Lecture 3: Functional linear regression models
SoF, FoS and FoF regression models

Valeria Vitelli

Oslo Centre for Biostatistics and Epidemiology (OCBE)
Department of Biostatistics, University of Oslo, Norway
valeria.vitelli@medisin.uio.no

UiO : University of Oslo

PhD course
An Introduction to Functional Data Analysis: Theory and Practice

University of Palermo, Italy; March 25^{th} – 28^{th}, 2024

Outline

Main Reference: Chapters 12-17, in R&S[1] (very limited selection!), Section 3 in Gertheiss et al. (2023)

[1] Ramsay & Silverman, 2005: Functional Data Analysis, 2^{nd} edt, *Springer*

Introduction and Motivations

Definition: Functional Regression

A regression problem in which at least one functional variable is found on the left and/or right-hand side of the model equation. This means that functional variables may be the **response**, **covariate(s)**, or **both**.

Here we focus on **linear associations**, where *linearity* is defined differently, depending on the specific setting

Functional Regression Models

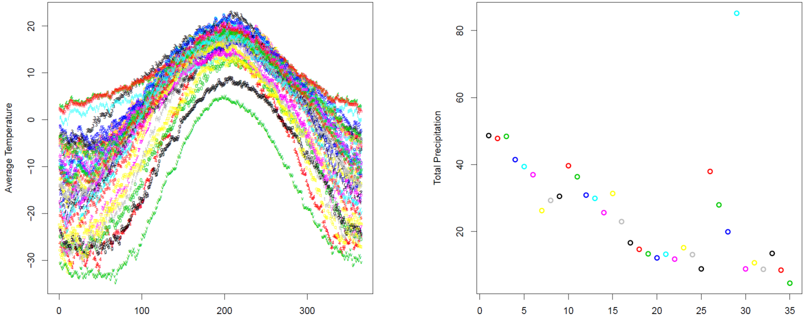
Main types of Functional Regression Models

To distinguish the different settings, we use the terms (respectively)

- "scalar-on-function(s)" regression (SoF)
- "function-on-scalar" regression (FoS)
- "function-on-function(s)" regression (FoF)

Scalar-on-function (SoF) regression

Motivating example: Canadian Weather data
Question of interest: predict precipitation from average daily temperature in Canada (replicates: the 35 Canadian stations)



In the figure: left, average daily temperature; right, total yearly precipitation. Colors link the corresponding stations in the two panels.

Does the model need to be functional? I

Alternative approach: Multiple linear regression

In the SoF regression case, a standard linear regression model for the observed discrete vectors would be possible

$$\text{precipitation}_i = \alpha + \sum_{j=1}^{365} \text{temperature}_{ij} \cdot \beta_j + \epsilon_i$$

where β_j is the effect of the temperature for day j on precipitation, and ϵ_i is the error term for the i-th station

Does the model need to be functional? II

Challenges

- $\mathbf{temperature}_i = (\text{temperature}_{i1}, \ldots, \text{temperature}_{i365})$ is a highly correlated vector for each i
- each value temperature_{ij} is usually very noisy
- estimating a smooth β_j over j has several advantages:
  - interpretation
  - borrow strength across j's

Bonus points

- Define better association models
- Infinite-dimensional spaces allow more parsimonious description
- Derivatives can be estimated

Notation

Data are pairs $\{y_i, x_i(\cdot) : t \in T\}_i$ for $i = 1, \dots, n$.

Common assumptions:

- $x_i(\cdot)$ is a functional covariate fully observed on the domain T
- perform pre-smoothing (deal with $x_i(\cdot)$ in functional form)
- $x_i(t) \in L^2(T)$
- $x_i(\cdot)$ i.i.d. zero mean curves with covariance function $\Sigma(\cdot, \cdot)$
- $\mathbb{E}[y_i] = 0$ for convenience

Objective

Develop association model to predict y_i from $x_i(\cdot)$

Functional Linear Model

We assume the following **functional linear model**, which is the *obvious generalization* of the standard linear regression model to functional spaces

$$y_i = \int_T x_i(t)\beta(t)dt + \epsilon_i \tag{1}$$

where

- $\beta(\cdot)$ is the **functional coefficient** that quantifies the effect of x_i on $\mathbb{E}[y_i]$
- $\beta(\cdot) : T \to \mathbb{R}$ is smooth (can be seen as a weighting function)
- $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$ i.i.d.

