Functional Covariates: Concurrent Linear Model

What if instead of just $\mathbf{z}_i$, I wanted to use $x_i(t)$ to predict $y_i(t)$?

There are many plausible models – we will see the most general next lecture.

A simple, and often useful, restriction is the *concurrent* model

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

That is, $y_i(t)$ is only dependent on the current value of $x_i(t)$.

Mechanics

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

write

$$\mathbf{b} = [\mathbf{c}_1^T \ \cdots \ \mathbf{c}_p^T]^T$$

and

$$\Psi_i(t) = [x_{i1}(t)\Phi_1(t) \ \cdots \ x_{ip}(t)\Phi_p(t)]$$

then

$$\hat{\mathbf{b}} = \left[\sum \int \Psi_i(t)\Psi_i(t)^T dt\right]^{-1} \left[\sum \int \Psi_i(t)^T y_i(t) dt\right]$$

Penalized Smoothing

As was the case for scalar covariates, penalized sum of errors is

$$\text{PENSSE}_{\lambda}(\beta) = \sum \int (y_i(t) - x_i(t)\beta(t))^2 \, dt + \sum_j \lambda_j \int [L_j \beta_j(t)]^2 \, dt$$

which can be cross-validated.

Functional Response Models in General

Consider functional-input functional-output regression

$$x(t) \to y(t)$$

So far we have considered the *concurrent linear model*

$$y(t) = \beta(t)x(t) + \epsilon(t)$$

but clearly this is unsatisfactory:

- $y(t)$ may depend on $x(t)$ at times other than the current
- $y(t)$ and $x(t)$ may be measured at different ranges

At Most General

Treat $y(t)$ as a scalar at each time t. The functional linear model is

$$y_t = \int \beta(s)x(s)ds + \epsilon$$

So that over all times t this becomes

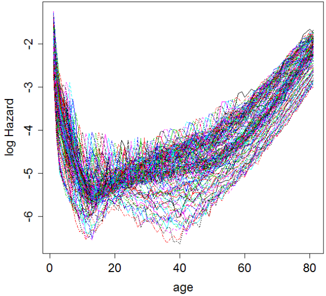
$$y(t) = \int \beta(s,t)x(s)ds + \epsilon(t)$$

As for the scalar response model, this is not identifiable without smoothing.

But we do know how to smooth bivariate functions!

Example: Swedish Lifetable Timeseries

Recall the Swedish mortality data: log hazard = instantaneous chance of death at each age.



Instead of treating time as a co-variate, we will consider a time-series model

$$y_i(t) = \int \beta(s,t) y_{i-1}(s) ds + \epsilon_i(t)$$

Estimating a Coefficient Function

We use an integrated squared error objective criterion:

$$\text{SISE} = \sum \left[\int \left(y_i(t) - \int \beta(s,t) x_i(s) ds \right)^2 dt \right]$$

with the usual bivariate roughness penalty.

Representing this by a bivariate basis expansion

$$\begin{aligned} \text{SISE} &= \sum \left[\int \left(y_i(t) - \psi(t) B \int \phi(s) x_i(s) ds \right)^2 dt \right] \\ &= \sum \left[\int \left(y_i(t) - \int \phi(s) x_i(s) ds \otimes \psi(t) \text{vec}(B) \right)^2 dt \right] \end{aligned}$$

Note vec(B) vectorizes B *column-wise*.

Estimating B

The minimizer of SISE is given by

$$\left[\sum \left[\int \phi(s)x_i(s)ds\right]\left[\int \phi(s)x_i(s)ds\right]^T \otimes \int \Psi(t)\Psi(t)^T dt\right]^{-1}$$
$$\left[\sum \int \phi(s)x_i(s)ds \otimes \int \Psi(t)y_i(t)dt\right]$$

Note separation into inner-products of basis defined w.r.t s and w.r.t. t.

Usual penalties result in additional penalty matrix inside the inverse.

With an Intercept

For simplicity, we have not considered an intercept.

