Smoothing sample extremes: the mixed model approach

Francesco Pauli
Dept of Statistical Sciences,
University of Padova
fpauli@stat.unipd.it

Fabrizio Laurini
Dept of Economics,
University of Parma
flaurini@stat.unipd.it

Outline of this talk

- Smoothing semipararametric tools & splines

Outline of this talk

- Smoothing semipararametric tools & splines
- Generalized linear mixed models (GLMM) for spline estimation

Outline of this talk

- Smoothing semipararametric tools & splines
- Generalized linear mixed models (GLMM) for spline estimation
- GLMM for smoothing sample extremes

Outline of this talk

- Smoothing semipararametric tools & splines
- Generalized linear mixed models (GLMM) for spline estimation
- GLMM for smoothing sample extremes
- Some bits of Bayesian inference and MCMC

Outline of this talk

- Smoothing semiparatametric tools & splines
- Generalized linear mixed models (GLMM) for spline estimation
- GLMM for smoothing sample extremes
- Some bits of Bayesian inference and MCMC
- Simulations and applications to pollutants

Parametric modeling

Popular models for response y_i, $i = 1, \ldots, n$ have form

$$y_i = f(x_i) + \text{error}_i$$

- f can be any parametric function
- many way to characterize error; examples:
  - Gaussian iid (simple regression)
  - Correlated errors
  - Exponential family (generalized linear models)

Take simple linear regression

$$f(x) = \beta_0 + \beta_1 x$$

Smoothing features:

- f is smooth; often unsuited to model real data
- f is a linear combination of basis functions

$$\begin{bmatrix} 1 & x \end{bmatrix}$$

Take simple linear regression

$$f(x) = \beta_0 + \beta_1 x$$

Smoothing features:

- f is smooth; often unsuited to model real data
- f is a linear combination of basis functions

$$\begin{bmatrix} 1 & x \end{bmatrix}$$

An extension is polynomial regression

$$f(x) = \beta_0 + \beta_1 x + \beta_2 x^2 + \ldots + \beta_p x^p$$

Smoothing features:

- f is smooth and suited to model non-linearity
- f is a linear combination of basis functions

$$\begin{bmatrix} 1 & x & x^2 & \cdots & x^p \end{bmatrix}$$

Broken linear stick modeling

Linear model for a "structural change" at time κ_τ (broken stick)

$f(x) = \beta_0 + \beta_1 x + b_\tau (x - \kappa_\tau)_+$ $\quad ()_+$ the positive part of $(x - \kappa_\tau)$

Smoothing features:

- f is rough and suited to explain structural changes
- f is a linear combination of basis functions

$$\begin{bmatrix} 1 & x & (x-\kappa_\tau)_+ \end{bmatrix}$$

$(x - \kappa_\tau)_+$ is a linear spline basis function

Linear spline modeling

Extension to linear spline regression by adding knots $\kappa_1, \ldots, \kappa_K$

$$f(x) = \beta_0 + \beta_1 x + \sum_{k=1}^{K} b_k (x - \kappa_k)_+$$

Smoothing features:

- f is piecewise linear, and more flexible than broken stick
- f is a linear combination of basis functions

$$\begin{bmatrix} 1 & x & (x-\kappa_1)_+ & \ldots & (x-\kappa_K)_+ \end{bmatrix}$$

Extension to linear spline regression by adding knots $\kappa_1, \dots, \kappa_K$

$$f(x) = \beta_0 + \beta_1 x + \sum_{k=1}^{K} b_k (x - \kappa_k)_+$$

Smoothing features:

- f is piecewise linear, and more flexible than broken stick
- f is a linear combination of basis functions

$$\begin{bmatrix} 1 & x & (x-\kappa_1)_+ & \dots & (x-\kappa_K)_+ \end{bmatrix}$$

