

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	3
2	Introduction	3
3	Theory	4
3.1	Inner Products and Hilbert Spaces	4
3.2	Hilbert Space of Square-Integrable Functions	5
3.3	Bases of \mathbb{L}^2	5
3.4	Approximation and Smoothing via Basis Truncation	7
3.5	Functional Data Sets	7
3.6	Karhunen-Loève Expansion and Empirical Eigenbases	8
3.7	Scalar-on-Function Regression	9
3.8	Estimation using Basis-Representation	10
3.9	Estimation using Functional Principal Components	11
3.10	Detailed Draft	11
3.11	Draft-Overview	14
3.12	Literature	14
4	Simulation Study	16
4.1	Draft-Overview	16
4.2	Motivation	16
4.3	Generating Similar Curves	16
4.4	Simulation setup	16
4.5	Results	18
4.5.1	Interpretation and Relevance for Application	18
4.6	Literature	18
5	Application	19
5.1	Draft-Overview	19
5.2	Literature	19
6	Outlook	19
6.1	Literature	19
7	Appendix	19
7.1	Basis Plots	19
7.2	Wiener Process	20
8	Bibliography	21

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 a relatively new field (**roots in the 1940s Grenander and Karhunen**) which is gaining more attention as researchers from different fields collect data that is functional in nature. Classical statistical methods can often process this data, but only FDA allows extracting information given by the smoothness of the underlying process (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).

To motivate scalar-on-function regression, consider the case of a data set containing a scalar response and observations of an underlying continuous process. In economics, one application could be the regression of stock market correlations on the Global Crisis Index (GCI), where the regression allows to assess the relationship between the correlation and the GCI at every point within a window (cf. Das et al. 2019).

The focus of this paper is to introduce Functional Linear Regression (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:

$$Y(\omega) = \alpha + \int_0^1 X(\omega)(s)\beta(s)ds + \epsilon(\omega), \quad i = 1, \dots, n \quad (1)$$

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 $X : \Omega \rightarrow \Omega'$ is an \mathcal{A} - \mathcal{A}' -measurable function, where (Ω, \mathcal{A}, P) is a probability space and (Ω', \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 $\langle \cdot, \cdot \rangle : \mathbb{V} \times \mathbb{V} \rightarrow \mathbb{F}$ 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 $\mathbb{F} = \mathbb{R}$.

$$\|v\| = \langle v, v \rangle^{\frac{1}{2}} \quad (2)$$

$$d(v_1, v_2) = \langle v_2 - v_1, v_2 - v_1 \rangle^{\frac{1}{2}} \quad (3)$$

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

$$\text{span}(S) = \left\{ \sum_{i=1}^k \lambda_i v_i \mid k \in \mathbb{N}, v_i \in S, \lambda_i \in \mathbb{R} \right\} \quad (4)$$

The closed span $\overline{\text{span}}(S)$ 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 $\overline{\text{span}}(S) = \mathbb{H}$.

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.

$$x = \sum_{j=1}^{\infty} \langle x, s_j \rangle s_j \quad (5)$$

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.

$$x \approx \sum_{j=1}^K \langle x, s_j \rangle s_j \quad (6)$$

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.

$$\|f\|^2 = \int_0^1 |f|^2 d\mu < \infty \quad (7)$$

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

$$\langle f_1, f_2 \rangle = \int_0^1 f_1 f_2 d\mu. \quad (8)$$

The Hilbert space of square integrable functions on $[0, 1]$ (instead, $\mathbb{L}^2[0, 1]$) is the function space that is most often used for theoretical considerations in functional data analysis without loss of generality. Trivially, it is possible to extend this to other closed intervals on the reals line, but there are also more complex generalizations. For the purpose of this paper we will focus on the typical case and assume that random functions are random variables realizing in \mathbb{L}^2 for some closed interval of the real numbers. The random function defined on $\mathbb{L}^2[0, 1]$ can be represented as

$$\{f(t, \omega) : t \in [0, 1], \omega \in \Omega\} \quad (9)$$

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 $\{b_i(t) \mid i \in \mathcal{I}\}$ be the basis used to express / approximate a realization $F(\omega_0) = f(t)$ of $F(\omega)$.

$$F(\omega_0) = f(t) = \sum_{j \in \mathcal{I}} \psi_j(\omega_0) b_j(t) \quad (10)$$

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

$$f(t) = \sum_{k=0}^K c_k (t - w)^k \quad (11)$$

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

$$S_k^{MB}(t) = (t - w)^k, k = 0, 1, \dots, K. \quad (12)$$

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

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 $[0, 1]$.

$$S_i^{FB}(x) = \begin{cases} 1 & \text{if } i = 1 \\ \sqrt{2} \cos(\pi i x) & \text{if } i \text{ is even} \\ \sqrt{2} \sin(\pi(i - 1)x) & \text{otherwise} \end{cases} \quad (13)$$

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.

