int_0^1 \mathbf{b}(s)\frac{1}{(n-1)^2}\sum_{i=2}^n\sum_{j=2}^n y_{i-1}(s)\Big[\int_0^1 y_i(r)\, y_j(r)\, dr\Big]y_{j-1}(s')\mathbf{b}(s')^T dsds' \\
&\approx \mathbf{Q} = \frac{1}{(n-1)^2}\sum_{i=2}^n\sum_{j=2}^n\Big[\sum_{k1=1}^l \mathbf{b}(s_{k1})y_{i-1}(s_{k1})\delta_{k1}\Big]\Big[\sum_{k2=1}^l y_i(r_{k2})y_j(r_{k2})\delta_{k2}\Big]\Big[\sum_{k3=1}^l \mathbf{b}^T(s'_{k3})y_{j-1}(s'_{k3})\delta_{k3}\Big].
\end{aligned}$$

Thus the numerator of the objective function (13) can be approximated by $\mathbf{c}^T Q\mathbf{c}$, where $\mathbf{c}$ is the coefficient vector of of the basis functions of $\xi(.)$. In similar way, the denominator of the objective function (13) can be approximated by $\mathbf{c}^T C\mathbf{c}$, with $C = H + \lambda(J + J^{(2)})$,

where the quantities H, J and $J^{(2)}$ is defined as follows:

$$\begin{aligned}
&\int_0^1\int_0^1 \xi(s)\hat{\mathbf{\Sigma}}(s,s')\xi(s')dsds' + \lambda\big[\int_0^1 \xi(s)^2 + \int_0^1 \xi''(s)^2 ds\big] \\
&= \int_0^1\int_0^1 \Big[\xi(s)\frac{1}{n-1}\sum_{i=2}^{n} y_{i-1}(s)\, y_{i-1}(s')\xi(s')\Big]dsds' + \lambda\big[\int_0^1 \xi(s)^2 ds + \int_0^1 \xi''(s)^2 ds\big] \\
&= \frac{1}{n-1}\sum_{i=2}^{n}\int_0^1\int_0^1 \Big[\xi(s) y_{i-1}(s)\, y_{i-1}(s')\xi(s')\Big]dsds' + \lambda\big[\int_0^1 \xi(s)^2 ds + \int_0^1 \xi''(s)^2 ds\big] \\
&\approx \mathbf{c}^T\mathbf{C}\mathbf{c} = \mathbf{c}^T\big[\mathbf{H} + \lambda(\mathbf{J} + \mathbf{J}^{(2)})\big]\mathbf{c};
\end{aligned}$$

$$\mathbf{c}^T\mathbf{H}\mathbf{c} = \frac{1}{n-1}\sum_{i=2}^{n}\mathbf{c}^T\Big[\sum_{k1=1}^{l}\mathbf{b}(s_{k1})y_{i-1}(s_{k1})\delta_{k1}\Big]\Big[\sum_{k2=1}^{l}\mathbf{b}^T(s'_{k2})y_{i-1}(s'_{k2})\delta_{k2}\Big]\mathbf{c},$$

$$\mathbf{c}^T\mathbf{J}\mathbf{c} = \mathbf{c}^T\big[\sum_{k3=1}^{l}\mathbf{b}(s_{k2})\mathbf{b}(s_{k3})^T\delta_{k3}\big]\mathbf{c} \text{ and } \mathbf{c}^T\mathbf{J}^{(2)}\mathbf{c} = \mathbf{c}^T\big[\sum_{k4=1}^{l}\mathbf{b}''(s_{k4})\mathbf{b}''(s_{k4})^T\delta_{k4}\big]\mathbf{c},$$

where $\mathbf{Q}, \mathbf{C}$ are symmetric non-negative definite matrices. Then the following problem is solved sequentially,

$$\hat{\mathbf{c}}_k = \max_{\mathbf{c}} \frac{\mathbf{c}^T\mathbf{Q}\mathbf{c}}{\mathbf{c}^T\mathbf{C}\mathbf{c}}, \quad \text{subject to} \quad \mathbf{c}^T\mathbf{H}\mathbf{c} = 1,\ \mathbf{c}^T\mathbf{H}\hat{\mathbf{c}}_l = 0,\ \forall 1 \le l \le k-1. \tag{16}$$

Then we obtain the corresponding $\hat{\xi}_k(s) = \hat{\mathbf{c}}_k\mathbf{b}(s)$ for $1 \le k \le K$. Now instead of solving the optimization problem (16) stated above, we focus on the following problem that differs only by a scalar constant,

$$\max_{\mathbf{c}} \frac{\mathbf{c}^T\mathbf{Q}\mathbf{c}}{\mathbf{c}^T\mathbf{C}\mathbf{c}}, \quad \text{subject to} \quad \mathbf{D}_{k-1}^T\mathbf{c} = 0, \tag{17}$$

where $\mathbf{D}_{k-1}^T = [\mathbf{H}\hat{\mathbf{c}}_1, \cdots, \mathbf{H}\hat{\mathbf{c}}_{k-1}]$. When $\lambda > 0$, $\mathbf{C}$ is of full rank. We can use Cholesky decomposition as $\mathbf{C} = \mathbf{R}^T\mathbf{R}$, where $\mathbf{R}$ is an invertible upper triangular matrix. Then $\tilde{\mathbf{c}}_k = \mathbf{R}^{-1}\mathbf{u_k}$ will be solution to optimization problem (17). Here $\mathbf{u}_k$ is the solution to following optimization problem:

$$\max_{\mathbf{u}} \frac{\mathbf{u}^T\mathbf{A}\mathbf{u}}{\|\mathbf{u}\|_2^2}, \quad \text{subject to} \quad \mathbf{E}_{k-1}^T\mathbf{u} = 0, \tag{18}$$

where $\mathbf{A} = (\mathbf{R}^{-1})^T\mathbf{Q}\mathbf{R}^{-1}$ and $\mathbf{E}_{k-1} = \mathbf{R}^{-1})^T\mathbf{D}_{k-1}$. Let us assume that the space spanned by the columns of $\mathbf{E}_{k-1}$ is $\mathcal{A}_{k-1}$ and $\mathbf{P}_{k-1}$ be the orthogonal projection matrix onto the space. For $k = 1$, $\mathbf{E}_{k-1} = \mathbf{P}_{k-1} = 0$. The solution of the optimization problem (18) is the first eigen vector of $(\mathbf{I} - \mathbf{P}_{k-1})\mathbf{A}(\mathbf{I} - \mathbf{P}_{k-1})$. We use the power method to find this largest eigenvector. Once we obtain $\mathbf{u_k}$, we can get the solution of optimization problem (17) as $\tilde{\mathbf{c}}_k = \mathbf{R}^{-1}\mathbf{c}_k$, which will lead to the solution of the original optimization problem (16) as $\hat{\mathbf{c}}_k = \tilde{\mathbf{c}}_k/\sqrt{\tilde{\mathbf{c}}_k^T\mathbf{H}\tilde{\mathbf{c}}_k}$.

4.2 Calculating $\hat{\phi}_k(t)$

We can compute the transformed predictors $\hat{\mathbf{z}}_k = \{\hat{z}_{ik};\ i = 1, \cdots, n\}$ using the estimates $\hat{\xi_k}(s)$ as follows:

$$\hat{z}_{ik} = \int_0^1 y_{i-1}(s)\hat{\xi_k}(s)ds \approx \sum_{k1=1}^{m} y_{i-1}(s_{k1})\hat{\xi}_k(s_{k1})\delta_{k1};\ 1 \le i \le n,\ 1 \le k \le K.$$

We can obtain estimated $\hat{\phi}_k(t)$ by solving an approximation of the optimization problem (15) using basis expansions. Let $\mathbf{\Phi}(t) = (\phi_1(t), \cdots, \phi_{mt}(t))^T$be the mt basis functions in $L^2[c,d]$ with the relation $\phi_k(t) = \mathbf{\Phi}(t)^T\mathbf{h}$, where $\mathbf{h}$ be unknown coefficients of the basis functions. We observe the l time points $t_1, \cdots, t_l$.

$$\min_{\mathbf{h}} \Big[\sum_{j=1}^{l} [\mathbf{\Phi}(t_j)^T\mathbf{h} - \hat{\phi}_k^0(t_j)]^2\delta_j + \eta\mathbf{h}^T\big(\int_0^1 \mathbf{\Phi}''(t)\mathbf{\Phi}''(t)dt\big)\mathbf{h}\Big], \tag{19}$$

where $\hat{\phi}_k^0(t) = \frac{1}{n}\hat{\mathbf{z}}_k^T\mathbf{y}(t)$. The solution of the above problem is,

$$\hat{\mathbf{h}}_k = \Big[\sum_{j=1}^{l} \mathbf{\Phi}(t_j)\mathbf{\Phi}(t_j)^T\delta_j + \eta \int_0^1 \mathbf{\Phi}''(t)\mathbf{\Phi}''(t)dt\Big]^{-1}\Big[\sum_{j=1}^{l} \mathbf{\Phi}(t_j)\hat{\phi}_k^0(t_j)\delta_j\Big],$$

and this leads to the estimate of $\phi_k(t)$ as $\hat{\phi}_k(t) = \mathbf{\Phi}(t)^T\hat{\mathbf{h}}_k$.

4.3 Choice of number of components and tuning parameters

We select the number of components K and the tuning parameters λ and η based on the cross validation procedure described in the following sub-section. Number of basis functions is chosen to be 50. Tuning parameters are chosen from the set of values $\{10^{-9}, 10^{-7}, 10^{-5}, 10^{-3}, 10^{-1}, 1, 10, 100, 10^3, 10^5, 10^7\}$. The maximum number of components is chosen to be 10.

4.4 Cross Validation for FAR

Cross validation porcedures are used to evaluate forecast accuracy of models. The procedure is also widely used to evaluate better performing models and in othter model fitting purposes. In general cross validation procedures, available data are randomly divided into training data and test data sets. The training data is used to fit the model that will be used for the purpose of interest and the test data is used to estimate the prediction error as well as forecast accuracy for the fitted model. In time-series analysis, random spliting of data set into training and test data sets is not valid as future data should not be used to fit the model that will be used to predict the past data. Possible non-stationarity and serial correlation in the data can also make the use of usual cross validation procedure problematic. In Bergmeir *et al.* (2015), Györfi *et al.* (2013), Burman and Nolan (1992) and Burman *et al.* (1994), approches like 'bias correction', 'h-block cross validation', etc. have been discussed for cross validation procedures in the dependent cases. For the time-series setup, a more sophisticated version of such procedures is 'time series cross-validation'. See Hyndman and Athanasopoulos (2018) for more details. In this procedure, test sets are consists of single observations and corresponding training sets are consists of all the prceding observations of the test data. This approach is also known as "evaluation on a rolling forecasting origin". This porcedure can also be modified to get multi-step ahead forecast, which is of more usefulness for time series analysis. The very first training set need to consists of sufficient number of observations to fit the model and the next observation can be used as the first test data for one-step forecast. Let us assume that we have N observations and we need atleast M observations to fit a time series model. then our first training set will consists of observations $\{1, 2, \cdots, M\}$ and the $(M+1)^{th}$ observation will be the first test data. Observations $\{1, 2, \cdots, (M+1)\}$ will build the second set of training data with $(M+2)^{th}$ observation as the test data. In this way, we will have $(N-M)$ different training and test data sets. We can calculate forecast errors for all of these $(N-M)$ sets. For functional