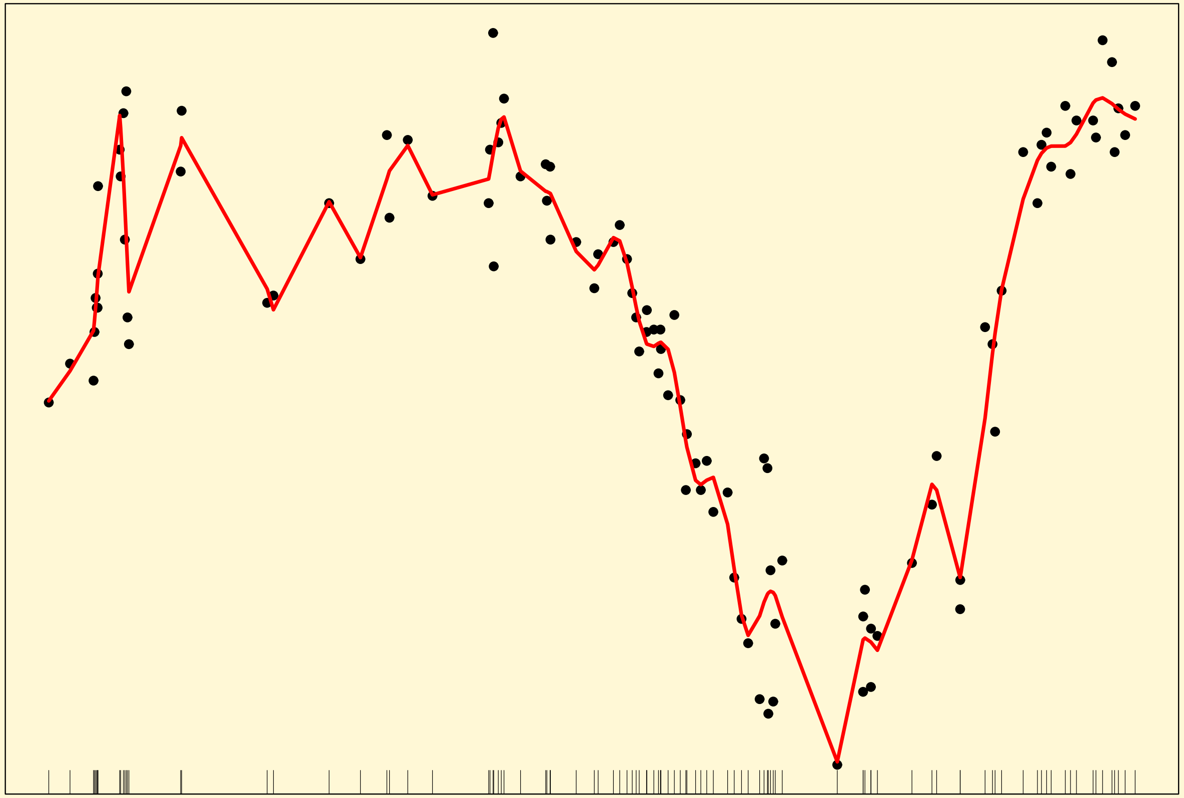
"Definition"

- The set of functions $\{(x - \kappa_j)_+\}$, $j = 1, \dots, K$ is a *linear spline basis*
- A linear combination of such basis functions is a piecewise linear function
- Commonly called spline with knots at $\kappa_1, \dots, \kappa_K$

How do we choose knots?

Knots selection and their placement have drawbacks

- Somehow *ad hoc* solution
- Overfitting
- Might be time consuming



Penalized spline regression

Let $\beta = (\beta_0, \beta_1, b_1, \ldots, b_K)$.
Wiggly fit is avoided by constraints on b_k such as

$$\sum_{k=1}^{K} b_k^2 < C \qquad \text{finite } C$$

Minimization is written as

$$\text{minimize } \|y - X\beta\|^2 \quad \text{subject to} \quad \beta^T D\beta \leq C$$

where D is a squared positive matrix with $K+2$ rows

$$D = \begin{bmatrix} \mathbf{0}_{2\times 2} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{K\times K} \end{bmatrix}$$

Smoothing features:

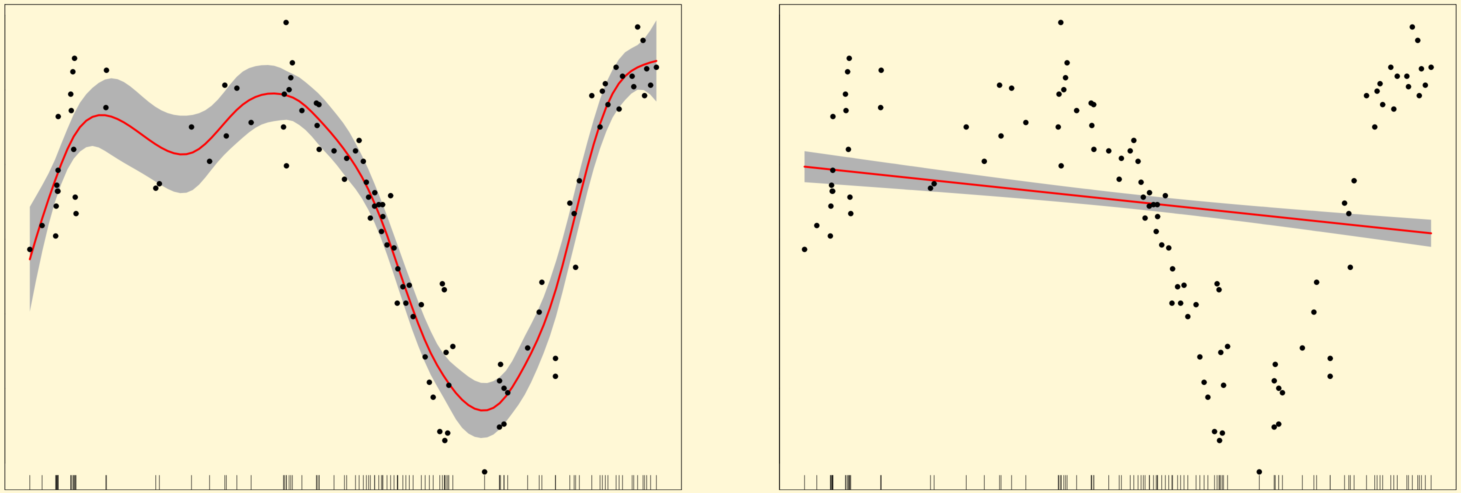
- f is smooth and pleasing
- The amount of smoothness is controlled by C, and does not depend on number/placement of knots

