VA Methods

Topic #1

Penalized splines as

Mixed Linear Models.

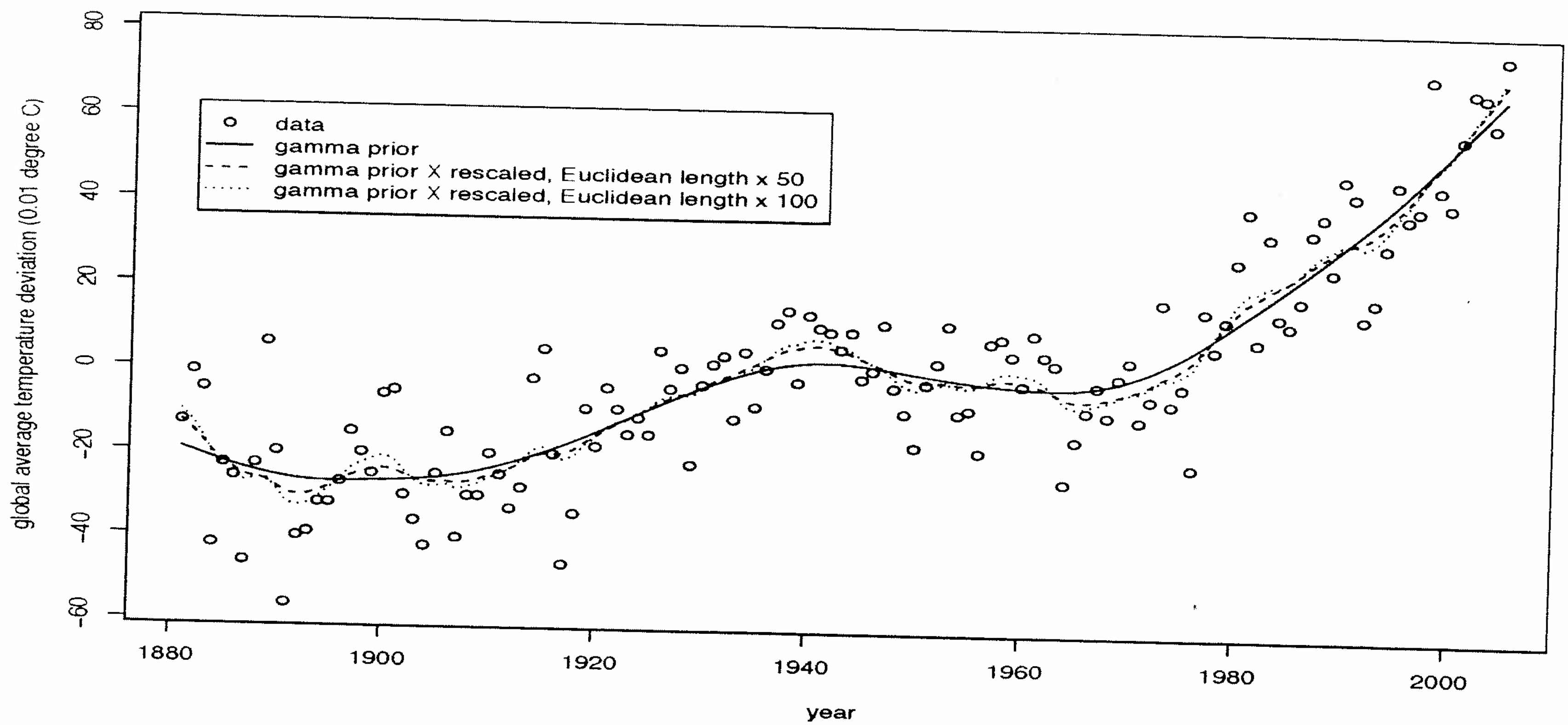


Figure 4: Example 2. Data and fitted smooths, gamma prior.

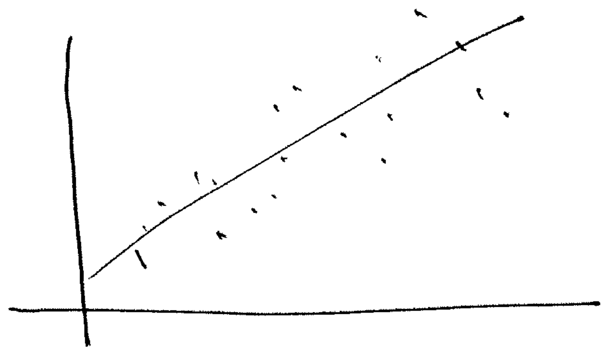
An xample of the sort of data for which we might want splines ...

I.C. Smoothing with penalized splines expressed as MLM

For now, one-dimensional (scatterplot) smoothing.

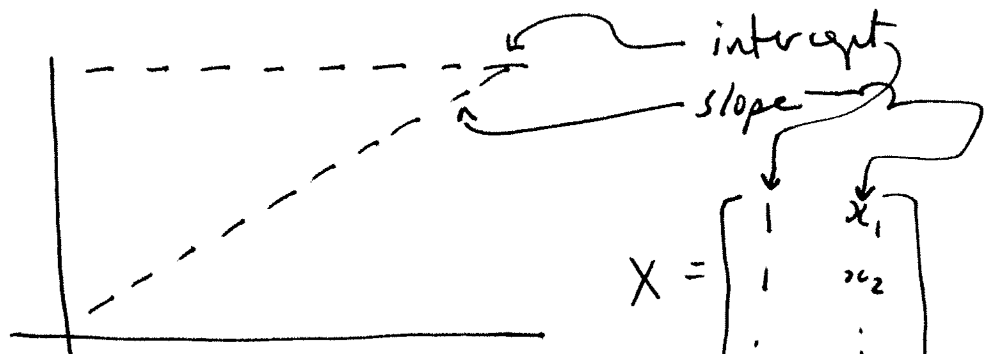
Basis (in the sense of linear algebra) RWC 3.2

straight-line model



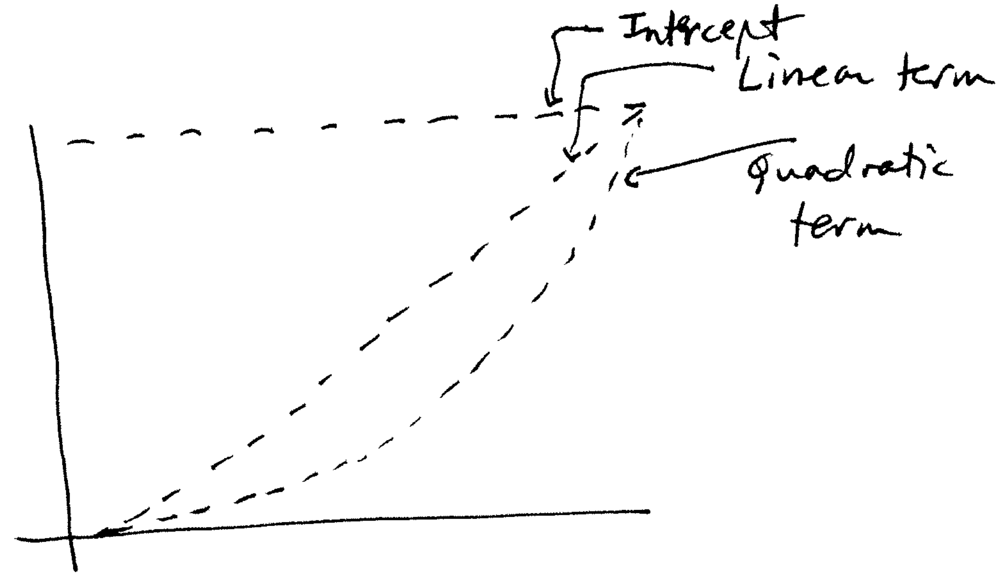
$$y = \beta_0 + \beta_1 x + \text{error}$$

