M7777 Applied Functional Data Analysis
7. Functional Linear Regression

Jan Koláček (kolacek@math.muni.cz)

Dept. of Mathematics and Statistics, Faculty of Science, Masaryk University, Brno

Functional Linear Regression

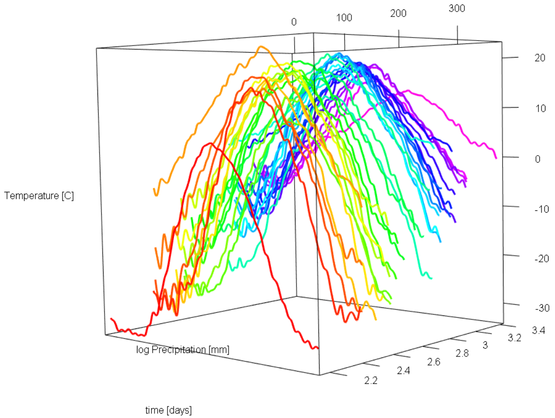
Three different scenarios

- **Scalar-on-function regression**: functional covariate, scalar response
- **Functional response models**
  - scalar covariate
  - functional covariate

We will deal with each in turn.

Scalar-on-function regression

Example: Log total Precipitation $\sim$ Temperature curve



We want to relate annual precipitation to the shape of the temperature profile.

A First Idea

- We observe $y_i, x_i(t)$
- Choose $t_1, \ldots, t_k$
- Then we set

$$
\begin{aligned}
y_i &= \alpha + \sum_{j=1}^{k} \beta_j x_i(t_j) + \varepsilon_i \\
&= \alpha + \mathbf{x}_i \boldsymbol{\beta} + \boldsymbol{\varepsilon}
\end{aligned}
$$

- And do linear regression.

But how many $t_1, \ldots, t_k$ and which ones? (it should be $k << n$!!!)

Scalar-on-function regression

In the Limit...

If we let $t_1, \ldots, t_k$ get increasingly dense (i.e. $k \to \infty$)

$$y_i = \alpha + \sum_{j=1}^{k} \beta_j x_i(t_j) + \varepsilon_i$$

becomes

$$y_i = \alpha + \int \beta(t) x_i(t) dt + \varepsilon_i \tag{1}$$

Minimize squared error:

$$\beta(t) = \arg\min \sum_{i=1}^{n} \left(y_i - \alpha - \int \beta(t) x_i(t) dt \right)^2$$

How to solve it? (3 approaches)

1. Estimation through a basis expansion

Expand the function β using basis functions

$$\beta(t) = \sum_{j=1}^{K} c_j \Phi_j(t).$$

Thus

$$\int \beta(t) x_i(t) dt = \sum_{j=1}^{K} c_j \underbrace{\int \Phi_j(t) x_i(t) dt}_{z_{ij}} = \mathbf{Zc}$$

and model (1) reduces to

$$\mathbf{y} = \alpha + \mathbf{Zc} + \varepsilon.$$

It is a classical linear regression model $\Rightarrow \hat{\mathbf{c}}$.

Scalar-on-function regression

The resulting estimate

$$\hat{\beta}(t) = \sum_{j=1}^{K} \hat{c}_j \Phi_j(t).$$

Disadvantages

- Assumption of $\beta(t)$ as a linear combination of basis functions $\boldsymbol{\Phi}(t)$
- Estimate $\hat{\beta}(t)$ depends on the shape of the basis functions and on their number K