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 $S_{l,m}^{BS}(x)$ $l = 1, \dots, L - 1$ be the B-spline of order m for an interval $[\tau_0, \tau_L]$ and knots $\{\tau_l \mid l = 1, \dots, L - 1\}$, then it is defined by the Cox-de Boor recursion

formula as follows.

$$S_{l,0}^{BS}(x) = \begin{cases} 1 & \text{if } x \in [\tau_l, \tau_{l+1}) \\ 0 & \text{otherwise} \end{cases} \quad (14)$$

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

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:

$$F(\omega_0) = f(t) = \sum_{j \in \mathcal{I}} \psi_j(\omega_0) b_j(t) = \sum_{j=1}^L \psi_j(\omega_0) b_j(t) + \delta(t) \approx \sum_{j=1}^L \psi_j(\omega_0) b_j(t) \quad (15)$$

where $\delta(t)$ is the truncation error. The number L is subjectively chosen but we can fix the number 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!

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 7.2.

3.6 Karhunen-Loève 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 $X(\omega)$ be a random vector realizing in \mathbb{R}^p with $\mathbb{E}(X(\omega)) = \mu_x$ and $Cov(X(\omega)) = \Sigma_X$. Then Σ_X can be expressed in terms of its Eigenvalues and orthonormal Eigenvectors as $\Sigma_X = \Gamma \Lambda \Gamma'$ using its Eigendecomposition.

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

$$X(\omega) = \mu_x + \Gamma \Lambda^{\frac{1}{2}} D(\omega) = \mu_x + \sum_{i=1}^p \sqrt{\lambda_i} D_i(\omega) \gamma_i \quad (16)$$

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 $F : \Omega \mapsto \mathbb{L}^2[0, 1]$ be such a random function. Then the mean and covariance functions of F are defined as follows.

$$\mu(t) = \mathbb{E}[F(\omega)(t)] \quad (17)$$

$$c(t, s) = \mathbb{E}[(F(\omega)(t) - \mu(t))(F(\omega)(s) - \mu(s))] \quad (18)$$

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

$$[Kv](t) = \int_0^1 c(t, s)v(s)ds \quad (19)$$

Then, the operator K has orthonormal basis functions $v_i \in L^2[0, 1]$ 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-dimentional.
3. The eigenvalues can be ordered in nonincreasing order as follows $\lambda_1 \geq \lambda_2 \geq \dots \geq 0$.

Therefore, we can approximate the function F with first few eigenfunctions. [to here](#). Then, we can justify why the function F can be explaiend enough well with few eigenfunctions and how to find the variance proporation with $\text{Var}(\xi) = \lambda$ Theoretical considerations lead to the result that F can be represented in the following form, called its Karhunen-Loève expansion.

$$F(\omega)(t) = \mu(t) + \sum_{j=1}^{\infty} \xi_j(\omega) \nu_j(t) \quad (20)$$

where the ν_j are defined by the countable set of solutions $\{(\lambda_i, \nu_i) \mid i \in \mathbb{N}\}$ of the following equation.

$$\int_0^1 c(t, s) \nu(s) ds = \lambda \nu(t) \quad (21)$$

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

$$\xi_j(\omega) = \langle F(\omega) - \mu, \nu_j \rangle = \int_0^1 (F(\omega)(s) - \mu(s)) \nu_j(s) ds \quad (22)$$

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.

