Robust splines and wavelets

Douglas Nychka,

`www.image.ucar.edu/~nychka`

- A penalized robust smoother
- Pseudo data.
- Empirical pseudo data
- Robust wavelets (and wavelets)

Supported by the National Science Foundation *L' Ecole D'Ete, Ovoronnaz Sep 2005*

The additive model

Given n pairs of observations (x_i, y_i), $i = 1, \ldots, n$

$$y_i = g(x_i) + \epsilon_i$$

ϵ_i's are random errors.

Assume that g is a smooth function or a realization of a Gaussian process and ϵ are symmetric but potentially heavy tailed.

Being tied to local averages makes it difficult to think about a robust method.

What is the model for the unknown function?

A univariate robust estimate as a minimization problem

Recall that

$$min_a \sum_{i=1}^{n} (y_i - a)^2$$

is a complicated way to characterize $\bar{y}$

Let $\rho(t)$ be a convex function symmetric about 0.

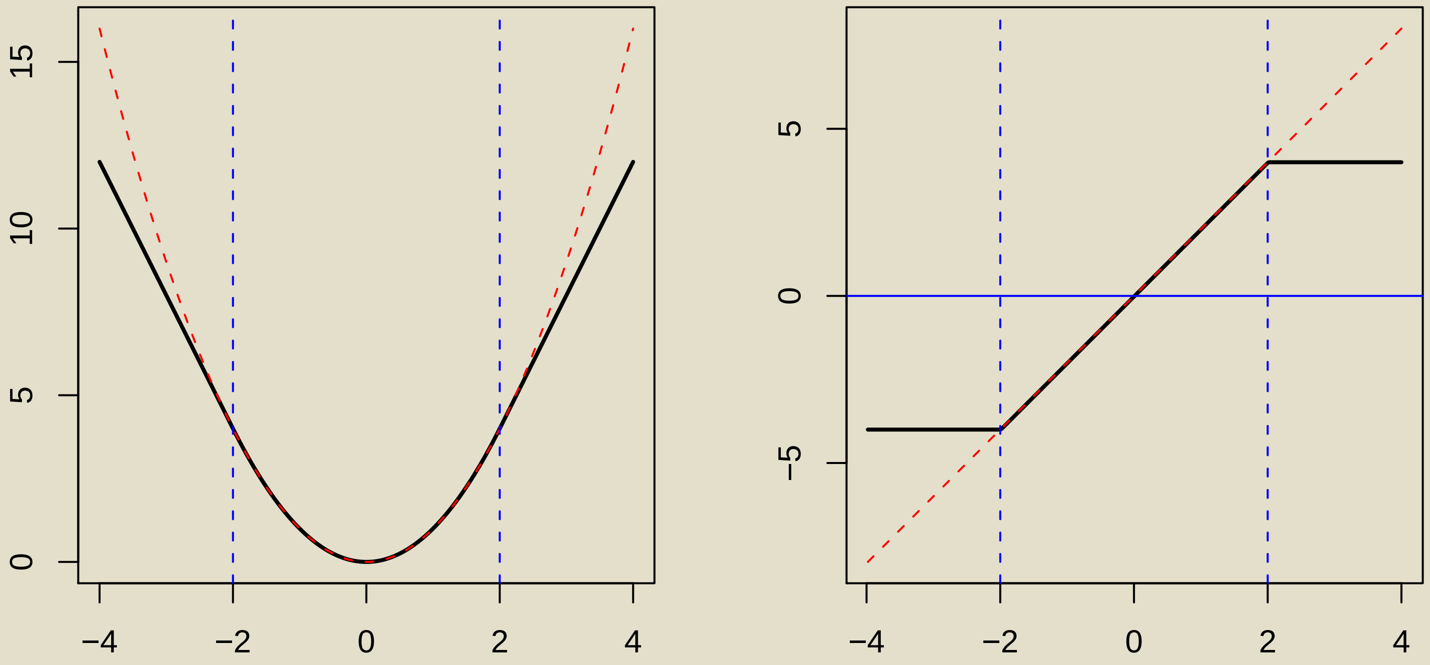
ρ is like t^2 close to 0 but grows like $|t|$ when t is large

$$min_a \sum_{i=1}^{n} \rho(y_i - a)$$

will give a robust estimator that down weights the effect of large values of Z.

An example of ρ and its derviative

Transition is at -2 and 2.



A smoother version is $\rho(t) = log(cosh(t))$ **...**

A reasonable penalized smoother

Just replace the sum of squares with a robust measure of fit.

