Model Selection for Scalar-on-Function Regression with Applications to Near-Infrared Spectroscopy

Jonghun Baek, Jakob R. Juergens, Jonathan Willnow

11.02.2022

Research Module in Econometrics and Statistics
Winter Semester 2021/2022

Contents

1 Colour Guide

- RED: is for general comments for your own text

- GREEN: is for Jona's comments

- ORANGE: is for Jonghun's comments

- BLUE: is for Jakob's comments

2 Introduction

- Describe the idea of regressing a scalar on functional data

- Describing the difference to multiple linear regression intuitively

- Giving an intuitive example

Functional Data Analysis (FDA) is (roots in the 1940s Grenander and Karhunen) gaining more attention as researchers from different fields collect data that is functional in nature. Although classical statistical methods can often process this data, but only FDA allows extracting the information given by the smoothness of the underlying process of the functional data (cf. Levitin et al. 2007). As Kokoszka and Reimherr 2017 describe, FDA should be considered when one can view variables or units of a given data set as smooth curves or functions and the interest is in analyzing samples of curves (cf. Kokoszka and Reimherr 2017, S. 17).

Consider the case of a dataset which is containing observations of an underlying continuous process, measured at some discrete points $t_{j,i}$:

$$x_i(t_{j,i}) \in \mathbb{R}, \quad i = 1, \dots, N, \ j = 1, \dots, J_i, \ t_{j,i} \in [T_1, T_2] \tag{1}$$

This measurement set can be viewed as set of discretized approximations of the underlying true process which could be the absorption of the Near-infrared spectrum which we will use for the simulation and application. This continuous underlying process for which $x_i(t)$ exists $\forall t \in [T_1, T_2]$ is only observed at the discrete measurement points. This dataset can then be used to perform Functional Linear Regression (FLR) on a scalar response variable. To conduct functional linear regression, several new concepts and methods need to be considered and explained that distinguish it multivariate linear regression. The focus of this paper is to introduce FLR in a scalar-on-function setting. We will be using the standard FLR framework, which relates functional predictors to a scalar response as follows:

$$Y(\omega) = \alpha + \int_0^1 X(\omega)(s)\beta(s)\mathrm{d}s + \epsilon(\omega), \qquad i = 1, ..., n \tag{2}$$

where the X_i are realizations of a random function $\mathbf{X}$, Y_i are the corresponding realizations of the response variable and $\beta(s)$ is the coefficient function. The distinct feature of this framework is that the regressor is a function, which necessitates a different approach to estimation. As in the well-known framework of scalar linear regression, this is motivated by an interest in $\beta(s)$ for prediction. For instance, fluctuation in $X_i(s)$ at a point s_0 will not have any effect on Y_i if $\beta(s_0) = 0$.

Estimation of $\beta(s)$ is inherently an infinite-dimensional problem. In Section 2, after introducing the necessary theoretical concepts, we describe three methods of estimating a scalar coefficient function

using a concept called truncated basis expansion. We report the results of the Monte-Carlo simulation regarding these three different methods in Section 3. Finally, in Section 4, we test the prediction of FLR in a real-world setting. (We may put some simple descriptions of results about each of MC and Application)

3 Theory

In multivariate regression, data is often observed in the form of elements from Euclidean space, $\mathbb{R}^p$. However, the statistics derived from infinite-dimensional random functions cannot be defined on a finite dimensional space. To understand functional linear regression and the differences between the methods presented in this paper, it is therefore necessary to introduce some concepts and extend known aspects of linear regression theory to include functional objects. One integral concept in inferential statistics are random variables. Paraphrasing a definition by Bauer 2020, a random variable $X : \Omega \to \Omega'$ is an $\mathcal{A}$-$\mathcal{A}'$-measurable function, where $(\Omega, \mathcal{A}, P)$ is a probability space and $(\Omega', \mathcal{A}')$ is a measure space. A typical case known to every undergraduate student of economics in less formal detail is $(\Omega', \mathcal{A}') = (\mathbb{R}, \mathcal{B})$, where $\mathcal{B}$ is the canonical σ-algebra on the real numbers. As a first intuition, it is possible to imagine a similar concept where a random variable does not realize as an element of the real numbers but as a function in a function space. A formalization of this idea makes some more theoretical considerations necessary. The following theoretical introduction closely follows chapters 2.3 and 2.4 from Hsing and Eubank 2015.

3.1 Inner Products and Hilbert Spaces

Let $\mathbb{V}$ be a vector space over some field of scalars $\mathbb{F}$. A function $\langle \cdot, \cdot \rangle : \mathbb{V} \times \mathbb{V} \to \mathbb{F}$ is called an inner product, if $\forall v, v_1, v_2 \in \mathbb{V}$ and $a_1, a_2 \in \mathbb{F}$ the following properties hold.

1. $\langle v, v \rangle \geq 0$

2. $\langle v, v \rangle = 0$ if $v = 0$

3. $\langle a_1 v_1 + a_2 v_2, v \rangle = a_1 \langle v_1, v \rangle + a_2 \langle v_2, v \rangle$

4. $\langle v_1, v_2 \rangle = \langle v_2, v_1 \rangle$

A vector space with an associated inner product is called an inner product space. [verbatim quote!] The inner product naturally defines a norm and an associated distance on the vector space as follows. In the following we restrict our analysis to the case of $\mathbb{F} = \mathbb{R}$.

$$||v|| = \langle v, v \rangle^{\frac{1}{2}} \tag{3}$$

$$d(v_1, v_2) = \langle v_2 - v_1, v_2 - v_1 \rangle^{\frac{1}{2}} \tag{4}$$

If the inner product space is complete with respect to the induced distance, it is called a Hilbert space, denoted $\mathbb{H}$ in the following. To extend the known concept of a basis in a finite dimensional space to the potentially infinite Hilbert spaces, it is necessary to define the closed span of a sequence of elements of $\mathbb{H}$. Recall that the span of a set of vectors $S \subseteq \mathbb{R}^P$ is given by

$$span(S) = \left\{ \sum_{i=1}^{k} \lambda_i v_i \;\middle|\; k \in \mathbb{N}, \, v_i \in S, \, \lambda_i \in \mathbb{R} \right\} \tag{5}$$

The closed span $\overline{span}(S)$ of a sequence S in $\mathbb{H}$ is defined as the closure of the span with respect to the distance induced by the norm. S is called a basis of $\mathbb{H}$ if $\overline{span}(S) = \mathbb{H}$.

It is called an orthonormal basis, if in addition the following properties hold.

1. $\langle v_i, v_j \rangle = 0 \quad \forall v_i, v_j \in S \quad i \neq j$
2. $||v|| = 1 \quad \forall v \in S$

As in the case of a Banach space, each element of a Hilbert space can be expressed in terms of a corresponding basis. This can be done using a Fourier expansion of an element $x \in \mathbb{H}$ w.r.t. a basis $S = \{s_n\}$ as follows.

$$x = \sum_{j=1}^{\infty} \langle x, s_j \rangle s_j \tag{6}$$

As can be seen, differing from the case of a Banach space, these representations can be limits of series as previously hinted at by using the closed span of the basis. As using an infinite number of basis functions is infeasible in applied contexts, an intuitive way to approximate elements of a Hilbert space, is to use a truncated series.

$$x \approx \sum_{j=1}^{K} \langle x, s_j \rangle s_j \tag{7}$$

3.2 Hilbert Space of Square-Integrable Functions

In functional data analysis, one Hilbert space of particular importance is the space of square-integrable functions on $[0,1]$ denoted $\mathbb{L}^2[0,1]$. To define it, look first at the measure space given by $([0,1], \mathcal{B}, \mu)$ where $\mathcal{B}$ is the Borel σ-algebra on $[0,1]$ and μ is the Lebesgue-measure. Then $\mathbb{L}^2[0,1]$ is the collection of all measurable functions f on $[0,1]$ that fulfill the following condition.

$$||f||^2 = \int_0^1 |f|^2 \mathrm{d}\mu < \infty \tag{8}$$

All functions f satisfying the above condition are called square-integrable functions. Moreover, it ensures that a random function has a finite second moment so that the variance and the covariance function can be defined. Its inner product is defined as

$$\langle f_1, f_2 \rangle = \int_0^1 f_1 f_2 \mathrm{d}\mu. \tag{9}$$

$\mathbb{L}^2[0,1]$ is the function space that is most often used for theoretical considerations in functional data analysis without loss of generality. We also focus on the interval $[0,1]$ for the purpose of this paper. The random function defined on $\mathbb{L}^2[0,1]$ can be represented as

$$\{f(t, \omega) : t \in [0,1], \omega \in \Omega\} \tag{10}$$

where Ω is a sample space. The realized $f(t)$ for every $t \in [0,1]$ is called a sample curve for the process. The collection of such sample curves constitutes a functional data set.