1. $\mathbb{E}[\xi_j(\omega)] = 0$
2. $Var(\xi_j(\omega)) = \lambda_i$
3. $Cov(\xi_j(\omega), \xi_k(\omega)) = 0$ 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 $\{f_1(t), \dots, f_n(t)\}$ be a set of i.i.d. realizations generated by a random function $F(\omega) \mapsto \mathbb{L}^2[0, 1]$. Define the following sample analogues for the mean and covariance functions.

$$\hat{\mu}(t) = \frac{1}{n} \sum_{j=1}^n f_j(t) \quad (23)$$

$$\hat{c}(t, s) = \frac{1}{n} \sum_{j=1}^n (f_j(t) - \hat{\mu}(t)) (f_j(s) - \hat{\mu}(s)) \quad (24)$$

With these it is possible to derive a set of sample analogs $\{(\hat{\lambda}_i, \hat{\nu}_i) \mid i \in \mathbb{N}\}$ for $\{(\lambda_i, \nu_i) \mid i \in \mathbb{N}\}$ as the solutions of the following equation. **I think the number of principal components has to be smaller than the number of observations. So I can give more information about \mathbb{N} .**

$$\int_0^1 \hat{c}(t, s) \hat{\nu}(s) ds = \hat{\lambda} \hat{\nu}(t) \quad (25)$$

This naturally leads to the following representation.

$$f_i(t) = \sum_{j=1}^{\text{!!!}} \hat{\xi}_{i,j} \hat{\nu}(t) \quad (26)$$

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.**

$$Y = X\beta + \epsilon \quad (27)$$

Where Y is the vector of response variables, X is the matrix containing the corresponding regressors in its columns and $\beta = (\beta_0, \beta_1, \dots, \beta_p)'$ 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.**

$$\hat{\beta}_{OLS} = (X'X)^{-1}X'Y \quad (28)$$

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.

$$Y(\omega) = \alpha + \Psi(F(\omega)) + \epsilon(\omega) \quad (29)$$

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**

$$Y(\omega) = \alpha + \int_0^1 \beta(s)F(\omega)(s)ds + \epsilon(\omega) \quad (30)$$

Where $f(t)$ 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.

3.8 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.

$$\beta(t) = \sum_{j=1}^{\infty} \psi_j b_j(t) \quad (31)$$

This enables us to write equation 30 with $\beta(t)$ represented in this way to obtain a formulation as a sum of scalar random variables $Z_j(\omega)$.

$$\begin{aligned} Y(\omega) &= \alpha + \int_0^1 \beta(s)X(\omega)(s)ds + \epsilon(\omega) = \alpha + \int_0^1 \left[\left(\sum_{j \in \mathcal{I}} c_j \phi_j(s) \right) X(\omega)(s) \right] ds + \epsilon(\omega) \\ &= \alpha + \sum_{j \in \mathcal{I}} \left[c_j \int_0^1 X(\omega)(s) \phi_j(s) ds \right] + \epsilon(\omega) = \alpha + \sum_{j \in \mathcal{I}} c_j Z_j(\omega) + \epsilon(\omega) \end{aligned} \quad (32)$$

This representation translates the original problem of regressing a scalar on a continuously observed function to a problem where a scalar is regressed on 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.

$$Y(\omega) \approx \alpha + \sum_{j=1}^L c_j Z_j(\omega) + \epsilon(\omega) \quad (33)$$

This fancy equation will be very useful to explain our simulation.

$$\begin{aligned} Y(\omega) &= \alpha + \int_0^1 \beta(s) X(\omega)(s) ds + \epsilon(\omega) = \alpha + \int_0^1 \left[\left(\sum_{j \in \mathcal{I}} c_j \phi_j(s) \right) \left(\sum_{k \in \mathcal{L}} d_k(\omega) \psi_k(s) \right) \right] ds + \epsilon(\omega) \\ &= \alpha + \sum_{j \in \mathcal{I}} c_j \left[\sum_{k \in \mathcal{L}} d_k(\omega) \int_0^1 \phi_j(s) \psi_k(s) ds \right] + \epsilon(\omega) = \alpha + \sum_{j \in \mathcal{I}} \left[c_j \sum_{k \in \mathcal{L}} Z_{j,k}(\omega) \right] + \epsilon(\omega) \\ &\approx \alpha + \sum_{j=1}^J \left[c_j \sum_{k=1}^K Z_{j,k}(\omega) \right] + \epsilon(\omega) = \alpha + \sum_{j=1}^J c_j \tilde{Z}_j(\omega) + \epsilon(\omega) \end{aligned} \quad (34)$$

A typical choice in this scenario is to use the same functional basis and the same truncation parameter L for both the coefficient function and the approximation of the observations. This leads to the following simplification of Equation 34.