In this context we have

$$y_i(t) = \beta_0(t) + \int \beta_1(s,t)x_i(s)ds + \epsilon_i(t)$$

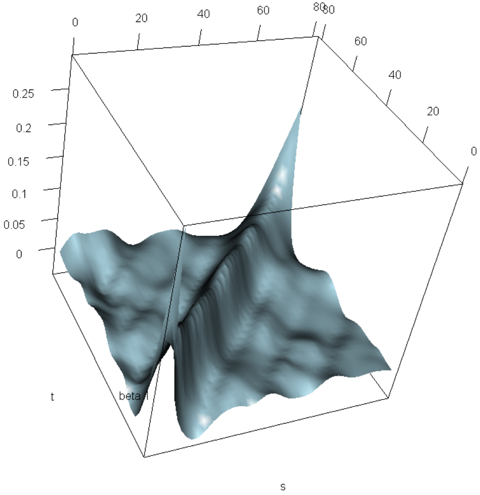
and we can estimate co-efficients for all terms by $(X+R)^{-1}Y$ for

$$X = \left[\begin{array}{cc} \int \Phi(t)\Phi(t)^T dt & \sum \left[\int \phi(s)x_i(s)ds\right]^T \otimes \int \Phi(t)\Phi(t)^T dt \\ \left[\int \phi(s)x_i(s)ds\right] \otimes \int \Phi(t)\Phi(t)^T dt & \left[\int \phi(s)x_i(s)ds\right]\left[\int \phi(s)x_i(s)ds\right]^T \otimes \int \Psi(t)\Psi(t)^T dt \end{array} \right]$$

and

$$Y = \left[\begin{array}{c} \int \Phi(t)y_i(t)dt \\ \sum \int \phi(s)x_i(s)ds \otimes \int \Psi(t)y_i(t)dt \end{array} \right]$$

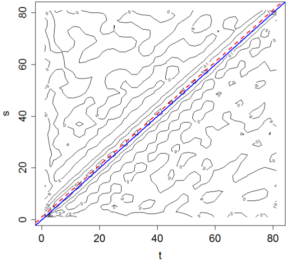
Obtaining an Estimate



Interpretation

Ridge in middle is not exactly diagonal (would imply *concurrent* model)

First off-diagonal $\Rightarrow$ events get passed one-year earlier.



Essentially, hazard-events happen to each cohort *at the same time*.

Confidence Intervals

As we had for the concurrent linear model, define B in terms of coefficients of $y(t)$.

$$\mathbf{y}(t) = \xi(t)^T C$$

This gives us

$$\hat{B} = X^{-1} \begin{bmatrix} \int \phi(t)\xi(t)^T dt & \cdots & \int \phi(t)\xi(t)^T dt \\ \int x_1(s)\Phi(s)ds \otimes \int \phi(t)\xi(t)^T dt & \cdots & \int x_n(s)\Phi(s)ds \otimes \int \phi(t)\xi(t)^T \end{bmatrix} \begin{bmatrix} \mathbf{c}_1 \\ \vdots \\ \mathbf{c}_n \end{bmatrix}$$

or

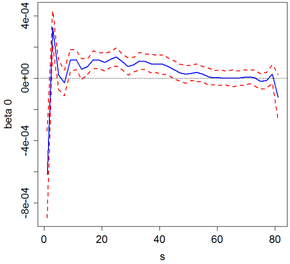
$$\hat{B} = \text{c2bmap} \circ \text{vec}(C)$$

Confidence Intervals

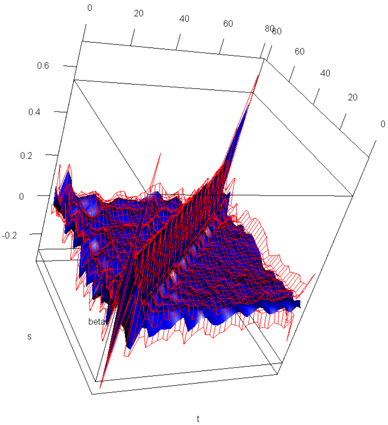
$$\mathrm{var}(\hat{B}) = \mathrm{c2bmap} \circ \begin{bmatrix} \Sigma & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \Sigma \end{bmatrix} \circ \mathrm{c2bmap}^T$$