3.3 Bases of $\mathbb{L}^2$

As previously described, a basis of a Hilbert space can be used to express elements of the space using the corresponding Fourier expansion. Two examples of bases that are often used in practice to express

/ approximate elements of $\mathbb{L}^2[0,1]$ are explained in the following.

Both of these can be used to express or in the case of the b-spline basis approximate elements of $\mathbb{L}^2[0,1]$ as a weighted sum of basis functions. Let therefore $\{\phi_i(t) \mid i \in \mathcal{I}\}$ be the basis used to express / approximate a realization $X(\omega_0) = x(t)$ of $X(\omega)$.

$$X(\omega_0) = x(t) = \sum_{j \in \mathcal{I}} A_j(\omega_0)\phi_j(t) \tag{11}$$

Monomial Basis   This is the basis that is needed to build polynomials, one of the best-known ways to expand a function. The polynomial of the form

$$x(t) = \sum_{i \in \mathcal{I}} c_i(t-w)^i \tag{12}$$

can be expressed as a finite linear combination of the monomials $(t-w)^i$ with coefficient c_i which form the basis

$$\phi_i^M(t) = (t-w)^k \quad i \in \mathcal{I} \tag{13}$$

The single basis function 1 within the monomial basis system, for $k = 0$, is called the constant basis system (cf. Horváth and Kokoszka 2012). Taylor expansion as motivation

The shift parameter w usually is specified to be the center of the interval subject to approximation. Problems of collinearity arise for higher degrees since the monomial basis becomes more correlated as the degrees increase. This restricts the number of degrees and therefore the use of this basis: For simple functions this may be useful, but the small number of degrees makes it impossible to capture pronounced local peculiarities and leads to undesirable behaviour at the tails. (cf. Ramsay and Silverman 2005)

Fourier Basis   The Fourier basis for $\mathbb{L}^2[0,1]$ is given by the following sequence of functions defined on $[0,1]$.

$$\phi_i^F(x) = \begin{cases} 1 & \text{if} \quad i = 1 \\ \sqrt{2}\cos(\pi i x) & \text{if} \quad i \quad \text{is even} \\ \sqrt{2}\sin(\pi(i-1)x) & \text{otherwise} \end{cases} \tag{14}$$

Therefore, its basis functions inherits a repeating behaviour which is useful to expand functions that represent an periodic or seasonal underlying process over the period T. Rephrasing from Ramsay and Silverman 2005, the Fourier basis functions are orthogonal when the values t_j are equally spaced within T. This basis is suitable to expand functions with a similar curvature order across the domain, resulting generally in uniformly smooth expansions. Motivate from fourier series

B-spline Basis   Following chapter 3.5 from Ramsay and Silverman 2005, splines are defined by first dividing the interval of interest $[\tau_0, \tau_L]$ into L subintervals of non-negative length divided by a non-decreasing sequence of points $(\tau_l)_{l=1,\dots,L-1}$ called knots. On each subinterval, a spline is a polynomial of chosen order $m = n+1$ where n is its degree. Additionally, at each τ_l the the polynomials on neighbouring subintervals must match derivatives up to order $m-2$. A B-spline is a spline belonging to a basis system developed by Boor 1978. Let $S_{l,m}^{BS}(x) \quad l = 1, \dots, L-1$ be the B-spline of order m

6

for an interval $[\tau_0, \tau_L]$ and knots $\{\tau_l \mid l = 1, \ldots, L-1\}$, then it is defined by the Cox-de Boor recursion formula as follows.

$$\phi_{l,0}^{BS}(x) = \begin{cases} 1 & \text{if } x \in [\tau_l, \tau_{l+1}) \\ 0 & \text{otherwise} \end{cases}$$

$$(15)$$

$$\phi_{l,m}^{BS}(x) = \frac{x - \tau_l}{\tau_{l+m} - \tau_l} \phi_{l,m-1}^{BS}(x) + \frac{\tau_{l+m+1} - x}{\tau_{l+m+1} - \tau_{l+1}} \phi_{l+1,m-1}^{BS}(x)$$

This, however, does not really yield a basis of $\mathbb{L}^2[0,1]$ as the closed span of this finite sequence of functions is not equal to $\mathbb{L}^2[0,1]$. To really obtain a basis of $\mathbb{L}^2[0,1]$ from B-splines, further theoretical considerations about, for example, infinite series of B-splines and specific knot choices would have to be made. As this is out of the scope of this paper, for the sake of simplicity, we will assume that a B-spline basis representation of a function in $\mathbb{L}^2[0,1]$ will serve as a sufficient approximation for an appropriately chosen B-spline basis. Even though, this approach is not theoretically exact, in practice, this is often a reasonable approach and yields satisfactory results in cases where the functional form of B-splines makes them an appropriate approximation tool.
Explain what happens when $l + m + 1 > L$!!!
Additionally, modify for multiple knots at the boundaries to get better behavior at the boundaries. Needs some more explanation.

3.4 Approximation and Smoothing via Basis Truncation

Write something about how to approximate functions by truncating the basis. As above-mentioned, the realized curves can be estimated with basis functions. For the basis expansion, it is technically possible to use all basis functions $b_{j \in \mathcal{I}}$. On the one hand, this way is not efficient since the technique with too many basis functions can even approximate a noise of the sample curves which possibly interrupts the analysis. On the other hand, the important information on the curves would be missed with a too-small number of basis functions. This discussion challenges the researcher to seek a point at which they truncate the basis function in order to remove noise and, at the same time, maintain significant fluctuation. The basis expansion with truncation is by:

$$X(\omega_0) = x(t) = \sum_{j \in \mathcal{I}} A_j(\omega_0) \phi_j(t) = \sum_{j=1}^{L} A_j(\omega_0) \phi_j(t) + \delta(t) \approx \sum_{j=1}^{L} A_j(\omega_0) \phi_j(t) \qquad (16)$$

where $\delta(t)$ is the truncation error. The number L is subjectively chosen but we can fix the nubmer through a more reasonable method, Cross-Validation, which minimizes Mean Squared Error (MSE).

3.5 Functional Data Sets

In functional data analysis the concept of a data set can include not only realizations of scalar random variables, but also realizations of random functions. In the following, all random functions are assumed to realize in $\mathbb{L}^2$. As in the finite dimensional setting the concept of identically distributed and independent data is important. This concept generalizes intuitively to the case of functional data using the concepts for general random variables. Explain this better!

7

An example of one such data set could be a set of n independent realizations of a Wiener-process and n associated scalar variables that could be subject of a regression analysis with respect to the corresponding realizations of the Wiener process. A definition and a plotted example for the Wiener process are given in 7.3.

3.6 Karhunen-Loéve Expansion and Empirical Eigenbases

Given a realization of a random function realizing in $\mathbb{L}^2[0,1]$, it is possible to represent this realization in terms of its generating stochastic process. To motivate it, recall the representation of a random vector derived from the spectral decomposition of its covariance matrix. Let therefore $V(\omega)$ be a random vector realizing in $\mathbb{R}^p$ with $\mathbb{E}(V(\omega)) = \mu_V$ and $Cov(V(\omega)) = \Sigma_V$. Then Σ_V can be expressed in terms of its Eigenvalues and orthonormal Eigenvectors as $\Sigma_V = \Gamma\Lambda\Gamma'$ using its Eigendecomposition.

This decomposition can be used to express the random vector $V(\omega)$ in the following way.

$$V(\omega) = \mu_V + \Gamma\Lambda^{\frac{1}{2}}D(\omega) = \mu_V + \sum_{i=1}^{p}\sqrt{\lambda_i}D_i(\omega)\gamma_i \tag{17}$$

where $D(\omega)$ is a random vector with $\mathbb{E}(D(\omega)) = 0_p$ and $Cov(D(\omega)) = \mathbb{I}_p$.