Penalized SoF regression: finite-dimensional representation

- Assume to have defined two basis systems in $L^2(T)$: $\{\varphi_l(\cdot), l \geq 1\}$ and $\{\theta_g(\cdot), g \geq 1\}$
- Expand $x_i(\cdot)$ using the $\varphi_l(\cdot)$'s, and $\beta(\cdot)$ using the $\theta_g(\cdot)$'s

$$x_i(t) = \sum_{l=1}^{L_x} \xi_{il}\varphi_l(t) \quad \beta(t) = \sum_{g=1}^{L} \beta_g\theta_g(t)$$

Then one can write the functional linear model (1) as

$$\int_T x_i(t)\beta(t)dt = \sum_{l=1}^{L_x}\sum_{g=1}^{L} \xi_{il}\left\{\int_T \varphi_l(t)\theta_g(t)dt\right\}\beta_g = \boldsymbol{\xi}_i'\boldsymbol{J}\boldsymbol{\beta}$$

where $\boldsymbol{\xi}_i = (\xi_{i1}, \ldots, \xi_{iL_x})'$, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_L)'$, and $\boldsymbol{J} \in \mathbb{R}^{L_x \times L}$ with $J_{lg} = \int_T \varphi_l(t)\theta_g(t)dt$

Penalized SoF regression: finite-dimensional representation

Then one can write the full model in matrix form as

$$y_i = \boldsymbol{\xi}_i' \boldsymbol{J} \boldsymbol{\beta} + \epsilon_i$$

Both $\boldsymbol{\xi}_i$ and $\boldsymbol{J}$ are known, therefore we only need to estimate $\boldsymbol{\beta}$.

First idea: use a simple sum-of-squared error criterion

$$min \sum_{i=1}^{n} (y_i - \boldsymbol{\xi}_i' \boldsymbol{J} \boldsymbol{\beta})^2$$

Problem: Estimation does not depend on the basis type but rather on the basis dimension:

- Larger basis $\rightarrow$ wigglier estimate $\rightarrow$ more variance
- Smaller basis $\rightarrow$ smoother estimate $\rightarrow$ more bias

Penalized SoF regression: Roughness Penalty

Instead of the simple sum-of-squared error criterion, use the penalized version

$$\sum_{i=1}^{n} \left\{ y_i - \int_T x_i(t)\beta(t)dt \right\}^2 + \lambda \int_T \{L\beta(t)\}^2 dt \qquad (2)$$

In matrix notation, this becomes

$$\sum_{i=1}^{n} (y_i - \boldsymbol{\xi}_i' \boldsymbol{J}\boldsymbol{\beta})^2 + \lambda \boldsymbol{\beta}' \boldsymbol{R} \boldsymbol{\beta} \qquad (3)$$

where $L\beta(t) = \sum_{g=1}^{L} \beta_g \{L\theta_g(t)\}$, and therefore $\boldsymbol{R} \in \mathbb{R}^{L \times L}$, with $R_{lg} = \int_T \{L\theta_l(t) L\theta_g(t)\} dt$

Penalized SoF regression: Roughness Penalty

The penalized criterion (3) is such that

- $\lambda \approx 0 \Rightarrow$ wiggly fit
- $\lambda >> 0 \Rightarrow$ smooth and biased fit
- λ controls the bias-variance trade-off, which means that it balances *smoothness* and *goodness* of fit

For fixed λ

- **Estimation:**

$$\hat{\boldsymbol{\beta}} = \left(\sum_{i=1}^{n} \boldsymbol{J}' \boldsymbol{\xi}_i \boldsymbol{\xi}_i' \boldsymbol{J} + \lambda \boldsymbol{R} \right)^{-1} \sum_{i=1}^{n} \boldsymbol{J}' \boldsymbol{\xi}_i y_i$$

- **Prediction:** $\hat{y}_i = \boldsymbol{\xi}_i' \boldsymbol{J} \hat{\boldsymbol{\beta}}$