Solution of constrained optimization

With Lagrange multiplier argument, for some $\lambda \geq 0$, choose β to minimize

$$\|y - X\beta\|^2 + \lambda^2 \beta^T D \beta \quad \text{with solution } \hat{\beta}_\lambda = (X^T X + \lambda^2 D)^{-1} X^T y.$$

- $\lambda^2 \beta^T D \beta$ is the roughness penalty term
- λ is the smoothing parameter
- Connections with ridge regression



λ small $\qquad\qquad$ λ big

Generalization of spline models

Generalization may involve

Use of different basis

- Use of B-spline (numerical stability)
- Use of natural cubic splines (arise as solution of optimization problem)

Use of different penalties

- Penalize "some difference" in spline coefficients
- Penalizing degree of spline function. Example: q-th derivative of f

Penalized splines as Linear Mixed Model

Take the Linear Mixed Model (LMM)

$$y = X\beta + Zu + \varepsilon$$

and assume

$$E\begin{bmatrix} u \\ \varepsilon \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \quad COV\begin{bmatrix} u \\ \varepsilon \end{bmatrix} = \begin{bmatrix} \mathbf{G} & \mathbf{0} \\ \mathbf{0} & \mathbf{R} \end{bmatrix} \quad \text{and} \quad \begin{matrix} \mathbf{G} = \sigma_u^2 \mathbf{I} \\ \mathbf{R} = \sigma_\varepsilon^2 \mathbf{I} \end{matrix}$$

- $X\beta$ is the fixed component
- u is the random component or random effect

Spline embedded in LMM

For LMM

$$
\begin{aligned}
y &= X\beta + Zu + \varepsilon \\
y_i &= \beta_0 + \beta_1 x_i + \sum_{k=1}^{K} u_k (x_i - \kappa_k)_+ + \varepsilon_i
\end{aligned}
$$

splines are the Best Linear Unbiased Predictor where u is a vector of random coefficients.
Details:

$$
X = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \quad \text{and} \quad Z = \begin{bmatrix} (x_1 - \kappa_1)_+ & \cdots & (x_1 - \kappa_K)_+ \\ \vdots & \ddots & \vdots \\ (x_n - \kappa_1)_+ & \cdots & (x_n - \kappa_K)_+ \end{bmatrix}
$$

- In general u_k i.i.d. $N(0, \sigma_u^2)$
- $\sigma_u = \infty$ leads to wiggly (over)fit
- Finite σ_u shrinks u_k (smooth fit)

Sketch of extensions to LMM (increasing complexity)

- Semiparametric models: one covariate is nonparametric

$$y_i = f(X_i^1) + X_i^{[-1]}\beta + \varepsilon_i, \qquad f \text{ smooth}$$

- Semiparamtric mixed models: add random effects u
- Additive models: covariates as penalized linear splines, e.g.

$$y_i = c + f_1(x_i) + f_2(t_i) + \varepsilon_i; \qquad f_1 \text{ and } f_2 \text{ are smooth}$$

- Generalized parametric regression with random effects (GLMM), e.g.

$$y_i \mid u \sim \text{ Ber}\left(\frac{\exp\{(X\beta + Zu)_i\}}{1 + \exp\{(X\beta + Zu)_i\}}\right) \quad u \sim MVN(\mathbf{0}, \mathbf{G}_\theta)$$

- Combine all above ingredients, eventually with a Bayesian approach.

Breath for a thought!

Obvious and less immediate features

- Model fit with (restricted) ML methods (S-PLUS, R, SAS, ?).
  1. Selection of smoothing parameter λ (cross validation)
  2. Degrees of freedom & model selection (AIC)
- Standard inference tools available
  1. Pointwise & simultaneous confidence bands
  2. Hypothesis testing
  3. Likelihood ratio and F-tests
- Inference on functional of splines (e.g. confidence bands for any derivative of f)
- Extensions to LMM are "quite" easy.