Cross validation set	Observation number								
	1	2	3	....	M	M+1	M+2	M+3	...
CV set 1			Training set			Test set	...	...	...
CV set 2				Training set			Test set	...	...
CV set 3					Training set			Test set	...
:	...	...	...	...	...	...	...	...	...
:	...	...	...	...	...	...	...	...	...

Figure 1: Selection of cross-validation sets for functional time series data.

autoregressive model fitting and calculating forecast accuracy, we have adapted the above described 'time series cross-validation' approach for functional observations. Here the first training set is consists of $\{1, 2, \cdots, M\}^{th}$ functional observations and the first test set is the $(M+1)^{th}$ functional observation. To have consistent error calculation, we keep the number of functional observations in each training data set fixed. Our second set of training data consists of $\{2, 3, \cdots, (M+1)\}^{th}$ functional observations and $(M+2)^{th}$ observation is the second test set and so on. The process is illustrated in figure (1).

5. Simulation studies

In this paper, we have considered functional autoregressive models of order two (FAR(2)) and functional autoregressive models of order three (FAR(3)). We generated data from these two types of model using three different sets of kernel functions. Our considered FAR(2) model is as follows:

$$Y_i(t) = \int \psi_1(t,s) Y_{i-1}(t) ds + \int \psi_2(t,s) Y_{i-2}(t) ds + \varepsilon_i(t), \tag{20}$$

The first set of Kernel functions we have considered was taken from Kokoszka and Reimherr (2013),

$$Setting1 : \begin{cases} \psi_1(t,s) = \frac{P1}{0.7468} e^{-(t^2+s^2)/2} \\ \psi_2(t,s) = \frac{P2}{0.7468} e^{-(t^2+s^2)/2} \end{cases}$$

We also have considered two additional sets of Kernel functions:

$$Setting2 : \begin{cases} \psi_1(t,s) = (P1 * 2.4) * \cos(2\pi t) * \cos(2\pi s) \\ \psi_2(t,s) = (P2 * 2.4) * \cos(2\pi t) * \cos(2\pi s) \end{cases}$$

$$Setting3 : \begin{cases} \psi_1(t,s) = \frac{P1}{1.6}\{2 - \log(0.5 + t + s)\} \\ \psi_2(t,s) = \frac{P2}{1.6}\{2 + \log(0.5 + t + s)\} \end{cases}$$

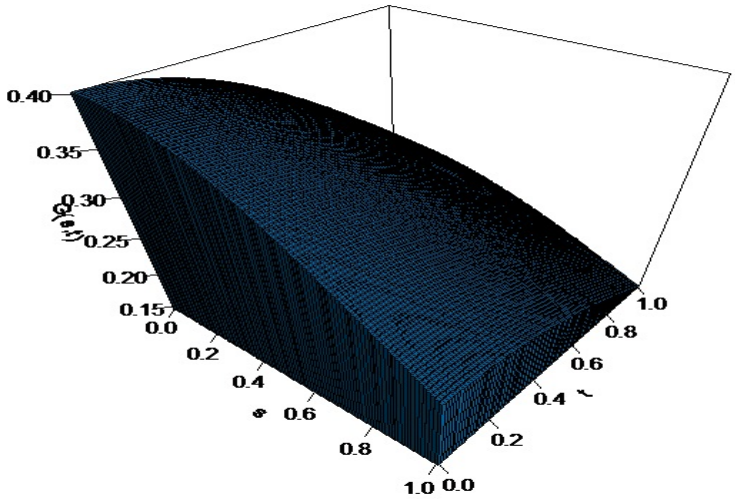


Figure 2: Graphical presentation of kernel function for setting 1.

The $\varepsilon_i' s$ are considered as standard brownian bridges. We have taken $P1 = 0.5$ and $P2 = 0.3$, 100 observations for s and t.

Similarly, our considered FAR(3) model is as follows:

$$Y_i(t) = \int \psi_1(t,s) Y_{i-1}(t) ds + \int \psi_2(t,s) Y_{i-2}(t) ds + \int \psi_3(t,s) Y_{i-2}(t) ds + \varepsilon_i(t), \tag{21}$$

with considered three different Kernel functions as,

$$Setting1: \begin{cases} \psi_1(t,s) = \frac{P1}{0.7} e^{-(t^2+s^2)/2} \\ \psi_2(t,s) = \frac{P2}{0.7} e^{-(t^2+s^2)/2} \\ \psi_3(t,s) = \frac{P3}{0.7} e^{-(t^2+s^2)/2} \end{cases}$$

$$Setting2: \begin{cases} \psi_1(t,s) = (P1 * 2.1) * \cos(2\pi t) * \cos(2\pi s) \\ \psi_2(t,s) = (P2 * 2.1) * \cos(2\pi t) * \cos(2\pi s) \\ \psi_3(t,s) = (P3 * 2.1) * \cos(2\pi t) * \cos(2\pi s) \end{cases}$$

$$Setting3: \begin{cases} \psi_1(t,s) = \frac{P1}{1.7} \{2 - \log(0.5 + t + s)\} \\ \psi_2(t,s) = \frac{P2}{2} \{2 + \log(0.5 + t + s)\} \\ \psi_3(t,s) = \frac{P3}{1.7} \{2 - \log(0.5 + t + s)\} \end{cases}$$

Where $\varepsilon_i' s$ are standard brownian bridges, considered $P1 = 0.4$, $P2 = 0.3$ and $P3 = 0.2$. 100 observations were taken for s and t.

