Function-on-Scalar Regression with the refund Package

Philip T. Reiss, New York University

Available at: https://works.bepress.com/phil_reiss/28/
Function-on-scalar regression
with the refund package

Philip T. Reiss
New York University and Nathan Kline Institute
phil.reiss@nyumc.org http://works.bepress.com/phil_reiss

Joint Statistical Meetings
San Diego, July 30, 2012

Joint work with Lei Huang (NYU / Johns Hopkins),
Lan Huo (NYU), Fabian Scheipl (LMU Munich)

Research supported in part by grant DMS-0907017,
U.S. National Science Foundation
Outline

• The refund package
• Function-on-scalar regression: raw vs. preprocessed responses
• Within-function dependence
• Hypothesis testing
Regression functions in the \textit{refund} package

(Crainiceanu et al., 2012)

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Scalar</th>
<th>Functional</th>
</tr>
</thead>
<tbody>
<tr>
<td>Responses</td>
<td></td>
<td></td>
</tr>
<tr>
<td>\textit{Scalar}</td>
<td>fpcr, (l)pfr, (l)peer, wcr, wnet</td>
<td></td>
</tr>
<tr>
<td>\textit{Functional}</td>
<td>fosr, pffr</td>
<td>pffr</td>
</tr>
</tbody>
</table>
A function-on-scalar regression example:
Midsagittal corpus callosum width
- We’ll consider corpus callosum width functions derived from
  - 98 controls
  - 70 subjects with very mild dementia, and
  - 28 with mild dementia,

  based on MRI scans from the Open Access Series of Imaging Studies (OASIS) database (Marcus et al., 2007).

- Image processing at Nathan Kline Institute:
  - each individual’s corpus callosum was extracted by an automated procedure developed by Babak Ardekani;
  - width at each of 99 points, from anterior to posterior, along the midsagittal plane was computed by Alvin Bachman.
We thereby obtain 196 corpus callosum width curves (functional responses) . . .

. . . which we are primarily interested in regressing on diagnostic group.
Function-on-scalar regression with raw-data responses

• Given response functions \(y_1, \ldots, y_n\) each observed at \(s_1, \ldots, s_T \in S\), and scalar predictor matrix \(X = (x_{im})_{1 \leq i \leq n, 1 \leq m \leq p}\) (usually \(x_{i1} \equiv 1\)), we consider the repeated-measures varying-coefficient model

\[
y_i(s_t) = \sum_{m=1}^{p} x_{im} \beta_m(s_t) + \varepsilon_i(s_t),
\]

where the error functions \(\varepsilon_i\) arise from a mean-zero process on \(S\).

• Representing the coefficient functions as \(\beta_m(s) = \theta_m^T b(s)\) for some \(\theta_m \in \mathbb{R}^K (m = 1, \ldots, p)\), where \(b(s) = [b_1(s), \ldots, b_K(s)]^T\) denotes basis functions (often \(B\)-splines), we reduce (1) to the \(n \times T\) matrix equation

\[
Y = X\Theta B^T + E
\]

where \(Y = [y_i(s_t)]_{1 \leq i \leq n, 1 \leq t \leq T}\), \(\Theta(p \times K)\) is the coefficient matrix, \(B(T \times K)\) is the basis function evaluation matrix, and \(E = [\varepsilon_i(s_t)]_{1 \leq i \leq n, 1 \leq t \leq T}\).

• Our estimate of \(\Theta\) is the minimizer of a penalized double SSE, say

\[
\sum_{i=1}^{n} \sum_{t=1}^{T} [y_i(s_t) - (X\Theta B^T)_{it}]^2 + \sum_{m=1}^{p} \lambda_m \int \left[\theta_m^T b''(s)\right]^2 ds.
\]
Criterion

$$
\sum_{i=1}^{n} \sum_{t=1}^{T} \left[ y_i(s_t) - (X\Theta B^T)_{it} \right]^2 + \sum_{m=1}^{p} \lambda_m \int \left[ \theta_m''(s) \right]^2 ds
$$

can be rewritten in the generalized ridge regression form

$$
\| \text{vec}(Y^T) - (X \otimes B)\text{vec}(\Theta^T) \|_2^2 + \text{vec}(\Theta^T)^T (\Lambda \otimes P) \text{vec}(\Theta^T)
$$

$$
= \| y - X_B \theta \|_2^2 + \theta^T P_\lambda \theta,
$$

where

$$
y = \text{vec}(Y^T), \quad \theta = \text{vec}(\Theta^T), \quad X_B = X \otimes B,
$$

$$
P_\lambda = \Lambda \otimes P = \text{diag}(\lambda_1, \ldots, \lambda_p) \otimes \left[ \int b_k''(s)b_\ell''(s)ds \right]_{1 \leq k, \ell \leq K}.
$$

For given $\lambda_1, \ldots, \lambda_p$, (2) is minimized by

$$
\hat{\theta} = (X_B^T X_B + P_\Lambda)^{-1} X_B^T y.
$$
From raw to preprocessed responses

- Sometimes (e.g., large data sets) one works with **preprocessed** response functions given by an $n \times K$ matrix $C$ of basis coefficients such that $y_i(s) = c_i^T b(s)$ for $i = 1, \ldots, n$.