To obtain the analogous concept for a random function, it is necessary to define the covariance function of a random function realizing in $\mathbb{L}^2[0,1]$. Therefore, let $X : \Omega \mapsto \mathbb{L}^2[0,1]$ be such a random function. Then the mean and covariance functions of X are defined as follows.

$$\mu(t) = \mathbb{E}\left[X(\omega)(t)\right] \tag{18}$$

$$c(t,s) = \mathbb{E}\left[\left(X(\omega)(t) - \mu(t)\right)\left(X(\omega)(s) - \mu(s)\right)\right] \tag{19}$$

suggestion from here where the $c(t,s)$ are Hilbert-Shmidt Kernels defined through c : $[0,1] \times [0,1] \to \mathbb{R}$. Let K be the integral operater on $\mathbb{L}^2[0,1]$ such that $K : \nu \to K\nu$ for $\nu \in \mathbb{L}^2[0,1]$, by

$$[K\nu](t) = \int_0^1 c(t,s)\nu(s)ds \tag{20}$$

Then, the operator K has orthonormal basis functions $\nu_i \in \mathbb{L}^2[0,1]$ corresponding to eigenvalues λ_i for all i since it is a positive compact self-adjoint operator. Moreover, K holds following traits:

1. The eigenspaces corresponding to distinct eigenvalues are mutually orthogonal.

2. The eigenspaces corresponding to non-zero eigenvalues are finite-dimentsional.

3. The eigenvalues can be ordered in nonincreasing order as follows $\lambda_1 \geq \lambda_2 \geq \cdots \geq 0$.

Therefore, we can approximate the function X with first few eigenfunctions. The functions X are approximated enough well by first few principal components since the order of them are sorted in descending order (e.g. $Var(\xi_j) \geq Var(\xi_k)$ for all $j > k$). Theoretical considerations lead to the result that X can be represented in the following form, called its Karhunen-Loéve expansion. The proofs are provided at 8.1 and 8.2.

$$X(\omega)(t) = \mu(t) + \sum_{m \in \mathcal{M}} \xi^m(\omega)\nu^m(t) \tag{21}$$

where the ν_j are defined by the countable set of solutions $\{(\lambda^m, \nu^m) \mid m \in \mathbb{N}\}$ of the following equation.

$$[K\nu](t) = \int_0^1 c(t,s)\nu(s)\mathrm{d}s = \lambda\nu(t) \tag{22}$$

The $\xi_j(\omega)$ are then given as

$$\xi^m(\omega) = \langle X(\omega) - \mu, \nu^m \rangle = \int_0^1 (X(\omega)(s) - \mu(s))\,\nu^m(s)\mathrm{d}s \tag{23}$$

and thereby random variables realizing in $\mathbb{R}$ and have the following properties akin to $\Lambda^{\frac{1}{2}}D(\omega)$ for the case of a random vector. We would need to assume that the process is centered, otherwise we may not derive below results(not sure though).

1. $\mathbb{E}\left[\xi_j(\omega)\right] = 0$
2. $Var\left(\xi_j(\omega)\right) = \lambda_i$
3. $Cov\left(\xi_j(\omega), \xi_k(\omega)\right) = 0$ if $j \neq k$

In the typical scalar setting, a similar consideration leads to the concept of principal components. This is also possible in a functional setting. Let $\{x_1(t), \ldots, x_n(t)\}$ be a set of i.i.d. realizations generated by a random function $X(\omega) \mapsto \mathbb{L}^2[0,1]$. Define the following sample analogues for the mean and covariance functions.

$$\hat{\mu}(t) = \frac{1}{n} \sum_{i=1}^n x_i(t) \tag{24}$$

$$\hat{c}(t,s) = \frac{1}{n} \sum_{i=1}^n (x_i(t) - \hat{\mu}(t))(x_i(s) - \hat{\mu}(s)) \tag{25}$$

With these it is possible to derive a set of sample analogs $\{(\hat{\lambda}^m, \hat{\nu}^m) \mid m \in \mathcal{M}\}$ for $\{(\lambda^m, \nu^m) \mid i \in \mathcal{M}\}$ as the solutions of the following equation. I think the number of principal components has to be smaller then the number of observations. So I can give more information about $\mathcal{M}$.

$$\int_0^1 \hat{c}(t,s)\hat{\nu}(s)\mathrm{d}s = \hat{\lambda}\hat{\nu}(t) \tag{26}$$

This naturally leads to the following representation.

$$x_i(t) = \hat{\mu}(t) + \sum_{j=1}^{!!!} \hat{\xi}_{i,j}\hat{\nu}(t) \tag{27}$$

where the $\hat{\xi}_i^m$ are derived as

$$\hat{\xi}_i^m(\omega) = \langle x_i - \hat{\mu}, \hat{\nu}^m \rangle = \int_0^1 (x_i(s) - \hat{\mu}(s))\,\hat{\nu}^m(s)\mathrm{d}s \tag{28}$$

In reality these calculations are often done using basis representations of both the functional principal components $\hat{\nu}^m$ and the observations $x_i(t)$ leading to the following representation.

$$\hat{\xi}_i^m = \int_0^1 (x_i(s) - \hat{\mu}(s))\hat{\nu}^m(s)\mathrm{d}s = \int_0^1 \left(\sum_{j\in\mathcal{I}} a_{i,j}\phi_j(s)\right)\left(\sum_{k\in\mathcal{L}} b_k^m\psi_k(s)\right)\mathrm{d}s$$
$$= \sum_{j\in\mathcal{I}}\left[a_{i,j}\sum_{k\in\mathcal{L}} b_k^m \int_0^1 \phi_j(s)\psi_k(s)\mathrm{d}s\right] \approx \sum_{j=1}^J\left[a_{i,j}\sum_{k=1}^K b_k^m \int_0^1 \phi_j(s)\psi_k(s)\mathrm{d}s\right] \tag{29}$$

In practice, a typical choice is to use the same basis $(\phi_j(t))_{j\in\mathcal{I}}$ and the same truncation parameter L for the basis expansion of both the demeaned observations $(x_i(t) - \hat{\mu}(t))$ and the functional principal components $\hat{\nu}^m$. This leads to the following simplification of Equation 29.

$$\hat{\xi}_i^m \approx \sum_{j=1}^L\left[a_{i,j}\sum_{k=1}^L b_k^m \int_0^1 \phi_j(s)\phi_k(s)\mathrm{d}s\right] =: \tilde{\xi}_i^{m,L} \tag{30}$$

3.7 Scalar-on-Function Regression

In the simple scalar setting one of the most important tools in econometrics is the linear regression. Its goal is to predict the value of a dependent variable given a set of associated variables. For reference assume a data generating process as follows. Structure is bad... This will be changed.

$$Y = X\beta + \epsilon \tag{31}$$

Where Y is the vector of response variables, X is the matrix containing the corresponding regressors in its columns and $\beta = (\beta_0, \beta_1, \ldots, \beta_p)'$ is the vector containing the unknown coefficients. In this finite dimensional setting one important question is how to estimate the unknown coefficients β. The most well known estimator in all of econometrics, the Ordinary Least Squares (OLS) estimator, fulfills this purpose under a set of assumptions. List Assumptions? Then we need to list the assumptions for functional linear regression as well I think.

$$\hat{\beta}_{OLS} = (X'X)^{-1}X'Y \tag{32}$$

The concept of linear regression can be extended to a setting of functional data, where a scalar response variable is supposed to be predicted from a functional variable. A general data generating process in this functional scenario could look like the following equation.

$$Y(\omega) = \alpha + \Psi\left(X(\omega)\right) + \epsilon(\omega) \tag{33}$$

Here Ψ is a functional that maps a realization of a random function in $\mathbb{L}^2[0,1]$ into $\mathbb{R}$. One simple example to illustrate the principle is the maximum $\Psi(f) = \max_{x\in[0,1]} f(x)$. However, the typical setup is often as follows mimicking the structure of the multivariate linear model extended from summation to integration. This structure is crucial for the extension of linear regression to the case of functional regressors. Therefore, in this paper we always implicitly assume that a data generating process has the following structure. maybe a better motivation is the one Jona gave: from very dense observations that run into problems of colinearity when using the original OLS estimator

$$Y(\omega) = \alpha + \int_0^1 \beta(s)X(\omega)(s)\mathrm{d}s + \epsilon(\omega) \tag{34}$$

Where $x(t)$ is the realization of a random function in $\mathbb{L}^2[0,1]$ and $\beta(t)$ is an unknown coefficient function. Similar to the finite dimensional setting, an interesting question is how to estimate the unknown function $\beta(t)$ given a data set containing realizations of a random function and associated scalar response variables. However, a simple extension of the OLS estimator to allow for infinite dimensional objects is not possible. Therefore, other options have to be considered.

3.7.1 Estimation using Basis-Representation

The most common way to make this problem tractable is via a basis representation of $\beta(t)$. Therefore, let $\{b_i(t) \mid i \in \mathbb{N}\}$ be a basis of $\mathbb{L}^2[0,1]$ and represent $\beta(t)$ in terms of this basis.

$$\beta(t) = \sum_{j=1}^{\infty} b_j \phi_j(t) \tag{35}$$

This enables us to write equation 34 with $\beta(t)$ represented in this way to obtain a formulation as a sum of scalar random variables $Z_j(\omega)$.

$$Y(\omega) = \alpha + \int_0^1 \beta(s)X(\omega)(s)\mathrm{d}s + \epsilon(\omega) = \alpha + \int_0^1 \left[\left(\sum_{j \in \mathcal{I}} b_j \phi_j(s) \right) X(\omega)(s) \right] \mathrm{d}s + \epsilon(\omega)$$

$$= \alpha + \sum_{j \in \mathcal{I}} \left[b_j \int_0^1 X(\omega)(s)\phi_j(s)\mathrm{d}s \right] + \epsilon(\omega) = \alpha + \sum_{j \in \mathcal{I}} b_j Z_j(\omega) + \epsilon(\omega) \tag{36}$$

This representation translates the original problem of regressing a scalar on a continuously observed function to a problem where a scalar is regressed on what is possibly a countably infinite sequence of regressors. Using a truncation of the basis at some parameter L can be used to make this problem tractable with typical theory from multivariate regression while staying reasonably accurate.

