ANALYSIS OF LONGITUDINAL DATA USING CUBIC SMOOTHING SPLINES

Tapio Nummi

Department of Mathematics, Statistics and Philosophy

33014 University of Tampere

Finland

1. Spline smoothing

Suppose that our aim is to model

$$y_i = d(x_i) + \epsilon_i, \ \ i = 1, \ldots, n,$$

where d is a smooth function and ϵ_i are iid with $E(\epsilon_i) = 0$ and $Var(\epsilon_i) = \sigma_\epsilon^2$.

The linear spline estimator is

$$d(x_i) = \beta_0 + \beta_1 x_i + \sum_{k=1}^{K} u_k (x - \kappa_k)_+,$$

$$(x - \kappa_k)_+ = \begin{cases} 0, & x \le \kappa_k \\ x - \kappa_k, & x > \kappa_k \end{cases}$$

and $\kappa_1, \ldots, \kappa_K$ are $knots$.

The curve d is now modeled by piecewise line segments tied together at knots $\kappa_1, \ldots, \kappa_K$.

We can generalize the above equation to a piecewise polynomial of degree p, but the most common choices in practice are quadratic ($p = 2$) and cubic ($p = 3$) splines.

For cubic splines we have

$$d(x_i; \boldsymbol{\beta}; \boldsymbol{u}) = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3 \\ + \sum_{k=1}^{K} u_k (x - \kappa_k)_+^3,$$

where $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3)'$, $\boldsymbol{u} = (u_1, \ldots, u_k)'$ and 1, x, x^2, x^3, $(x - \kappa_1)_+^3, \ldots, (x - \kappa_K)_+^3$ are called basis functions. Other possible choices of **basis functions** include B-splines, wavelet, Fourier Series and polynomial bases etc.

A **natural cubic spline** is obtained by assuming that the function is linear beyond the boundary knots.

The number (K) and location of knots $\kappa_1, \ldots, \kappa_K$ must be specified in advance.

Coefficients $\boldsymbol{\beta}$ and $\boldsymbol{u}$ can be estimated using standard least squares procedures.

However, in some cases the estimated curve tends to be a very rough estimate.

Our approach is to apply smoothing splines, where the smoothing is controlled by a smoothing parameter α.

Smoothing splines have a knot at each unique value of x and the fitting is carried out by least squares with a roughness penalty term.

2. Penalized smoothing

If $x_1, \ldots, x_n$ are points in $[a, b]$ satisfying $a < x_1, \ldots, x_n < b$ the penalized sum of squares (PSS) is given as

$$\sum_{i=1}^{n} \{y_i - d(x_i)\}^2 + \alpha \int_a^b \{d''(x)\}^2 dx,$$

where

$$\alpha \int_a^b \{d''(x)\}^2 dx$$

is the roughness penalty (RP) term with $\alpha > 0$.

Note that here α represents the rate of exchange between residual error and local variation.

If α is very large the main component of PSS will be RP and the estimated curve will be very smooth.

If α is relatively small the estimated curve will track the data points very closely.

If we define a non-negative definite matrix

$$\boldsymbol{K} = \boldsymbol{\nabla}\boldsymbol{\Delta}^{-1}\boldsymbol{\nabla}',$$

where $\boldsymbol{\nabla}$ and $\boldsymbol{\Delta}$ are certain functions of the points $x_1, \ldots, x_n$ the PSS becomes as

$$PSS(\boldsymbol{K}) = (\boldsymbol{y} - \boldsymbol{d})'(\boldsymbol{y} - \boldsymbol{d}) + \alpha \boldsymbol{d}'\boldsymbol{K}\boldsymbol{d}$$

and its minimum is obtained at

$$\widehat{\boldsymbol{d}} = (\boldsymbol{I} + \alpha \boldsymbol{K})^{-1}\boldsymbol{y}.$$

It can be shown (e.g. Green and Silverman, 1994) that $\widehat{\boldsymbol{d}}$ is a natural cubic smoothing with knots at the points $x_1, \ldots, x_n$.

Note that the special form $\widehat{\boldsymbol{d}}$ follows from the chosen RP term

$$\alpha \int_a^b \{d''(x)\}^2 dx.$$

If we, for example, would use a discrete approximation

$$\mu_{i+1} - 2\mu_i + \mu_{i-1}$$

of the second derivative the PSS would be (Demidenko, 2004)

$$PSS(\boldsymbol{QQ}') = (\boldsymbol{y} - \boldsymbol{d})'(\boldsymbol{y} - \boldsymbol{d}) + \alpha \boldsymbol{d}'\boldsymbol{QQ}'\boldsymbol{d},$$

where ($n = 6$)

$$\boldsymbol{Q} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ -2 & 1 & 0 & 0 \\ 1 & -2 & 1 & 0 \\ 0 & 1 & -2 & 1 \\ 0 & 0 & 1 & -2 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Then the minimizer is

$$\tilde{\boldsymbol{d}} = (\boldsymbol{I} + \alpha \boldsymbol{QQ}')^{-1}\boldsymbol{y}.$$

Note that for fixed α the spline fit

$$\widehat{\boldsymbol{d}} = (\boldsymbol{I} + \alpha \boldsymbol{K})^{-1}\boldsymbol{y} = \boldsymbol{S}_\alpha \boldsymbol{y}$$

is linear in $\boldsymbol{y}$ and the matrix $\boldsymbol{S}_\alpha$ is known as the *smoother matrix*.

The smoother matrix $\boldsymbol{S}_\alpha$ has many interesting properties discussed e.g. in Hastie, Tibshirami and Friedman (2001), but here I briefly mention only the following :

1. Choosing the smoothing parameter:

   $$CV(\alpha) = \sum_{i=1}^{n} \left(\frac{y_i - \widehat{d}_\alpha(x_i)}{1 - S_\alpha(i,i)}\right)^2,$$

   where $S_\alpha(i,i)$ are diagonal elements of $\boldsymbol{S}_\alpha$.