We have generated a total of $280 + 50$, $400 + 50$ or $500 + 50$ observations. First 200 observations were discarded as burn-in samples for each of the situations and a total of 50 functional observations were used as test data set. The remaining 80, 200 or 300 observations were used to build the Functional autoregressive model. We have repeated the whole procedure 1000 timese and recorded averages and standard deviations of observed Mean

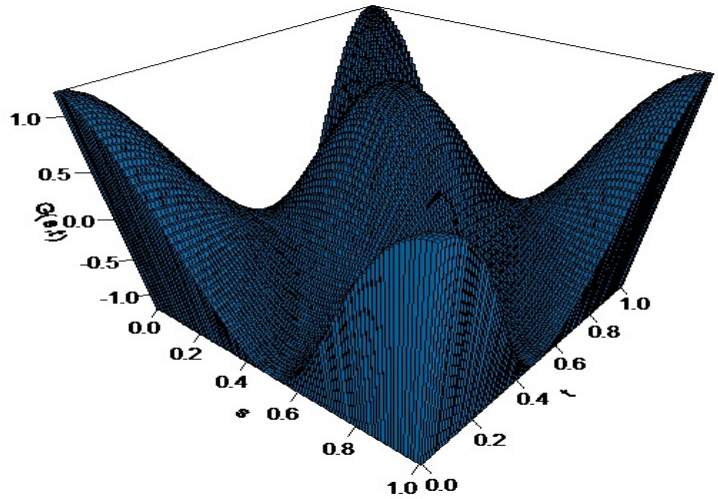


Figure 3: Graphical presentation of kernel function for setting 2.

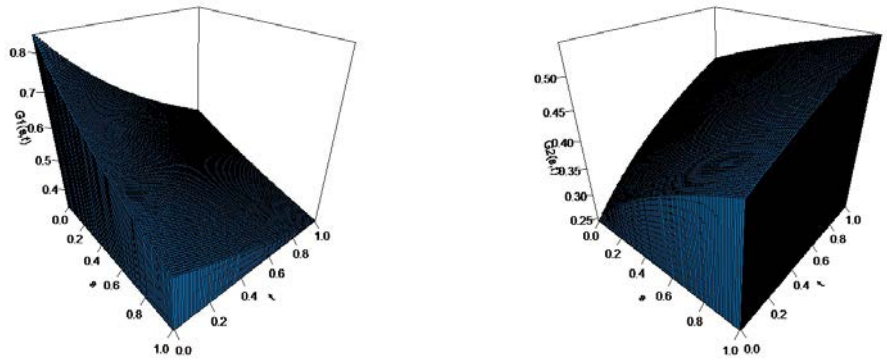
Squared Errors (MSE) and the Relative Estimation Errors (REE). We also have applied autoregressive forecast procedure of "FTSA" package of R software for the same simulated data sets and recorded averages and standatd deviations of observed mean squared errors. The results are listed in table 1 and 2.

Table 1: The averages (and standard deviations) of MSEs of 80, 200 and 300 sample sizes for FAR(2) model.

Settings for $\psi_1(t,s)$,$\psi_2(t,s)$	Sample size	Mean SigComp REE (S.E.)	Mean MSE (S.E.)	
			SigComp	FTSA
$\psi_1(t,s) \propto e^{-(t^2+s^2)/2}$ $\psi_2(t,s) \propto e^{-(t^2+s^2)/2}$	80	1.347 (6.691)	0.228 (0.086)	0.287 (0.113)
	200	0.515 (2.498)	0.202 (0.052)	0.267 (0.087)
	300	1.141 (3.853)	0.203 (0.058)	0.262 (0.087)
$\psi_1(t,s) \propto \cos(2\pi t)\cos(2\pi s)$ $\psi_2(t,s) \propto \cos(2\pi t)\cos(2\pi s)$	80	1.424 (4.008)	0.351 (0.247)	0.420 (0.370)
	200	0.743 (1.862)	0.261 (0.195)	0.363 (0.239)
	300	1.113 (2.394)	0.241 (0.152)	0.360 (0.210)
$\psi_1(t,s) \propto \{2-\log(0.5+t+s)\}$ $\psi_2(t,s) \propto \{2+\log(0.5+t+s)\}$	80	1.033 (4.379)	0.311 (0.374)	0.767 (0.895)
	200	0.367 (1.559)	0.220 (0.155)	.616 (0.519)
	300	0.772 (3.518)	0.221 (0.140)	0.630 (0.525)

We have observed that, for different settings, SigComp procedure provides smaller average of Mean Squared Errors compared to that of "FTSA" package. In terms of observed standard deviations, we also observe smaller values for SigComp procedure compared with "FTSA" package.

We also have listed our selected optimal orders with corresponding percentages using SigComp procedure for the 80, 200 and 300 sample sizes for all the three choice of Kernel functions and two different settings of FAR model. We have observed that the procedure is choosing the actual order or closer values as optimal order in satisfactory number of times. The results are listed in table 3 and 4.