- $\rightarrow$ *fda* package (Ramsay and Silverman, 2005; Ramsay et al., 2009) chooses $\Theta$ to minimize the penalized integrated sum of squared errors

$$\int \sum_{i=1}^{n} [c_i^T b(s) - x_i^T \Theta b(s)]^2 ds + \sum_{m=1}^{p} \lambda_m \int [\theta_m^T b''(s)]^2 ds.$$ 

- Solution can be derived (Reiss et al., 2010) as limit of raw-response fit $\theta = (X_B^T X_B + P_\Lambda)^{-1} X_B^T y$ for uniform grid $s_1, \ldots, s_T$ as $T \rightarrow \infty$:

$$\frac{1}{T} X_B^T X_B = \frac{1}{T} (X \otimes B)^T (X \otimes B) = \frac{1}{T} (X^T X) \otimes (B^T B)$$

$$= (X^T X) \otimes \left[ \frac{1}{T} \sum_{t=1}^{T} b_k(s_t) b_\ell(s_t) \right]_{1 \leq k, \ell \leq K}$$

$$\rightarrow \infty (X^T X) \otimes \left[ \int b_k(s) b_\ell(s) ds \right]_{1 \leq k, \ell \leq K}.$$ 

- *fosr* works with either raw or preprocessed responses.
We simulated 3 groups $\times$ 50 functional responses from

$$y_i(s) = \mu(s) + \beta_{gp(i)}(s) + \underbrace{\eta_i(s)}_{GP(0, \sigma^2_{\eta}\rho|s-t|)} + \underbrace{\xi_i(s)}_{WN(\sigma^2_\xi=0.2^2)}$$

$\sigma_\eta = 0.2$

$\sigma_\eta = 0.5$

... and estimated $\mu$ and $\beta_1, \beta_2, \beta_3$ using raw vs. preprocessed responses.
1000 × mean integrated squared error: raw vs. preprocessed

\[ \sigma_{\eta} = 0.2 \]

\[ \sigma_{\eta} = 0.5 \]

\[ T = 30 \]

\[ T = 120 \]
Extension to function-on-concurrent-function regression?

- The raw-response design matrix $X_{(n \times p)} \otimes B_{(T \times K)}$ has $(i, m)$ block
  
  $$x_{im}B = [x_{im}b_k(s_t)]_{1 \leq t \leq T, 1 \leq k \leq K}$$

  ($1 \leq i \leq n, 1 \leq m \leq p$).

- For the “concurrent” function-on-function model

  \[ y_i(s) = \sum_{m=1}^{p} x_{im}(s) \beta_m(s) + \varepsilon_i(s) \]

  (i.e., time-varying predictors), the $(i, m)$ block becomes

  $$[x_{im}(s_t)b_k(s_t)]_{1 \leq t \leq T, 1 \leq k \leq K}.$$ 

- For preprocessed responses, the extension is more complex!
Callosum thickness example

Intercept

Very mild

Mild

Brain size

Female – Male

Age
Residual dependence

• When minimizing the penalized OLS criterion

\[ \sum_{i=1}^{n} \left\| \mathbf{y}_i - \sum_{m=1}^{p} x_{im} \mathbf{\beta}_m \right\|^2 + \sum_{m=1}^{p} \lambda_m \int \mathbf{\beta}_m''(s)^2 \, ds, \]

where \( \mathbf{y}_i = [y_i(s_1), \ldots, y_i(s_T)]^T \), \( \mathbf{\beta}_m = [\beta_m(s_1), \ldots, \beta_m(s_T)]^T \), selection of \( \lambda_1, \ldots, \lambda_p \) by REML or GCV is not straightforward since in general \( \mathbf{\Sigma} = \text{Cov}(\mathbf{y}_i|x_{i1}, \ldots, x_{ip}) \neq \mathbf{I} \).

• Three ways to take this within-function dependence into account in smoothness selection:
  1. leave-one-function-out cross-validation
  2. penalized generalized least squares*
  3. functional mixed-effects models*

* extensions of repeated-measures methodology to varying-coefficient models.
Method 1: Leave-one-function-out cross-validation

- This is the “classical” method (Rice and Silverman, 1991; Ramsay and Silverman, 2005): minimize

\[
\sum_{i=1}^{n} (y_i - \hat{y}_i(-i))^2.
\]  

(3)

- For given \( \lambda \), \texttt{forsr} uses the “leave-one-function-out” lemma

\[
y_i - \hat{y}_i(-i) = (I - H_{ii})^{-1}(y_i - \hat{y}_i)
\]

to compute (3) without refitting.