- Hall and Tajvidi (2000)

  motivation, scope : exploratory analysis, assessment of trend

  method and model : local likelihood smoothing on GEV and GPD models

  smoothing parameter : CV for bandwith choice

  error assessment : goodness of fit evaluated using probability plots

  application(s) :
  - intensities of windstorms ($N = 45$) $\rightarrow$ GPD
  - Australian temperature extrema $\rightarrow$ GEV

- Davison and Ramesh (2000) - Ramesh and Davison (2002)

  motivation, scope : exploratory analysis, assessment of trend

  method and model : local likelihood smoothing on GEV model

  smoothing parameter :
  - bandwith chosen “by eye”
  - likelihood cross validation

  error assessment : bootstrap

  application(s) :
  - central England temperatures (r-largest)
  - athletic records (r-largest)
  - extreme sea level
  - river flow

Smoothing and extremes

- Pauli and Coles (2001)

  motivation, scope : exploratory analysis, more flexible than previous approaches, allow for multiple series

  method and model : penalized likelihood spline smoothing on GEV model

  smoothing parameter : bandwith chosen “by eye”

  error assessment : bayesian credibility intervals

  application(s) :

  - temperature at Oxford and Worthing (r-largest)
  - athletic records (r-largest)

- Chavez-Demoulin and Davison (2005)

  motivation, scope : this approach can be applied to dataset with numerous series

  method and model : generalized additive models estimated by penalized likelihood approach

  smoothing parameter : AIC

  error assessment : differences of deviances and bootstrap

  application(s) : daily minimum temperature at 21-weather stations in Switzerland

Our choices

We use a mixed model approach to estimate the spline function

- Bayesian approach
- Smoothing coefficient is a parameter to be estimated
- Error bands arise naturally as part of the procedure
- Use of GEV, GPD and Poisson processes (instead of exponential family of GLM)
- The model extends to multiple series

Standard assumptions

- Extreme observations (t, Y_t) (where $t \in [0, 1]$)
- follow a non-stationary Poisson process with intensity

$$\psi_t^{-1}\left(1 + \xi_t \frac{y - \mu_t}{\psi_t}\right)_+^{-1/\xi_t - 1}, \qquad \xi_t \neq 0$$

$$\psi_t^{-1} \exp\left(\frac{y - \mu_t}{\psi_t}\right), \qquad \xi_t = 0$$

- μ_t and ψ_t are for location-scale
- ξ_t shape parameter
- Parameter are assumed time-varying

Some conditions

- Define $N_u = \sum_{t\in[T_1,T_2]} \mathbb{I}\{Y_t > u\}$
-
$$N_u \sim \text{ Poi}\Big(\int_{T_1}^{T_2} \lambda(t)dt\Big)$$
-
$$\lambda(t) = \{1 + \xi(t)\frac{u-\mu(t)}{\psi(t)}\}_+^{-1/\xi(t)}$$
- Fix $N_u = n$; then data $\{Y_t - u \mid Y_t > u\} \sim GPD(\xi_t, \sigma_t)$, $\sigma_t = \psi_t + \xi_t(u - \mu_t)$
- log-likelihood decomposes as follows
$$l(\lambda, \sigma, \xi) = l_N(\lambda) + l_{Y-u}(\sigma, \xi)$$

Model description: first glance

Poisson process parameters are assumed as

$$\log(\lambda_t) = \beta_1 + \beta_2 t + f_\lambda(t)$$

$$\log(\nu_t) = \beta_5 + \beta_6 t + f_\nu(t)$$

$$\xi_t = \beta_3 + \beta_4 t + f_\xi(t)$$

- Separate inference for λ and (σ, ξ) due to likelihood factorization

$$l(\lambda, \sigma, \xi) = l_N(\lambda) + l_{Y-u}(\sigma, \xi)$$

- We consider orthogonal reparametrization

$$(\sigma, \xi) \to (\nu = \sigma(1 + \xi), \xi)$$

- Estimation of Poisson process intensity $\lambda(t)$
  - Divide interval $[0, 1]$ into $n_\delta = 1/\delta$ intervals of length δ
  - d_i = count of observations in $[(i-1)\delta, i\delta]$
  - $\lambda(t)$ is assumed constant in $[(i-1)\delta, i\delta]$
  - i-th interval contribution is given by

$$L_i(\lambda) = \lambda(i\delta)^{d_i} e^{-\delta\lambda(i\delta)}$$

Bayesian model specification

$$d_i \sim \text{Poisson}(\lambda_i)$$
$$\log(\lambda_i) = \beta_1 + \beta_2 t + Z_d b_\lambda$$
$$b_\lambda \sim \mathcal{N}(0, \mathbf{I}\tau_\lambda^{-1})$$

$$Y_j \sim \text{GPD}(\nu_j, \xi_j)$$
$$\xi_j = \beta_3 + \beta_4 t_j + Z_y b_\xi$$
$$\log(\nu_j) = \beta_5 + \beta_6 t_j + Z_y b_\nu$$
$$b_\xi \sim \mathcal{N}(0, \mathbf{I}\tau_\xi^{-1})$$
$$b_\nu \sim \mathcal{N}(0, \mathbf{I}\tau_\nu^{-1})$$

$$\beta_i \sim \mathcal{N}(0, 10^6)$$
$$\tau_\lambda, \tau_\nu, \tau_\xi \sim Gamma(10^{-3}, 10^{-3})$$