2. Estimation of the effective degrees of freedom

$$df_\alpha = tr(\boldsymbol{S}_\alpha).$$

This can be compared to matrix

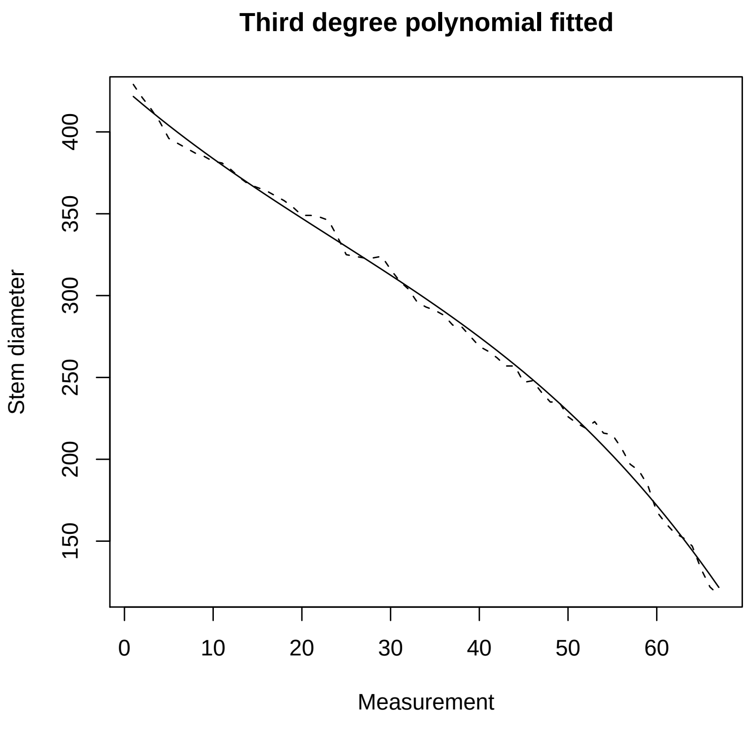
$$\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$$

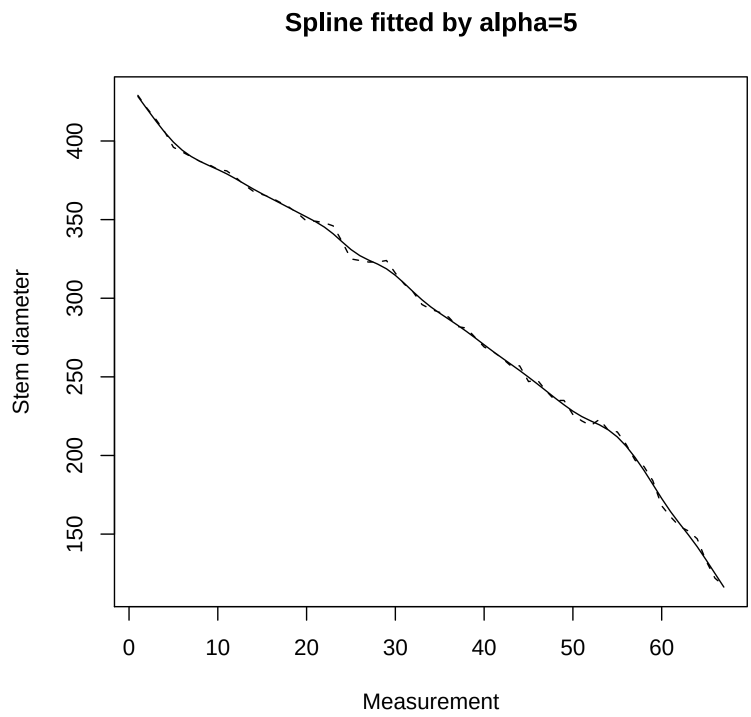
in regression analysis (or in regression splines) in a sense that

$$tr(\boldsymbol{H})$$

gives the number of estimated parameters (or the number of basis functions utilized).

Example: Stem curve model - modelling the degrease of stem diameter as a function stem height.





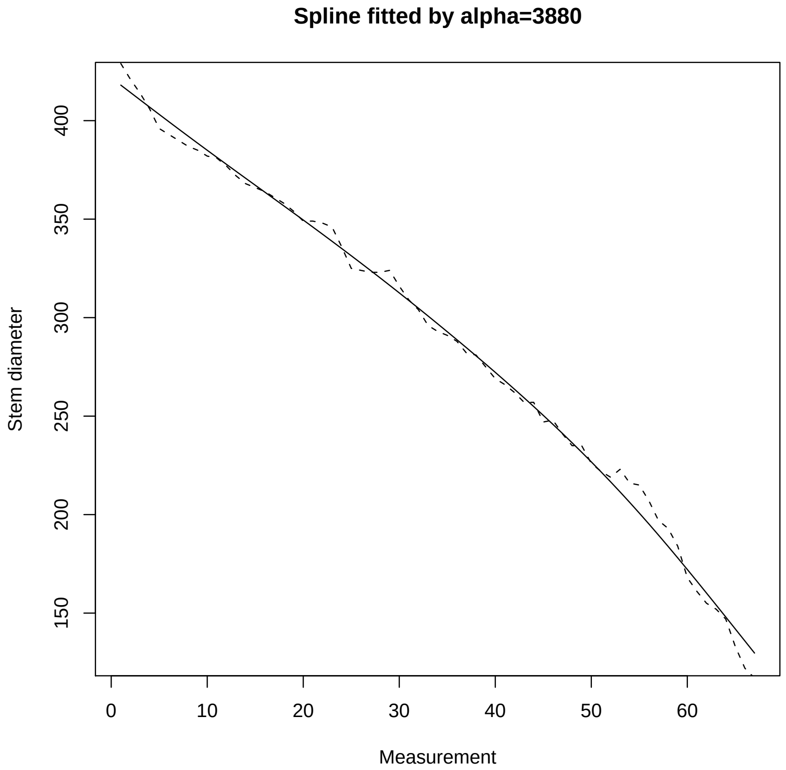
The effective number of degrees of freedom

$df_\alpha = tr(S_{\alpha=5}) = 16.79628.$

Note that if

$$\alpha \to 0, \; df_\alpha \to n$$

$$\alpha \to \infty, \; df_\alpha \to 2$$



Since $df_\alpha = tr(S_\alpha)$ is monotone in α, we can invert the relationship and specify α by fixing df. For $df = 4$ this gives $\alpha = 3880$.

This yields to model selection with different values for df, where more traditional criteria developed for regression models maybe used.

3. Connection to mixed models

If we let

$$\boldsymbol{X} = [\mathbf{1}, \boldsymbol{x}],$$

where $\boldsymbol{x} = (x_1, \ldots, x_n)'$ and by the special form of $\boldsymbol{\nabla}$ we note that

$$\boldsymbol{X}'\boldsymbol{\nabla} = \mathbf{0}$$

and

$$(\boldsymbol{I}+\alpha\boldsymbol{K})^{-1} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'+\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z}+\alpha\Delta^{-1})\boldsymbol{Z}',$$

where $\boldsymbol{Z} = \boldsymbol{\nabla}(\boldsymbol{\nabla}'\boldsymbol{\nabla})^{-1}$.