uses this basis for fitted value space

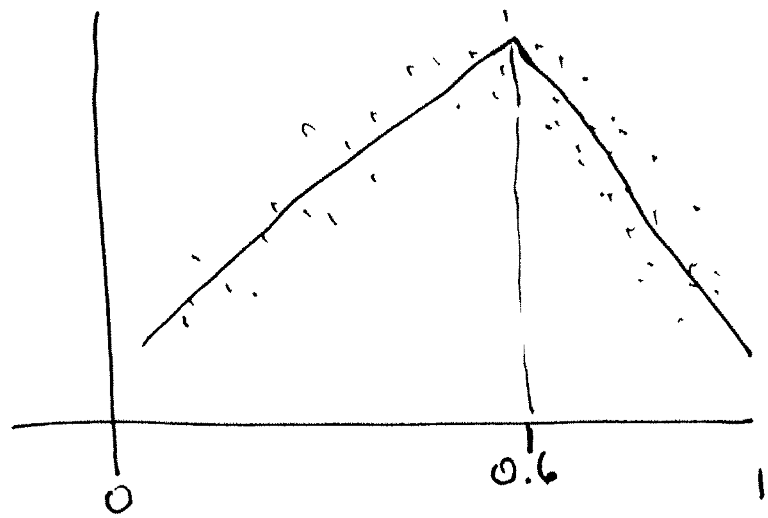


$$X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}$$

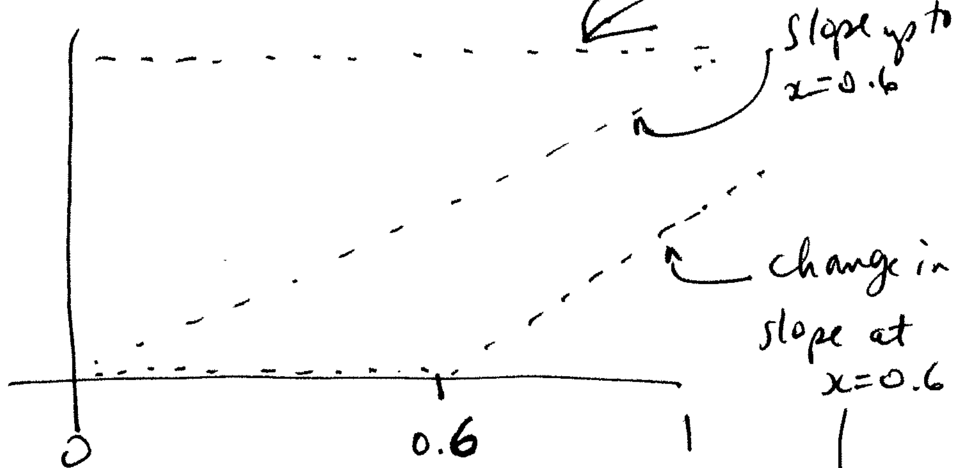
Quadratic model uses this basis



A "broken stick" model like this



Uses this basis for the fitted value space



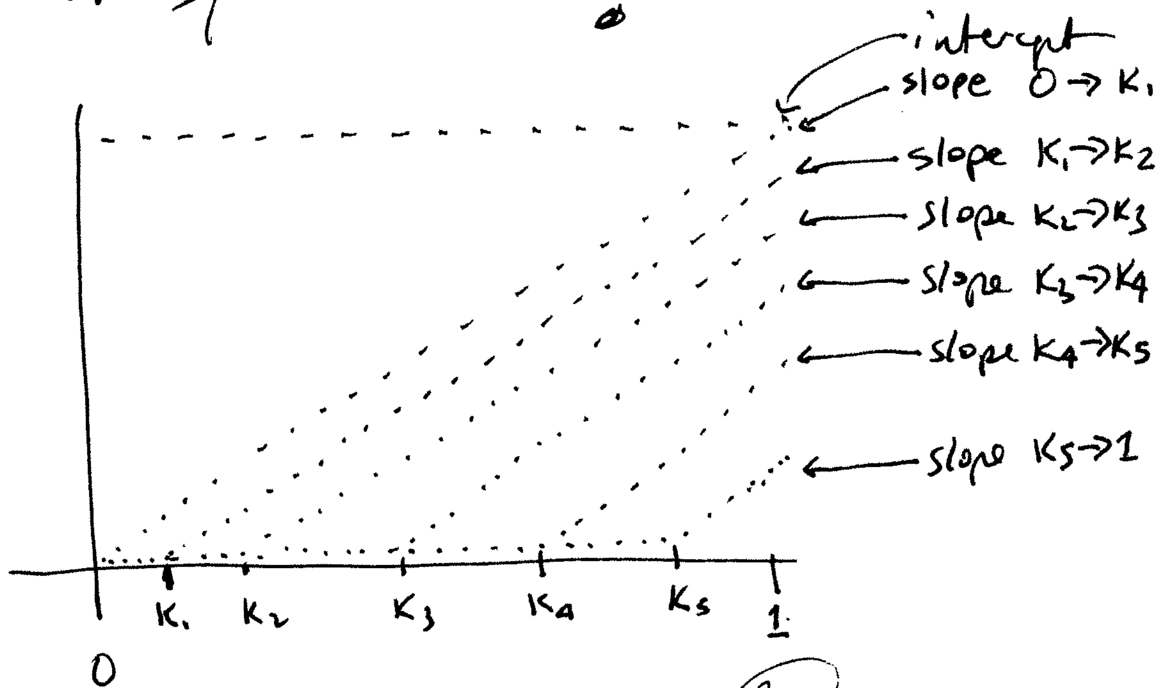
$$y_i = \beta_0 + \beta_1 x_i + \beta_{11}(x_i - 0.6)_+ + \text{error}.$$

The obvious generalization of the broken stick model is to allow the line to change slope in several places, giving the ...

I C1/2 2/3/08

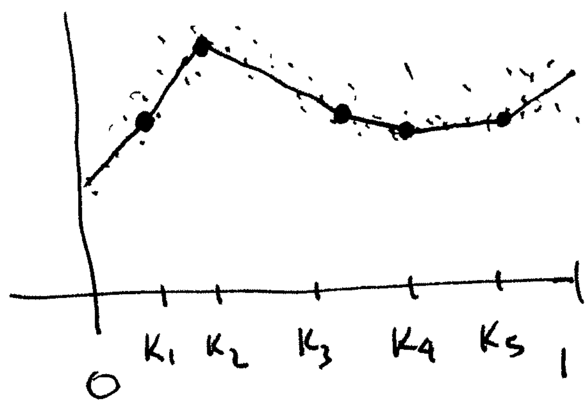
Linear spline model (Not too smart)

Which has this basis:



$$y_i = \beta_0 + \beta_1 x_i + \beta_{11}(x_i - K_1)_+ + \beta_{12}(x_i - K_2)_+ + \cdots + \beta_{15}(x_i - K_5)_+ + \text{error}$$

fit looks like this:



Basic idea of a spline:

- Choose a basis
- Choose knots.

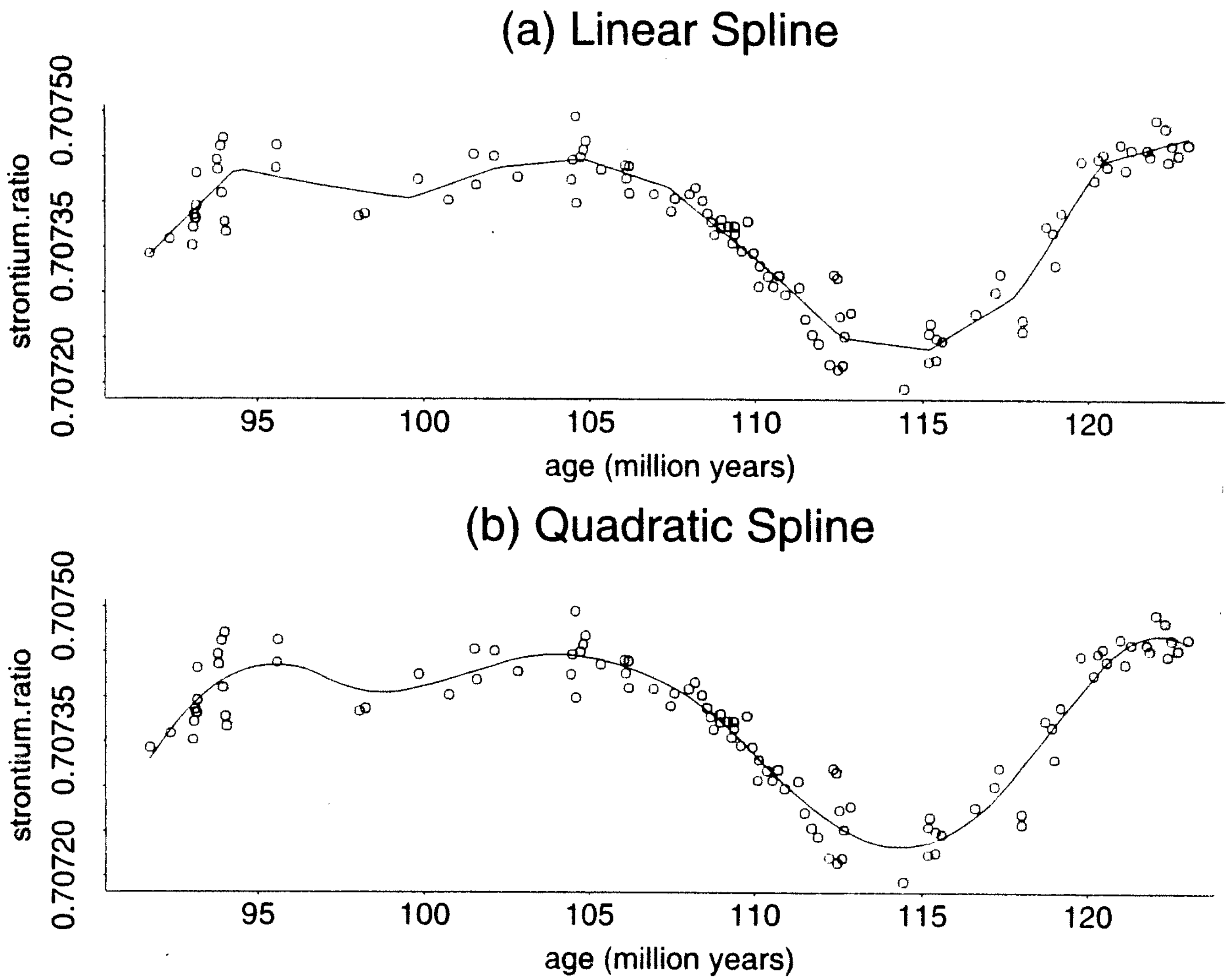
I c 1/3   2/3/08

Other possible bases (two of many)

Quadratic (RWC 3.6)