Confidence intervals

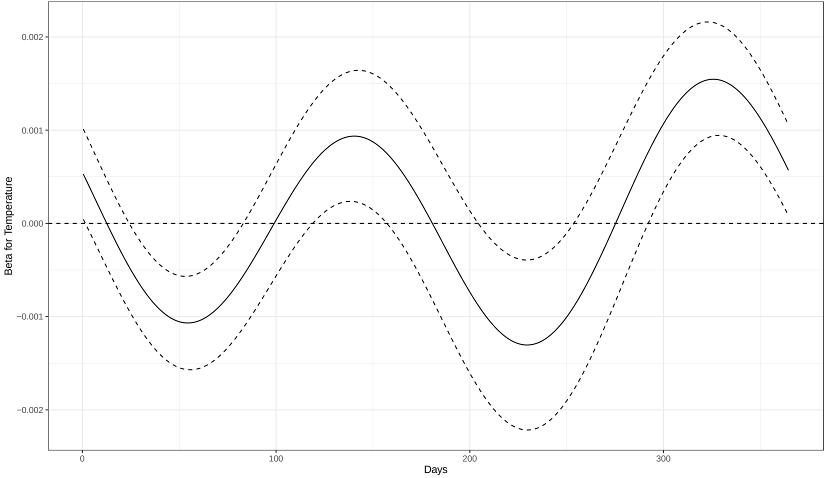
Assuming normality of errors, 95% confidence interval for $\beta(t)$:

$$\hat{\beta}(t) \pm 1.96 \sum_{j=1}^{K} \hat{\sigma}_j \Phi_j(t),$$

where $\hat{\sigma}_j$ is j-th diagonal entry of $\hat{\sigma}_\varepsilon \left(\mathbf{X}'\mathbf{X}\right)^{-1}$; $\mathbf{X} = [\mathbf{1}_n|\mathbf{Z}]$, $\hat{\sigma}_\varepsilon$ is the sample variance of $\mathbf{y} - \hat{\mathbf{y}}$.

Scalar-on-function regression

The estimate of $\beta(t)$



2. Estimation with a roughness penalty

Main idea

- The same expansion for $\beta(t)$, but K is taken to be some large value (often K is the number of t_i) $\Rightarrow$ no longer sensitivity to K
- The control of smoothness is shifted from K to the smoothing parameter λ and a differential operator L (a penalty term)

$$P_\lambda(\alpha, \beta) = \sum_{i=1}^{n} \left(y_i - \alpha - \int \beta(t) x_i(t) dt \right)^2 + \lambda \int [(L\beta)(t)]^2 dt.$$

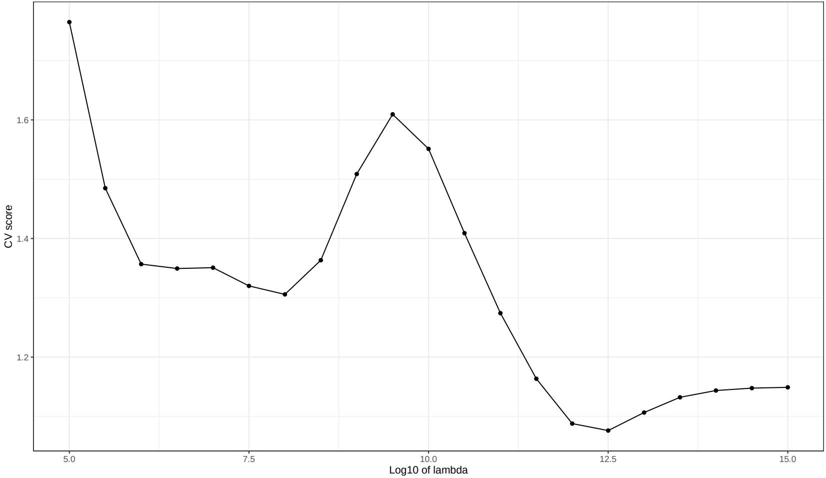
Thus

$$P_\lambda(\alpha, \beta) = \sum_{i=1}^{n} \left(y_i - \alpha - \sum_{j=1}^{K} c_j z_{ij} \right)^2 + \lambda \int \left[\sum_{j=1}^{K} c_j (L\Phi_j)(t) \right]^2 dt.$$

The optimal λ is selected by cross-validation.