How to select λ? Cross-Validation

Extensions / Generalizations

- Ideas can be immediately extended to multiple functional covariates
- When the response is **not continuous** (e.g. binary or count)

$$\mathbb{E}\{y_i|x_i(\cdot)\} = g^{-1}\{\alpha + \int_T x_i(t)\beta(t)dt\}$$

  Estimation via the same penalized criterion, with sum-of-square term replaced by the model likelihood for y_i
- When dealing with sparsely/irregularly sampled data:
  - joint (Bayesian) modeling of y_i and $x_i(\cdot)$ (see McLean et al. (2013) and further developments)

SoF regression in practice

- Observed data are not functions but noisy discretized versions of the $x_i(\cdot)$'s, so the *actual data* are the pairs

$$(y_i, \{(x_i(t_{ij}) + \epsilon_{ij}, t_{ij}) : j = 1, \ldots, m_i\})$$

- Model same as before
- First smooth the functional covariates using a smoothing technique (Lecture 1!)
  Denote by $\hat{x}_i(t)$ the estimated curve
- Fit the model as before by simply taking $\hat{x}_i(t)$ as if it was the true signal

Function-on-scalar (FoS) regression

In this case, *actual data* are the vectors

$$(\{(y_{ij}, t_{ij}) : j = 1, \ldots, m_i\}, x_{i1}, \ldots, x_{ip}) \quad \text{with } t_{ij} \in T$$

The *functional linear model* for $y_{ij} = y_i(t_{ij})$ is

$$y_i(t) = \beta_0(t) + \sum_{m=1}^{p} \beta_m(t) x_{im} + \epsilon_i(t)$$

- $\beta_0(\cdot) \rightarrow$ marginal mean of the response
- $\beta_m(\cdot) \rightarrow$ effect of the covariate x_m on the mean response at t
- $\epsilon_i(\cdot) \rightarrow$ residual process (zero mean, covariance function usually non-trivial)

Objective: prediction + inference of regression functions

Penalized FoS regression: finite-dimensional representation

Similar approach as seen for SoF regression:

- **model** the smooth effects $\beta_m(\cdot)$ for $m = 1, \ldots, p$ using basis expansions
  - consider the basis in $L^2(T)$ $\{\theta_l(\cdot), l \geq 1\}$
  - expand all $\beta_m(\cdot)$'s wrt the basis: $\beta_m(t) = \sum_{l=1}^{L} \beta_{ml}\theta_l(t)$
- **control** the smoothness of the $\beta_m(\cdot)$'s using a penalty, for ex

$$||D^2\beta_m(t)||_2 = \int_T \{D^2\beta_m(t)\}^2 dt = \beta_m' \boldsymbol{R} \beta_m$$

  (in the usual matrix notation)
- **estimate** the $\beta_m(\cdot)$'s using the usual *penalized criterion*

$$\sum_{i=1}^{n}\sum_{j=1}^{m_i}\{y_{ij} - \sum_{m=1}^{p}\sum_{l=1}^{L}\beta_{ml}\theta_l(t_{ij})x_{im}\}^2 + \sum_{m=1}^{p}\lambda_m\beta_m'\boldsymbol{R}\beta_m$$

Penalized FoS regression: Practicalities

- each λ_m controls the smoothness of the corresponding $\beta_m(\cdot)$ (often fixed to $\lambda_m = \lambda$)
- tuning of λ_m via CV or GCV
- closed form solution $(\boldsymbol{\beta_0}, \boldsymbol{\beta_1} \ldots, \boldsymbol{\beta_p})$ exists for fixed λ_m, where we have used the notation $\boldsymbol{\beta_m} = (\beta_{m1}, \ldots, \beta_{mL})'$
- **predict** $y_i(\cdot)$ as $\hat{y}_i(t) = \hat{\beta}_0(t) + \sum_{m=1}^{p} \hat{\beta}_m(t) x_{im}$
- **assess the goodness-of-fit** by using the *functional version* of the usual $R^2 = \int_T R^2(t)dt$, estimated as