Then the solution of $PSS(\boldsymbol{K})$ can be written as

$$\widehat{\boldsymbol{d}} = \boldsymbol{X}\widehat{\boldsymbol{\beta}} + \boldsymbol{Z}\widehat{\boldsymbol{u}},$$

where

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}\boldsymbol{y}$$

and

$$\widehat{\boldsymbol{u}} = (\boldsymbol{Z}'\boldsymbol{Z} + \alpha\boldsymbol{\Delta}^{-1})^{-1}\boldsymbol{Z}'\boldsymbol{y}.$$

These estimates can be seen as (BLUP) solutions of the mixed model

$$\boldsymbol{y} = \boldsymbol{X}\beta + \boldsymbol{Z}\boldsymbol{u} + \boldsymbol{\epsilon},$$

where $\boldsymbol{X}$ and $\boldsymbol{Z}$ are defined before and

$$\boldsymbol{u} \sim N(\boldsymbol{0}, \sigma_u^2\boldsymbol{\Delta}) \text{ and } \boldsymbol{\epsilon} \sim N(\boldsymbol{0}, \sigma^2\boldsymbol{I})$$

with smoothing parameter as a variance ratio $\alpha = \frac{\sigma^2}{\sigma_u^2}$.

Note that we may always rewrite

$$\boldsymbol{y} = \boldsymbol{X}\beta + \boldsymbol{Z}_*\boldsymbol{u}_* + \boldsymbol{\epsilon},$$

where $\boldsymbol{Z}_* = \boldsymbol{Z}\Delta^{1/2}$ and $\boldsymbol{u}_* = \Delta^{-1/2}\boldsymbol{u}$ with

$$\boldsymbol{u}_* \sim N(\mathbf{0}, \sigma_u^2\boldsymbol{I}) \text{ and } \boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2\boldsymbol{I}).$$

We can now use standard statistical software for parameter estimation (e.g. LME in R or Proc Mixed in SAS).

4. Application 1: Harvesting

4.1 Introduction Forest Harvesting

- The general objective in harvesting is to maximize the value of timber obtained for further processing.

- The optimization requires that several phases in the production chain are successfully combined.

- Trees are converted into smaller logs immediately at harvest (Nordic countries).

- A great portion of the annual cut in Scandinavia is nowadays accomplished by computerized forest harvesters.

- Optimization of crosscutting is based:

  a) on the assessment of the stem curve (degrease in diameter) and

  b) on the given targets (price, volume, demand etc.)

Some references

Kivinen V.-P., Uusitalo, J. & Nummi, T. (2005). Comparison of four measures designed between the demand and output distributions of logs. *Canadian Journal of Forest Research* 35, pp. 693-702.

Koskela, L., Nummi, T., Wentzel, S. and Kivinen, V. (2006), On the Analysis of Cubic Smoothing Spline-Base Stem Curve Prediction for Forest Harvesters, *Canadian Journal of Forest Research*, 36, pp. 2909-2920.

Liski, E. & Nummi, T. (1995). Prediction of tree stems to improve efficiency in automatized harvesting of forest. *Scandinavian Journal of Statistics* 22, pp. 255-269.

Nordhausen and Nummi (2006). Estimation of the diameter distribution of a stand marked for cutting using finite mixtures. Accepted to *Canadian Journal of Forest Research*.

Nummi, Tapio and Möttönen, Jyrki (2004), Estimation and Prediction for Low-Degree Polynomial Models under Measurement Errors with an Application to Forest Harvester, *Journal of the*

Royal Statistical Society, Series C, vol. 53, part 3, pages 495-505.

Nummi, Tapio and Möttönen, Jyrki (2004), Prediction of Stem Measurements of Scots Pine, *Journal of Applied Statistics*, Vol. 31, No. 1, 105-114.

Nummi, T., Sinha, B. and Koskela, L. (2005), Statistical properties of the apportionment degree and alternative measures in bucking outcome, *Revista Investigacion Operational*, Vol. 26, No. 3, pp. 1-7 .

Sinha, B., Koskela, L. and Nummi, T. (2005), On a Family of Apportionment Indices and Its Limiting Properties, *IAPQR Transactions*, Vol. 30, No. 2, pp. 65-85.

Sinha B., Koskela, L. and Nummi, T. (2005), On some statistical properties of the apportionment index, *Revista Investigacion Operational*, Vol 26, No. 2, pp. 169-179.

Uusitalo, J., Puustelli, A., Kivinen, V-P, Nummi, T. and Sinha, B.K. (2006). Bayesian estimation of diameter distribution during harvesting, *Silva Fennica*, 40(4), pp. 663-671.

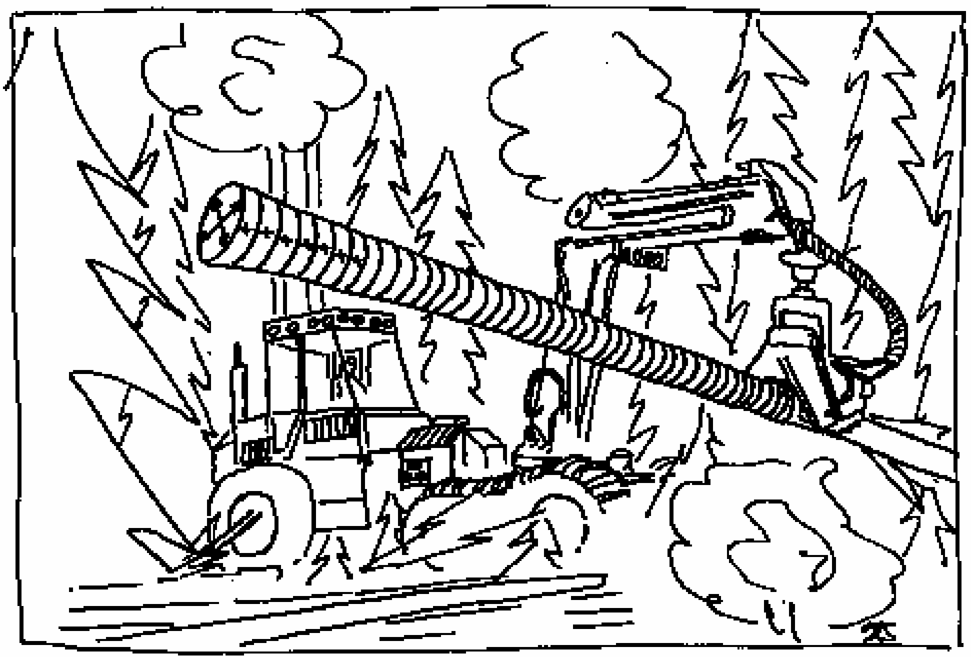
4.2 Prediction of stem curves

- If the whole stem curve were known we may apply techniques discussed e.g. in Nasberg (1985) to find the optimal cutting patterns of a stem.