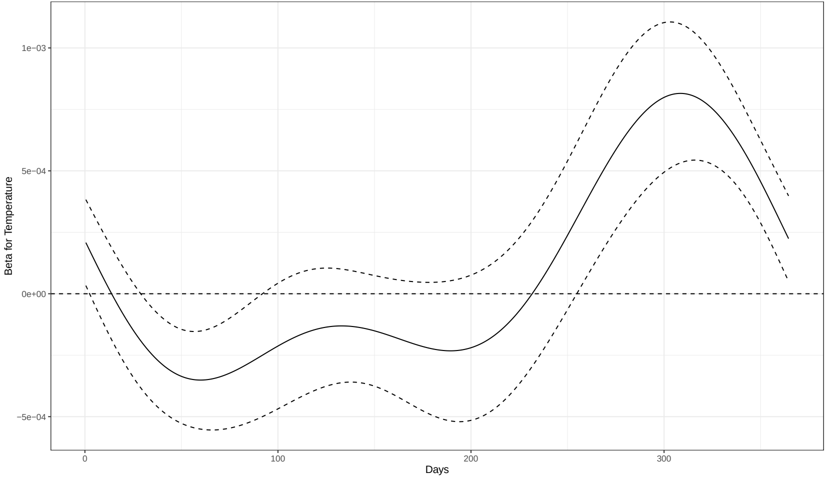
Scalar-on-function regression

Cross-validation scores



Scalar-on-function regression

The estimate of $\beta(t)$



Scalar-on-function regression

3. Regression on functional principal components

Let us consider an approximation $\hat{x}_i(t)$ of $x_i(t)$ by K principal components

$$\hat{x}_i(t) = \bar{x}(t) + \sum_{j=1}^{K} c_{ij}\xi_j(t),$$

where $\xi_j(t)$ is the j-th principal component, $c_{ij} = \int \xi_j(t)[x_i(t) - \bar{x}(t)]dt$ is its score. By plugging it in the model (1), it reduces to

$$\begin{aligned} y_i &= \alpha + \int \beta(t) \left(\bar{x}(t) + \sum_{j=1}^{K} c_{ij}\xi_j(t) \right) dt + \varepsilon_i \\ &= \beta_0 + \sum_{j=1}^{K} c_{ij}\beta_j + \varepsilon_i, \end{aligned}$$

where $\beta_0 = \alpha + \int \beta(t)\bar{x}(t)dt$, $\beta_j = \int \beta(t)\xi_j(t)dt$.

Scalar-on-function regression

It is a "classic" regression model

$$\mathbf{y} = \mathbf{\Xi}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

with $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_K)'$ and $\mathbf{\Xi} = [\mathbf{1}_n|\mathbf{C}]$, $\mathbf{C}$ is the score matrix. Denoting the estimates thus obtained by $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_K$ the estimates of the parameters in (1) are

$$\hat{\beta}(t) = \sum_{j=1}^{K} \hat{\beta}_j \xi_j(t), \qquad \hat{\alpha} = \hat{\beta}_0 - \sum_{j=1}^{K} \hat{\beta}_j \int \xi_j(t)\bar{x}(t)dt.$$

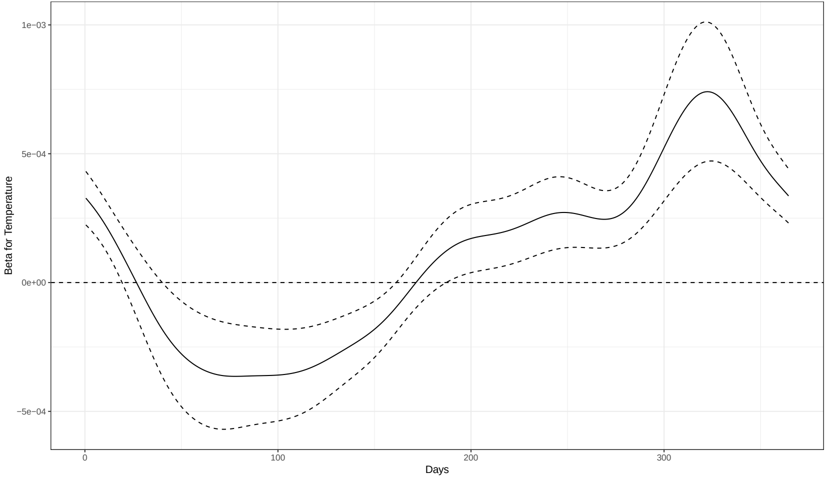
- first K components explain 85 or 90 percent of cumulative variance