(a) The first kernel function (b) The second kernel function

Figure 4: Graphical presentation of kernel functions for setting 3.

Table 2: The averages (and standard deviations) of MSEs of 80, 200 and 300 sample sizes for FAR(3) model.

Settings for $\psi_1(t,s)$, $\psi_2(t,s)$, $\psi_3(t,s)$	Sample size	Mean SigComp REE (S.E.)	Mean MSE (S.E.)	
			SigComp	FTSA
$\psi_1(t,s) \propto e^{-(t^2+s^2)/2}$	80	1.666 (8.097)	0.292 (0.478)	0.647 (0.689)
$\psi_2(t,s) \propto e^{-(t^2+s^2)/2}$	200	1.479 (5.514)	0.233 (0.254)	0.545 (0.492)
$\psi_3(t,s) \propto e^{-(t^2+s^2)/2}$	300	1.790(5.694)	0.246(0.737)	0.530(0.422)
$\psi_1(t,s) \propto \cos(2\pi t)\cos(2\pi s)$	80	3.345 (6.978)	0.274 (0.238)	0.304 (0.268)
$\psi_2(t,s) \propto \cos(2\pi t)\cos(2\pi s)$	200	2.694 (9.820)	0.203 (0.094)	0.268 (0.113)
$\psi_3(t,s) \propto \cos(2\pi t)\cos(2\pi s)$	300	2.735(6.978)	0.196 (0.090)	0.263(0.121)
$\psi_1(t,s) \propto \{2-\log(0.5+t+s)\}$	80	1.612 (7.376)	0.247 (0.487)	0.566 (2.785)
$\psi_2(t,s) \propto \{2+\log(0.5+t+s)\}$	200	2.009 (8.812)	0.207 (0.124)	0.441 (0.331)
$\psi_3(t,s) \propto \{2-\log(0.5+t+s)\}$	300	2.068 (7.305)	0.200(0.110)	0.421(0.293)

6. Application to Australia Fertility Data

We have applied the functional autoregressive model to Australia fertility data set and compared the resulted MSE with that of procedure using FTSA package of R. The data were obtained from the Australian Bureau of Statistics (Cat.No.3105.0.65.001, Table 38). It is also available in "rainbow" package of R software. The observed data were smoothed using the penalized regression spline with concave constraint. The procedure is described in Hyndman and Ullah (2007). We have used the smoothed australlia fertility rates of women aged 15- 49 for the years 1921 - 2006 from R package "rainbow".

We have used fertility rates of the yars 1921 to 1994 as trainee data, fitted both the Functional Autoregressive Models (FAR) using Sigcomp procedure and Functional Autoregressive Procedures available in "FTSA" package of $\mathbb{R}$ software. After fitting the models, we predicted age specific fertility rates using both of the procedure for a total of 12 forecast horizons for the years 1995 to 2006 and calculated Mean Squared Errors. We have calculated average of observed Mean Squared Errors over all the twelve years and observe that SigComp procedure provides an average MSE as about 64.48 whereas forecast procedure of "FTSA" package provides average MSE as about 1557.37. We also have calculated average of the MSEs for each of the forecasted years. Based on MSE, we have obtainde very

Table 3: The observed frequency (percent) of selected optimal orders using cross validation and SigComp procedure for FAR(2) model.

$\psi_1(t,s)$, $\psi_2(t,s)$	Sample size	Observed Optimal Order				
		$P=1$	$P=2$	$P=3$	$P=4$	$P=5$
$\psi_1(t,s) \propto e^{-(t^2+s^2)/2}$ $\psi_2(t,s) \propto e^{-(t^2+s^2)/2}$	80	285 (28.5)	399 (39.9)	138 (13.8)	99 (9.9)	79 (7.9)
	200	113 (11.3)	575 (57.5)	140 (14.0)	94 (9.4)	78 (7.8)
	300	132 (13.2)	448 (44.8)	163 (16.3)	136 (13.6)	121 (12.1)
$\psi_1(t,s) \propto \cos(2\pi t)\cos(2\pi s)$ $\psi_2(t,s) \propto \cos(2\pi t)\cos(2\pi s)$	80	784 (78.4)	178 (17.8)	28(2.8)	7(0.7)	3 (0.3)
	200	383 (38.3)	508 (50.8)	73 (7.3)	26(2.6)	10 (1.0)
	300	301 (30.1)	469 (46.9)	112 (11.2)	75 (7.5)	43 (4.3)
$\psi_1(t,s) \propto \{2-\log(0.5+t+s)\}$ $\psi_2(t,s) \propto \{2+\log(0.5+t+s)\}$	80	135 (13.5)	571 (57.1)	140 (14.0)	86 (8.6)	68 (6.8)
	200	37 (3.7)	704 (70.4)	112 (12.2)	71(7.1)	76 (7.6)
	300	63 (6.3)	587 (58.7)	130 (13.0)	99 (9.9)	121 (12.1)

Table 4: The observed frequency (percent) of selected optimal orders using cross validation and SigComp procedure for FAR(3) model.