$$\begin{aligned} Z_{j,k}(\omega) &= d_k(\omega) \int_0^1 \phi_j(s) \phi_k(s) ds \quad j = 1, \dots, L \quad k = 1, \dots, L \\ \tilde{Z}_j(\omega) &= \sum_{k=1}^L \left[d_k(\omega) \int_0^1 \phi_j(s) \phi_k(s) ds \right] \quad j = 1, \dots, L \end{aligned} \quad (35)$$

3.9 Estimation using Functional Principal Components

$$\begin{aligned} \hat{\xi}_{j,i} &= \int_0^1 (f_i(s) - \hat{\mu}(s)) \hat{\nu}(s) ds = \int_0^1 \left(\sum_{j \in \mathcal{I}} a_{i,j} \phi_j(s) \right) \left(\sum_{k \in \mathcal{L}} b_k \psi_k(s) \right) ds \\ &= \sum_{j \in \mathcal{I}} \sum_{k \in \mathcal{L}} a_{i,j} b_k \int_0^1 \phi_j(s) \psi_k(s) ds = \sum_{j \in \mathcal{I}} \sum_{k \in \mathcal{L}} Z_{i,j,k} \approx \sum_{j=1}^J \sum_{k=1}^K Z_{i,j,k} = \sum_{j=1}^J \tilde{Z}_{i,j} \end{aligned} \quad (36)$$

Again using the same basis and the same truncation parameter L leads to the following simplifications of Equation 36.

$$\begin{aligned} Z_{i,j,k} &= a_{i,j} b_k \int_0^1 \phi_j(s) \phi_k(s) ds \quad j = 1, \dots, L \quad k = 1, \dots, L \\ \tilde{Z}_{i,j} &= a_{i,j} \sum_{k=1}^L \left[b_k \int_0^1 \phi_j(s) \phi_k(s) ds \right] \quad j = 1, \dots, L \end{aligned} \quad (37)$$

3.10 Detailed Draft

- Motivate random functions from introduction and the general concept of random variables
- Formalize random function in this context as random variables realizing in a Hilbert space

- Introduce $\mathbb{L}^2[0, 1]$ as the Hilbert space of square integrable functions on $[0, 1]$
- Specialize to Hilbert space being $\mathbb{L}^2[0, 1]$ for this context
- Define mean and covariance function of a random function realizing in $\mathbb{L}^2[0, 1]$
- Introduce the concept of a basis of a Hilbert space and specialize to $\mathbb{L}^2[0, 1]$
- Introduce b-spline and Fourier bases
- Introduce eigenfunctions and FPCA on the basis of covariance function (Karhunen-Lo  ve expansion)
- explain similarities to Eigenvalues and Eigenvectors of matrix + PCA (fraction of explained variance etc...)
- Introduce functional observations in this context as realizations of a random variable realizing in $\mathbb{L}^2[0, 1]$
- Explain the concept of iid data in a functional setting
- Define point-wise mean (sample), point-wise standard deviation (sample) and sample covariance function
- Explain approximations of functional observations using truncated basis representations
- Introduce linear operator L_1 and sufficient condition associated with it
- Motivate Scalar-on-function regression from multivariate linear regression with a scalar response variable

There are several important aspects of functional regression in this functional setting that separate it from usual multiple regression according to Kokoszka and Reimherr 2017: In functional linear regression, the aim is not only to obtain an estimate of the function $\beta(s)$ — this estimate also needs to have a useful interpretation. Without it, there might be prediction, but the increase in understanding of the underlying question will be minimal. One aspect of a useful interpretation is that the estimate $\beta(s)$ should not jump in a seemingly random fashion, because an interpretation of this erratic behavior will often be impossible.