Confidence intervals

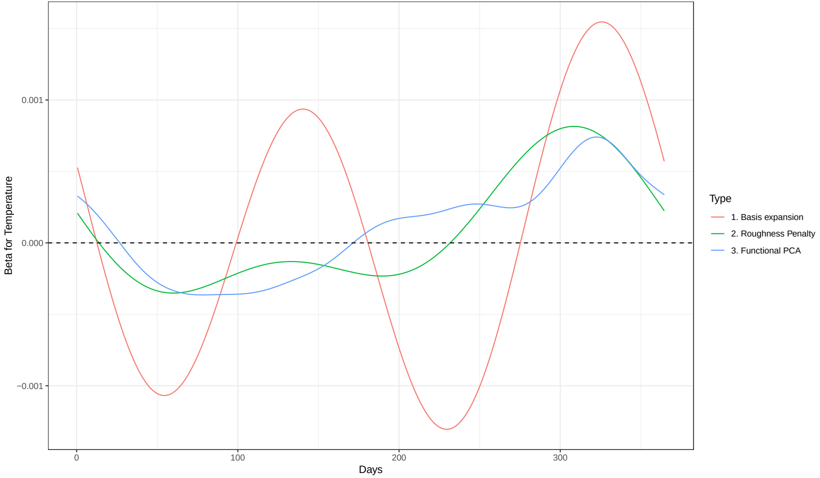
$$\mathrm{Var}\hat{\beta}(t) = \sum_{j=1}^{K} \mathrm{Var}(\hat{\beta}_j)\xi_j^2(t) \Rightarrow \textbf{CI:}\hat{\beta}(t) \pm 1.96 \left(\sum_{j=1}^{K} \mathrm{Var}(\hat{\beta}_j)\xi_j^2(t) \right)^{\frac{1}{2}}$$

Scalar-on-function regression

The estimate of $\beta(t)$



Comparison of estimates of $\beta(t)$



Scalar-on-function regression

Assessing the quality

Set

$$SSE_0 = \sum_{i=1}^{n} (y_i - \bar{y})^2, \qquad SSE_1 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

- Squared Multiple Correlation

$$RSQ = \frac{SSE_0 - SSE_1}{SSE_0}$$

- F-ratio

$$F = \frac{\frac{SSE_0 - SSE_1}{k-1}}{\frac{SSE_1}{n-k}},$$

  where $k \dots$ degrees of freedom (usually No. of parameters)

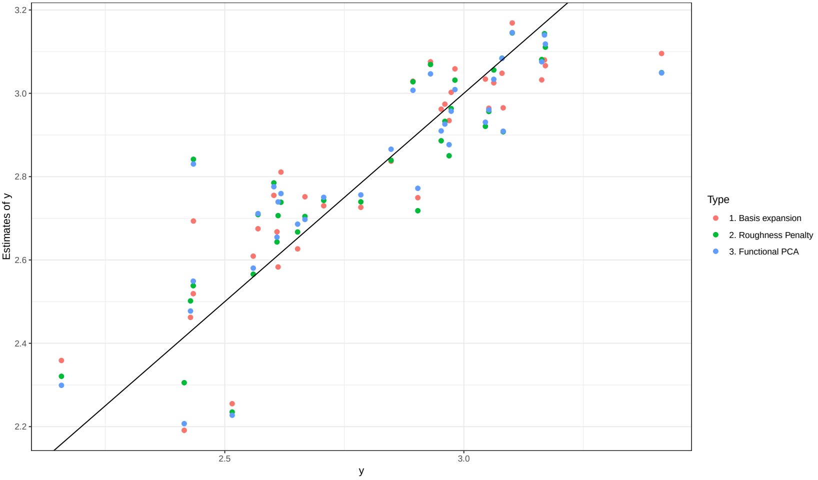
- Plotting $\hat{y}$ vs. y
- Cross-validation