$1,\ x,\ x^2,\ (x-\kappa_1)_+^2,\ \ldots,\ (x-\kappa_K)_+^2$

- Does not have ~~sharp~~ "corners" like $(x-\kappa_j)_+$

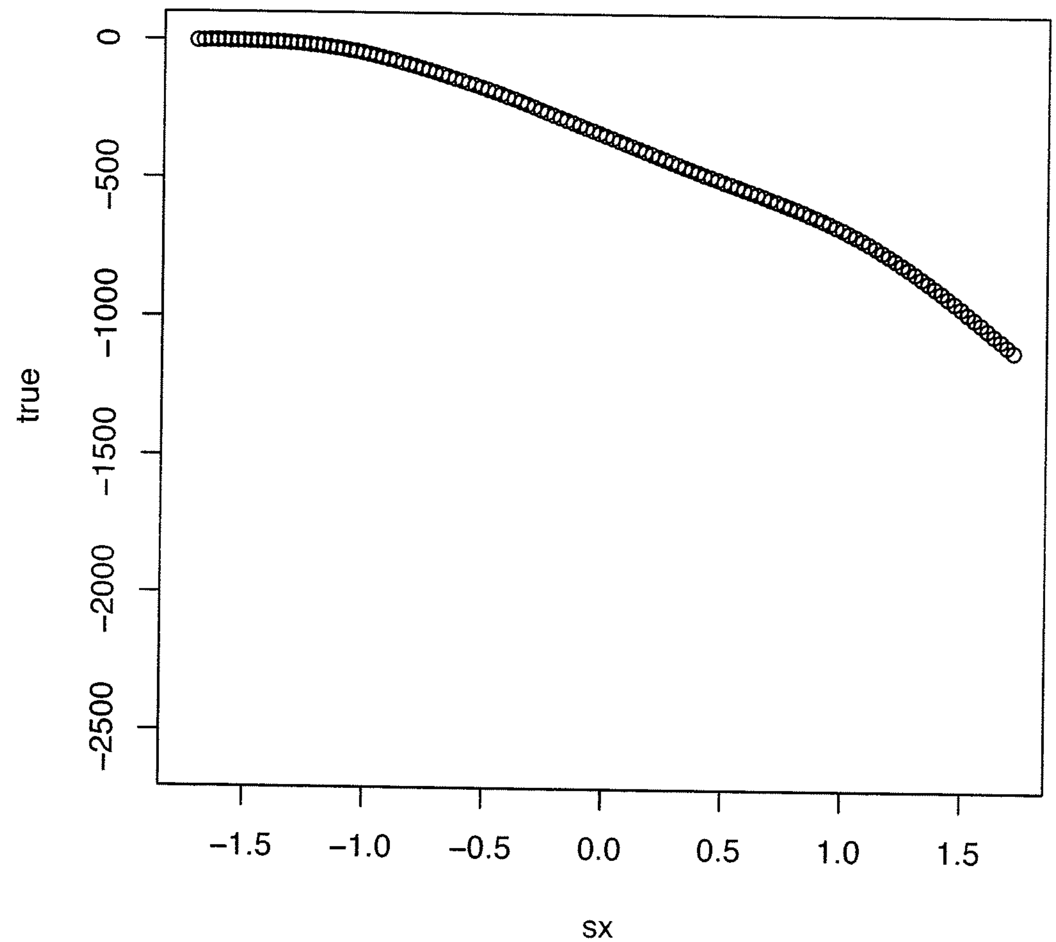


IC 1/4 2/3/08

IC 1/4a 7/3/08

Here's a curve computed as

$\beta_0 + \beta_1 x + \beta_2 x^2 + \sum_{i=1}^{30} \beta_{1i} (x - K_i)_+^2$ for K_i equally spaced among 125 x values.



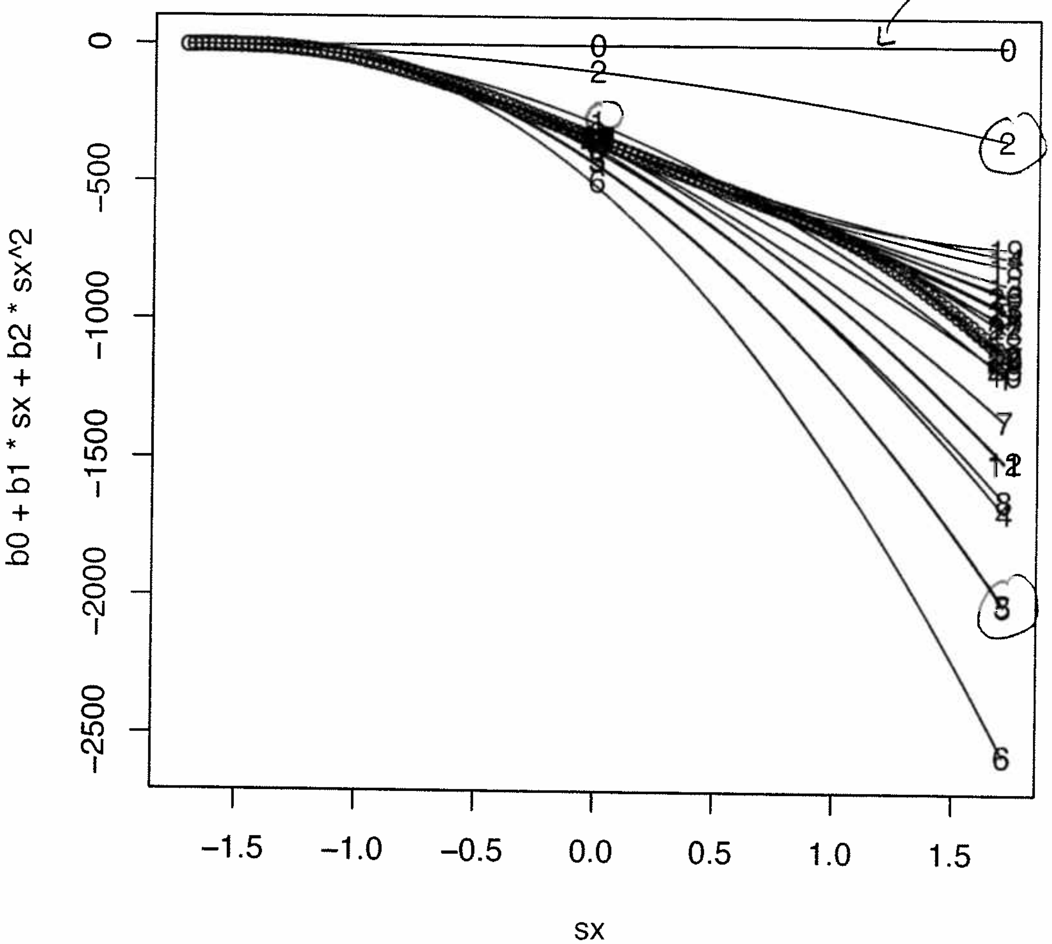
$\beta_0 = \frac{1}{2}$ $\beta_1 = 1$ $\beta_2 = -\frac{1}{2}$

ICI/46 3/3/08

Here's how the "increments" to the curve effect it

$\beta_0 + \beta_1 x + \beta_2 x^2 + \sum_{i=1}^{30} (x - K_i)_+^2 \beta_i$

$\beta_0 + \beta_1 x + \beta_2 x^2$



each line shows the effect of adding one more $\beta_i (x - K_i)_+^2$

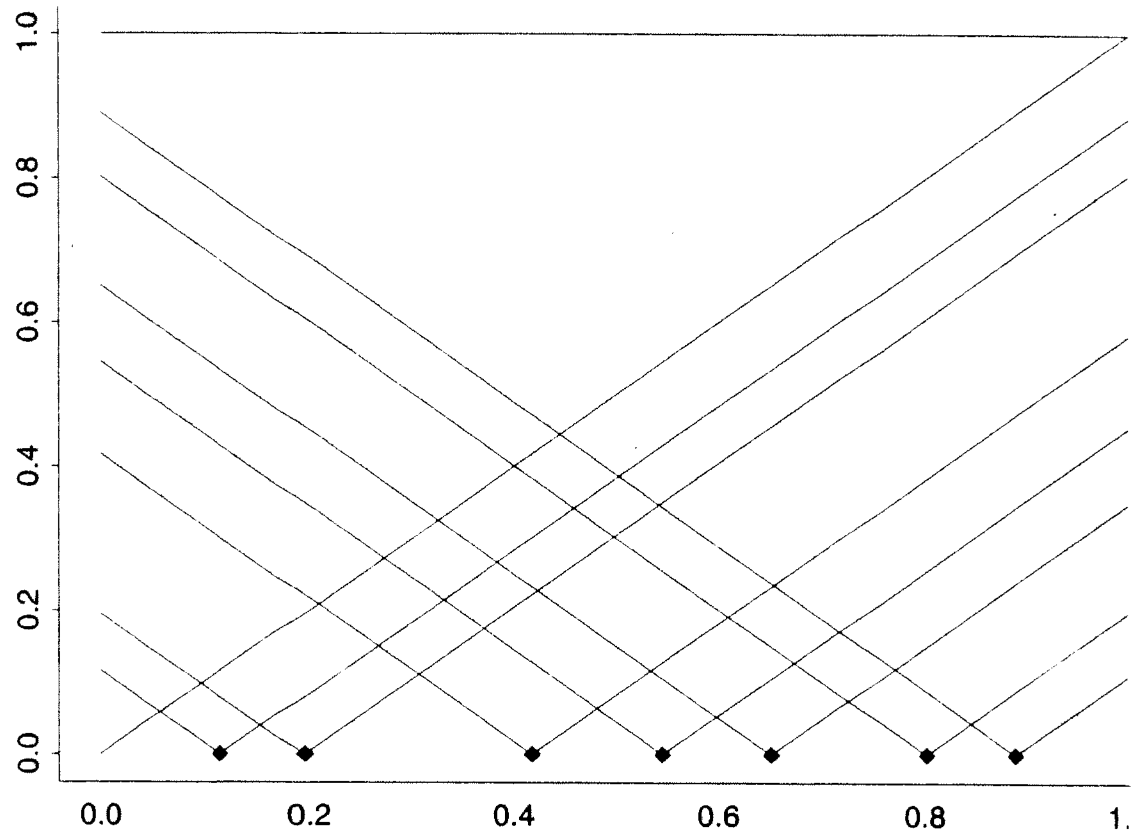
~~ICI/46 3/7/08~~

Radial Basis (RWC 3.7.3)

$$1,\ x,\ \ldots,\ x^p,\ \underbrace{|x-\kappa_1|^p,\ \ldots,\ |x-\kappa_K|^p}$$

– function only of the distance $|x-\kappa_j|$

$p=1$ linear radial basis functions.