$$Y(\omega) \approx \alpha + \sum_{j=1}^{L} b_j Z_j(\omega) + \epsilon(\omega) \tag{37}$$

In practice it is common to not only express the coefficient function in terms of a basis but also the observations. Therefore two bases $((\phi_j(t))_{j \in \mathcal{I}}$ and $(\psi_k(t))_{k \in \mathcal{L}})$ and two corresponding truncation parameters (J and K) can be chosen. This leads to the following representation.

$$Y(\omega) = \alpha + \int_0^1 \beta(s)X(\omega)(s)\mathrm{d}s + \epsilon(\omega) = \alpha + \int_0^1 \left[\left(\sum_{j \in \mathcal{I}} b_j \phi_j(s) \right) \left(\sum_{k \in \mathcal{L}} d_k(\omega)\psi_k(s) \right) \right] \mathrm{d}s + \epsilon(\omega)$$

$$= \alpha + \sum_{j \in \mathcal{I}} b_j \left[\sum_{k \in \mathcal{L}} d_k(\omega) \int_0^1 \phi_j(s)\psi_k(s)\mathrm{d}s \right] + \epsilon(\omega) \approx \alpha + \sum_{j=1}^{J} \left[b_j \sum_{k=1}^{K} d_k(\omega) \int_0^1 \phi_j(s)\psi_k(s)\mathrm{d}s \right] + \epsilon(\omega) \tag{38}$$

A typical choice in this scenario is to use the same functional basis $(\phi_j(t))_{j \in \mathcal{I}}$ and the same truncation parameter L for both the coefficient function and the approximation of the observations. Defining the following notation

$$\tilde{Z}_j(\omega) = \sum_{k=1}^{L} \left[d_k(\omega) \int_0^1 \phi_j(s)\phi_k(s)\mathrm{d}s \right] \quad j = 1, \dots, L \tag{39}$$

This leads to a considerable simplification of Equation 38.

$$Y(\omega) \approx \alpha + \sum_{j=1}^{J} b_j \tilde{Z}_j(\omega) + \epsilon(\omega) \tag{40}$$

A model in the form of Equation 40 lends itself to be estimated using theory from multivariate linear regression. Define therefore the following objects

$$Y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \quad Z = \begin{pmatrix} 1 & \tilde{Z}_{1,1} & \dots & \tilde{Z}_{1,J} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \tilde{Z}_{n,1} & \dots & \tilde{Z}_{n,J} \end{pmatrix} \tag{41}$$

Then an OLS estimator can be calculated in the usual way to obtain an estimate for the values of α and b_j and an estimate of the coefficient function can be derived accordingly.

$$b^L = \left(Z'Z \right)^{-1} Z'Y \in \mathbb{R}^{L+1} \quad \hat{\alpha} = b_1^L \quad \hat{\beta}^L(t) = \sum_{j=1}^{J} b_{j+1}^L \phi_j(t) \tag{42}$$

3.7.2 Estimation using Functional Principal Components

Using the Karhunen-Loéve Expansion to represent $X(\omega)$, it is also possible to express the data generating process in a slightly different way.

$$Y(\omega) = \alpha + \int_0^1 X(\omega)(s)\beta(s)\mathrm{d}s + \epsilon(\omega) = \alpha + \int_0^1 \left(\mu(s) + \sum_{m \in \mathcal{M}} \xi^m(\omega)\nu^m(s) \right) \beta(s)\mathrm{d}s + \epsilon(\omega)$$
$$= \alpha + \int_0^1 \mu(s)\beta(s)\mathrm{d}s + \sum_{m \in \mathcal{M}} \xi^m(\omega) \int_0^1 \nu^m(s)\beta(s)\mathrm{d}s + \epsilon(\omega) = \bar{\alpha} + \sum_{m \in \mathcal{M}} \xi^m(\omega)\beta^m + \epsilon(\omega) \tag{43}$$

As these theoretical Eigenfunctions and Eigenvalues are typically unknown, the corresponding equation in sample analogues is more interesting as a representation of an observation.

$$y_i = \alpha + \int_0^1 x_i(s)\beta(s)\mathrm{d}s + \epsilon_i = \alpha + \int_0^1 \left(\hat{\mu}(s) + \sum_{m \in \mathcal{M}} \hat{\xi}_i^m \hat{\nu}^m(s) \right) \beta(s)\mathrm{d}s + \epsilon_i$$
$$= \alpha + \int_0^1 \hat{\mu}(s)\beta(s)\mathrm{d}s + \sum_{m \in \mathcal{M}} \hat{\xi}_i^m \int_0^1 \hat{\nu}^m(s)\beta(s)\mathrm{d}s + \epsilon_i = \bar{\alpha} + \sum_{m \in \mathcal{M}} \hat{\xi}_i^m \hat{\beta}^m + \epsilon_i \tag{44}$$

This, however, is a simplification for the purposes of real-world estimation as in most implementations the coefficient function and the principal components are also expressed or derived in terms of a basis that can be chosen freely. Introducing both concepts one step at a time leads to the following complication if we first introduce an expansion of the coefficient function.

$$y_i = \alpha + \int_0^1 x_i(s)\beta(s)\mathrm{d}s + \epsilon_i = \alpha + \int_0^1 \left(\hat{\mu}(s) + \sum_{m \in \mathcal{M}} \hat{\xi}_i^m \hat{\nu}^m(s)\right)\left(\sum_{j \in \mathcal{I}} b_j \phi_j(s)\right)\mathrm{d}s + \epsilon_i$$

$$= \alpha + \int_0^1 \left[\sum_{j \in \mathcal{I}} b_j \phi_j(s)\hat{\mu}(s) + \sum_{m \in \mathcal{M}}\left[\hat{\xi}_i^m \sum_{j \in \mathcal{I}} b_j \hat{\nu}^m(s)\phi_j(s)\right]\right]\mathrm{d}s + \epsilon_i \tag{45}$$

$$= \alpha + \sum_{j \in \mathcal{I}} b_j \int_0^1 \phi_j(s)\hat{\mu}(s)\mathrm{d}s + \sum_{m \in \mathcal{M}}\left[\hat{\xi}_i^m \sum_{j \in \mathcal{I}} b_j \int_0^1 \hat{\nu}^m(s)\phi_j(s)\mathrm{d}s\right] + \epsilon_i$$

If we additionally derive and approximate the principal components and corresponding scores using a truncated basis representation as in Equation 30 we obtain the following. To not complicate things more than necessary, the following equation assumes that the same basis $(\phi_j(t))_{j \in \mathcal{I}}$ was used in the derivation of the principal components and the expansion of the coefficient function. Additionally, the following approximation also truncates the basis for the expansion of the coefficient function at the same parameter L that was used for the approximation of the principal components and scores. The parameter M corresponds to the chosen number of principal components and therefore constitutes another choice in the approximation.

$$y_i \approx \alpha + \sum_{j=1}^{L} b_j \int_0^1 \phi_j(s)\hat{\mu}(s)\mathrm{d}s + \sum_{m=1}^{M}\left[\tilde{\xi}_i^{m,L} \sum_{j=1}^{L} b_j \int_0^1 \hat{\nu}^{m,L}(s)\phi_j(s)\mathrm{d}s\right] + \epsilon_i$$

$$\approx \tilde{\alpha} + \sum_{m=1}^{M}\left[\tilde{\xi}_i^{m,L} \sum_{j=1}^{L} b_j \int_0^1 \left(\sum_{k=1}^{L} n_k^m \phi_k(s)\right)\phi_j(s)\mathrm{d}s\right] + \epsilon_i \tag{46}$$

$$\approx \tilde{\alpha} + \sum_{m=1}^{M}\left[\tilde{\xi}_i^{m,L} \sum_{j=1}^{L}\left[b_j \sum_{k=1}^{L} n_k^m \int_0^1 \phi_k(s)\phi_j(s)\mathrm{d}s\right]\right] + \epsilon_i = \tilde{\alpha} + \sum_{m=1}^{M} \tilde{\xi}_i^{m,L} \tilde{\beta}^{m,L} + \epsilon_i$$

As in the previous section, this equation again lends itself for estimation with OLS. Define the following objects:

$$Y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \quad Z = \begin{pmatrix} 1 & \tilde{\xi}_1^{1,L} & \cdots & \tilde{\xi}_1^{M,L} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \tilde{\xi}_n^{1,L} & \cdots & \tilde{\xi}_n^{M,L} \end{pmatrix} \tag{47}$$

We can then derive the following estimators.

$$\tilde{b}^{L,M} = \left(Z'Z\right)^{-1} Z'Y \in \mathbb{R}^{M+1} \quad \hat{\tilde{\alpha}} = \tilde{b}_1^{L,M} \quad \hat{\tilde{\beta}}^{m,L} = \tilde{b}_{m+1}^{L,M} \tag{48}$$

3.8   Literature

- Alexanderian 2015

- Kokoszka and Reimherr 2017

- Hsing and Eubank 2015

- Ramsay and Silverman 2005

- Horváth and Kokoszka 2012

- Cai and Hall 2006

- Levitin et al. 2007

4 Simulation Study

4.1 Motivation

For the simulation study, we deviate from the standard simulation setting. Instead of generating data by ourselves, we use the gasoline data which consists of 60 samples of Near-infrared (NIR) spectra measured by 2-nm from 900 to 1,700 nm, and a response variable, the octane rating. NIR- spectroscopy allows to analyse samples of gasoline much faster, cheaper and with the same reproducibility as standard tests while not destroying the sample (cf. Gy. Bohács, Z. Ovádi, A. Salgó 1998). We chose this setup to improve the approach towards the application in which we predict the octane ratings from the gasoline dataset.