WinBUGS code (part of) for MCMC

Estimation is carried with WinBUGS to produce MCMC output

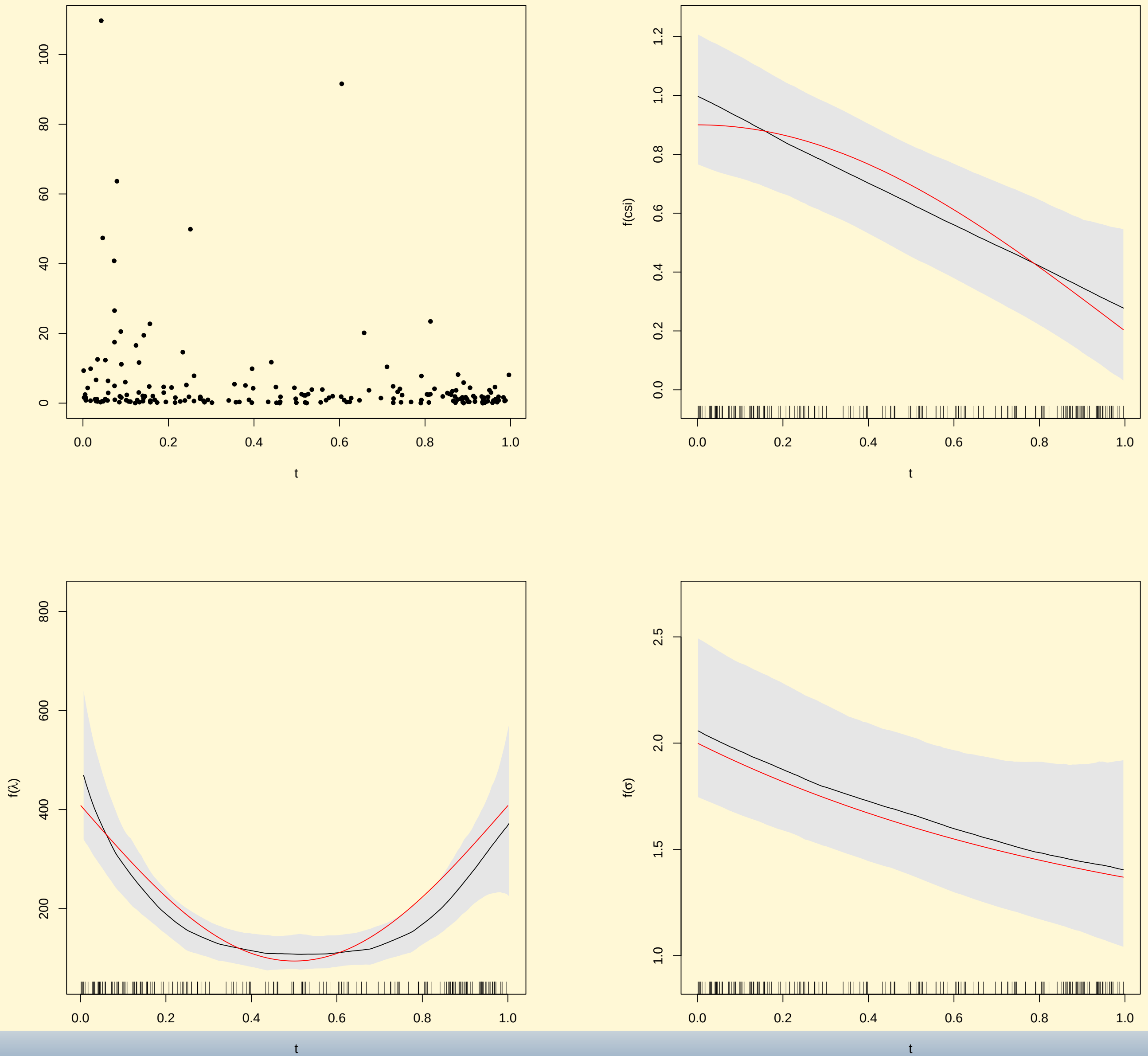
One issue in using WinBUGS is that GPD is not among the built-in distributions, so

```
for (i in 1:noss) {
    ones[i] <- 1 # fictitious observations
    ones[i] ~ dbern(p[i])
    Lik[i] <- (likelihood of i-th observation)
    p[i] <- Lik[i]/C
    csi[i] <- beta[3]+beta[4]*(t[i])+
                        inprod(bcsi[],Z[i,])
    vu[i] <- exp(beta[5]+beta[6]*(t[i])+
                        inprod(bvu[],Z[i,])
}
```