Recall

$$\widehat{g}(x) = \sum_{l=1}^{n} \widehat{\theta}_k \psi_k(x)$$

$$\min_{\boldsymbol{\theta}} \sum_{i=1}^{n} \rho(\boldsymbol{y} - [W\boldsymbol{\theta}]_i) + \lambda \boldsymbol{\theta}^T B \boldsymbol{\theta}$$

This is now a nonlinear problem to find θ

But in some sense we are done ...

Some details of actually finding $\hat{g}(x)$

For convenience parametrize this as $\boldsymbol{f} = W\boldsymbol{\theta}$ $W^{-1}\boldsymbol{f} = \boldsymbol{\theta}$

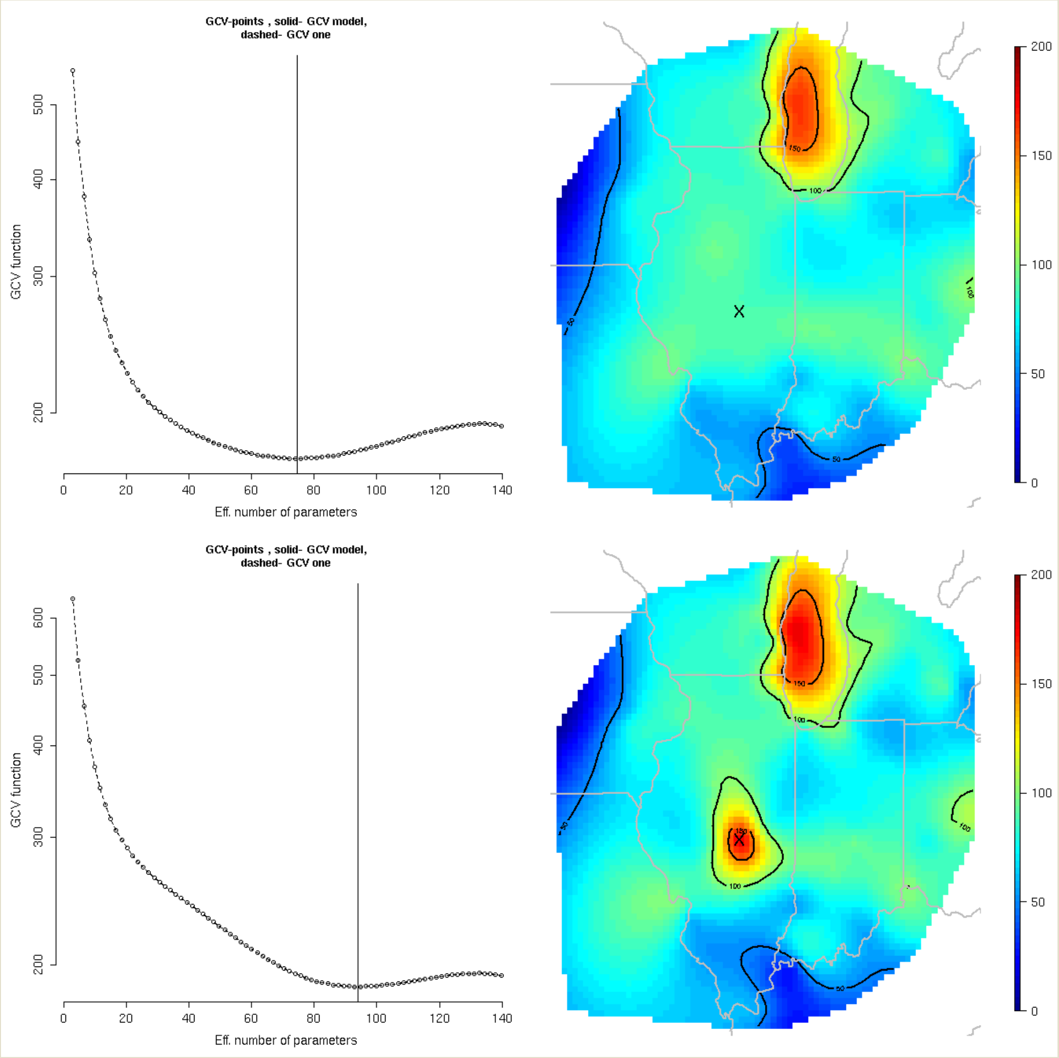
$$\min_{\boldsymbol{f}} \sum_{i=1}^{n} \rho(\boldsymbol{y}_i - \boldsymbol{f}_i) + \lambda \boldsymbol{f}^T W^{-T} B W^{-1} \boldsymbol{f}$$

Differentiate, set equal to zero and Solve for f.