4.2 Generating Similar Curves

To avoid small sample problems, we generated 200 similar curves from the gasoline dataset, motivated by Karhunen-Loéve Expansion. First, the initial curves are expressed in terms of a generated bspline basis which is created using 50 knots. These smooth curves are then centered, before applying the Karhunen-Loéve Expansion. It is assumed that the scores follow a normal distribution, thus, the new realizations for the scores are drawn from a multivariate normal $\tilde{\xi} = \left(\tilde{\xi}_1, \ldots, \tilde{\xi}_J\right)' \sim \mathcal{N}(0_J, \; diag(\hat{\lambda}_1, \ldots, \hat{\lambda}_J))$. Finally, we obtain the generated curves NIR_{sim}

$$\tilde{X}(\omega)(t) = \hat{\mu}(t) + \sum_{j=1}^{J} \tilde{\xi}_j(\omega)\hat{\nu}_j(t)$$

where $\tilde{X}(\omega)(t)$, $\hat{\mu}(t)$ and $\hat{\nu}_j(t)$ are approximated as vectors in $\mathbb{R}^{401}$.

4.3 Simulation setup

The simulation study follows Reiss and Ogden 2007 as a guideline. Two different true coefficient functions, $f_1(t)$ and $f_2(t)$, are created that differ in their smoothness, to compare the introduced methods with differing true coefficient functions:

$$f_1(t) = 2\sin(0.5\pi t) + 4\sin(1.5\pi t) + 5\sin(2.5\pi t) \tag{49}$$

$$
\begin{aligned}
f_2(t) =& 1.5\exp\left(\frac{-0.5(t-0.3)^2}{0.02^2}\right) - 4\exp\left(\frac{-0.5(t-0.45)^2}{0.015^2}\right) \\
&+ 8\exp\left(\frac{-0.5(t-0.6)^2}{0.02^2}\right) - \exp\left(\frac{-0.5(t-0.8)^2}{0.03^2}\right)
\end{aligned}
\tag{50}
$$

The bumpy function, $f_2(t)$, was generated by referring to Cardot 2002. Both functions were created such that the inner product $\langle NIR_{sim}, f \rangle, f \in \{f_1(t), f_2(t)\}$ creates responses similar to the original octane ratings of the gasoline dataset.

Two different error-terms ϵ were created by first generating an *i.i.d.* standard normal error term and then multiplying it by two error variations σ_e. The error variations represent different signal-to-noise ratios to test the methods both with a low and a high amount of noise. They are created such that the
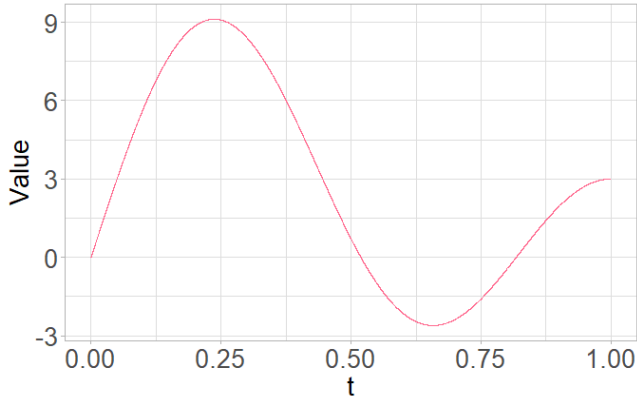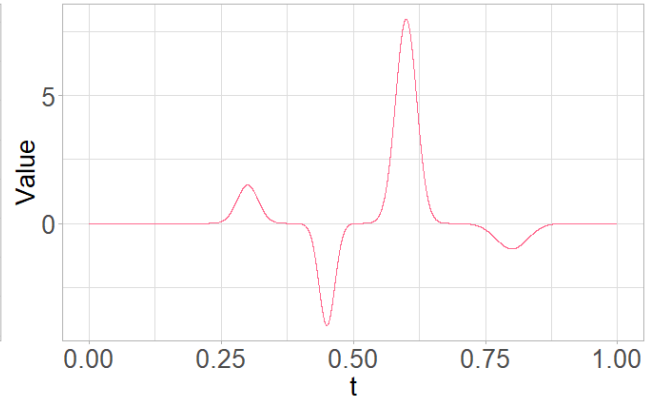
Figure 1: $f_1(t)$, smooth function

Figure 2: $f_2(t)$, bumpy function

squared multiple correlation coefficient $R^2 = var(Xf)/(var(Xf) + \sigma_e^2)$ is equal to 0.9 and 0.6. The two error-terms are then used to generate two sets of responses for $f \in \{f_1(t), f_2(t)\}$

$$Y_{1,f} = \langle NIR, f \rangle + Z \left[\frac{var(\langle NIR, f \rangle)}{0.9} - var(\langle NIR, f \rangle) \right]$$
$$Y_{2,f} = \langle NIR, f \rangle + Z \left[\frac{var(\langle NIR, f \rangle)}{0.6} - var(\langle NIR, f \rangle) \right]$$

(51)

where $Z \sim \mathcal{N}(0,1)$. In total, we created four combinations for the simulations, using the two true coefficient functions and the two sets of responses. These four combinations are then used with a different number of monomial basis functions $\in \{1, 2, \ldots, 6\}$, cubic bspline basis-function $\{5, 6, ..., 18\}$ and fourier functions $\{1, 3, ..., 25\}$ to predict the generated responses using the basis expansion approach and the FPCR approach. For the evaluation, we used the prediction RMSE calculated by 10 fold cross-validation. To obtain valid out of sample properties for the FPCR, within each of the ten 10 fold cross-validation splits, we first calculate the Functional Principal Components of the trainig-set $\mathcal{T}$ for each curve. These scores are then used to estimate the scores of the holdout set $\mathcal{H}$, $\hat{\xi}_i^{m,\mathcal{H}}$ by the equation:

$$\hat{\xi}_i^{m,\mathcal{H}} = \int_0^1 \left(X_i^{\mathcal{H}}(s) - \hat{\mu}^{\mathcal{T}}(s) \right) \hat{\nu}^{m,\mathcal{T}} \mathrm{d}s) = \int_0^1 \left(\sum_{j \in \mathcal{I}} a_{i,j}^{\mathcal{H}} \phi_j(s) \right) \left(\sum_{k \in \mathcal{L}} b_k^{m,\mathcal{T}} \psi_k(s) \right) \mathrm{d}s$$
$$= \sum_{j \in \mathcal{I}} \left[a_{i,j}^{\mathcal{H}} \sum_{k \in \mathcal{L}} b_k^{m,\mathcal{T}} \int_0^1 \phi_j(s) \psi_k(s) \mathrm{d}s \right]$$

(52)

The simulation was done with R (version...). In total, 5000 repetitions were done for each set of simulations.

4.4 Results

DESCRIBE RESULTS: make paragraph for each setup? so 6 paragraphs? Like this:

16

4.4.1 Basis Expansion Regression

The following results origin from the **Estimation using Basis-Representation**, in which we transform the observed functions to perform regression of a scalar on a countable sequence of regressors, which is then tractable with typical multivariate regression theory.

Monomial Basis   Due to the high collinearity of these basis, simulations were conducted up until the sixth monomial basis, excluding the first one since this just represents a constant. For reasons outlined in Link chapter, they are suited for the smooth function f_1 and show a better performance than bsplines for this coefficient function, but show the weakest performance for the coefficient function f_2. For f_1, the simulation selects 5(3) and for f_2 5(5) monomial basis functions for the high(low) signal-to-noise ratios. This weakness is especially pronounced in the setup f_2, Y_1; for the high noise setup, f_2, Y_2 this weakness is still visible, but less strongly pronounced.

Bspline Basis   Simulations with bspline basis functions were possible from 5 to 18, since from 18 onwards the simulations were running into problems concerning what happend here exactly? The smooth function f_1 requires only 5(4) Bspline basis functions, for the high(low)-signal-to-noise ratio to obtain the best fit for the bspline basis, which performs the worst for f_1. For f_2, 11(6) bspline basis functions are needed for the high(low) signal-to-noise ratios. For f_2, the bspline basis functions outperform the monomial basis, but come second to the fourier basis.