For $p=1$, $y_i = \beta_0 + \beta_1 x + \sum_{k=1}^{K} \beta_{1k} |x-\kappa_k| + \text{error}$

so the slope at x is $\left(\sum_{j:x>\kappa_j} \beta_{1k}\right) - \left(\sum_{j:x<\kappa_j} \beta_{1k}\right) + \beta_1$

(This generalizes nicely to 2, 3, + dimensions)

IC 1/5 2/3/08

How do we choose a basis?

- NOT CLEAR
- power $p \Rightarrow p-1$ continuous first derivatives.
- ?? Numerical / computing considerations.

How do we choose knots?

- How many? Where?
- RWC (5.5.3, pp. 125-126) default:
  - $K = \min\left(35, \frac{1}{4} \times \{\#\text{unique } x_i\}\right)$
  - $\kappa_k = \left(\frac{k+1}{K+2}\right)^{th}$ sample quantile of unique x_i
- Massage this for specific problems, e.g., if f appears flat in one area and especially bumpy in another area.

Penalized spline regression

We have two conflicting goals in fitting a spline:

(1) Flexibility ⇒ use many knots

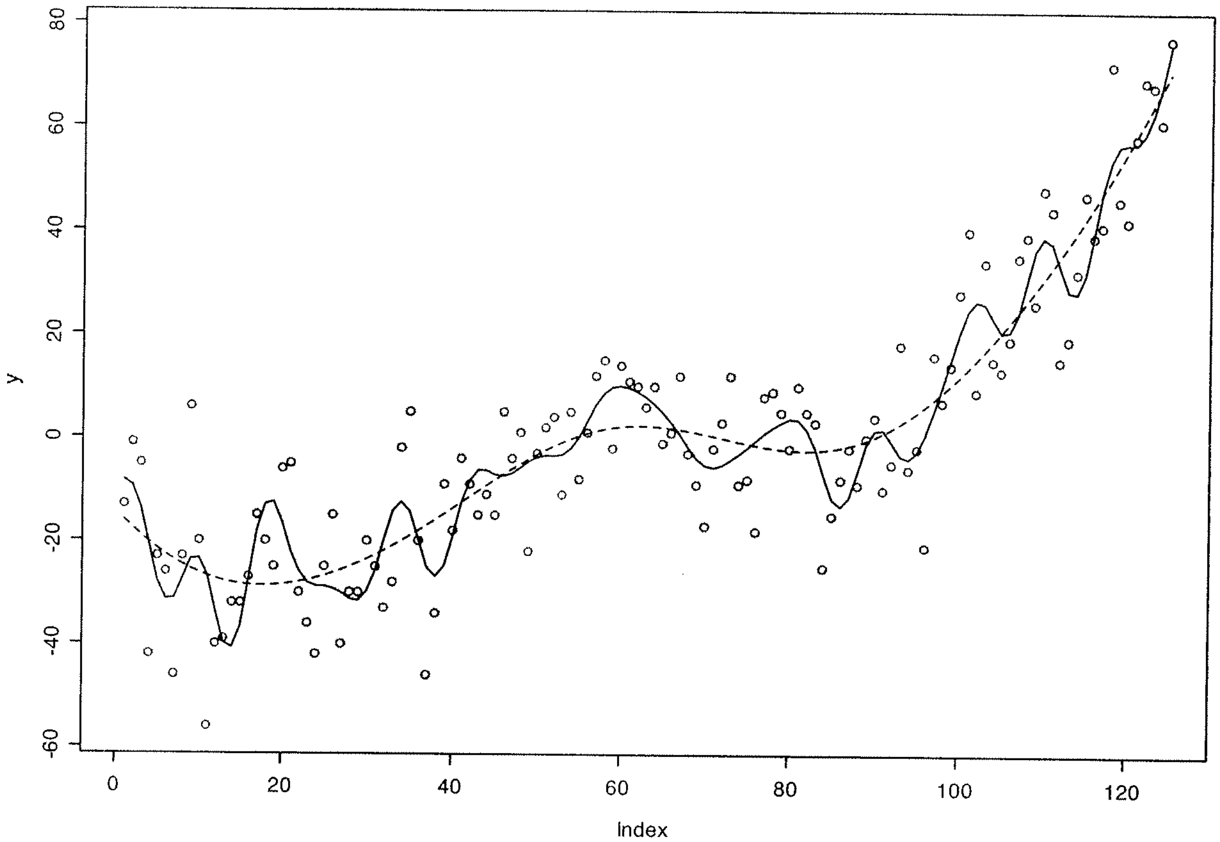
(2) Avoid overfitting – using fewer knots risks missing interesting features

Solution is a compromise:

- use many knots but
- "penalize" large β_{1k}

RWC (p. 65): - Constrain the influence of the knots
- reduce overfitting and variability in the fit.

IC 1/7 2/3/08



- GMST data - circles o o o o o

- Unsmoothed least squares fit of quadratic basis, 30 knots equally spaced ———

- Shrink the 30 coefficients corresponding to knots - - - -

(DF in fit: 6.7)

IC1/7a 7/3/08

Example: Linear spline

basis	1	x	$(x-\kappa_1)_+$	$(x-\kappa_2)_+$	$\cdots$	$(x-\kappa_K)_+$
parameters	β_0	β_1	β_{11}	β_{12}	$\cdots$	β_{1K}

How to constrain β? Many ways, but RWC choose $C > \sum_{j=1}^{K} \beta_{1j}^2$ for convenience

Define $D = \left[\begin{array}{c|c} 0_{2\times 2} & 0_{2\times K} \\ \hline 0_{K\times 2} & I_K \end{array}\right]$

The minimization problem is to choose β to

minimize $(y-X\beta)'(y-X\beta)$ subject to $\beta' D \beta \le C$

$\iff$ minimize $(y-X\beta)'(y-X\beta) + \lambda^2 \beta' D \beta$

for a particular $\lambda \ge 0$.

IC1/8 2/13/08

$\rightarrow$

minimize $(y-X\beta)'(y-X\beta) \ni \beta' D\beta \le c$

$\iff$ minimize $(y-X\beta)'(y-X\beta) + \lambda^2 \underbrace{\beta' D\beta}_{\text{"Roughness penalty"}}$ for $\lambda^2 \geq 0$

Solution: $\hat{\beta}_\lambda = (X'X + \lambda^2 D)^{-1} X'y$

fitted $\hat{y} = X\hat{\beta}_\lambda = X(X'X + \lambda^2 D)^{-1} X'y$

with $DF_\lambda = tr\left(X(X'X + \lambda^2 D)^{-1} X'\right)$

This should look familiar:

- Mixed linear model

OR - Bayesian: $\beta'(\lambda^2 D)\beta \rightsquigarrow \beta \sim N(0, \lambda^{-2} D^{-1})$ a priori

I:C 1/9 2/3/08

<u>General definition of a penalized spline</u> (RWC 3.9)

$\hat{y} = B(\underset{\sim}{x})\hat{\beta}$ for some basis $B(\underset{\sim}{x})$ of observed $\underset{\sim}{x}$

$\hat{\beta}$ minimizes $(y - B(\underset{\sim}{x})\beta)'(y - B(\underset{\sim}{x})\beta) + \alpha\beta' D\beta$

for a scalar $\alpha > 0$ and pos. semi-def D

To specify a penalized spline, you must choose:

- The degree (linear, quadratic, etc)
- Knot locations
- Penalty

- The specific basis $\Rightarrow B(\underset{\sim}{x})$
  $\Rightarrow D$ (given the penalty)

RWC prefer (Sec 3.7)

Basis $1, x, \ldots, x^p, (x-\kappa_1)_+^p, \ldots, (x-\kappa_K)_+^p$

$$D = \left(\begin{array}{c|c} 0_{(p+1)\times(p+1)} & 0 \\ \hline 0 & I_K \end{array}\right)$$

I:C 10/10 2/3/08

→

I. Mixed Linear Models, per RWC

A. Mixed linear models in the standard formulation

1. The mixed linear model, standard form

$$\underset{n\times 1}{y} = \underset{n\times p}{X}\,\underset{p\times 1}{\beta} + \underset{n\times q}{Z}\,\underset{q\times 1}{u} + \underset{n\times 1}{\varepsilon}$$

$$E\begin{pmatrix} u \\ \varepsilon \end{pmatrix} = \begin{pmatrix} 0_n \\ 0_n \end{pmatrix} \qquad \mathrm{cov}\begin{pmatrix} u \\ \varepsilon \end{pmatrix} = \begin{bmatrix} G & 0 \\ 0 & R \end{bmatrix}$$

Eliminate Zu and it's the familiar linear model

All the novelty, including new modeling possibilities, arises from the term Zu.

One of RWC's main points is the breadth of models that can be written this way.

$$y = X\beta + Zu + \varepsilon, \quad \varepsilon \sim N_n(0, R)$$

$$u \sim N_q(0, G)$$

β is referred to as "the fixed effects"
u is referred to as "the random effects"

- Older usage: - Random effects describe individuals sampled from a population
  - these individuals are not of interest themselves, only their variation is of interest

- Newer usage: "$u \sim N_q(0, G)$" is a model for variation among elements of u and in the elements of Zu, which may be the entire population of interest and may be of interest themselves.

The three examples that follow all fit the older usage; later models fit the newer usage

IA1/2 1/24/08

Penalized splines as Mixed Linear Models

Splines developed in their own little universe, and have a huge literature of their own.

This is an example of merging this large, diverse class of models into the MLM framework, at the (modest?) cost of throwing away part of the spline-model universe.