- But, there’s no standard algorithm for minimizing (3) wrt multiple \( \lambda \)’s.
Method 2: Penalized generalized least squares

- Reiss et al. (2010) consider replacing the penalized OLS criterion

\[
\sum_{i=1}^{n} \left\| y_i - \sum_{m=1}^{p} x_{im} \beta_m \right\|^2 + \sum_{m=1}^{p} \lambda_m \int \beta_m''(s)^2 ds
\]

with the penalized GLS criterion

\[
\sum_{i=1}^{n} \left[ y_i - \sum_{m=1}^{p} x_{im} \beta_m \right]^T \hat{\Sigma}^{-1} \left[ y_i - \sum_{m=1}^{p} x_{im} \beta_m \right] + \sum_{m=1}^{p} \lambda_m \int \beta_m''(s)^2 ds,
\]

where \( \hat{\Sigma} \) is an estimate of \( \text{Cov}(y_i|x_{i1}, \ldots, x_{ip}) \) (cf. Krafty et al., 2008).

- Intuition: Equivalent to penalized OLS for “prewhitened” responses \( \hat{\Sigma}^{-1/2} y_1, \ldots, \hat{\Sigma}^{-1/2} y_n \). Viewing these as approximately IID lets us use \text{mgcv} (Wood, 2006, 2011) for multiple smoothing parameter selection.

- Can iterate between updated estimates of \( \Sigma \) and of \( \beta_1, \ldots, \beta_p \).
Method 3: Functional mixed effects

• Idea:

\[ y_i(s) = \sum_{m=1}^{p} x_{im} \beta_m(s) + \eta_i(s) + \xi_i(s), \]

i.e., decompose residual function \( \varepsilon_i(s) \) into

random effect function \( \eta_i(s) \) + noise function \( \xi_i(s) \).

• Express \( \eta_i(s) \) as a linear combination of the leading functional principal components.

• This is implemented for \texttt{pffr}, but not yet in \texttt{f_osr} (for version 0.1-6).
Details for functional mixed effects

The penalized OLS criterion can be written as

\[
\sum_{i=1}^{n} \left\| y_i - (x_{i1} B \ldots x_{ip} B) \begin{pmatrix} \theta_1 \\ \vdots \\ \theta_p \end{pmatrix} \right\|^2 + \sum_{m=1}^{p} \lambda_m \theta_m^T P \theta_m.
\]

Consider instead the doubly penalized criterion

\[
\sum_{i=1}^{n} \left\| y_i - \begin{pmatrix} \beta \\ \theta_1 \\ \vdots \\ \theta_p \end{pmatrix} - \begin{pmatrix} \sqrt{\rho_1} \phi_1 \\ \vdots \\ \sqrt{\rho_C} \phi_C \end{pmatrix} \begin{pmatrix} u_{i1} \\ \vdots \\ u_{iC} \end{pmatrix} \right\|^2 + \sum_{m=1}^{p} \lambda_m \theta_m^T P \theta_m + \lambda_u \sum_{i=1}^{n} \| u_i \|^2,
\]

where \( \phi_j = [\phi_j(s_1), \ldots, \phi_j(s_T)]^T \) is the \( j \)th eigenfunction of the residual covariance operator, and \( \rho_j \) is the \( j \)th eigenvector (\( \rightarrow u_{ij} \)’s iid).
Hypothesis testing

Test $\mathcal{M}_0$ vs. $\mathcal{M}_A$ via pointwise $F$-statistics (Ramsay and Silverman, 2005)

$$F(s) = \frac{[SSE_{\mathcal{M}_0}(s) - SSE_{\mathcal{M}_A}(s)]/(p_A - p_0)}{SSE_{\mathcal{M}_A}/(n - p_A)}.$$  

Declare significance at $t$ such that

$$F(t) > 100(1 - \alpha) \text{ percentile, under } \mathcal{M}_0, \text{ of } \max_s F(s).$$

estimated by permutation

Diagnostic group effect
Great Books
Thank you!

And many thanks to

- Eva Petkova, Lassell Lu and Aaron Chen (NYU), Todd Ogden (Columbia), Ciprian Crainiceanu (Johns Hopkins) and Giles Hooker (Cornell), for valuable discussions;
- the NIH, for supporting refund via grant 5R01EB009744-02 (PI: Ogden) and for funding the OASIS project via grants P50 AG05681, P01 AG03991, R01 AG021910, P50 MH071616, U24 RR021382, R01 MH56584;
- Babak Ardekani, Al Bachman and Sang Han Lee (Nathan Kline Institute), for providing the corpus callosum data derived from the OASIS database.
References