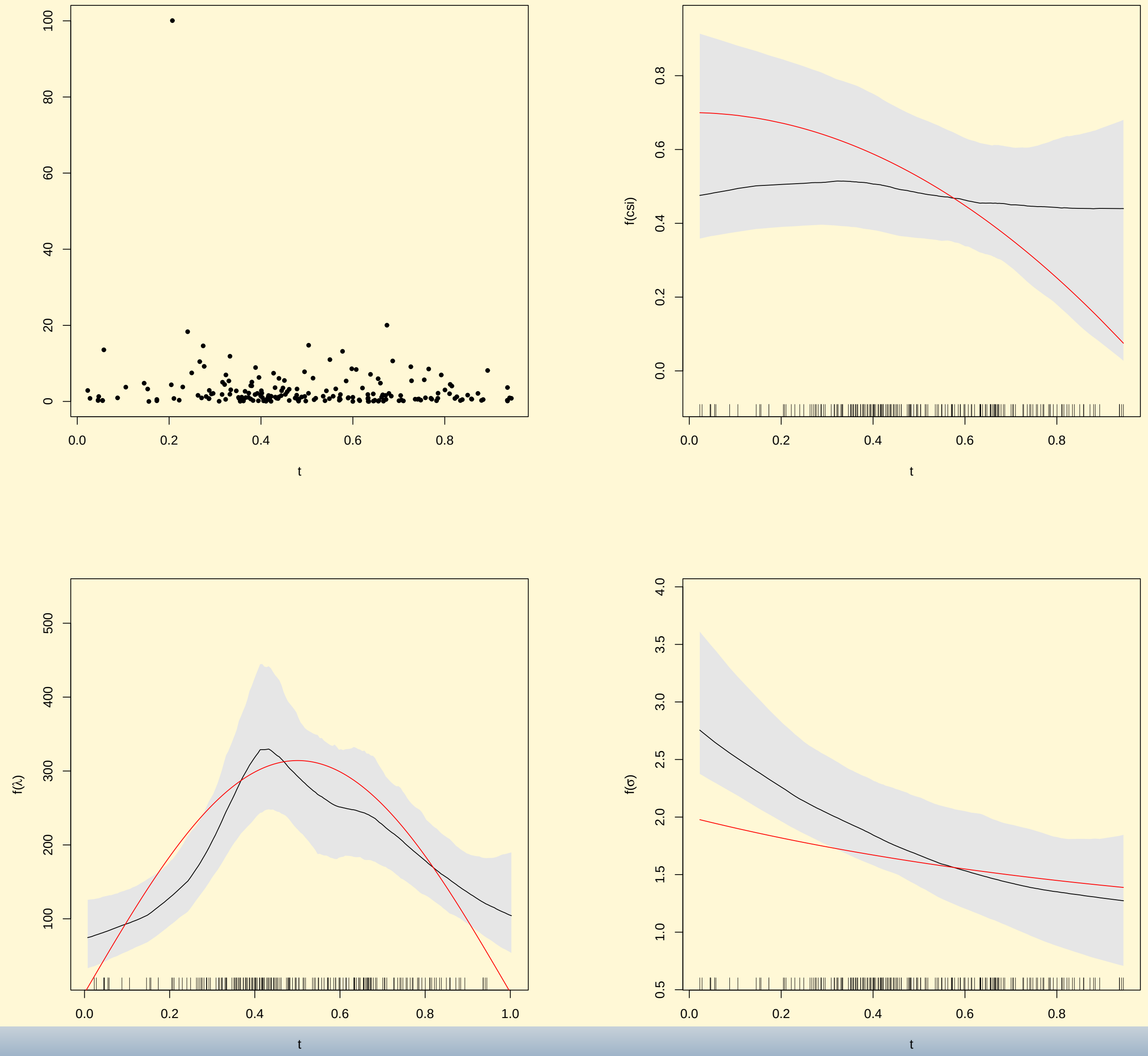
Some estimates on simulated samples

- we simulate a Poisson point process model with parameters $\lambda(t)$, $\xi(t)$, $\sigma(t)$
- samples are simulated involving approximately 200 observation (the exact number is random)
- in what follows we compare estimates and true values
  - red lines in the plot represent true values of the parameters
  - bands are pointwise 95% credibility intervals

Simulation 1



Simulation 2



Ozone data

We consider daily maxima of O_3 concentration (ppm) in Milan measured

- at three different sites (Juvara, Parco Lambro, Verziere)
- from 1995 to 2004 (10 years)

We consider only observations from June to September (included) since O_3 concentration is high only if temperatures is high.

The aim is to assess wether a time trend exist for the extremes of the series

Seasonality must be taken into account, semiparametric regression is used for this.