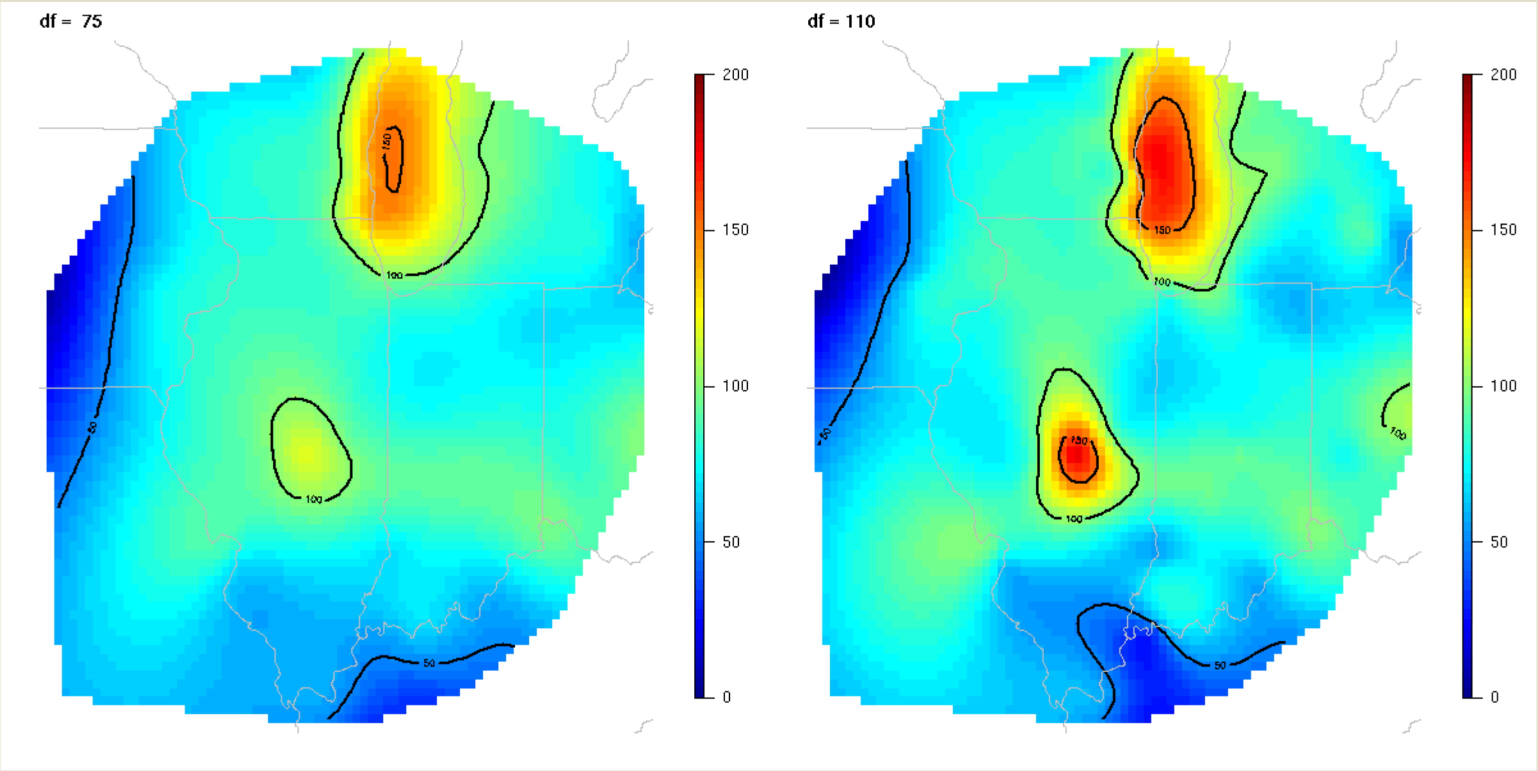
$\eta = \rho'$ and $R = W^{-T} B W^{-1}$

$$\eta(\boldsymbol{y}_i - \boldsymbol{f}_i) - \lambda [R\boldsymbol{f}]_i = 0$$

Adding an outlier to the ozone data



Some examples of robust smooths



Analysis of the robust estimator based on pseudo-data

This is theory – not yet practical. Create alternative observations: pseudo-data

$$\tilde{y}_i = g(\boldsymbol{x}_i) + \frac{\eta(\epsilon_i)}{2}$$

Recall that $\eta = \rho$ and $\eta(\epsilon_i)$ will be bounded.