For the spline model

$$y_i = \underbrace{\beta_0 + \beta_1 x_i + \cdots + \beta_p x_i^p}_{\text{Fixed Effects}} + \underbrace{\sum_{k=1}^{K} \beta_{1k} (x_i - \kappa_k)_+^p}_{\text{Random Effects}} + \varepsilon_i$$

$$X = \begin{bmatrix} 1 & x_1 & & x_1^p \\ 1 & x_2 & \cdots & x_2^p \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ 1 & x_n & & x_n^p \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_p \end{bmatrix} \qquad Z = \begin{bmatrix} (x_1 - \kappa_1)_+^p & \cdots & (x_1 - \kappa_K)_+^p \\ (x_2 - \kappa_1)_+^p & \cdots & (x_2 - \kappa_K)_+^p \\ \vdots & & \vdots \\ \vdots & & \vdots \\ (x_n - \kappa_1)_+^p & \cdots & (x_n - \kappa_K)_+^p \end{bmatrix} \quad u = \begin{bmatrix} \beta_{11} \\ \vdots \\ \beta_{1K} \end{bmatrix}$$

$$\text{cov}(\varepsilon) = R = \sigma_\varepsilon^2 I_n \qquad \text{cov}(u) = G = D^{-1} \sigma_u^2$$

IC 2/1 2/3/08

2. Doing statistics with the mixed linear model

(a) Conventional (likelihood-based) analyses

Mean structure estimates

Preliminary: lots of jargon here

- estimate fixed effects
- predict random effects ← reflects old usage
- (Estimated) Best linear unbiased prediction, (E)BLUP

I will avoid this as much as I can, in favor of . . .

Unified approach (based on likelihood)

- Assume normal errors & random effects
  - The same approach works for non-normal models.

- Use as estimates the ~~maximum likelihood~~ likelihood maximizing values given the variance components.

We have: $y = X\beta + Zu + \varepsilon$ $\qquad \varepsilon \sim N_n(0, R)$

$\qquad\qquad\qquad\qquad\qquad\qquad u \sim N_g(0, G)$

In the conventional view, y and u are the random variables, so we can write their joint density as

$$f(y, u \mid \beta, R, G) = f(y \mid u, \beta, R)\, f(u \mid G)$$

Take the log and write the two terms on the left side as

$$\log f(y, u \mid \beta, R, G) = K - \tfrac{1}{2}\log|R| - \tfrac{1}{2}\log|G|$$
$$-\tfrac{1}{2}\left\{(y - X\beta - Zu)' R^{-1} (y - X\beta - Zu) + u' G^{-1} u\right\}$$

If we treat R and G as known, β and u are estimated (BLUPs) by minimizing

$$\left[y - (X|Z)\begin{pmatrix}\beta \\ u\end{pmatrix}\right]' R^{-1} \left[y - (X|Z)\begin{pmatrix}\beta \\ u\end{pmatrix}\right] - (\beta, u)\begin{pmatrix}0 & 0 \\ 0 & G^{-1}\end{pmatrix}\begin{pmatrix}\beta \\ u\end{pmatrix}$$

jargon: "likelihood" "penalty"

It is easy to show that this is maximized by

$$\begin{bmatrix} \tilde{\beta} \\ \hline \tilde{u} \end{bmatrix}_{R,G} = \left[\begin{pmatrix} X' \\ \hline Z' \end{pmatrix} R^{-1} (X|Z) + \begin{pmatrix} 0 & 0 \\ 0 & G^{-1} \end{pmatrix} \right]^{-1} \begin{pmatrix} X' \\ \hline Z' \end{pmatrix} R^{-1} y$$

- Omit the penalty term, and this is the generalized least squares (GLS) estimate.

- Alternatively: We have an extra piece of information about u : $u \sim N_q(0, G)$

- This is how it affects the point estimates of both u and β

(<u>Note</u>: This is the conditional posterior mean of $\begin{pmatrix} \beta \\ u \end{pmatrix} \Big| R, G$)

Fitted values (BLUP): $[X|Z] \begin{pmatrix} \tilde{\beta} \\ \tilde{u} \end{pmatrix}_{R,G}$

$$[X|Z] \left[\begin{bmatrix} X' \\ \hline Z' \end{bmatrix} R^{-1} (X|Z) + \begin{pmatrix} 0 & 0 \\ 0 & G^{-1} \end{pmatrix} \right]^{-1} \begin{pmatrix} X' \\ \hline Z' \end{pmatrix} R^{-1} y$$

substitute in estimates for R and G to give EBLUPs

I A 2/3 1/21/08

Conventional continued: "Standard errors" for FEs & REs.

Warning: the conventional approach handles this poorly (in my view - RWC seem to agree)

Approach: • Derive $\mathrm{cov}(\tilde{\beta})$ or $\mathrm{cov}\begin{pmatrix}\tilde{\beta}\\ \tilde{u}\end{pmatrix}$ with R, G known, i.e. ignore uncertainty about R, G

Alternatives: • Approximations that, in general, are too hideous to contemplate
• Bayesian analysis (Markov chain Monte Carlo)

$\mathrm{cov}(\tilde{\beta})$: Using the model $y \sim N(X\beta, V)$

for $V = ZGZ' + R$ assumed known,

$\tilde{\beta}$ has MLE $(X'V^{-1}X)^{-1}X'V^{-1}y$ with $\mathrm{cov}(\tilde{\beta}) = (X'V^{-1}X)^{-1}$

so $SE(\hat{\beta}_i) = \sqrt{i^{th} \text{ diagonal entry of } (X'\hat{V}^{-1}X)^{-1}}$

Note!

This is what SAS gives in PROC MIXED.

IA 2/8 1/21/08

$$\begin{bmatrix} \tilde{\beta} \\ \tilde{u} \end{bmatrix}_{R,G} = \left[\left(\frac{X'}{Z'} \right) R^{-1} (X|Z) + \begin{pmatrix} 0 & 0 \\ 0 & G^{-1} \end{pmatrix} \right]^{-1} \left(\frac{X'}{Z'} \right) R^{-1} y$$

has (assuming R, G known)

$$\text{cov} \left(\begin{matrix} \tilde{\beta} \\ \tilde{u} - u \end{matrix} \middle| R, G \right) = \left[\left(\frac{X'}{Z'} \right) R^{-1} (X|Z) + \begin{pmatrix} 0 & 0 \\ 0 & G^{-1} \end{pmatrix} \right]^{-1}$$

which I'll prove in a later lecture (not too useful)

<u>we'll also use</u>

$$\text{cov} \left(\begin{matrix} \tilde{\beta} \\ \tilde{u} \end{matrix} \middle| u \right) = \left[\left(\frac{X'}{Z'} \right) R^{-1} (X|Z) + \begin{pmatrix} 0 & 0 \\ 0 & G^{-1} \end{pmatrix} \right]^{-1} \left(\frac{X'}{Z'} \right) R^{-1} (X|Z) \left[\left(\frac{X'}{Z'} \right) R^{-1} (X|Z) + \begin{pmatrix} 0 & 0 \\ 0 & G^{-1} \end{pmatrix} \right]^{-1}$$

RWC notation $= (C'R^{-1}C + B)^{-1} \; C'R^{-1}C \; [C'R^{-1}C + B]^{-1}$

(Derivation is trivial)

RWC extend an idea from linear regression and smoothers for scatter plots

Linear Regression: $\hat{y} = X(X'X)^{-}X'y \equiv Hy$

df in fit is $tr(H) = rank(X)$

Linear Smoothers: For a linear smoother S_λ,

$\hat{y} = S_\lambda y$ Note! S_λ is a function of λ; this is for given λ.

Analogy: df in fit is $tr(S_\lambda)$ (e.g. Hastie & Tibshirani, Gen'l'd Additive Mod)

$\varepsilon \sim N(0, R)$
$u \sim N(0, G)$

Extend to Mixed LINEAR Model $y = X\beta + Zu + \varepsilon$

$$\hat{y} = \left\{ [X|Z] \left[\begin{bmatrix} X' \\ \hline Z' \end{bmatrix} R^{-1} (X|Z) + \begin{bmatrix} 0 & 0 \\ 0 & G^{-1} \end{bmatrix} \right]^{-1} \begin{bmatrix} X' \\ \hline Z' \end{bmatrix} R^{-1} \right\} y$$

df in fit is $tr\{\ \}$, a function of R, G.

a big, square matrix

I B 3/2 1/28/08

RWC 3.13, 8.3 – this is a bit more general

Apply this to penalized splines:

The fitted (β, u) minimize

$$\frac{1}{\sigma_\varepsilon^2}\left((y - \underbrace{X\beta - Zu}_{\text{p-spline fit}})'(y - X\beta - Zu) + \frac{\sigma_\varepsilon^2}{\sigma_u^2}\underbrace{u'u}_{\text{p-spline coefficients}}\right).$$

σ_ε^2 by itself does not affect this; $\frac{\sigma_\varepsilon^2}{\sigma_u^2}$ does

The solution is (for fixed $\sigma_\varepsilon^2/\sigma_u^2$)

$$\begin{pmatrix}\tilde\beta \\ \tilde u\end{pmatrix}_{\sigma_\varepsilon^2/\sigma_u^2} = \left[\begin{pmatrix}X' \\ Z'\end{pmatrix}(X|Z) + \frac{\sigma_\varepsilon^2}{\sigma_u^2}\begin{pmatrix}0_{p+1} & 0 \\ 0 & I_K\end{pmatrix}\right]^{-1}\begin{pmatrix}X' \\ Z'\end{pmatrix} y$$

Fitted values for observed $\underset{\sim}{x}$

$$\tilde y = \underbrace{(X|Z)\left[\begin{pmatrix}X' \\ Z'\end{pmatrix}(X|Z) + \frac{\sigma_\varepsilon^2}{\sigma_u^2}\begin{pmatrix}0_{p+1} & 0 \\ 0 & I_K\end{pmatrix}\right]^{-1}\begin{pmatrix}X' \\ Z'\end{pmatrix}}_{\text{symmetric square matrix } S, \text{ depends on } \frac{\sigma_\varepsilon^2}{\sigma_u^2}} y$$