A common setting in non-functional regression is akin to the following. Assume a model as follows:

$$Y = X'\beta + \epsilon \quad (38)$$

where $X \in \mathbb{R}^{n \times J}$ is a matrix containing the regressors, $\beta \in \mathbb{R}^J$ is a coefficient vector and ϵ is a vector containing the error term. For simplicity, assume that the data generating process fulfills the Markov assumptions. Then the famous OLS-estimator is given by:

$$\hat{\beta}_{OLS} = (X'X)^{-1}X'Y \quad (39)$$

A naive approach to FLR would be to try to generalize this to the functional setting. Assuming a data generating process of the form:

$$Y = \int \beta(s)X(s) dt + \epsilon \quad (40)$$

it becomes clear that we cannot compute the estimate of $\beta(t)$ as we would do in a classical multivariate setup because of the infinite dimensionality of the underlying objects. Where in the finite dimensional setting the OLS estimator can be derived as a method of moments estimator by solving a system of equations of sample moment restrictions, this leads to a system of infinitely many equations in the functional setting. In practice, functional observations are never truly continuously observed. If we assume that the functional observations are observed at a finite set of points $\{t_1, \dots, t_J\}$ this makes the derivation of an OLS estimator possible as before.

$$Y_i = \sum_{j=1}^n \beta(t_j)X_i(t_j) + \epsilon_i, \quad (41)$$

However, this often still results in a large and difficult to solve system of equations. Even if solved, the result is often a noisy function $\hat{\beta}(s)$ that is not useful for interpretation since it does not use the intuition of smooth functions. Another reason why estimation is not feasible using this approach is colinearity. Looking at equation 41 and assuming continuous functions X_i it becomes clear that if t_j is close to $t_{j'}$, $X_i(t_j)$ is close to $X_i(t_{j'})$. Thereby, there will be vectors $X_i = (X_i(t_1), \dots, X_i(t_J))'$ that are highly correlated and thus lead to large variances of β . (cf. Kokoszka and Reimherr 2017)

A different approach is necessary.

Define therefore

$$c_{\mathbf{X}}(t, s) = E[\mathbf{X}(t)\mathbf{X}(s)], \quad c_{\mathbf{XY}}(t) = E[\mathbf{X}(t)\mathbf{Y}], \quad (42)$$

Under the assumption that X is independent from ϵ we obtain

$$c_{\mathbf{XY}}(t) = E[\mathbf{X}(t) \int \beta(s)\mathbf{X}(s) ds + \epsilon] \quad (43)$$

$$c_{\mathbf{XY}}(t) = E\left[\int \beta(s)\mathbf{X}(s)\mathbf{X}(t) ds \mid X\right] + E[\epsilon|\mathbf{X}] \quad (44)$$

$$c_{\mathbf{XY}}(t) = \int c_{\mathbf{X}}(t, s)\beta(s) ds \quad (45)$$

In practice, this results in a large and often difficult to solve system of equations. If solved, a perfect fit is possible, but results in a noisy and erratic function $\hat{\beta}(s)$ that is not useful for interpretation since it does not utilize the intuition of smooth functions (cf. Horváth and Kokoszka 2012). Another reason why estimation is not feasible using this approach is colinearity. If we approximate the scalar-on-functional regression by assuming a set of discrete observation points for all realizations of the data generating process as

$$Y_i = \sum_{j=1}^n \beta(t_j)X_i(t_j) + \epsilon_i, \quad (46)$$

it becomes clear that if t_j is close to $t_{j'}$, $X_i(t_j)$ is close to $X_i(t_{j'})$ there will be vectors $X_i = (X_i(t_1), \dots, X_i(t_J))'$ that are highly correlated and thus lead to large variances of β . (cf. Kokoszka and Reimherr 2017) **isn't something missing here? Like "and employ standard multivariate linear**

regression” // Will be done (Jona)

- Solution: estimation using truncated basis expansion to approximate data (theoretical description)

The simplest approach to regularize the noisy estimate of $\hat{\beta}(s)$ is to expand it with deterministic basis function. Assume

$$\beta(t) = \sum_{k=1}^K c_k B_k(t) \quad (47)$$

to expand

$$\int \beta(s) X(s) dt = \sum_{k=1}^K c_k \int B_k(t) X_i(t) dt =: \sum_{k=1}^K x_{ik} c_k \quad (48)$$

to the linear model of equation 2 with $\mathbf{c} = [\alpha, c_1, c_2, \dots, c_K]^T$ (corresponding to β) estimated by $\hat{\mathbf{c}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}$. Hence, the estimate $\hat{\beta}$ depends on the basis function B_k and its corresponding shape, where K is a subjective choice of the researcher. This subjective choice highly depends on the researchers intuition of the smoothness of the estimate. It is best practice to use the same basis functions as in [add link to smoothing](#) (cf. Kokoszka and Reimherr 2017)

include Part about CI's? its about inference and not about prediction

- Problem: truncation error δ and how to deal with it?
- Explain how to address truncation error in standard errors
- Motivate three estimation procedures
 1. truncated b-spline basis expansion without addressing truncation error
 2. truncated b-spline basis expansion WITH addressing truncation error
 3. truncated Eigenbasis expansion (advantages: low number of basis functions get low approximation error)