Fourier Basis   For f_1, the simulation chooses a smaller number of fourier basis functions, 5(3) and a higher number for the bumpy function f_2, 9(7) for the setup with the high(low) signal-to-noise ratio. With the low signal-to-noise ratio, the simulation chooses a smaller number to account for the higher noise in the response variable. The fourier basis functions performs the best in each setup for the basis expansion regression. When recalling f_1 and f_2, both, but especially f_1 shows similar curvature order across the domain which contributes to the strong performance of this basis.

For all specifications, the effect of the Bias-Variance tradeoff can be observed:: The function f_1 is rather smooth and not much bias is introduced here when choosing a small number of basis functions since f_1 lacks pronounced peculiarities. For f_2, higher number of basis functions are needed, resulting from the inherent peculiarities of f_2: We can choose a higher number of basis functions since the amount of bias is deceasing much faster in the number of basis functions than the variance is increasing, compared to f_1.

4.4.2 Functional Principal Component Regression

The model which is used for the FPCR is described in **Estimation using Functional Principal Components**, in which we transform the observed functions to perform regression of a scalar on a countable sequence of regressors, which is then tractable with typical multivariate regression theory.

Monomial Basis

Bspline Basis

Fourier Basis

4.4.3 Interpretation and Relevance for Application

4.5 Literature

- Shonkwiler and Mendivil 2009

- R-packages: fda, refund, mgcv

5   Application

The application uses the insights from the previous sections to predict the octane ratings of the introduced gasoline dataset. Following the results from the simulation study,...

- incorporate results of simulation study: number of basis, components, etc...

- point out difficulties estimating sderror

- describe setup and results

5.1   Literature

- Carey et al. 2002

6   Outlook

6.1   Literature

- James, Wang, and Zhu 2009 (shape-restrictions)
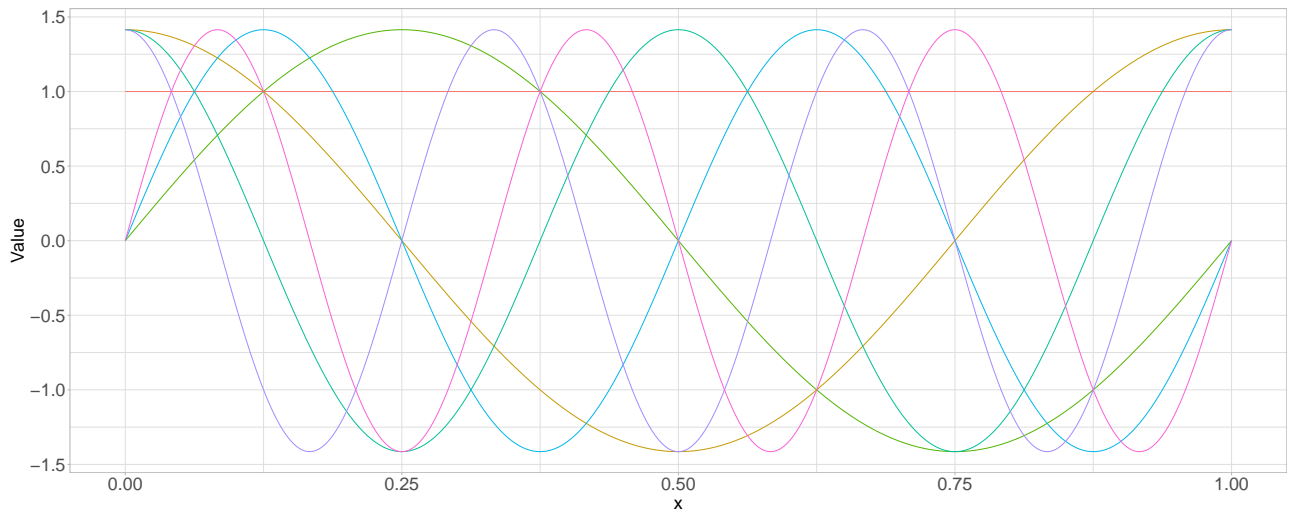
7 Appendix

7.1 Basis Plots

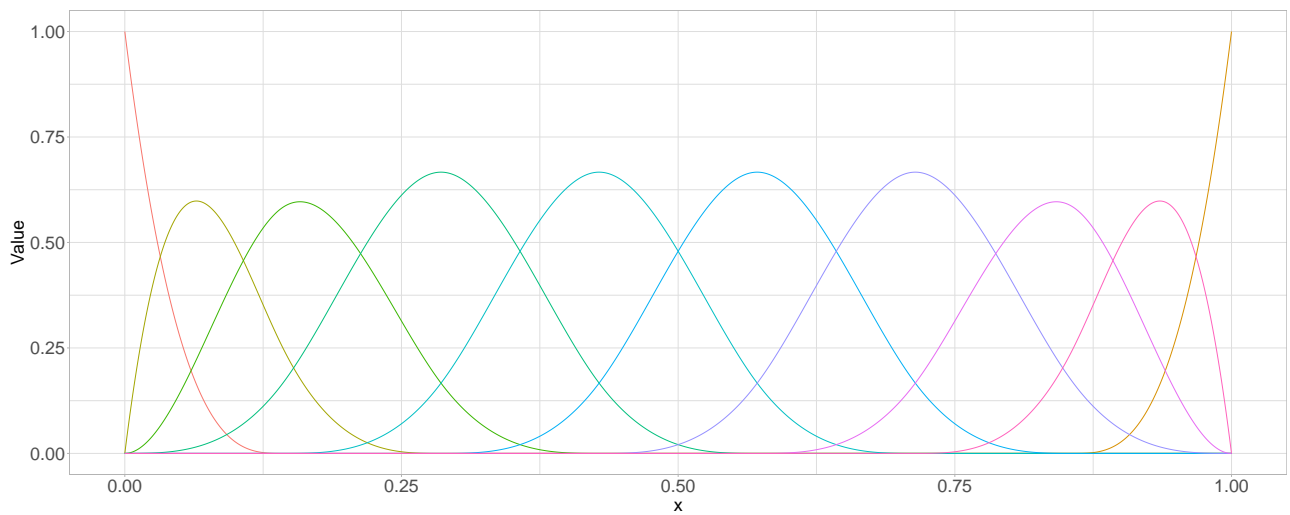Figure 3: Fourier basis functions for $i = 1, \ldots, 7$

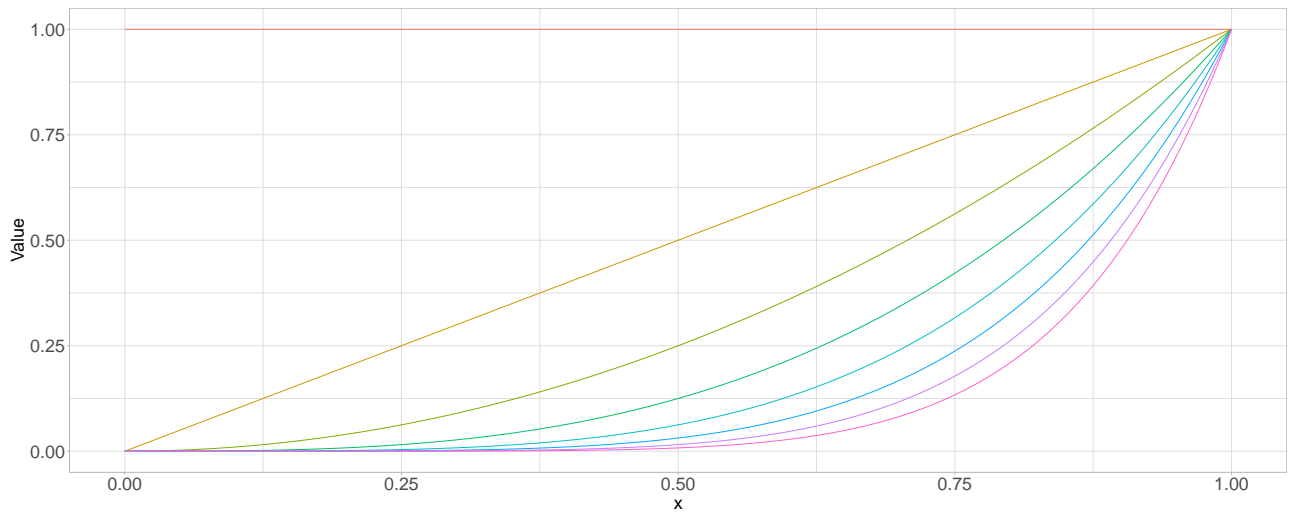Figure 4: B-spline basis functions of order 4 for 8 equidistant knots on $[0, 1]$