Confidence Intervals $\beta_0(t)$

Based on the first entries in $\hat{B}$



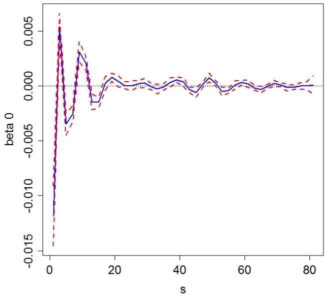
Confidence Intervals $\beta_1(s, t)$

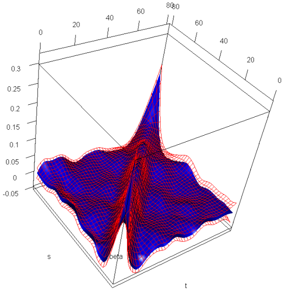


Effects of Smoothing Parameters

Making one λ larger can reduce confidence intervals for other components.

Set $\lambda_s = 10^3$, $\lambda_t = 10^3$:

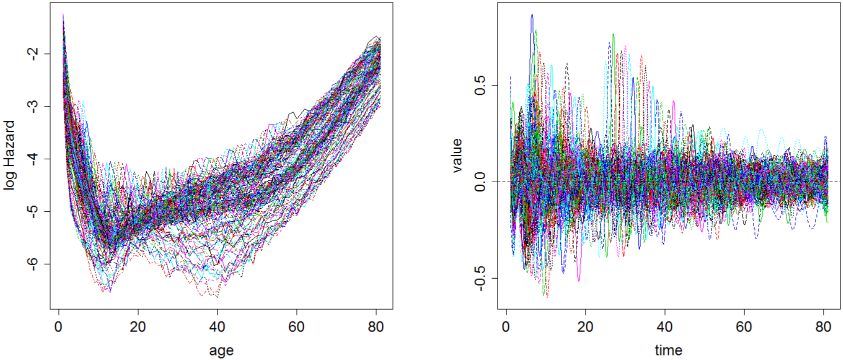




Cross-validation possible, but now three smoothing parameters.

Exploring Residuals

First, no clear trends along age



Yet More Models

fAR(p) Processes:

$$y_i(t) = \beta_0(t) + \sum_{j=1}^{k} \int \beta_j(s,t) y_{i-j}(t) dt + \epsilon_i(t)$$

fARMA(p,q) Processes:

$$y_i(t) = \beta_0(t) + \sum_{j=1}^{k} \int \beta_j(s,t) y_{i-j}(t) dt + \epsilon_i(t) + \sum_{k=1}^{q} \int \theta_k(s,t) \epsilon_{i-k}(t) dt$$

These are models on a *mixed continuous-discrete domain*.

Bivariate continuous analogue:

$$\frac{d}{dw} y(t,w) = \beta_0(t) + \int \beta(s,t) y(s,w) ds + \epsilon(s,w)$$

essentially a partial differential equation: dream up your own.

Some Useful Restrictions

Historical Linear Model: frequently $y_i(t)$ should only depend on $x_i(t)$ at times *before* t:

$$y_i(t) = \beta_0(t) + \int_0^t \beta_1(s,t)x_i(s)ds + \epsilon_i(t)$$

sets $\beta_1(s,t) = 0$ for $s > t$. Requires triangular bases.

Functional Convolution Model: Also restrict dependence to a short time window.

$$y_i(t) = \beta_0(t) + \int_{t-\delta}^t \beta_1(s,t)x_i(s)ds + \epsilon_i(t)$$

- Can be implemented with a kronecker product basis.
- Frequently, set $\beta_0(t) = 0$

Summary

- Most general functional response linear model $\Rightarrow$ bivariate coefficient function.
- Note some pathological cases: $y_i(t) = \beta(t)x_i(T) + \epsilon_i(t)$ for some fixed number T.
- Smooth cases efficiently computed with `inprod` and `eval.basis`
- Confidence intervals follow from concurrent linear model
- Direct extensions to time-series
- Restricted models can also be useful, but harder to code.