3.11 Draft-Overview

- Motivate Karhunen-Loeve-Expansion and Eigenbasis from PCA
- Explain Scalar-on-Function Regression
- Estimation through basis-expansion (incl. Eigenbasis) [and estimation with roughness penalty]
- Address approximation error due to basis-truncation

3.12 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 Draft-Overview

- Motivate Simulation for some data generating process from application
- Describe Simulation Setting from technical standpoint (DGP, set-up for replication, ...)
- Compare estimation with
 1. b-spline basis without addressing approximation error
 2. ... including proper treatment of approximation error
 3. Eigenbasis constructed from observations
- Prediction not Inference (Alternative: Focused on a testing procedure motivated by the application)
- Present Results
- Explain relevance for application

4.2 Motivation

For the simulation study, we deviate from the standard simulation setting. Instead of generating data by ourselves, we imposed the gasoline data which consists of 60 samples of Near Infrared (NIR) spectroscopy measured by 2-nm from 900 to 1,700 nm, and a response variable, octane rate. NIR-spectroscopy allows to analyze samples of gasoline much faster and with the same reproducibility as standard tests (cf. Gy. Bohács, Z. Ovádi, A. Salgó 1998). This setup was chosen to improve the approach towards the application, which involves predicting the octane ratings from the gasoline dataset.

4.3 Generating Similar Curves

To avoid small sample problems, we generated 200 similar curves from the gasoline dataset. This is motivated by Karhunen-Loève 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 performing Karhunen-Loève Expansion. It is assumed that the scores follow a normal distribution, thus, the new realizations for the scores are drawn from a multivariate normal $\tilde{\xi} = (\tilde{\xi}_1, \dots, \tilde{\xi}_J)' \sim \mathcal{N}(0_J, \text{diag}(\hat{\lambda}_1, \dots, \hat{\lambda}_J))$ 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} .

4.4 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 methods with differing true coefficient functions:

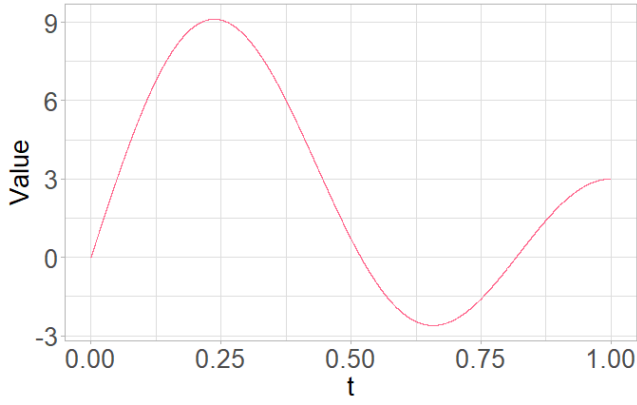


Figure 1: $f_1(t)$, smooth function

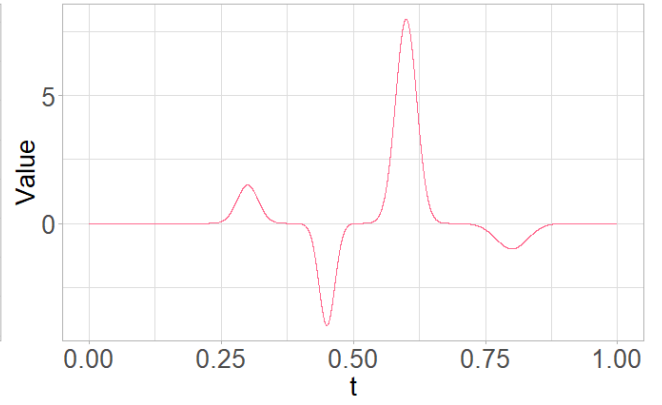


Figure 2: $f_2(t)$, bumpy function

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

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

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

$$\begin{aligned} Y_{1,f} &= \langle NIR, f \rangle + Z \left[\frac{\text{var}(\langle NIR, f \rangle)}{0.9} - \text{var}(\langle NIR, f \rangle) \right] \\ Y_{2,f} &= \langle NIR, f \rangle + Z \left[\frac{\text{var}(\langle NIR, f \rangle)}{0.6} - \text{var}(\langle NIR, f \rangle) \right] \end{aligned} \quad (51)$$