Figure 5: Monomial basis functions of degree 0 to 7

7.2 Simulation Study Results

Table 1: Monomial Basis Expansion Regression

f_1, Y_1	f_1, Y_2	f_2, Y_1	f_2, Y_2	n_basis
7.177234352026647	131.92716520680838	0.8936725847175324	2.5839516380614462	2
4.170391190343292	129.64258934345224	0.8122408608804015	2.511772278865247	3
3.912749103345826	130.1170783435205	0.3807927171910713	2.090474762002978	4
3.638501968909747	130.59817144950233	0.09216751463887264	1.8134256416939782	5
6.016442343989896	201.15534584991525	0.7620805051836907	3.3937342332820926	6

Table 2: BSpline Basis Expansion Regression

f_1, Y_1	f_1, Y_2	f_2, Y_1	f_2, Y_2	n_basis
3.9127490963955247	130.1170783090359	0.3807927168700936	2.090474761101881	4
3.6430480624640396	130.61094959896604	0.09511820431115069	1.8164314387816713	5
3.6542611008674575	131.35354539843397	0.07775281510299467	1.8091346529425483	6
3.6770545937377634	132.14205183557064	0.07518488287219192	1.8168036577708813	7
3.710740869697218	133.3653658811635	0.05810108378740701	1.8164984632128844	8
3.7192146351418724	133.6814928169343	0.05640282340664383	1.8193083539013821	9
3.742010447810594	134.5121710907784	0.0521785231194299	1.8257576297312081	10
3.7643964092462974	135.29727347111702	0.051902252769905684	1.8360966075573522	11
3.80109308638683	136.5758146586355	0.052040315671138586	1.8531479959118535	12
3.834360204971099	137.78406681481994	0.052706573106132304	1.869969847032031	13
3.862166917165765	138.73431326927349	0.053068654178390816	1.8826933364948355	14
3.8685638723557956	138.97427074660192	0.0525955164519272	1.8848221385552892	15
3.88505750612787	139.57418873526626	0.05283448721173191	1.8933474257957223	16
3.912671952971778	140.51490301138338	0.053220205231306356	1.9061942087104604	17
3.948125713522542	141.80885020010606	0.053584434348457485	1.9231184762987383	18

Table 3: Fourier Basis Expansion Regression

f_1, Y_1	f_1, Y_2	f_2, Y_1	f_2, Y_2	n_basis
3.6975216821108177	129.19133735032568	0.6952364931614831	2.394399533437222	3
3.6347035142043147	130.59281554182246	0.07418129892625985	1.795816800944426	5
3.6762259845089287	132.0824820138016	0.05147130962678385	1.7934269021325526	7
3.718852537544315	133.67575428625497	0.05104970622270979	1.812908100982769	9
3.764506721475903	135.26218648523314	0.05146451789575423	1.8346302450265894	11
3.810949927900835	136.90281800923086	0.05197380387819929	1.8572425102431276	13
3.8567441840723165	138.57257529224665	0.0525212098897837	1.8802107534919845	15
3.9061852571557436	140.29178124765565	0.05304327165297495	1.9028253007284568	17
3.9551703628111876	142.057266882177	0.053647142499864804	1.9271845824346567	19

7.3 Wiener Process

A Wiener process W_t is a real-valued continuous-time stochastic process ... This is wikipedia, look for right citation! It is characterized by the following properties.

1. $W_0 = 0$

2. $\forall t > 0 \, W_{t+u} - W_t \perp\!\!\!\perp W_s \, \forall s \leq t$

3. $W_{t+u} - W_t \sim \mathcal{N}(0, u)$

4. W_t is continuous in t
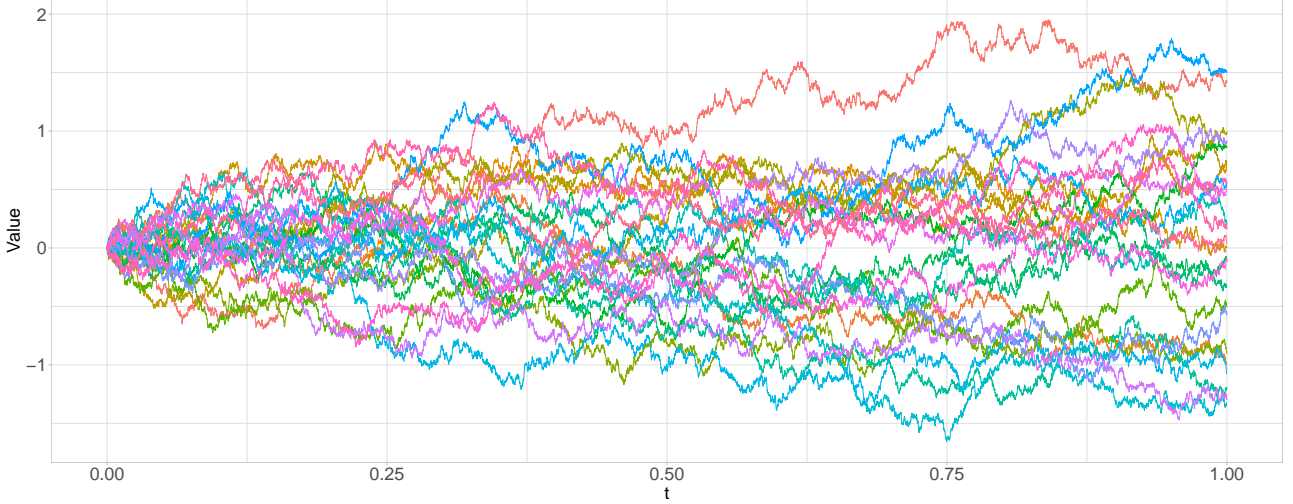

Figure 6: 25 i.i.d. realizations of a Wiener Process on $[0, 1]$

8 Proofs

8.1 Lemma

We might put the below theorems for KL. The deterministic basis expansions are pretty clear about why we can use the truncation version by itself. In contrast, I feel like that for the case of KL it does not. And in the centered case? The coefficients ξ_j satisfy the following:

1. $\mathbb{E}[\xi_j] = 0$

2. $\mathbb{E}[\xi_j \xi_k] = \lambda_j$, if $j = k$, otherwise 0

3. $\mathrm{Var}[\xi_j] = \lambda_j$

Proof. To obtain the first result, we can show that

$$
\begin{aligned}
\mathbb{E}[\xi_j] &= \mathbb{E}\left[\int_0^1 F(t)\nu_j(t)dt\right] \\
&= \int_\Omega \int_0^1 F(t)\nu_j(t)dtdP(\omega) \\
&= \int_0^1 \int_\Omega F(t)\nu_j(t)dP(\omega)dt \quad \text{(Fubini)} \\
&= \int_0^1 \mathbb{E}[F(t)]\nu_j(t)dt \\
&= \mu(t)\int_0^1 \nu_j(t) \\
&= \mu(t)\langle \nu_j(t), 1\rangle = 0
\end{aligned}
\tag{53}
$$

where the constant function is always an eigenfunction of the operater K. It leads the inner product to 0 due to orthonormality of eigenfunctions. Another way to prove or use centered mean For the case of centered process

$$
\begin{aligned}
\mathbb{E}[\xi_j] =& \mathbb{E}\left[\int_0^1 F(t)\nu_j(t)dt\right] \\
=& \int_\Omega \int_0^1 F(t)\nu_j(t)dt dP(\omega) \\
=& \int_0^1 \int_\Omega F(t)\nu_j(t)dP(\omega)dt \quad \text{(Fubini)} \\
=& \int_0^1 \mathbb{E}[F(t)]\nu_j(t)dt = 0
\end{aligned}
\tag{54}
$$

where $\mathbb{E}[F(t)]$ is 0 since $F(t)$ is a centered process. The second claim is proved as:

$$
\begin{aligned}
\mathbb{E}[\xi_j \xi_k] =& \mathbb{E}\left[\int_0^1 F(s)\nu_j(s)ds \int_0^1 F(t)\nu_k(t)dt\right] \\
=& \mathbb{E}\left[\int_0^1 \int_0^1 F(s)\nu_j(s)F(t)\nu_k(t)dsdt\right] \quad \text{(Fubini)} \\
=& \int_0^1 \int_0^1 \mathbb{E}[F(s)F(t)]\nu_j(s)\nu_k(t)dsdt \\
=& \int_0^1 \left(\int_0^1 c(s,t)\nu_j(s)ds\right)\nu_k(t)dt \\
=& \int_0^1 [K\nu_j](t)\nu_k(t)dt \\
=& \langle K\nu_j, \nu_k \rangle \\
=& \langle \lambda_j \nu_j, \nu_k \rangle = \lambda_j \quad \text{if } j = k, \text{ otherwise } 0
\end{aligned}
\tag{55}
$$

where the result is produced from orthonormaility of the eigenfunctions. The last assertion is confirmed from the other two properties.

$$
Var[\xi_i] = \mathbb{E}\left[(\xi_i - \mathbb{E}[\xi_i])^2\right] = \mathbb{E}[\xi_i^2] = \lambda_i
\tag{56}
$$

$\square$

8.2   Karhunen-Loéve expansion

Let $F : [0,1] \to \mathbb{R}$ be a centered mean-square continuous stochastic process with $F \in \mathbb{L}^2[0,1]$. Then there exists a basis ξ_i of $\mathbb{L}^2[0,1]$ such that for all $t \in [0,1]$,

$$
F(t) = \sum_{j=1}^\infty \xi_j \nu_j(t),
\tag{57}
$$

where coefficients ξ_j are given by $\xi_j(\omega) = \int_0^1 F(t)(\omega)\nu_j(t)dt$ and satisfy the following conditions.