DF in fit = tr(S), a function of $\frac{\sigma_\varepsilon^2}{\sigma_u^2}$

Fitted values at a new x, x_0:

Define $X_x = [1, x_0, \dots, x_0^p]$

$Z_x = [(x_0 - \kappa_1)_+^p \quad \cdots \quad (x_0 - \kappa_K)_+^p]$

$$\tilde{f}(x_0) = X_x \tilde{\beta} + Z_x \tilde{u}$$

"$\sim$" means taking σ_ε^2, σ_u^2 as given

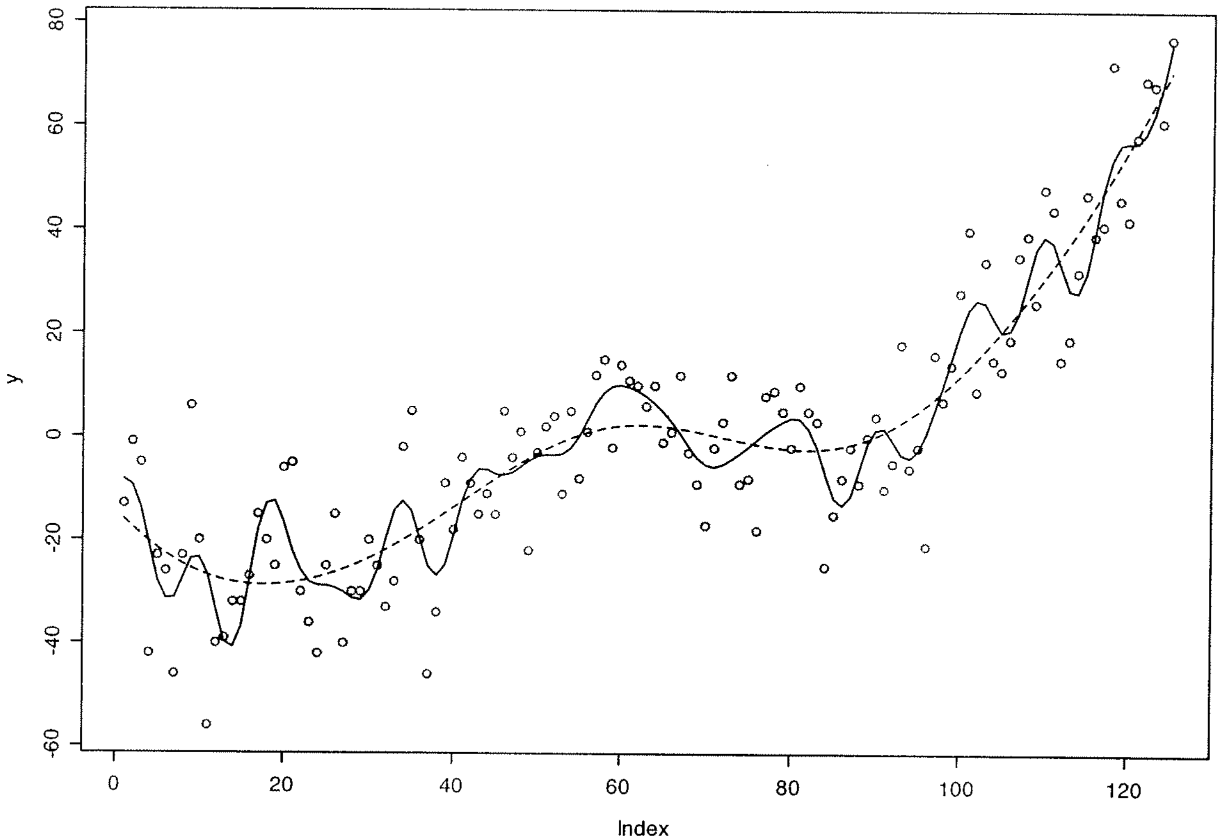
replace $\sigma_\varepsilon^2, \sigma_u^2$ by $\hat{\sigma}_\varepsilon^2, \hat{\sigma}_u^2 \Rightarrow$ replace "$\sim$" with "$\wedge$"

$$= \underbrace{(X_x | Z_x)\left[\begin{pmatrix} X' \\ \hline Z' \end{pmatrix}(X|Z) + \frac{\sigma_\varepsilon^2}{\sigma_u^2}\begin{pmatrix} 0_{p+1} & 0 \\ 0 & I_K \end{pmatrix}\right]^{-1}\begin{pmatrix} X' \\ \hline Z' \end{pmatrix}} y$$

$= \ell_x' y$ for a column vector ℓ_x

so, given $\sigma_\varepsilon^2/\sigma_u^2$, $\tilde{f}(x_0)$ is a linear function of the observed y.

VAM pg 2/26 7/3/08



Here are the two fits: ———— unpenalized
- - - - penalized.

VAM ~~~ 2/3a 2/3/08

- For penalized splines, we don't care about doing inference on β and u
- This is fortunate because the collinearity is <u>terrible</u>
- For the Global Mean Surface Temp data:
  - quadratic basis, 30 knots

	Unpenalized		Penalized	
	Estimate	Std. Error	Estimate	Std. Error
Intercept	-5924.8	10661.9	85.35	131.16
Linear	-6930.8	12906.3	176.40	174.82
Quadratic	-2029.7	3903.7	68.46	58.69
Spline terms				
H4	4387.7	5258.6	-0.53	30.78
H5	-4688.3	2817.7	-3.48	30.69
H6	5679.9	2519.4	-6.62	30.36
H7	-6310.3	2465.8	-8.41	29.79
H8	4015.4	2455.8	-5.68	29.15
H9	-1305.9	2453.9	-3.93	28.63
H10	1662.1	2453.5	-5.80	28.32
H11	-3786.6	2453.5	-9.19	28.18
H12	4810.5	2453.5	-12.12	28.14
H13	-3868.4	2453.4	-15.85	28.14
H14	1859.5	2453.4	-16.96	28.13
H15	-704.7	2453.4	-15.37	28.13
H16	921.5	2453.4	-11.44	28.12
H17	-1280.9	2453.4	-4.55	28.11
H18	430.3	2453.4	5.42	28.11
H19	19.8	2453.4	14.97	28.11
H20	776.0	2453.4	20.25	28.11
H21	-500.1	2453.4	20.77	28.12
H22	-222.0	2453.4	19.35	28.12
H23	-977.6	2453.4	17.33	28.13
H24	3185.8	2453.4	12.59	28.14
H25	-3738.0	2453.4	4.25	28.14
H26	2975.1	2453.5	-3.46	28.15
H27	-1127.1	2453.5	-8.83	28.23
H28	-1763.1	2453.5	-8.76	28.45
H29	3538.6	2453.6	-4.15	28.87
H30	-4306.1	2454.3	0.20	29.47
H31	5013.2	2458.1	2.69	30.10
H32	-4427.8	2482.3	1.58	30.56
H33	3008.7	2650.0	0.57	30.76

$F = 17.3$ on $(32, 92)$ DF
$P < 2 \times 10^{-16}$

$|\text{coef}| < SE$ for all spline terms

VAM np 2/3ai 7/3/08

- Instead, we care about the <u>fitted smooth</u>, fitted values at observed x_i or a new x_0.
- This seems simple, but at this point conventional (non-Bayesian) theory begins to trip over itself.

<u>The issue:</u> - In computing SEs for $\tilde{f}(x_0)$, should you condition on the true, unknown $\underset{\sim}{u}$?

<u>Yes</u>: The model is $\underset{\sim}{y} = \underset{\sim}{f} + \varepsilon$ for $f = X\beta + Zu$,

- f is a fixed but unknown smooth function
- $u \sim N(0, \sigma_u^2 I_K)$ is just a device to shrink the u and give a smooth fit
- <u>In truth</u>, u is not random, so we should treat it as fixed and unknown, like any other parameter

<u>No</u>:
(a) $\begin{cases} \text{- } u \text{ is a random draw} \\ \text{- In frequentist theory, we don't condition on} \\ \quad \text{randomly-drawn things} \end{cases}$

(b): $\begin{cases} \text{- conditional on } u,\ \tilde{f}(x_0|u) \text{ is biased} \\ \text{- unconditionally, } \tilde{f}(x_0) \text{ is unbiased} \\ \text{- To account for bias, we must use the} \\ \quad \text{unconditional approach.} \end{cases}$

VAM nyp 2136 7/3/08

- This is strange and obscure (to me) and arises because we are using the random effects <u>form</u> for effects that aren't <u>old-fashioned</u> random effects
- NOTE: Bayes theory does not have this problem.

Let's have a look at this and see.

Define $C = (X \mid Z)$ $D = \begin{pmatrix} 0_{p+1} & 0 \\ 0 & I_K \end{pmatrix}$ $C_{x_0} = (X_{x_0} \mid Z_{x_0})$

Then $\begin{pmatrix} \tilde{\beta} \\ \tilde{u} \end{pmatrix} = \left(C'C + \frac{\sigma_\varepsilon^2}{\sigma_u^2} D\right)^{-1} C'y$, $\tilde{f}(x_0) = C_{x_0} \begin{pmatrix} \tilde{\beta} \\ \tilde{u} \end{pmatrix}$

Thus $var(\tilde{f}(x_0) \mid u) = C_{x_0} \operatorname{cov}\left(\left[\begin{matrix} \tilde{\beta} \\ \tilde{u} \end{matrix} \middle| u\right]\right) C'_{x_0}$

from earlier slide, $\operatorname{cov}\left(\left[\begin{matrix} \tilde{\beta} \\ \tilde{u} \end{matrix} \middle| u\right]\right) = \sigma_\varepsilon^2 \left(C'C + \frac{\sigma_\varepsilon^2}{\sigma_u^2} D\right)^{-1} C'C \left(C'C + \frac{\sigma_\varepsilon^2}{\sigma_u^2} D\right)^{-1}$

so

$$\widehat{SD}\{\hat{f}(x_0) \mid u\} = \hat{\sigma}_\varepsilon \sqrt{C_{x_0}\left(C'C + \frac{\hat{\sigma}_\varepsilon^2}{\hat{\sigma}_u^2} D\right)^{-1} C'C \left(C'C + \frac{\hat{\sigma}_\varepsilon^2}{\hat{\sigma}_u^2} D\right)^{-1} C'_{x_0}}$$

Note all the hats: - in practice, you estimate $\sigma_\varepsilon^2, \sigma_u^2$ by $\hat{\sigma}_\varepsilon^2, \hat{\sigma}_u^2$
- then forget they're estimates, plug them in.