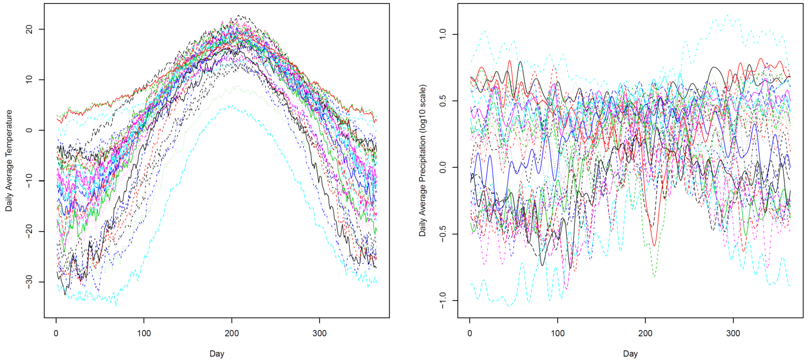
$$R^2(t) \approx 1 - \frac{\sum_{i=1}^{n} \sum_{j=1}^{m_i} (y_{ij} - \hat{y}_i(t_{ij}))^2}{\sum_{i=1}^{n} \sum_{j=1}^{m_i} (y_{ij} - \bar{y}(t_{ij}))^2}$$

where $\bar{y}(t_{ij})$ is the point-wise mean function

Function-on-function (FoF) regression

Motivating example: Canadian Weather data

Question of interest: How is the daily precipitation affected by the daily temperature in Canada? (replicates: the 35 Canadian stations)



In the figure: left, average daily temperature; right, average daily precipitation. Colors link the corresponding stations in the two panels.

Notation

In this case, *actual data* are the vectors of pairs

$$(\{(y_{ij}, t_{ij}) : j = 1, \dots, m_i\}, \{(x_{il}, t_{il}) : l = 1, \dots, l_i\}) \quad \text{with } t_{ij}, t_{il} \in T$$

Common assumptions:

- $x_i(\cdot)$ is a functional covariate fully observed on the domain T
- perform pre-smoothing (deal with $x_i(\cdot)$ in functional form)
- $x_i(t) \in L^2(T)$
- $x_i(\cdot)$ i.i.d. zero mean curves with covariance function $\Sigma(\cdot, \cdot)$
- $\mathbb{E}[y_{ij}] = 0$ for convenience
- From now on $y_{ij} = y_i(t_{ij})$ (slight abuse of notation)

Objective

Develop association model to predict $y_i(\cdot)$ from $x_i(\cdot)$

Functional Concurrent Model

Assumptions

- response and predictor are defined on the same domain
- the response at t is affected by the covariate at the same t

Functional Concurrent Model

$$y_i(t) = \beta(t)x_i(t) + \epsilon_i(t)$$

Modeling and Estimation: as before

- Basis expansion for $\beta(\cdot)$: $\beta(t) = \sum_{l=1}^{L} \beta_l \theta_l(t)$
- Estimate $\beta(\cdot)$ using a *penalized criterion*

$$\sum_{i=1}^{n} ||y_i(\cdot) - \sum_{l=1}^{L} \beta_l \theta_l(\cdot) x_i(\cdot)||^2 + \lambda \boldsymbol{\beta}' \boldsymbol{R} \boldsymbol{\beta}$$

Alternative 1: FoF Linear Model

$$y_i(t) = \int_{T_x} x_i(s)\beta(s,t)ds + \epsilon_i(t)$$

where $t \in T$ and T_x denotes the domain of the functional covariate (no need to assume same domains here)

The model above has also been relaxed to (Scheipl et al. 2015)

$$y_i(t) = \int_{T_x} F(x_i(s), s, t)ds + \epsilon_i(t)$$

Alternative 2: FoF Historical Model

$$y_i(t) = \int_{t-t_x}^{t} x_i(s)\beta(s,t)ds + \epsilon_i(t)$$

where $t \in T = [a,b]$ and $T_x := [a - t_x, b - t_x]$

References

Jan Gertheiss, David Rügamer, Bernard XW Liew, and Sonja Greven. Functional data analysis: An introduction and recent developments. *arXiv preprint arXiv:2312.05523*, 2023.

Mathew W McLean, Fabian Scheipl, Giles Hooker, Sonja Greven, and David Ruppert. Bayesian functional generalized additive models with sparsely observed covariates. *arXiv preprint arXiv:1305.3585*, 2013.

Fabian Scheipl, Ana-Maria Staicu, and Sonja Greven. Functional additive mixed models. *Journal of Computational and Graphical Statistics*, 24(2): 477–501, 2015.