Scalar-on-function regression

Assessing the quality

Model	degrees of freedom	*RSQ*	*F*-ratio
Basis expansion	6	0.796	22.58
Roughness penalty	4.6	0.754	25.42
Functional PCA	5	0.757	23.33

Comparison of fits $\hat{y}$



Cross-validation

- Divide y to 2 groups, **training** and **testing** data $y = [\tilde{y}, y^*]$
- Construct model based on $\tilde{y}$
- Use the model to predict $\hat{y}^*$
- Compare $\hat{y}^*$ against y^*

Scalar-on-function regression

4. Nonparametric regression

The model (1)

$$y_i = \alpha + \int \beta(t) x_i(t) dt + \varepsilon_i$$

with no parameters assumption becomes to a general model

$$y_i = m(x_i(t)) + \varepsilon_i,$$

where $m : L^2 \to \mathbb{R}$ is a functional that must be estimated.

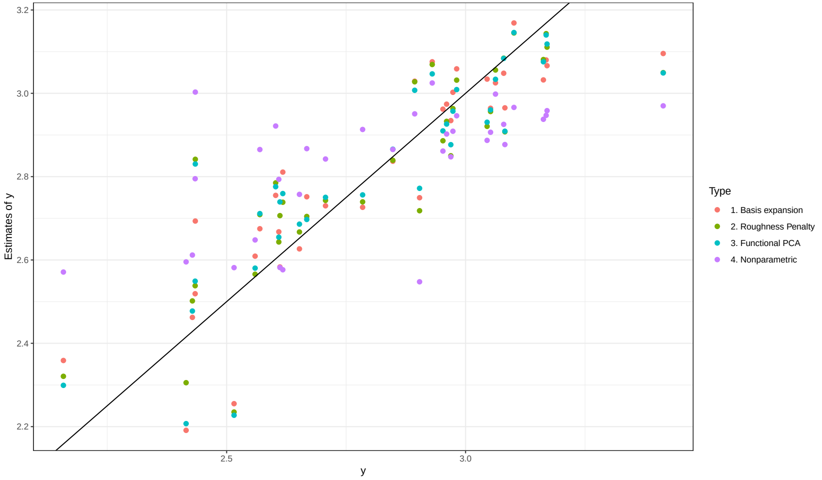
Kernel smoothing

$$\hat{m}(x) = \sum_{i=1}^{n} w_i(x) y_i, \qquad w_i(x) = \frac{K\left(h^{-1} d(x, x_i)\right)}{\sum_{j=1}^{n} K\left(h^{-1} d(x, x_j)\right)},$$

where h is a **smoothing parameter**, K is a **kernel function** and $d(f, g)$ is a measure of the **distance** between functions f and g.

Scalar-on-function regression

Comparison of fits $\hat{y}$



Problems to solve

1. Medfly Data
   - Load the variable `medfly` from the `medfly.RData` file.
   - Perform a functional linear regression to predict the total lifespan of the fly from their egg laying. Choose a smoothing parameter by cross validation, and plot the coefficient function along with confidence intervals (see Figure 1).
   - Plot the estimated values of lifespan against the measured values (see Figure 2). Calculate the R^2 for your regression.
   - Try a linear regression of lifespan on the principal component scores from your analysis (the previous lesson). What is the R^2 for this model? Does `lm` find that the model is significant? Reconstruct and plot the coefficient function for this model along with confidence intervals (see Figure 3).
   - Conduct the nonparametric regression. How does it compare to the model obtained through functional linear regression and to the model obtained through PCA? Plot estimated values of lifespan against the measured values for all three cases (see Figure 4).

