BayesX: Analysing Bayesian semiparametric regression models

Andreas Brezger, Thomas Kneib and Stefan Lang
Institut für Statistik, Universität München

Workshop AG-Bayes, 6. Dezember 2002

Outline of the Talk

- What is BayesX?
- Bayesian semiparametric regression
- Example(s)

What is BayesX?

BayesX is a tool for Bayesian inference via MCMC simulation techniques.

available as a Windows (NT, 95, 98, 2000) based application at

`http://www.stat.uni-muenchen.de/~lang/`

Features of the current version

- Functions for handling and manipulating data
- Functions for handling spatial data
- Functions for drawing geographical maps, scatterplots, etc.
- Bayesian semiparametric regression
- Model selection for DAG's (by Eva-Maria Fronk)

Features of the regression tool

- Estimation of any generalized additive model
- Response: Gaussian, Poisson, Gamma, Binomial, Multinomial

BayesX includes as special cases . . .

- Generalized linear models
- Generalized additive models
- Dynamic or state space models
- Varying coefficient models
- Mixed models
- BYM model for disease mapping

Observation models

- Distributional and structural assumptions, given covariates and parameters, are based on generalized linear models.

- Response: Gaussian, Gamma, Poisson, Binomial, Multinomial

- Replace the linear predictor

$$\eta = z'\gamma$$

by a semiparametric additive predictor

$$\eta = f_1(x_1) + \cdots + f_p(x_p) + z'\gamma$$

$f_1, ..., f_p$ are unknown functions of the covariates

γ parameter vector for fixed effects

Extensions

Varying coefficient terms

$$\eta = \cdots + f(x)z + \cdots$$

Surface smoothing

$$\eta = \cdots + f(x_1, x_2) + \cdots$$

Priors for a function f

$$\eta = f_1(x_1) + \cdots + f_p(x_p) + z'\gamma$$

$$f = X\beta$$

X design matrix

β are unknown parameters

$$\eta = \cdots + X\beta + \cdots$$

The general prior

$$\beta|\tau^2 \propto \exp(-\frac{1}{2\tau^2}\beta' K \beta)$$

$$\tau^2 \sim IG(a, b)$$

- K is a penalty matrix that penalizes too rough functions f
- structure of K depends on type of covariate and on prior beliefs on smoothness of f
- amount of smoothness is controlled by τ^2