1. $\mathbb{E}[\xi_j] = 0$

2. $\mathbb{E}[\xi_j \xi_k] = \lambda_j$, if $j = k$, otherwise 0

3. $Var[\xi_j] = \lambda_j$

Proof. We know that K has a complete set of eigenvectors ν_j in $\mathbb{L}^2[0,1]$ and non-negative eigenvalues λ_j. With the fact that $\xi_j(\omega) = \int_0^1 F(\omega)(t)\nu_j(t)dt$ satisfy the three conclusions by Lemma. We prove it by considering

$$\epsilon_n(t) := \mathbb{E}\left[\left(F(t) - \sum_{j=1}^n \xi_j\nu_j(t)\right)^2\right] \tag{58}$$

Once it is shown that $\lim_{n\to\infty} \epsilon_n(t) = 0$ uniformly in [0,1], the proof is completed.

$$\begin{aligned}
\epsilon_n(t) =& \mathbb{E}\left[\left(F(t) - \sum_{j=1}^n \xi_j\nu_j(t)\right)^2\right] \\
=& \mathbb{E}[F(t)^2] - 2\mathbb{E}\left[F(t)\sum_{j=1}^n \xi_j\nu_j(t)\right] + \mathbb{E}\left[\sum_{j=1}^n\sum_{k=1}^n \xi_j\xi_k\nu_j(t)\nu_k(t)\right]
\end{aligned} \tag{59}$$

Now, $\mathbb{E}[F(t)^2] = c(t,t)$ as in (19),

$$\begin{aligned}
\mathbb{E}\left[F(t)\sum_{j=1}^n \xi_j\nu_j(t)\right] =& \mathbb{E}\left[F(t)\sum_{j=1}^n\left(\int_0^1 F_s\nu_i(s)ds\right)\xi_j(t)\right] \\
=& \sum_{j=1}^n\left(\int_0^1 \mathbb{E}[F(t)F(s)]\nu_j(s)ds\right)\nu_j \\
=& \sum_{j=1}^n\left(\int_0^1 c(t,s)\nu_j(s)ds\right)\nu_j \\
=& \sum_{j=1}^n [K\nu_j](t)\nu_j = \sum_{j=1}^n \lambda_j\nu_j^2
\end{aligned} \tag{60}$$

We derive from (55) that

$$\begin{aligned}
\mathbb{E}\left[\sum_{j=1}^n\sum_{k=1}^n \xi_j\xi_k\nu_j(t)\nu_k(t)\right] =& \sum_{j=1}^n\sum_{k=1}^n \mathbb{E}[\xi_j\xi_k]\nu_j(t)\nu_k(t) \\
=& \sum_{j=1}^n\sum_{k=1}^n \delta_{jk}\lambda_j\nu_j(t)\nu_k(t) = \sum_{j=1}^n \lambda_j\nu_j(t)^2
\end{aligned} \tag{61}$$

where $\delta_{jk} = 1$ if $j = k$, otherwise 0. Therefore, by (59), (60), and (61) we obtain

$$\epsilon_n(t) = c(t,t) - \sum_{j=1}^n \lambda_j\nu_j(t)\nu_j(t) \tag{62}$$

implementing Mercer's Theorem this proof is concluded by

$$\lim_{n\to\infty} \epsilon_n(t) = \lim_{n\to\infty} \mathbb{E}\left[\left(F(t) - \sum_{j=1}^n \xi_j\nu_j(t)\right)^2\right] = 0 \tag{63}$$

$\square$

9 Bibliography

Alexanderian, Alen (2015). "A brief note on the Karhunen-Loève expansion". In: *arXiv: Probability*. URL: https://arxiv.org/abs/1509.07526.

Bauer, Heinz (May 2020). *Wahrscheinlichkeitstheorie und Grundzüge der Maßtheorie*. de. Publication Title: Wahrscheinlichkeitstheorie und Grundzüge der Maßtheorie. De Gruyter. ISBN: 978-3-11-231316-9. DOI: 10.1515/9783112313169. URL: https://www.degruyter.com/document/doi/10.1515/9783112313169/html (visited on 11/13/2021).

Boor, Carl de (Jan. 1978). *A Practical Guide to Spline*. Vol. Volume 27. Journal Abbreviation: Applied Mathematical Sciences, New York: Springer, 1978 Publication Title: Applied Mathematical Sciences, New York: Springer, 1978. DOI: 10.2307/2006241.

Cai, T. Tony and Peter Hall (Oct. 2006). "Prediction in functional linear regression". In: *The Annals of Statistics* 34.5. Publisher: Institute of Mathematical Statistics, pp. 2159–2179. ISSN: 0090-5364, 2168-8966. DOI: 10.1214/009053606000000830. URL: https://projecteuclid.org/journals/annals-of-statistics/volume-34/issue-5/Prediction-in-functional-linear-regression/10.1214/009053606000000830.full (visited on 10/24/2021).

Cardot, Hervé (2002). "Spatially Adaptive Splines for Statistical Linear Inverse Problems". In: *Journal of Multivariate Analysis* 81.1, pp. 100–119. ISSN: 0047-259X. DOI: https://doi.org/10.1006/jmva.2001.1994. URL: https://www.sciencedirect.com/science/article/pii/S0047259X01919943.

Carey, James R. et al. (2002). "Life history response of Mediterranean fruit flies to dietary restriction". en. In: *Aging Cell* 1.2. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1046/j.1474-9728.2002.00019.x, pp. 140–148. ISSN: 1474-9726. DOI: 10.1046/j.1474-9728.2002.00019.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1046/j.1474-9728.2002.00019.x (visited on 10/24/2021).

Gy. Bohács, Z. Ovádi, A. Salgó (1998). "Prediction of Gasoline Properties with near Infrared Spectroscopy". In: *Journal of near infrared spectroscopy.* 6, pp. 341–348.

Horváth, Lajos and Piotr Kokoszka (May 2012). *Inference for Functional Data with Applications*. en. Google-Books-ID: OVezLB_ZpYC. Springer Science & Business Media. ISBN: 978-1-4614-3655-3.

Hsing, Tailen and Randall Eubank (Mar. 2015). *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators*. en. Google-Books-ID: om9uBwAAQBAJ. John Wiley & Sons. ISBN: 978-1-118-76256-1.

James, Gareth M., Jing Wang, and Ji Zhu (2009). "Functional linear regression that's interpretable". In: *The Annals of Statistics* 37.5A. ISSN: 0090-5364. DOI: 10.1214/08-AOS641.

Kokoszka, Piotr and Matthew Reimherr (Aug. 2017). *Introduction to Functional Data Analysis*. Englisch. 1st ed. Boca Raton: Chapman and Hall/CRC. ISBN: 978-1-4987-4634-2.

Levitin, Daniel et al. (Aug. 2007). "Introduction to Functional Data Analysis". In: *Canadian Psychology/Psychologie canadienne* 48, pp. 135–155. DOI: 10.1037/cp2007014.

Ramsay, James and B. W. Silverman (2005). *Functional Data Analysis*. en. 2nd ed. Springer Series in Statistics. New York: Springer-Verlag. ISBN: 978-0-387-40080-8. DOI: 10.1007/b98888. URL: https://www.springer.com/de/book/9780387400808 (visited on 10/23/2021).

Reiss, Philip T. and R. Todd Ogden (2007). "Functional Principal Component Regression and Functional Partial Least Squares". In: *Journal of the American Statistical Association* 102.479, pp. 984–996. ISSN: 0162-1459. DOI: 10.1198/016214507000000527.

Shonkwiler, Ronald W. and Franklin Mendivil (2009). *Explorations in Monte Carlo Methods*. en. Undergraduate Texts in Mathematics. New York: Springer-Verlag. ISBN: 978-0-387-87836-2. DOI: 10.1007/978-0-387-87837-9. URL: https://www.springer.com/gp/book/9780387878362 (visited on 10/23/2021).

10   Affidavit

"I hereby confirm that the work presented has been performed and interpreted solely by myself except for where I explicitly identified the contrary. I assure that this work has not been presented in any other form for the fulfillment of any other degree or qualification. Ideas taken from other works in letter and in spirit are identified in every single case."

Bonn, 11.02.2021

Jonghun Baek

Bonn, 11.02.2021

Jakob R. Juergens

Bonn, 11.02.2021

Jonathan Willnow