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 $\in \{1, 2, \dots, 5\}$, cubic bspline basis-function $\{5, 6, \dots, 25\}$ and fourier functions $\{1, 3, \dots, 25\}$ to predict the generated responses using the basis expansion approach and the FPCR approach. For the evaluation, we used 10 fold cross-validation. The simulation was done with R (version...). In total we conducted out 2500 repetitions for each set of simulations. [Information about version and link to repo / package in footnote??](#)

4.5 Results

DESCRIBE RESULTS

4.5.1 Interpretation and Relevance for Application

4.6 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 Draft-Overview

- Prediction not Inference (Alternative: Focused on a testing procedure motivated by the data set)
- IID data set (no dependence between the curves, don't want to do functional time series)
- Not necessarily data from economics (like biology, sports, whatever)
- Smooth curves or random walk (both fine)
- <https://functionaldata.wordpress.ncsu.edu/resources/>

5.2 Literature

- Carey et al. 2002

6 Outlook

6.1 Literature

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

7 Appendix

7.1 Basis Plots

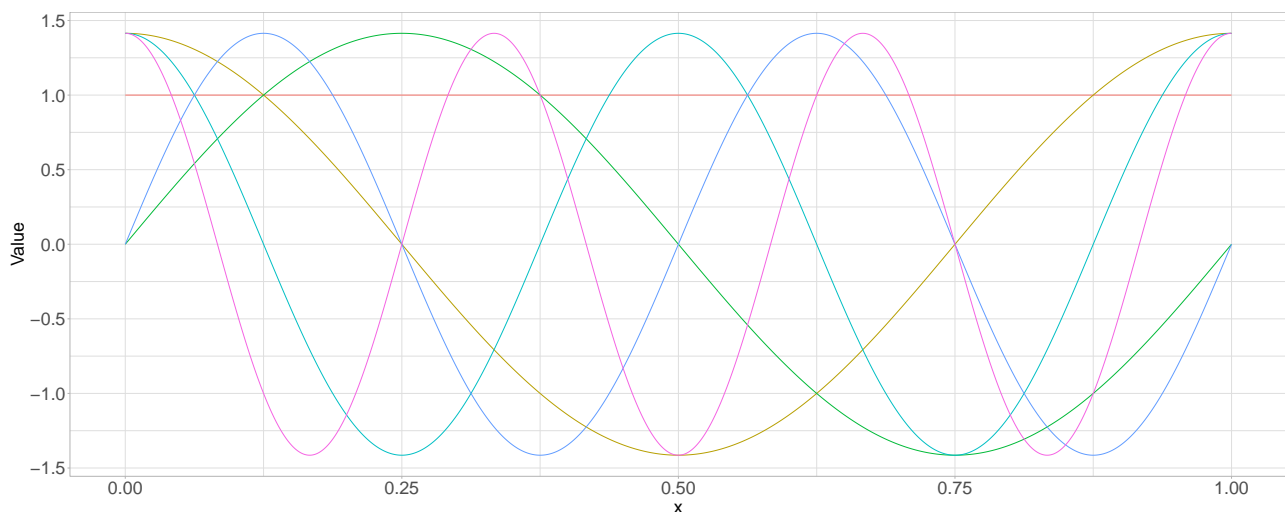


Figure 3: Fourier basis functions for $i = 1, \dots, 6$

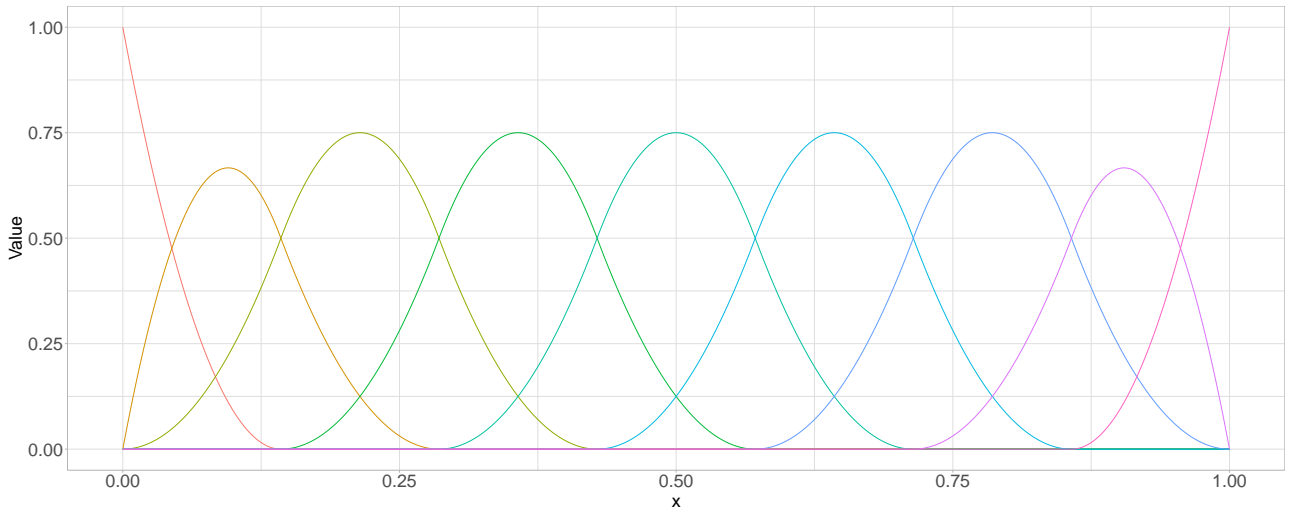


Figure 4: B-spline basis functions of order 3 for 6 equidistant knots on $[0, 1]$

7.2 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

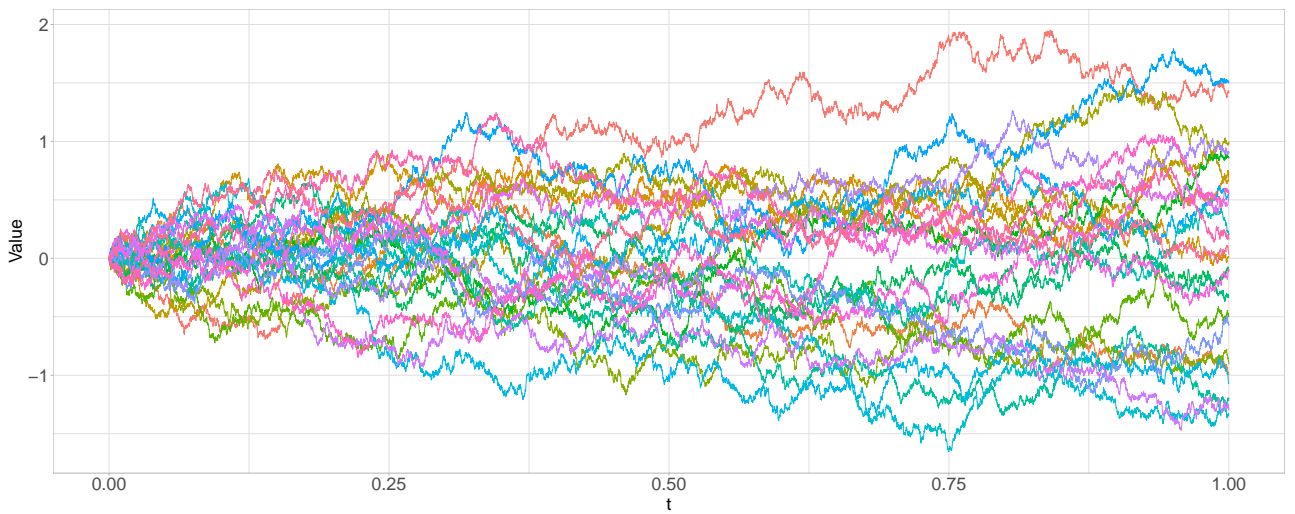


Figure 5: 25 i.i.d. realizations of a Wiener Process on $[0, 1]$

8 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).
- Das, Sonali et al. (2019). “The effect of global crises on stock market correlations: Evidence from scalar regressions via functional data analysis”. In: *Structural Change and Economic Dynamics* 50, pp. 132–147. ISSN: 0954-349X. DOI: 10.1016/j.strueco.2019.05.007. URL: <https://www.sciencedirect.com/science/article/pii/S0954349X19301407>.
- 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).