Problems to solve

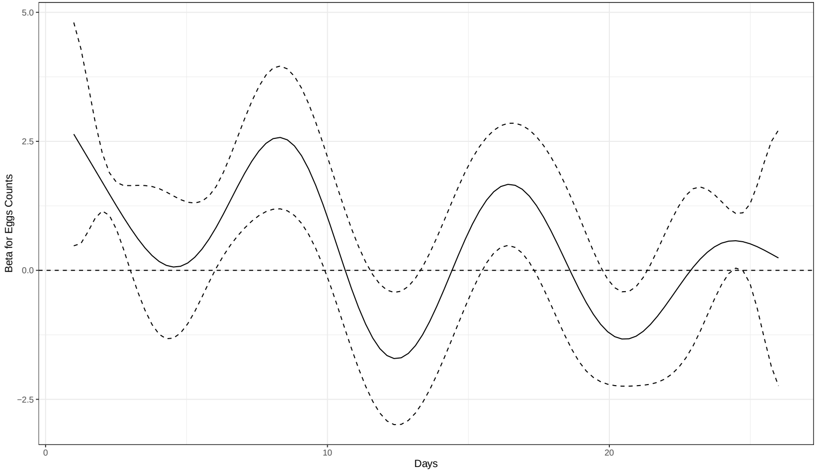


Figure 1.

Problems to solve

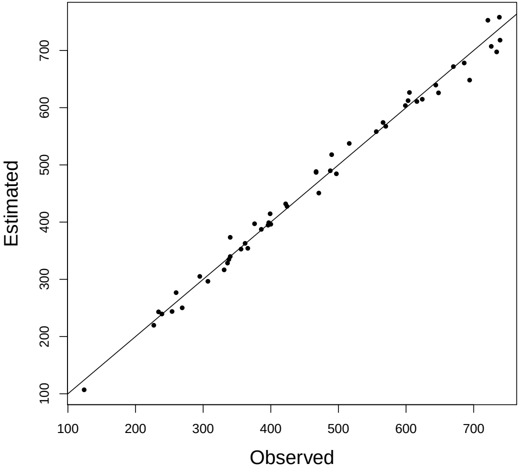


Figure 2.

Problems to solve

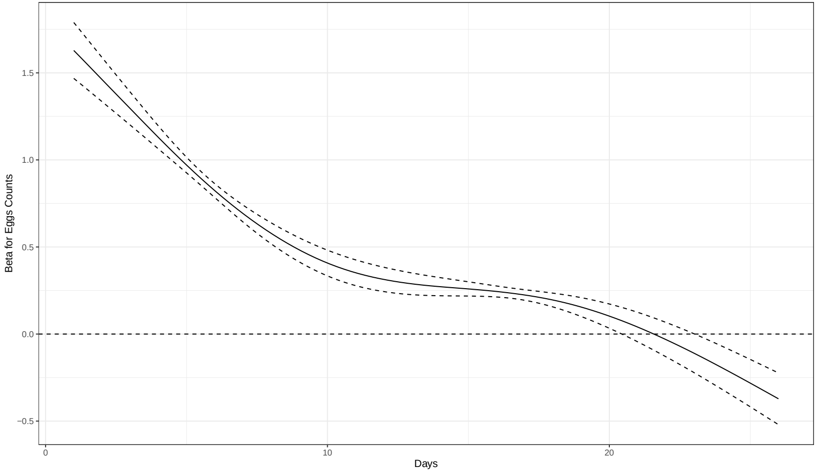


Figure 3.

Problems to solve

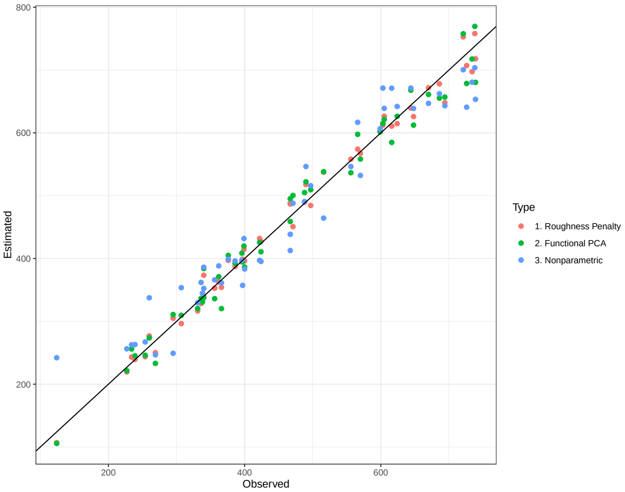


Figure 4.