$\psi_1(t,s)$, $\psi_2(t,s)$, $\psi_3(t,s)$	Sample size	Observed Optimal Order				
		$P=1$	$P=2$	$P=3$	$P=4$	$P=5$
$\psi_1(t,s) \propto e^{-(t^2+s^2)/2}$	80	110 (11.0)	287 (28.7)	389 (38.9)	94(9.4)	120 (12.0)
$\psi_2(t,s) \propto e^{-(t^2+s^2)/2}$	200	77 (7.7)	235 (23.5)	402 (40.2)	122(12.2)	164(16.4)
$\psi_3(t,s) \propto e^{-(t^2+s^2)/2}$	300	73 (7.3)	187 (18.7)	433 (43.3)	167(16.7)	140 (14.0)
$\psi_1(t,s) \propto \cos(2\pi t)\cos(2\pi s)$	80	757 (75.7)	165 (16.5)	36(3.6)	28(2.8)	14 (1.4)
$\psi_2(t,s) \propto \cos(2\pi t)\cos(2\pi s)$	200	261 (26.1)	384 (38.4)	261 (26.1)	66(6.6)	28(2.8)
$\psi_3(t,s) \propto \cos(2\pi t)\cos(2\pi s)$	300	197 (19.7)	329 (32.9)	308 (30.8)	100(10.0)	66 (6.6)
$\psi_1(t,s) \propto \{2-\log(0.5+t+s)\}$	80	117 (11.7)	317 (31.7)	353 (35.3)	123 (12.3)	90 (9.0)
$\psi_2(t,s) \propto \{2+\log(0.5+t+s)\}$	200	63(6.3)	250(25.0)	427 (42.7)	131(13.1)	129(12.9)
$\psi_3(t,s) \propto \{2-\log(0.5+t+s)\}$	300	72 (7.2)	208 (20.8)	437 (43.7)	137(13.7)	146(14.6)

impressive performance by the FAR using SigComp procedure over that of FTSA package. Eventhough for smaller forecast horizons, FAR procedure in FTSA package performs satisfactory, for larger forecast horizons, FAR using Sigcomp procedure outperforms the FAR in FTSA package. The FAR in FTSA package gets very high MSE for higher forecast horizons. The results are displayed in Figure 8.

7. Discussion

We proposed functional autoregressive model of general order using signal compression approach for function on function regression model proposed by Luo and Qi (2017). The singal compression approach takes account of possible dependency of covariate functions of FAR model of general order where dependency exists among the lag functions. Signal compression procedure makes finite dimesional approximation of coefficient functions that also reduce computation cost.

We also have proposed order detection procedure for FAR model. We used proposed extension of timeseries cross-validation procedure to functional time series data. Based on minimum cross-validation error, the optimal order of FAR model was selected. The selected model with optimal order was furter used for forecasating purpose.

We performed simulation studies considering different orders and structures of coefficient or kernel functions. We observed performeces of our porposed procedure with exist-

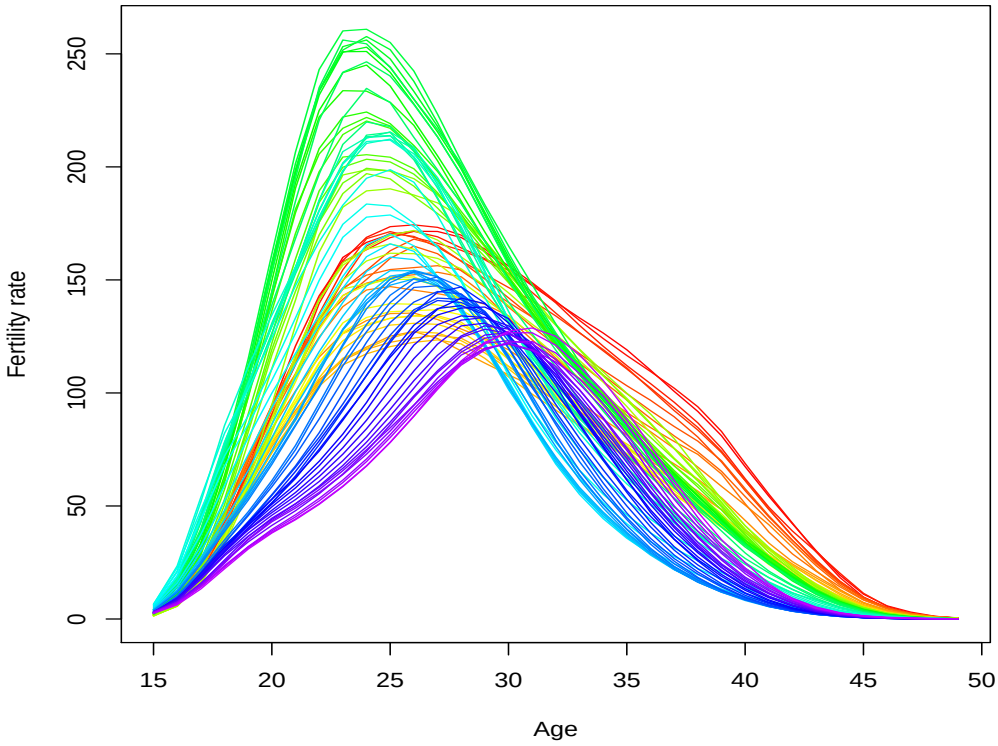


Figure 5: Smoothed fertility rates' curve observed in different years during 1921 to 2006 over women's age 15-49 in Australia.

ing "FTSA" package of R software.

We have observed that, on average there are comparatively smaller mean squared error for SigComp procedure compared to the existing procedure of "FTSA" package. We also have observed satisfactory performance of the procedure in terms of determining order of underlying FAR model.