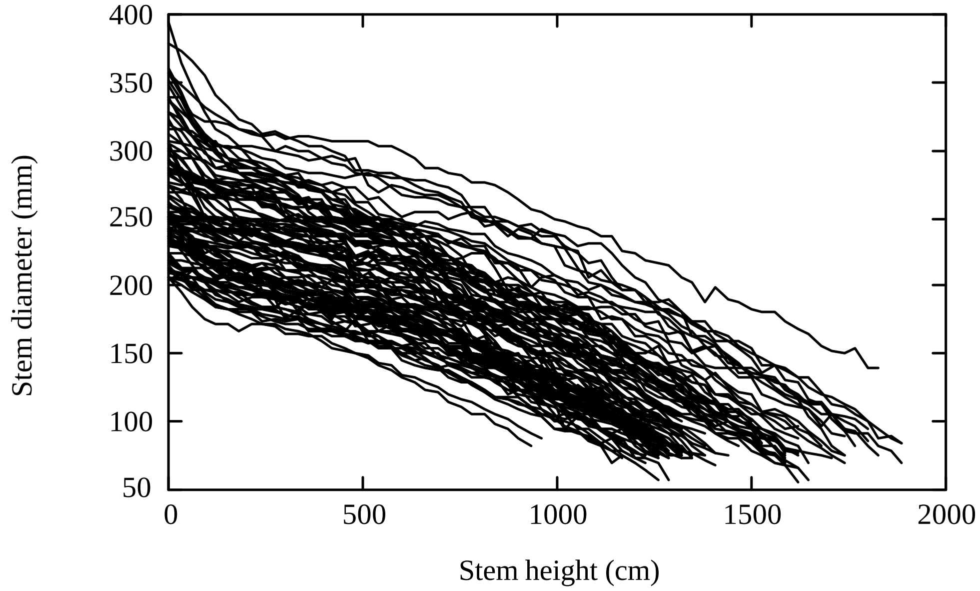
- In practise stem is only partly known and we must compensate the unknown part of the stem by predictions.

- In the first cutting decision only about 4 meters of the stem known.

Figure: A forest harvester at work



Stem curves for 100 Spruces in Finland



- Factors affecting to form of stem curve (site type, climate, genetical factors etc.) difficult or impossible to measure in a harvesting situation.

- In forestry stem curve models are often presented for relative heights (e.g Laasasenaho, 1982 and Kozak, 1988).

  However:

  - height is unknown for a harvester.

  - height has influence to the form of the complete curve.

- if measurement errors $\rightarrow$ model parameters can not be unbiasedly estimated by standard methods (see e.g. Nummi and Möttönen, 2004a).

- do not account for individual variation.

- Low degree polynomial models (e.g. Liski and Nummi, 1995)

  - The stem curve model (Spruce)

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

  fits well to individual curves (butt measurements dropped).

- Great variation of the estimates $\widehat{\beta}_0$, $\widehat{\beta}_1$ and $\widehat{\beta}_2$ between different stems.

- The second degree mixed model

  $$y_{ij} = (b_{i0}+\beta_0)+(b_{i1}+\beta_1)x_j+\beta_2 x_j^2+\epsilon_{ij}$$

  provides good predictions for Spruce stem data.

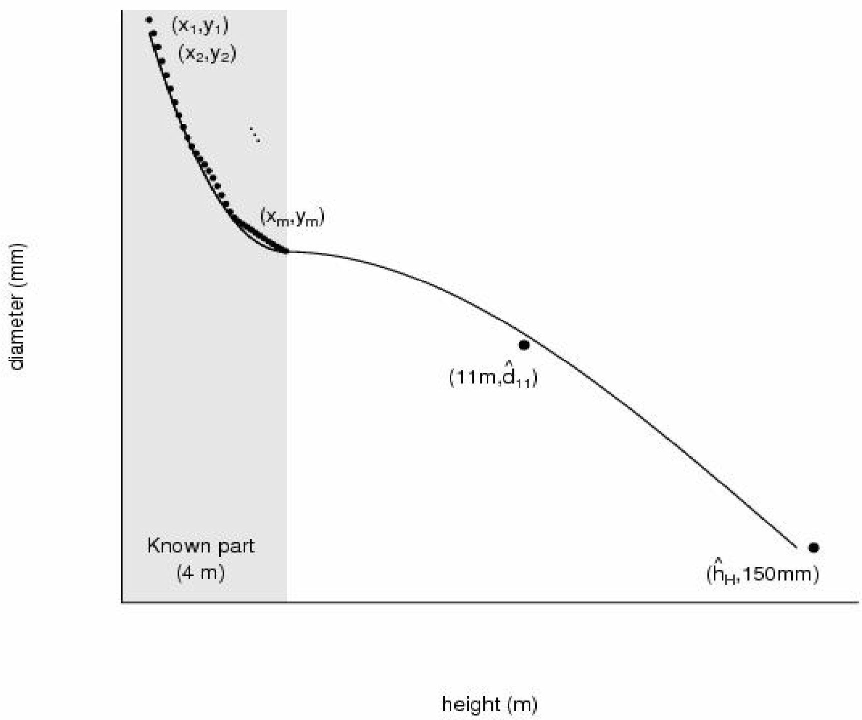
• Spline-functions.

A Method for Stem Curve Prediction

Two phases:

- First predict diameter at 11 m and height at 15 cm (the most valuable part of the stem is predicted with high accuracy).

- Fit smoothing spline through known part and predicted points.



(x_1,y_1)
(x_2,y_2)
(x_m,y_m)
$(11m,\hat{d}_{11})$
Known part
(4 m)
$(\hat{h}_H,150mm)$
diameter (mm)
height (m)

- For hight (Schumacher model)

$$h = 1.3 + exp(\beta_0 + \beta_1 \frac{1}{d_{bh}}) + \epsilon$$

  and for diameter the linear regression

$$d = \beta_0 + \beta_1 d_{bh} + \beta_2 s + \epsilon.$$

- Models estimated from 50 earlier stems (see Liski and Nummi, 1995).

- The length of the known part was 4 m.

- According to trail data and visual inspections we set $\alpha = 1$

Test Data

- For details see (Koskela, Nummi, Wentzel, Kivinen, 2006).

- Five stands from Southern Finland.