Use a least squares type smoother applied to the pseudo data

Theoretical Result for pseudo data

Consider smoothers that are based on a reproducing kernel (i.e. spatial statistics type estimates)

$\widehat{g}$ robust estimate

$\tilde{g}$ LS estimate of g based on pseudo data.

$\|f\|_n^2 = (1/n)\sum_{i=1}^n f(x_i)^2$ **and** $C_n = E\|\tilde{g} - g\|_n^2$

$$\|\widehat{g} - \tilde{g}\|_n / \sqrt{(C_n)} \to 0$$

as $n \to \infty$.

Under several assumptions , e.g. on λ, $A(\lambda)$

Empirical pseudo data

Use an estimate in place of g.

$$\tilde{y}_i = \hat{g}(\boldsymbol{x}_i) + \eta(y_i - \hat{g}(x_i))/2$$

Empirical pseudo data (EPD) and fixed points

Apply a LS smoother to the EPD to get a robust estimate of g. Our goal is a fixed point where the estimate obtained is the same as that used to construct the EPD

An algorithm:

- **Start with initial $\hat{g}^0$.**
- **Repeat:**

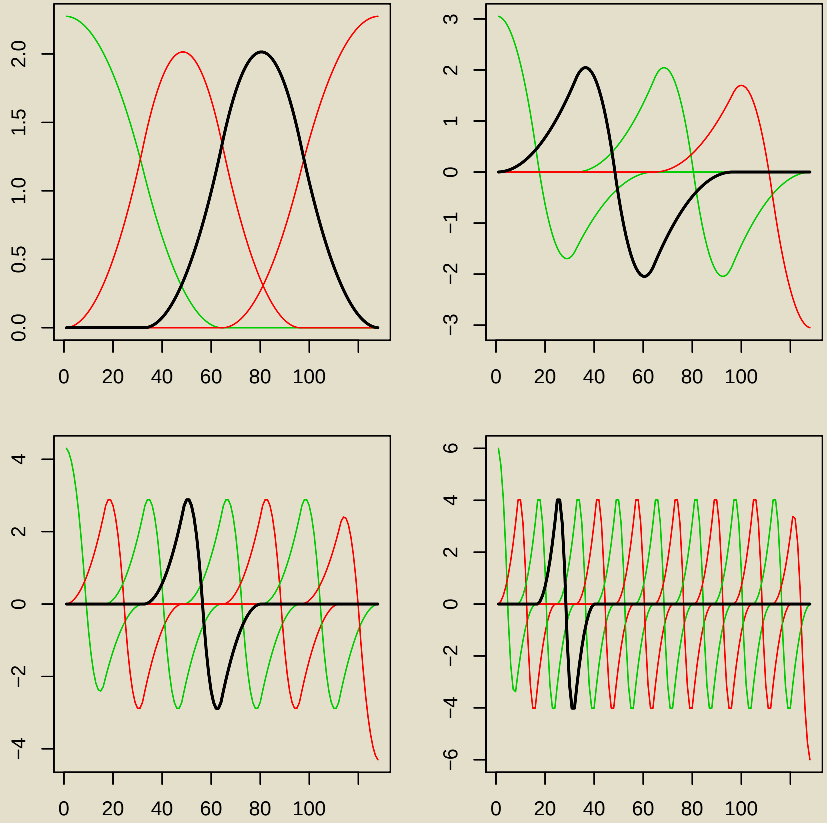
  1. **Form EPD:** $\tilde{y}_i = \hat{g}^J(\boldsymbol{x}_i) + \eta(y_i - \hat{g}^J(x_i))/2$
  2. $\hat{g}^{J+1}$ **LS smoother based on** $\tilde{\boldsymbol{y}}$

Comments

- If the algorithm converges it will give the robust estimator!
- In the smoothing step 2 one can also use Cross validation to choose the smoothing parameter.
- One need not use a spline type smoother any penalized LS smoother can work.

Wavelets

Choose a multiresolution basis: members are similiar shape but different sizes of supports at different locations.