Ozone, model specification

- we employ a threshold of 170ppm chosen "by eye" (and do not discuss this choice further)
- a poisson point process model is estimated in which
  - a spline model is employed to allow for seasonality
  - random effects are employed to allow for site and year effect

Ozone, model specification

Consider observations (t, Y_{si}) where Y_{si} is ozone concentration measured

- at site s ($s = 1, 2, 3$)
- in year i ($i = 1995, \ldots, 2004$)
- at time of year (actually, of summer) t (renormalized so that $t \in [0, 1]$)

Poisson intensity is given by

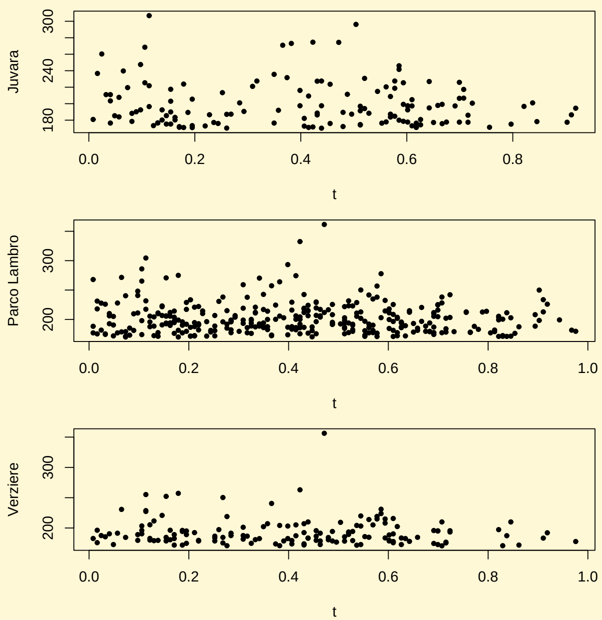
$$\log(\lambda_{tis}) = \beta_1 + \beta_2 t + f_\lambda(t) + \gamma_i^{(\lambda)} + \delta_s^{(\lambda)}$$

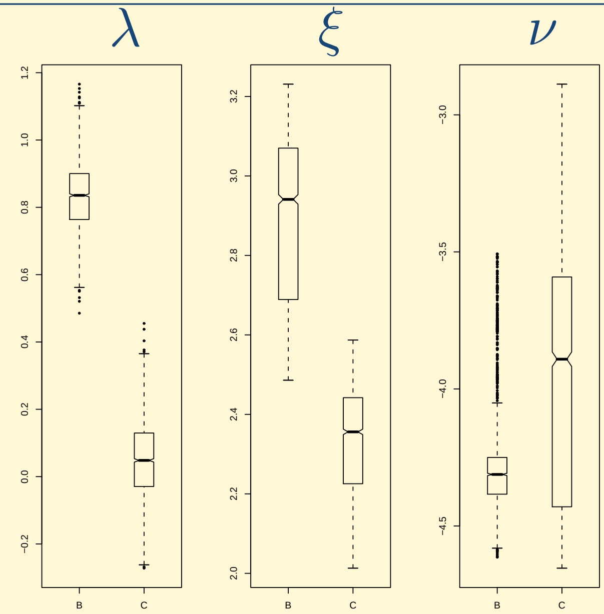
Parameters of the generalized Pareto for excesses is

$$\xi_{tis} = \beta_3 + \beta_4 t + f_\xi(t) + \gamma_i^{(\xi)} + \delta_s^{(\xi)}$$

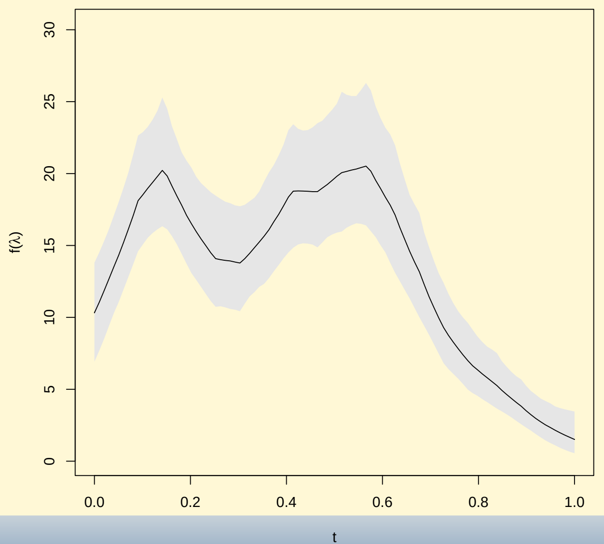
$$\log(\nu_{tis}) = \beta_5 + \beta_6 t + f_\nu(t) + \gamma_i^{(\nu)} + \delta_s^{(\nu)}$$