- Two species

  - Pine: A(n=1226), B(n=565), C(n=185)

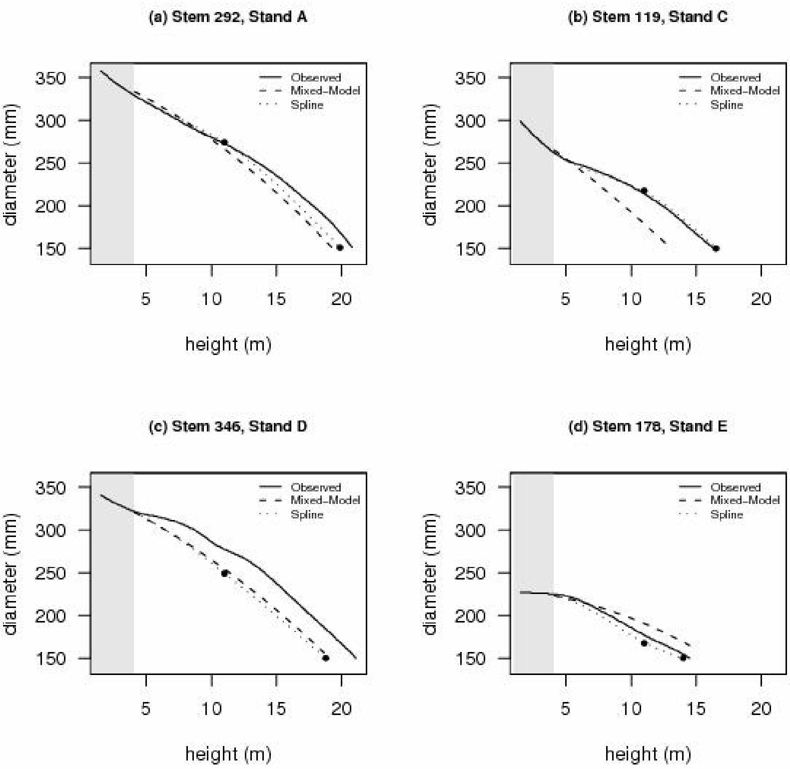
  - Spruce: D(n=544), E(n=613).

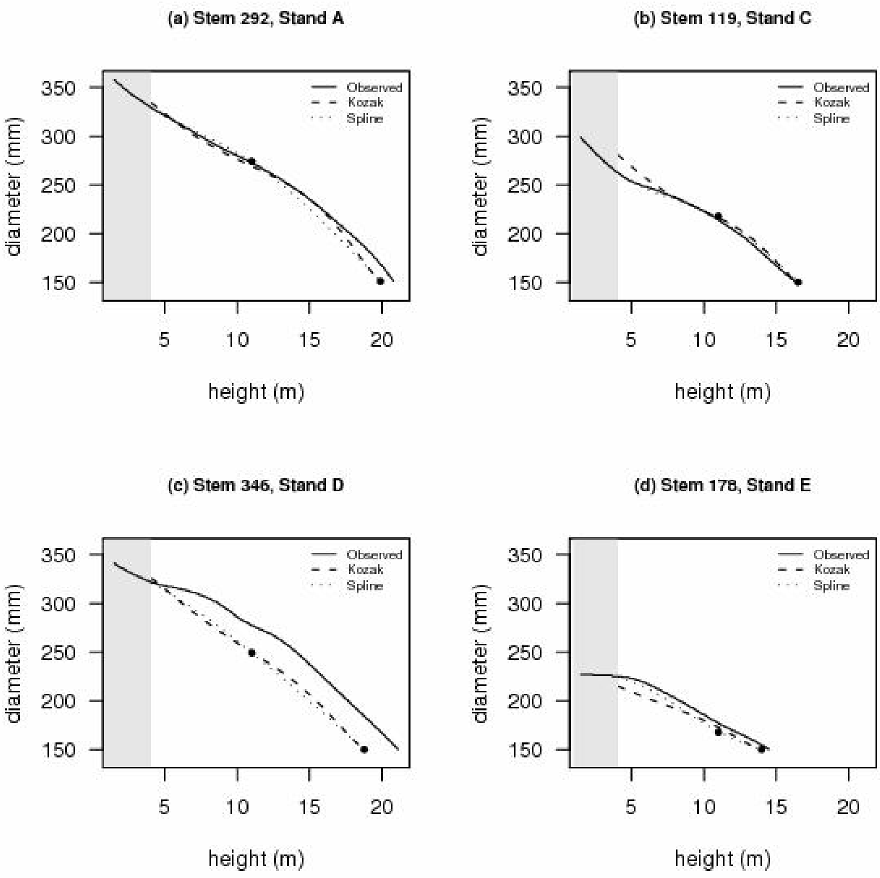
Evaluation of prediction errors

$$\begin{aligned} \text{MAE} &= (1/n)\sum \mid y_i - \hat{y}_i \mid \\ \text{RMSE} &= \sqrt{(1/n)\sum (y_i - \hat{y}_i)^2} \\ \text{MAPE} &= (1/n)\sum \mid (y_i - \hat{y}_i)/y_i \mid \end{aligned}$$

	Met.	Criter.	Stand A	B	C	D	E
(a)	Spl.	MAE	8.35	5.80	4.08	6.68	9.22
		RMSE	13.16	9.21	6.68	11.12	14.58
		MAPE	0.036	0.028	0.021	0.031	0.040
(b)	Mix.	MAE	15.23	11.48	8.65	14.09	14.87
		RMSE	23.68	17.66	13.71	25.25	26.11
		MAPE	0.070	0.058	0.046	0.069	0.070
(c)	Koz.	MAE	9.24	7.20	5.38	8.29	10.76
		RMSE	13.73	10.67	8.46	12.68	16.13
		MAPE	0.040	0.035	0.027	0.038	0.047

Also the sign test based on the individual MAE values indicated that Spline method is superior over Mixed model and Kozak model.


(a) Stem 292, Stand A
(b) Stem 119, Stand C
(c) Stem 346, Stand D
(d) Stem 178, Stand E
Observed
Mixed-Model
Spline
diameter (mm)
height (m)


(a) Stem 292, Stand A
(b) Stem 119, Stand C
(c) Stem 346, Stand D
(d) Stem 178, Stand E
Observed
Kozak
Spline
diameter (mm)
height (m)
350
300
250
200
150
5
10
15
20

Some comments

- Kozak model and mixed model strictly tied to certain functional forms.
  - The form may not be flexible enough to describe the stem curve
  - Possibly discontinuity point after the known section.
- Irregular butt degrades predictions especially for mixed models. Longer known part $\rightarrow$ better predictions.
- The form of the curve is determined by the stem height in Kozak model. Biased parameter estimates if the height is measured with error.
- Kozak model do not perceive the individual form variation.