A simple shrinkage estimate

Usually the wavelet basis is orthogonal, locations eqaully spaced.

$$W^T \boldsymbol{y} = \boldsymbol{\theta}^*$$

These are the empirical coefficents for g. (i.e. $\boldsymbol{y} = W\boldsymbol{\theta}^*$)

Now shrink and decimate

For some $C > 0$

$$\widehat{\theta}_k = sign(\boldsymbol{\theta}_k^*)(|\boldsymbol{\theta}_k^*| - C)_+$$

$(u)_+ = u$ if $u > 0$ and 0 if $u \leq 0$

So what about penalized LS?

For orthogonal wavelets. $W^T W = I$.

$$\min_{\boldsymbol{\theta}} \sum_{i=1}^{n} (\boldsymbol{y}_i - [W\boldsymbol{\theta}]_i)^2 + \lambda \sum_{i=1}^{n} |\boldsymbol{\theta}_i|$$

So what about penalized LS?

For orthogonal wavelets. $W^T W = I$.

$$\min_{\boldsymbol{\theta}} \sum_{i=1}^{n} (\boldsymbol{y}_i - [W\boldsymbol{\theta}]_i)^2 + \lambda \sum_{i=1}^{n} |\boldsymbol{\theta}_i|$$

$$\min_{\boldsymbol{\theta}} \|(\boldsymbol{y} - W\boldsymbol{\theta}\|^2 + \lambda \sum_{i=1}^{n} |\boldsymbol{\theta}_i|$$

$$\min_{\boldsymbol{\theta}} \|\boldsymbol{\theta}^* - \boldsymbol{\theta}\|^2 + \lambda \sum_{i=1}^{n} |\boldsymbol{\theta}_i|$$

So what about penalized LS?

For orthogonal wavelets. $W^T W = I$.

$$\min_{\boldsymbol{\theta}} \sum_{i=1}^{n} (\boldsymbol{y}_i - [W\boldsymbol{\theta}]_i)^2 + \lambda \sum_{i=1}^{n} |\boldsymbol{\theta}_i|$$

$$\min_{\boldsymbol{\theta}} \|(\boldsymbol{y} - W\boldsymbol{\theta}\|^2 + \lambda \sum_{i=1}^{n} |\boldsymbol{\theta}_i|$$

$$\min_{\boldsymbol{\theta}} \|\boldsymbol{\theta}^* - \boldsymbol{\theta}\|^2 + \lambda \sum_{i=1}^{n} |\boldsymbol{\theta}_i|$$

$$\min_{\boldsymbol{\theta}} \sum_{i=1}^{n} (\boldsymbol{\theta}_i^* - \boldsymbol{\theta}_i)^2 + \lambda \sum_{i=1}^{n} |\boldsymbol{\theta}_i|$$

So what about penalized LS?

For orthogonal wavelets. $W^T W = I$.

$$\min_{\boldsymbol{\theta}} \sum_{i=1}^{n} (\boldsymbol{y}_i - [W\boldsymbol{\theta}]_i)^2 + \lambda \sum_{i=1}^{n} |\boldsymbol{\theta}_i|$$

$$\min_{\boldsymbol{\theta}} \|(\boldsymbol{y} - W\boldsymbol{\theta}\|^2 + \lambda \sum_{i=1}^{n} |\boldsymbol{\theta}_i|$$

$$\min_{\boldsymbol{\theta}} \|\boldsymbol{\theta}^* - \boldsymbol{\theta}\|^2 + \lambda \sum_{i=1}^{n} |\boldsymbol{\theta}_i|$$

$$\min_{\boldsymbol{\theta}} \sum_{i=1}^{n} (\boldsymbol{\theta}_i^* - \boldsymbol{\theta}_i)^2 + \lambda \sum_{i=1}^{n} |\boldsymbol{\theta}_i|$$

For each i

$$\min_{\boldsymbol{\theta}_i} (\boldsymbol{\theta}_i^* - \boldsymbol{\theta}_i)^2 + \lambda |\boldsymbol{\theta}_i|$$

We know how to do this! And it is the same as shrinkage.

A robust wavelet

Choose a multiresolution basis ....

$$\min_{\boldsymbol{\theta}} \sum_{i=1}^{n} \rho(\boldsymbol{y} - [W\boldsymbol{\theta}]_i) + \lambda \sum_{i=1}^{n} |\boldsymbol{\theta}_i|$$

The absolute value penalty really changes the estimator!

EPD also works here.

John Lennon

with outlier noise

LS denoising

Robust version

Summary

We have substituted a robust measure of fit for the sum of squares to get a robust smoother.

There is some theory for the robust smoother based on pseudo data.

Empirical pseudo data provides a computational algorithm and a generalization to more exotic penalized estimators.