In this paper, we did not have the scope to compare the performance of order selection procedure with other porcedures, rather than reportign the correct order detection rates. We consider the comparison for future scope. We want to perform more comarisons of our model with other existing procedures both in terms of the performence of forecast accuracy and optimal order selection.

Incorporating exogenous predictors can imporve the performance of the model, it can also take care of missing value problems. Our future goal would be to extend the model to incorporate exogenous predictors. Prediction intervals provide better understanding of the estimation accuracy. In future, we would like to develop prediction intervals for the forecasted curves of our proposed procedure.

Our future interest would also be on comparison of machine learning porcedures in such area in terms of performance. The SigComp procedure is computationally intensive, there can be some scope of improvements in the computations and optimization porblems. We would also like to apply the procedure to recent datasets of different areas of interest.

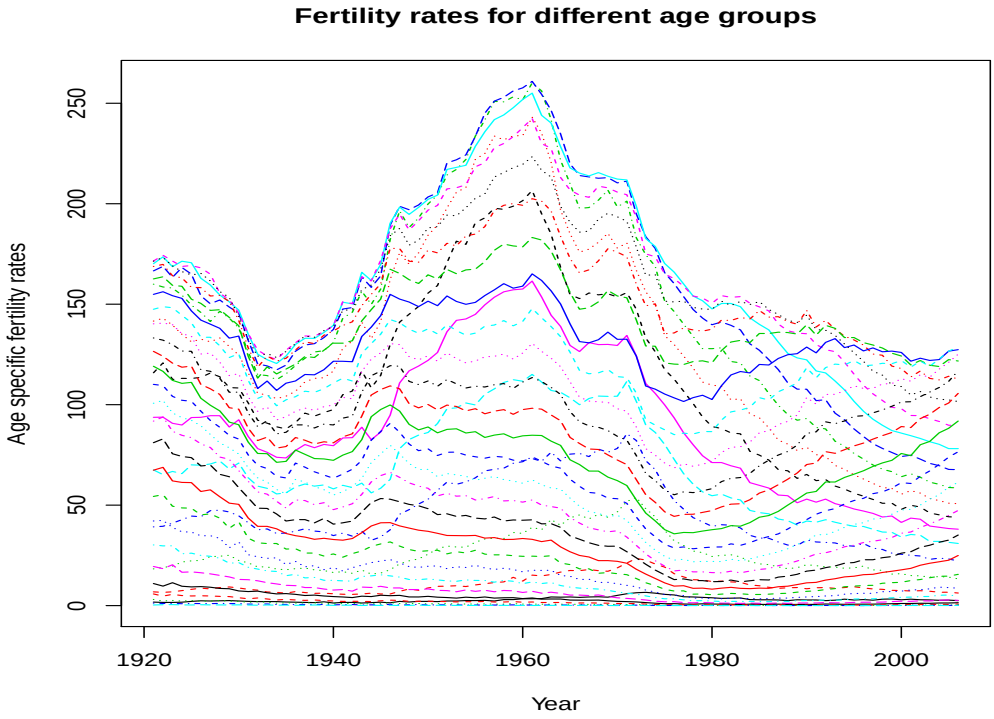


Figure 6: Age specific fertility rates of Australian women of ages 15- 49 over the years 1921 - 2006.

References

Aue, A., Norinho, D. D. and Hörmann, S. (2015) On the prediction of stationary functional time series. *Journal of the American Statistical Association*, **110**, 378–392. URL https://doi.org/10.1080/01621459.2014.909317.

Beguin, J.-M., Gourieroux, C. and Monfort, A. (1980) Identification of a mixed autoregressive-moving average process: The corner method. *Time series*, 423–436.

Bergmeir, C., Hyndman, R. J., Koo, B. *et al.* (2015) A note on the validity of cross-validation for evaluating time series prediction. *Monash University Department of Econometrics and Business Statistics Working Paper*, **10**, 15.

Bosq, D. (2000) *Linear Processes in Function Spaces, Theory and Applications*. Lecture Notes in Statistics, 149, Springer-Verlag, New York.

Box, G. E. and Jenkins, G. M. (1976) *Time series analysis: forecasting and control, revised ed*. Holden-Day.

Box, G. E., Jenkins, G. M., Reinsel, G. C. and Ljung, G. M. (2015) *Time series analysis: forecasting and control*. John Wiley & Sons.

Burman, P., Chow, E. and Nolan, D. (1994) A cross-validatory method for dependent data. *Biometrika*, **81**, 351–358.

Burman, P. and Nolan, D. (1992) Data-dependent estimation of prediction functions. *Journal of Time Series Analysis*, **13**, 189–207.