5. Application 2: Growth Curves
5.1 Model and estimation

- The growth curve model (GCM) of Potthoff & Roy (1964)

$$\boldsymbol{Y} = \boldsymbol{T}\boldsymbol{B}\boldsymbol{A}' + \boldsymbol{E},$$

  where $\boldsymbol{Y} = (\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_n)$ is a matrix of obs.,

  $\boldsymbol{T}$ and $\boldsymbol{A}$ are design matrices (within and between individual),

  $\boldsymbol{B}$ is a matrix of unknown parameters, and

  $\boldsymbol{E}$ is a matrix of random errors.

- The columns of $\boldsymbol{E}$ are independently distributed as

$$\boldsymbol{e}_i \sim N(\mathbf{0}, \boldsymbol{\Sigma}).$$

- Here I assume that

$$\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{R},$$

  where $\boldsymbol{R}$ takes certain parsimonious covariance structure with covariance parameters $\boldsymbol{\theta}$.

- Now we may write

$$\boldsymbol{Y} = \boldsymbol{G}\boldsymbol{A}' + \boldsymbol{E},$$

  where $\boldsymbol{G} = (\boldsymbol{g}_1, \ldots, \boldsymbol{g}_m)$ is the matrix of mean curves.

- The GCM is a linear approximation

$$\begin{aligned} \boldsymbol{G} &= (\boldsymbol{g}_1, \ldots, \boldsymbol{g}_m) \\ &= (\boldsymbol{T\beta}_1, \ldots, \boldsymbol{T\beta}_m) \\ &= \boldsymbol{TB}. \end{aligned}$$

- The aim here is to develop the methods needed when $\boldsymbol{G}$ is approximated by more flexible cubic smoothing splines.

- Penalized log-likelihood function

$$2l = -\frac{1}{\sigma^2}\mathrm{tr}[(\boldsymbol{Y}' - \boldsymbol{A}\boldsymbol{G}')\boldsymbol{R}^{-1}(\boldsymbol{Y}' - \boldsymbol{A}\boldsymbol{G}')' +$$

$$\alpha(\boldsymbol{A}\boldsymbol{G}')\boldsymbol{K}(\boldsymbol{A}\boldsymbol{G}')'] - n\log|\sigma^2\boldsymbol{R}| - c.$$

- For given α, σ^2 and $\boldsymbol{R}$, the maximum is obtained at

$$\tilde{\boldsymbol{G}} = (\boldsymbol{R}^{-1} + \alpha\boldsymbol{K})^{-1}\boldsymbol{R}^{-1}\boldsymbol{Y}\boldsymbol{A}(\boldsymbol{A}'\boldsymbol{A})^{-1}.$$

- If $\boldsymbol{R}$ satisfies

$$\boldsymbol{R}\boldsymbol{K} = \boldsymbol{K},$$

this simplifies to

$$\widehat{\boldsymbol{G}} = (\boldsymbol{I} + \alpha\boldsymbol{K})^{-1}\boldsymbol{Y}\boldsymbol{A}(\boldsymbol{A}'\boldsymbol{A})^{-1}.$$

- It is easily seen that

  $\boldsymbol{R} = \boldsymbol{I}$ (Independent),

  $\boldsymbol{R} = \boldsymbol{I} + \sigma_d^2 \mathbf{1}\mathbf{1}'$ (Uniform),

  $\boldsymbol{R} = \boldsymbol{I} + \sigma_{d'}^2 \boldsymbol{X}\boldsymbol{X}'$ (Linear1),

  $\boldsymbol{R} = \boldsymbol{I} + \boldsymbol{X}\boldsymbol{D}\boldsymbol{X}'$ (Linear2)

  satisfies the condition $\boldsymbol{R}\boldsymbol{K} = \boldsymbol{K}$.

- This result can be compared to estimation in linear models, when BLUE coinsides with OLSE.

- We can write $\widehat{\boldsymbol{G}}$ as

$$\widehat{\boldsymbol{g}} = [(\boldsymbol{A}'\boldsymbol{A})^{-1}\boldsymbol{A}' \otimes (\boldsymbol{I} + \alpha\boldsymbol{K})^{-1}]\boldsymbol{y},$$

where $\widehat{\boldsymbol{g}} = \text{vec}(\widehat{\boldsymbol{G}})$ and $\boldsymbol{y} = \text{vec}(\boldsymbol{Y})$.

- Further

$$\widehat{\boldsymbol{g}} = (\boldsymbol{I}_m \otimes \boldsymbol{X})\widehat{\boldsymbol{\beta}}_* + (\boldsymbol{I}_m \otimes \boldsymbol{Z})\widehat{\boldsymbol{u}}_*,$$

where

$$\widehat{\boldsymbol{\beta}}_* = [(\boldsymbol{A}'\boldsymbol{A})^{-1}\boldsymbol{A}' \otimes \boldsymbol{I}_q]\widehat{\boldsymbol{\beta}}$$

and

$$\widehat{\boldsymbol{u}}_* = [(\boldsymbol{A}'\boldsymbol{A})^{-1}\boldsymbol{A}' \otimes \boldsymbol{I}_q]\widehat{\boldsymbol{u}}.$$

- The spline solution of the model $\boldsymbol{Y} = \boldsymbol{G}\boldsymbol{A}' + \boldsymbol{E}$ can be expressed as the BLUP of the mixed model

$$\boldsymbol{Y} = (\boldsymbol{X}\boldsymbol{B}_* + \boldsymbol{Z}\boldsymbol{U}_*)\boldsymbol{A}' + \boldsymbol{E},$$

where

$$\begin{pmatrix} vec(\boldsymbol{u}_*) \\ \boldsymbol{e}_i \end{pmatrix} \sim N\left(\boldsymbol{0}, \begin{pmatrix} \sigma_u^2(\boldsymbol{A}'\boldsymbol{A})^{-1} \otimes \boldsymbol{\nabla} & \boldsymbol{O} \\ \boldsymbol{O} & \sigma^2\boldsymbol{R} \end{pmatrix}\right).$$

- For large α the mean spline approaches $\boldsymbol{X}\boldsymbol{B}_*\boldsymbol{A}'$ and it is not influenced by any particular choice of α.
- We may utilize "the fixed part"$\boldsymbol{X}\boldsymbol{B}_*\boldsymbol{A}'$ to extract rough features of the curves.

- In fact since $\boldsymbol{X} = (\mathbf{1}, \boldsymbol{x})$

  $$E(\boldsymbol{y}_i) = a_{i1}(b_{01}\mathbf{1} + b_{11}\boldsymbol{x}) + \cdots + a_{im}(b_{0m}\mathbf{1} + b_{1m}\boldsymbol{x})$$

  is a sum of straight lines and if

  $$a_i = (0, \ldots, 1, 0, \ldots, 0)'$$

  we have

  $$b_{0j}\mathbf{1} + b_{1j}\boldsymbol{x}.$$

- For smooth curves we may assume that these lines roughly reflects the average development of individuals summarized by splines.