This is the <u>conditional</u> SE, treating u as fixed & unknown

VAM rep 2/3c 7/3/08

BUT! $\tilde{f}(x)$ is a biased estimate of $f(x) = X\beta + Zu$, treating u as fixed but unknown.

$$E\left(\tilde{f}(x_0) - f(x_0) \mid u\right) = E\left(C_x\left(\begin{pmatrix}\tilde{\beta}\\ \tilde{u}\end{pmatrix} - \begin{pmatrix}\beta\\ u\end{pmatrix}\right) \Big| \, u\right)$$

$$\begin{pmatrix}\tilde{\beta}\\ \tilde{u}\end{pmatrix} = \left(C'C + \frac{\sigma_\varepsilon^2}{\sigma_u^2}D\right)^{-1} C'y \qquad \text{recall } C = (X \mid Z)$$

$$D = \begin{pmatrix}0 & 0\\ 0 & I_K\end{pmatrix}$$

and $y = C\begin{pmatrix}\beta\\ u\end{pmatrix} + \varepsilon$

so
$$E\left(\begin{pmatrix}\tilde{\beta}\\ \tilde{u}\end{pmatrix} - \begin{pmatrix}\beta\\ u\end{pmatrix} \Big| \, u\right) = \left(C'C + \frac{\sigma_\varepsilon^2}{\sigma_u^2}D\right)^{-1} C'C\begin{pmatrix}\beta\\ u\end{pmatrix} - \begin{pmatrix}\beta\\ u\end{pmatrix}$$

$$= \underbrace{\left(C'C + \frac{\sigma_\varepsilon^2}{\sigma_u^2}D\right)^{-1}\left(C'C + \frac{\sigma_\varepsilon^2}{\sigma_u^2}D\right)\begin{pmatrix}\beta\\ u\end{pmatrix} - \begin{pmatrix}\beta\\ u\end{pmatrix}} - \left(C'C + \frac{\sigma_\varepsilon^2}{\sigma_u^2}D\right)^{-1}\left(\frac{\sigma_\varepsilon^2}{\sigma_u^2}D\right)\begin{pmatrix}\beta\\ u\end{pmatrix}$$

$$= 0 - \frac{\sigma_\varepsilon^2}{\sigma_u^2}\left(C'C + \frac{\sigma_\varepsilon^2}{\sigma_u^2}D\right)^{-1}\begin{pmatrix}0\\ u\end{pmatrix} \qquad D\begin{pmatrix}\beta\\ u\end{pmatrix} = \begin{pmatrix}0\\ u\end{pmatrix}$$

$E(\text{bias in fit} \mid u) = C_x(\text{bias in } \tilde{\beta}, \tilde{u})$

This is obscure in general; in practice, smoothers flatten peaks, fill valleys, and round corners.

Note: $E(\text{bias}) = 0$ because $E(u) = 0$

VAM ~p213d 7/3/08

Because of this bias

$$\tilde{f}(x_0|u) \pm 1.96\ SD(\tilde{f}(x_0)|u)$$

is <u>only</u> an (approx) 95% CI for $E(\tilde{f}(x_0)|u)$,

<u>not</u> for $f(x_0)$

<u>Naive (?) reaction 1:</u> So estimate the bias using $\hat{\sigma}_\varepsilon^2/\hat{\sigma}_u^2$, $\hat{u}$, and adjust the center of the confidence interval.

(Not discussed in RWC) (see next page "rp 2/3ei")

<u>Naive (?) reaction 2:</u>

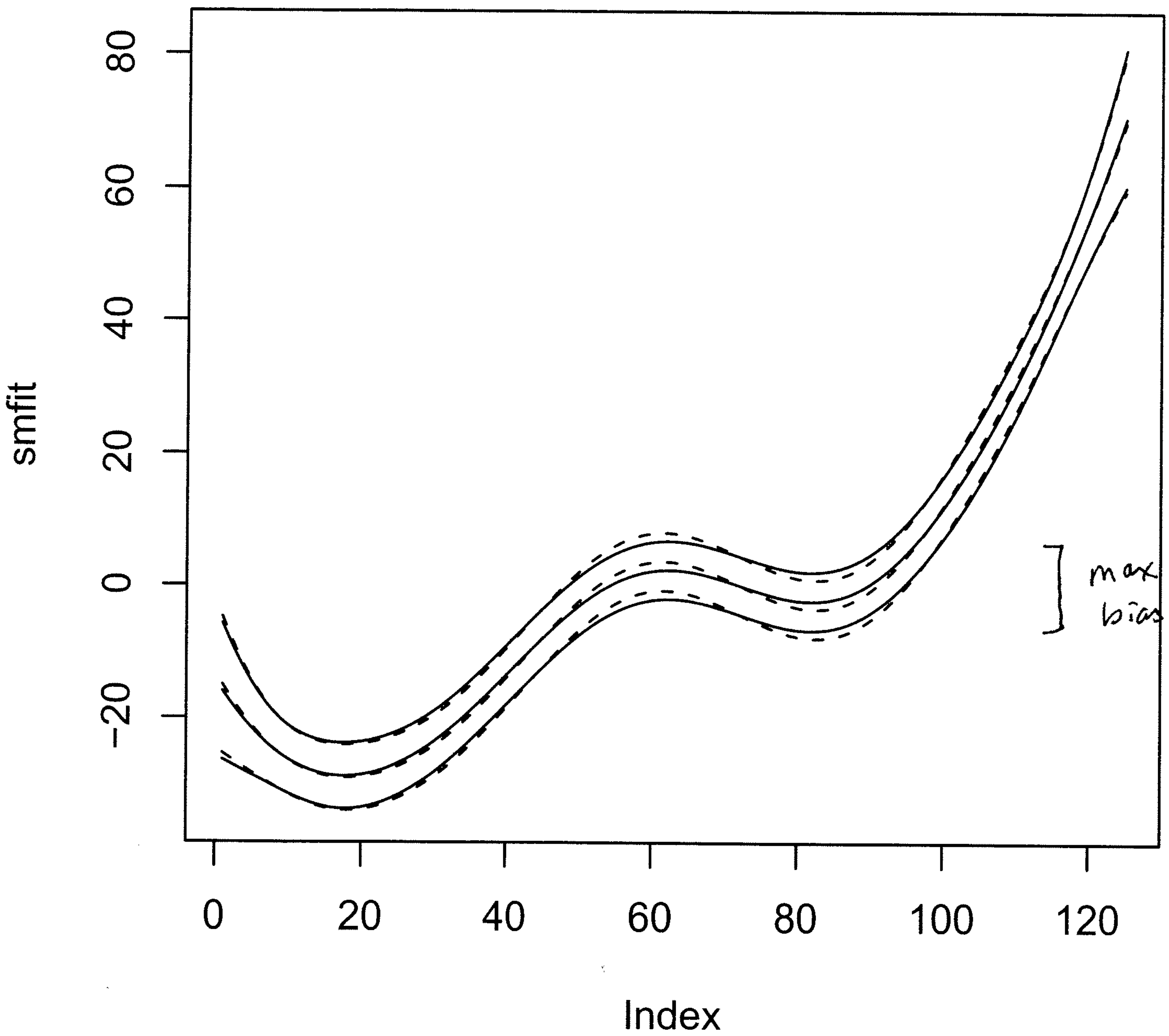
so replace $SD(\tilde{f}(x_0)|u)$ by $\sqrt{[SD(\tilde{f}(x_0)|u)]^2 + (\text{bias})^2}$ in the CI above. (see rp 2/3eii)

- RWC bottom p. 139 say to do this <u>AND</u> take the expectation over $u \sim N(0, \sigma_u^2 I_K)$
- I don't understand the "AND" part.