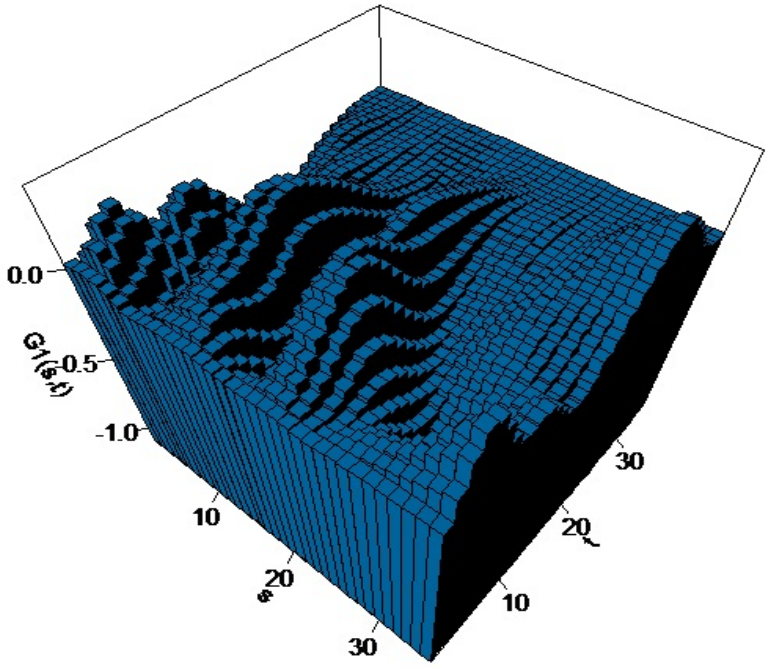


Figure 7: Estimated coefficient function obtained using Functional Autoregressive Models using SigComp procedure for forecasting age specific Australian fertility rates.

Cryer, J. D. and Chan, K.-S. (2008) *Time series analysis with Application in R*. New York: Springer-Verlag.

Damon, J. and Guillas, S. (2002) The inclusion of exogenous variables in functional autoregressive ozone forecasting. *Environmetrics*, **13**, 759–774.

Györfi, L., Härdle, W., Sarda, P. and Vieu, P. (2013) *Nonparametric curve estimation from time series*, vol. 60. Berlin: Springer Verlag.

Hörmann, S., Kidziński, Ł. and Hallin, M. (2015) Dynamic functional principal components. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **77**, 319–348.

Hörmann, S. and Kokoszka, P. (2012) 7 - functional time series. In *Time Series Analysis: Methods and Applications* (eds. T. S. Rao, S. S. Rao and C. Rao), vol. 30 of *Handbook of Statistics*, 157 – 186. Elsevier. URL http://www.sciencedirect.com/science/article/pii/B9780444538581000077.

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

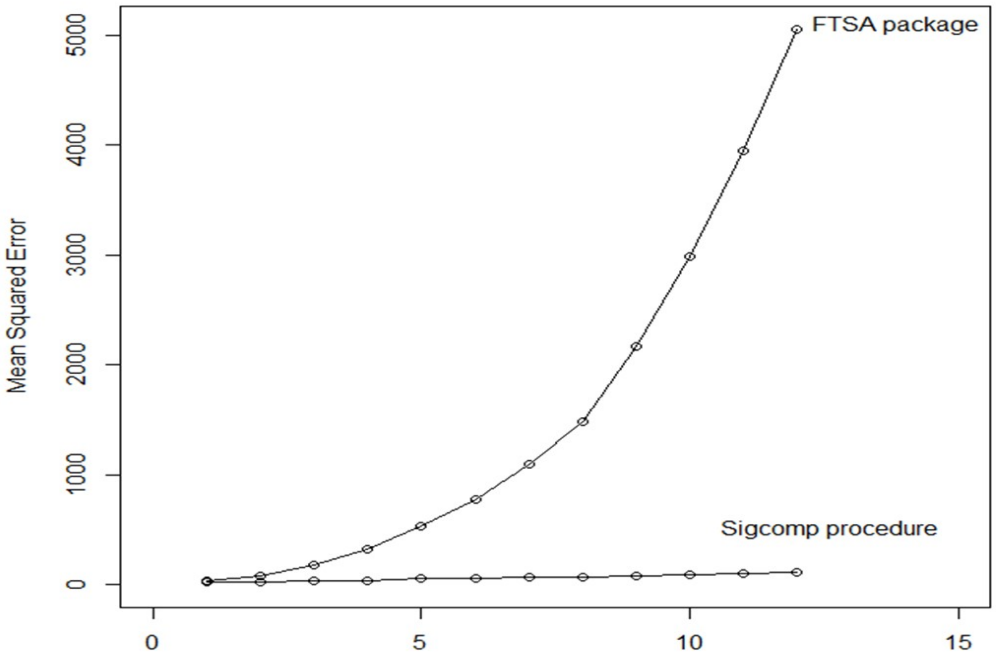


Figure 8: Mean Squared Errors obtainde using Functional Autoregressive Models with Sigcomp procedure and Functional Autoregressive procedure in FTSA package for forecasting Australian age specific fertility rates.

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

Hyndman, R. J. and Athanasopoulos, G. (2018) *Forecasting: principles and practice*. OTexts.

Hyndman, R. J. and Shang, H. L. (2009) Forecasting functional time series. *Journal of the Korean Statistical Society*, **38**, 199–211.

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

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

Luo, R. and Qi, X. (2017) Function-on-function linear regression by signal compression. *Journal of the American Statistical Association*, **112**, 690–705.

Montgomery, D. C., Johnson, L. A. and Gardiner, J. S. (1990) *Forecasting and time series analysis*. McGraw-Hill Companies.

Panaretos, V. M. and Tavakoli, S. (2013) Cramér–karhunen–loève representation and harmonic principal component analysis of functional time series. *Stochastic Processes and their Applications*, **123**, 2779–2807.

Ramsay, J. O. and Dalzell, C. (1991) Some tools for functional data analysis. *Journal of the Royal Statistical Society. Series B (Methodological)*, 539–572.

Ramsay, J. O. and Silverman, B. W. (2005) *Functional Data Analysis. 2nd Edition.* New York: Springer.

Shang, H. L. (2013) ftsa: An r package for analyzing functional time series. *R Journal*, **5**.

Tsay, R. S. and Tiao, G. C. (1985) Use of canonical analysis in time series model identification. *Biometrika*, **72**, 299–315.