5.2. Some ideas of testing

- We restrict our attention to the "fixed part" $\boldsymbol{X}\boldsymbol{B}_*\boldsymbol{A}'$.
- The variance-covariance matrix of $\widehat{\boldsymbol{\beta}}_*$ is

$$\mathrm{Cov}(\widehat{\boldsymbol{\beta}}_*) = [(\boldsymbol{A}'\boldsymbol{A})^{-1} \otimes \sigma^2(\boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{X})^{-1}],$$

  which does not depend on the spline features of the mean curve.
- Consider the general linear hypothesis

$$\mathrm{H}_0: \ \boldsymbol{C}\boldsymbol{B}_*\boldsymbol{L} = \boldsymbol{O},$$

  where $\boldsymbol{C}$ and $\boldsymbol{L}$ are $r \times 2$ and $m \times c$ matrices with ranks r and c, respectively.

- It can be shown that under H_0

$$Q = \mathrm{tr}[\{\sigma^2 \boldsymbol{C}(\boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{X})^{-1}\boldsymbol{C}'\}^{-1} \cdot \boldsymbol{C}\widehat{\boldsymbol{B}}_*\boldsymbol{L} \cdot$$

$$\{\boldsymbol{L}'(\boldsymbol{A}'\boldsymbol{A})^{-1}\boldsymbol{L}\}^{-1} \cdot (\boldsymbol{C}\widehat{\boldsymbol{B}}_*\boldsymbol{L})'] \sim \chi^2_{cr}.$$

- Parameters σ^2 and $\boldsymbol{R}$ unknown (estimated) $\rightarrow$ distribution of Q is only approximate.

6. Application 3: Covariance Modelling

- In modified Cholesky decomposition (MCD) we decompose

$$\boldsymbol{H}\boldsymbol{\Sigma}\boldsymbol{H}' = \boldsymbol{W}$$

or

$$\boldsymbol{\Sigma}^{-1} = \boldsymbol{H}'\boldsymbol{W}\boldsymbol{H},$$

$\boldsymbol{H}$ is a uniq. lower dg with 1's as dg and $\boldsymbol{W}$ is a uniq. dg with positive dg.

- $\boldsymbol{H}$ and $\boldsymbol{W}$ have easy interpretation.

- The below-diagonal entries of $\boldsymbol{H}$ can be interpreted as negatives of the autoregressive coefficients, ϕ_{jk}, in

$$\widehat{y}_j = \mu_j + \sum_{k=1}^{q-1} \phi_{jk}(y_k - \mu_k).$$

- Diagonal entries of $\boldsymbol{W}$ are innovation variances

$$\sigma_j^2 = \mathsf{Var}(y_j - \widehat{y}_j),$$

where $j = 1, \ldots, q$.

Note that

$$\boldsymbol{Hy} =$$

$$\begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ -\phi_{21} & 1 & 0 & \dots & 0 \\ -\phi_{31} & -\phi 32 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \dots \\ -\phi_{q1} & -\phi_{q2} & -\phi_{q3} & \dots & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_q \end{pmatrix} = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \vdots \\ \epsilon_q \end{pmatrix}$$

where $(\epsilon_1, \dots, \epsilon_q)' = \boldsymbol{\epsilon}$ is a vector of prediction errors and

$$\text{Var}(\boldsymbol{\epsilon}) = diag(\sigma_1^2, \dots, \sigma_q^2) = \boldsymbol{W}.$$

So the matrix $\boldsymbol{H}$ diagonalises the covariance matrix $\boldsymbol{\Sigma}$.

- When $\boldsymbol{\Sigma}$ is unstructured, the non-redundant entries of $\boldsymbol{H}$ and log $\boldsymbol{W}$ are unconstrained and the dimension of the parameters space can be reduced.
- New estimate is positive definite.
- Let