VAM rp 2/3e 7/3/08

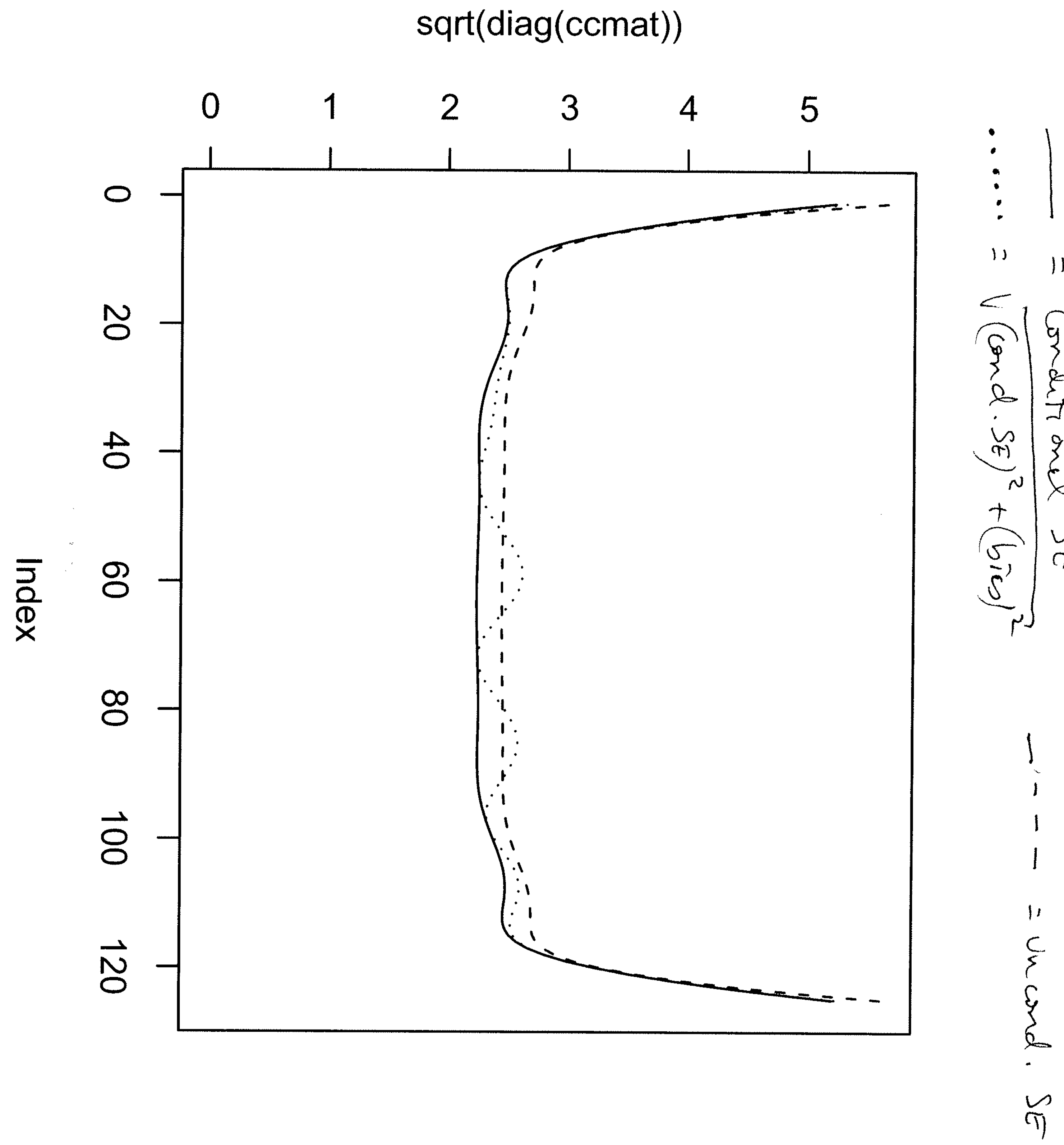
—— = $\tilde{f}$, $\tilde{f} \pm 1.96 \times$ conditional SE

- - - - = $\tilde{f} - \widehat{bias}$, $\tilde{f} - \widehat{bias} \pm 1.96 \times$ conditional SE



max shift from bias is about ± 1.5

VAM rep 2/3ei 7/7/08



VAM rp 243eii 7/4/08

If you take expectations,

$$E(\tilde{f}(x_0) - f(x_0)) = 0$$

and $$MSE(\tilde{f}(x_0)) = \operatorname{var}(\tilde{f}(x_0) - f(x_0))$$

$$= C_{x_0} \operatorname{Cov}\begin{pmatrix}\tilde{\beta} \\ \tilde{u}-u\end{pmatrix} C_{x_0}' = C_{x_0}\left(C'C + \frac{\sigma_\varepsilon^2}{\sigma_u^2} D\right)^{-1} C_{x_0}$$

so the unconditional SE is

$$\widehat{SD}(\hat{f}(x_0)) = \hat{\sigma}_\varepsilon \sqrt{C_{x_0}\left(C'C + \frac{\sigma_\varepsilon^2}{\sigma_u^2} D\right)^{-1} C_{x_0}}$$

This is bigger than

$$\widehat{SD}(\hat{f}(x_0) \mid u) = \hat{\sigma}_\varepsilon \sqrt{C_{x_0}\underbrace{\left(C'C + \frac{\sigma_\varepsilon^2}{\sigma_u^2} D\right)^{-1} C'C \left(C'C + \frac{\sigma_\varepsilon^2}{\sigma_u^2} D\right)^{-1}}_{\text{extra part}} C_{x_0}}$$

(see pp 213eii, pp 213fi)

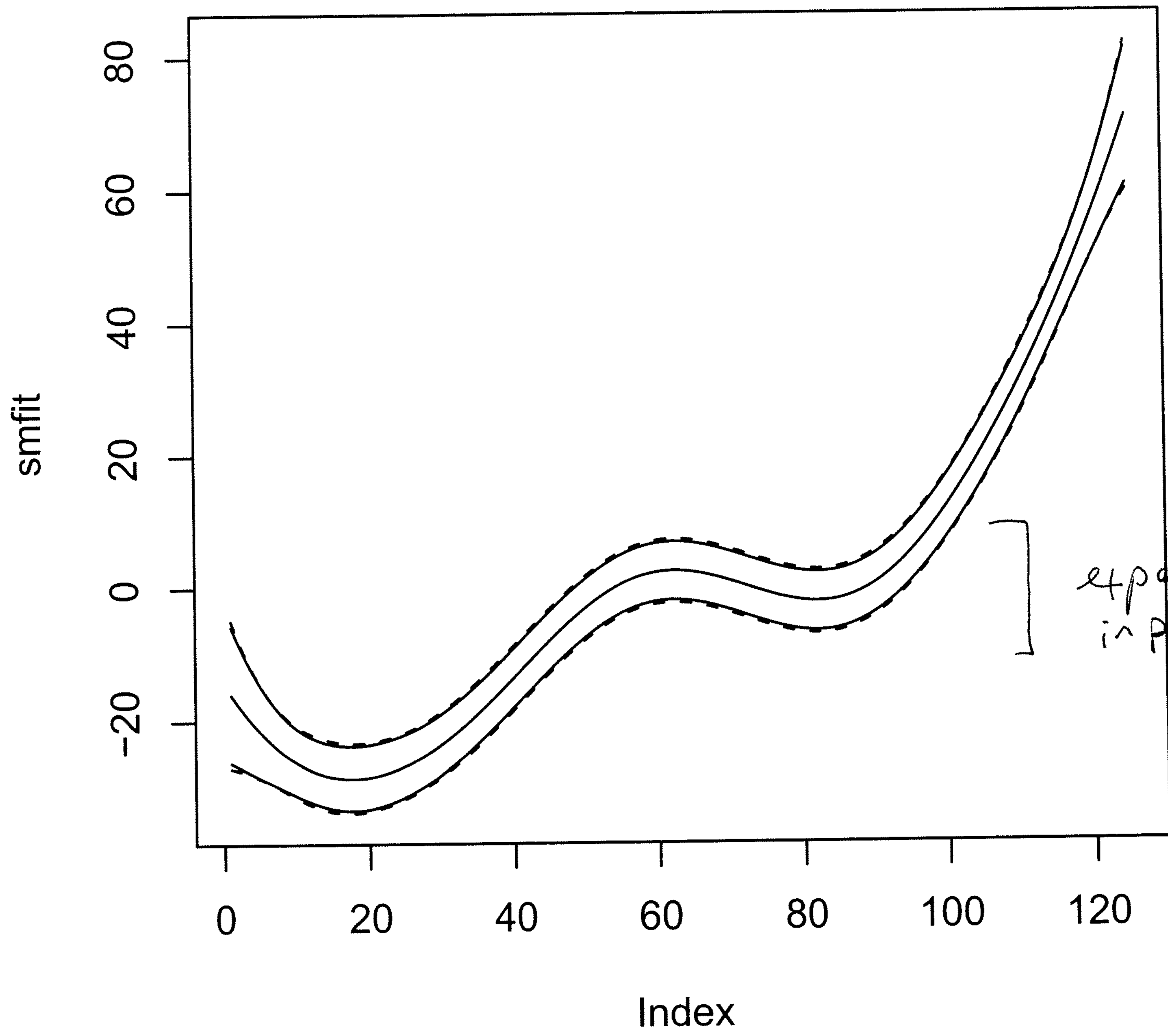
<u>Claim</u>:- Nychka (1988) $\widehat{SD}(\tilde{f}(x_0))$ give coverage 95%
averaging over x_0

x_0

<u>RWC</u>:- But coverage is a lot lower in regions of high curvature & a lot higher in regions that are flat.

—— = $\ddot{f}$, $\ddot{f} \pm 1.96$ conditional SE

- - - = $\hat{f} \pm 1.96$ unconditional SE



expands $\approx \pm 0.3$-0.4 in peak parts.

JAM [illegible]3fi 7/7/08

Doing statistics : Bayes

- Straight forward application of Bayes theory
- Re complication in conventional:
  - in Bayes, all unknowns are treated as random variables.

  - Posterior SD of $f(x_0)$ corresponds to unconditional SE in conventional analysis.

(NOTE: RWC p. 141 : Local bias is possible always because of oversmoothing, and coverage can be poor anyway even if the unconditional SE is used.)

More analysis stuff in RWC

- Simultaneous confidence intervals:
  - conventional: RWC sec 6.5 (simulation)
  - Bayes: expand pointwise intervals until 95% of MCMC draw curves are contained.

- Test vs. parametric fit, i.e. $\sigma_u^2 = 0$ vs $\sigma_u^2 > 0$
  - conventional: - RLRT, F-tests
    (RWC 6.6) - RWC recommend simulating P-values
  - Bays: - DIC; prior for σ_u^2 with a positive probability on zero

- Testing for no effect: $\beta_1 = \sigma_u^2 = 0$ vs. $\beta_1 \neq 0$ or $\sigma_u^2 > 0$
  - conventional (RWC 6.7): F-tests, simulate
  - Bays: - ditto

IC2/5 2/3/08