Ozone, results

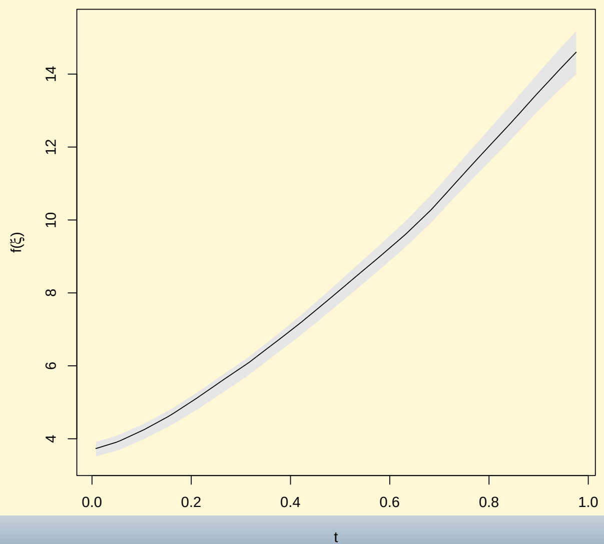




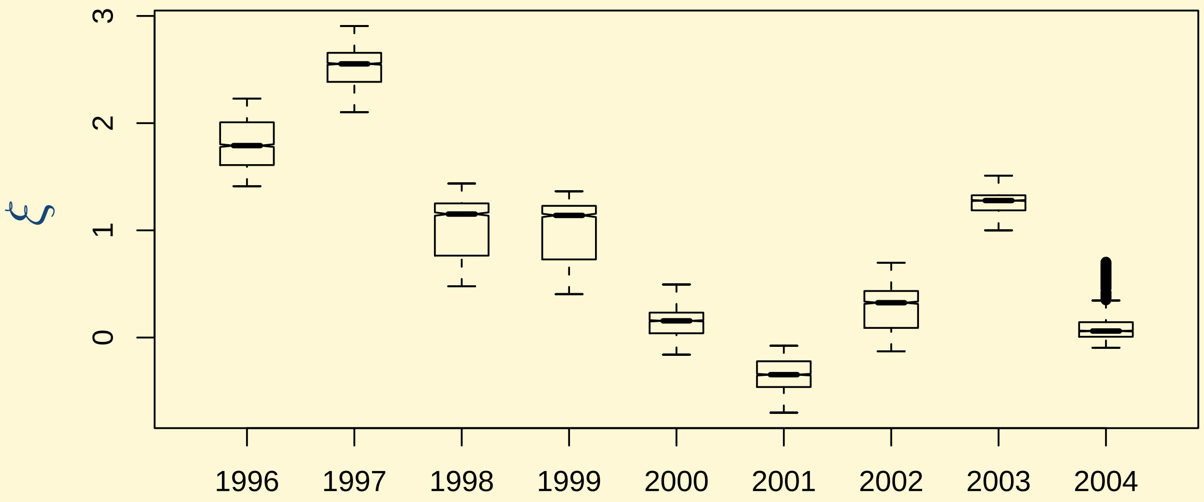
$\lambda(t)$

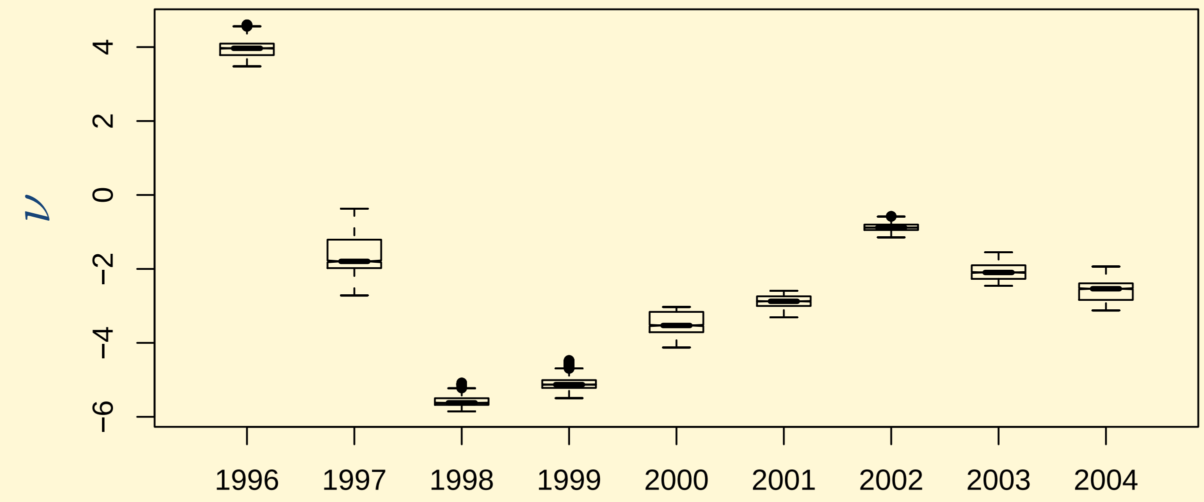


$\xi(t)$



Ozone, results on year effects





Concluding remarks

- Flexible set of tools to make inference for non-stationary extreme value models

What next?

- Make inference on Poisson process directly (no reparametrization)

- Compare results with existing approaches (GCV/AIC smoothness)