$$\text{log}\sigma_j^2 = \nu(z_j, \boldsymbol{\lambda})$$

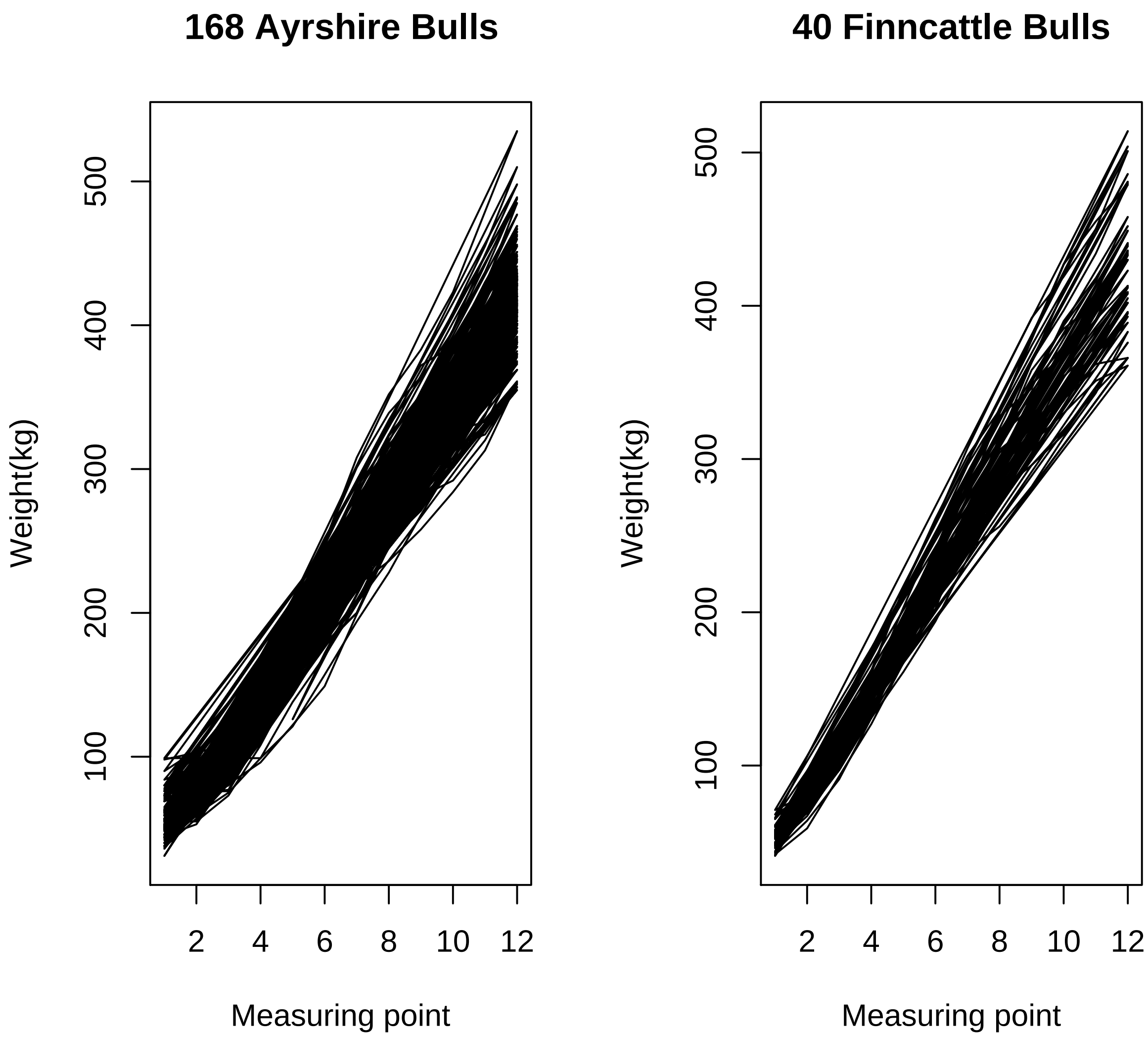
and

$$\phi_{jk} = \eta(z_{jk}, \boldsymbol{\gamma}),$$

where $\nu(.,.)$ and $\eta(.,.)$ are functions of covariates z_j and z_{jk} and $\boldsymbol{\lambda}$ and $\boldsymbol{\gamma}$ are parameters.

Example: Growth curves of bulls

168 and 40 Ayrshire and Finncattle bulls measured once per month during one year.



The sample covariance matrix gives MCD for

Finncattle bulls

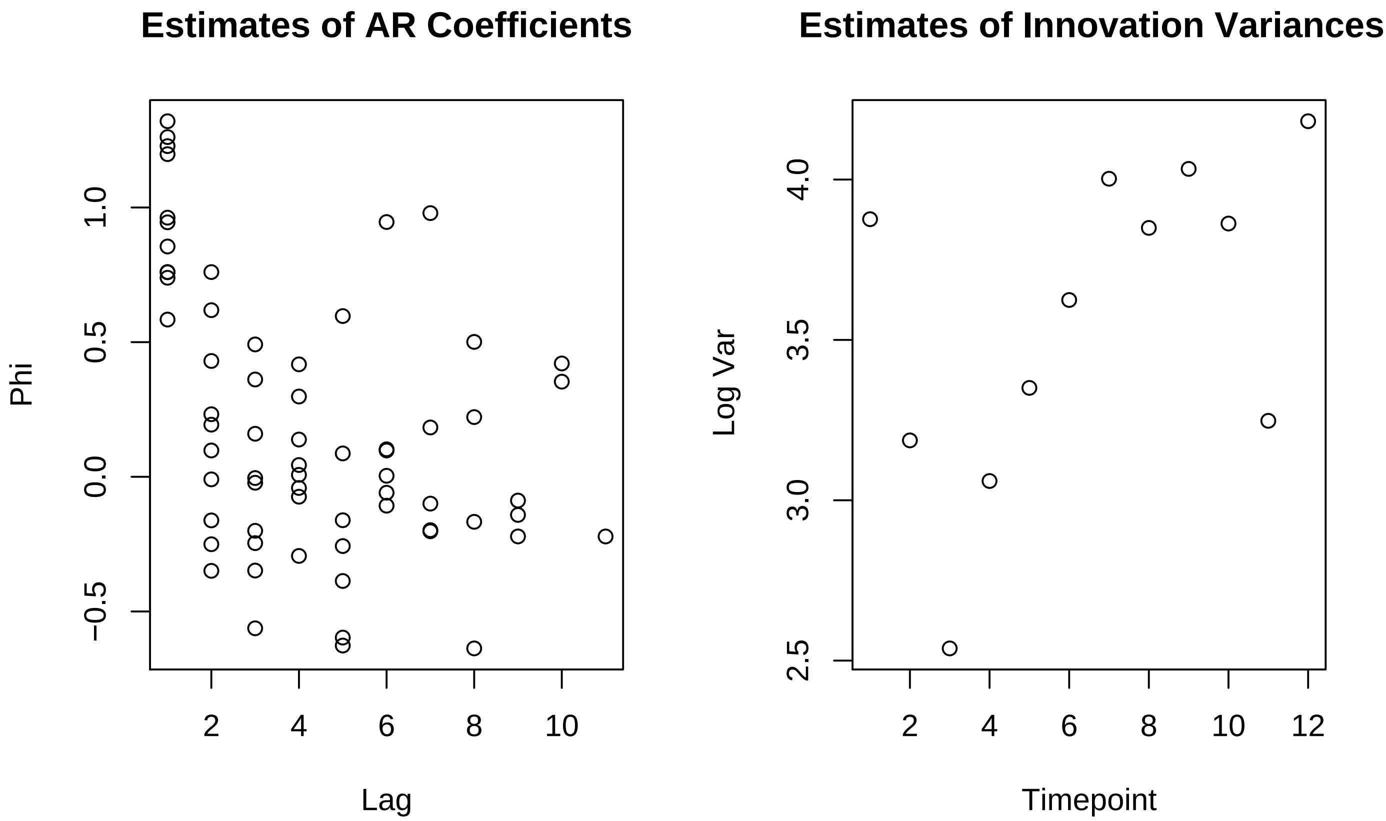
$$\widehat{\boldsymbol{H}} = \begin{pmatrix} 1 & 0 & 0 & \ldots & 0 \\ -0.584 & 1 & 0 & \ldots & 0 \\ -0.194 & -0.945 & 1 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ldots & \vdots \\ 0.221 & -0.421 & 0.088 & \ldots & 1 \end{pmatrix}$$

and

$$diag(\widehat{\boldsymbol{W}}) = (48.254, 24.209, 12.656 \ldots, 65.497)$$

plots of non-redundant elements of $\widehat{\boldsymbol{H}}$ and

$log(\widehat{\sigma}_1^2), \ldots, log(\widehat{\sigma}_q^2)$ give


Estimates of AR Coefficients
Phi
1.0
0.5
0.0
−0.5
2
4
6
8
10
Lag
Estimates of Innovation Variances
Log Var
4.0
3.5
3.0
2.5
2
4
6
8
10
12
Timepoint