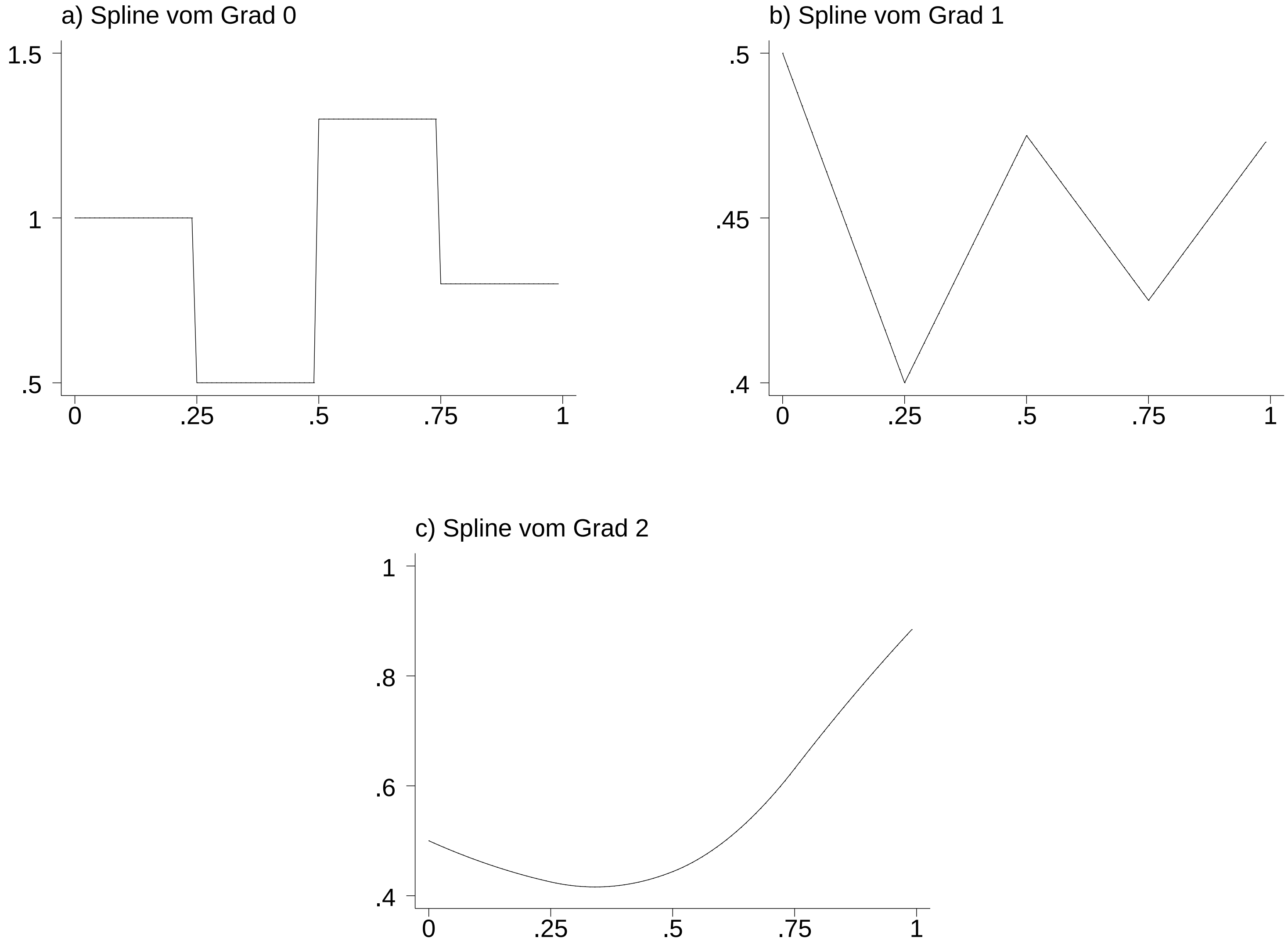
Example 1: P-splines

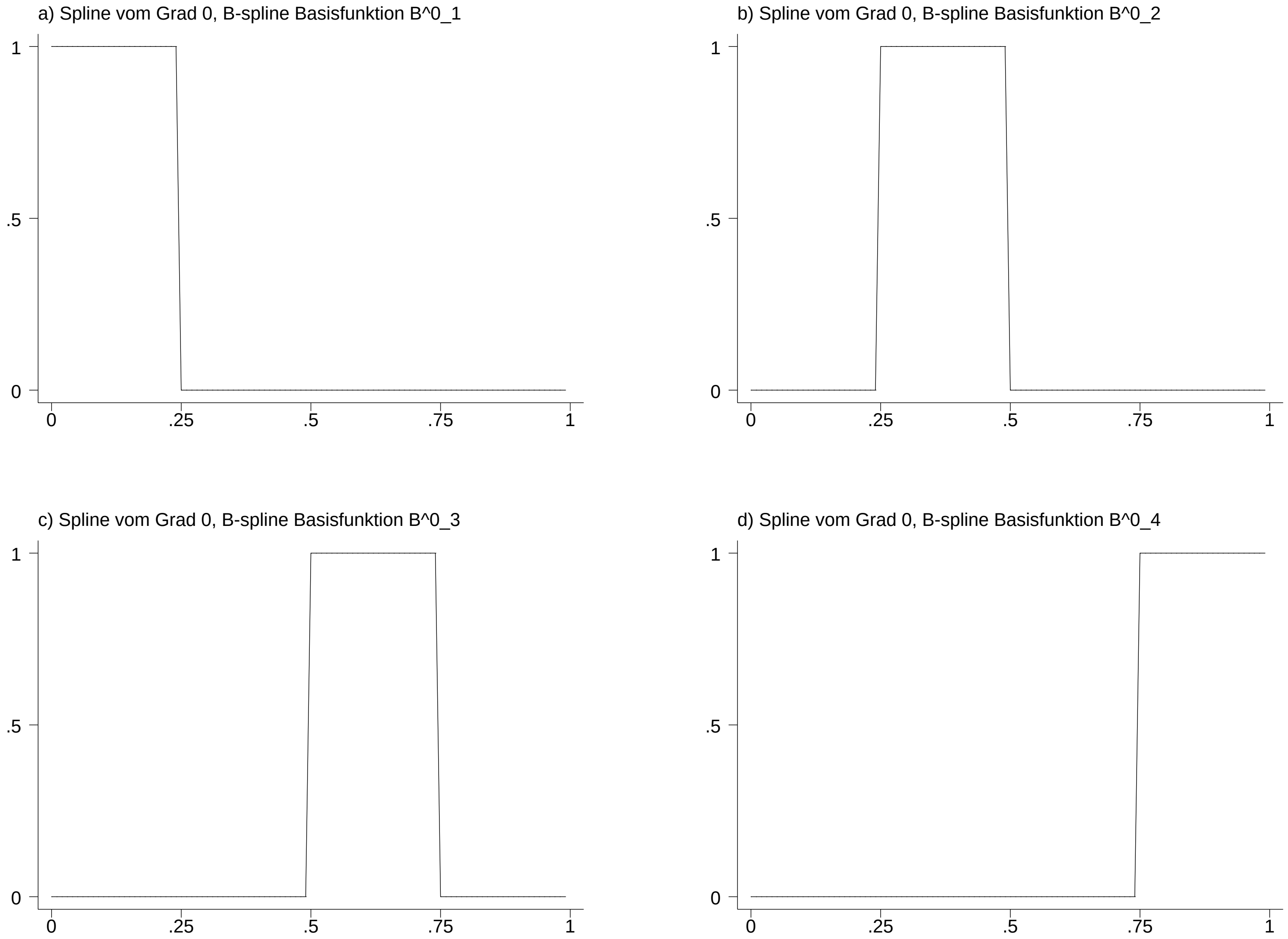
(Eilers and Marx, 1996; Lang and Brezger, 2002)

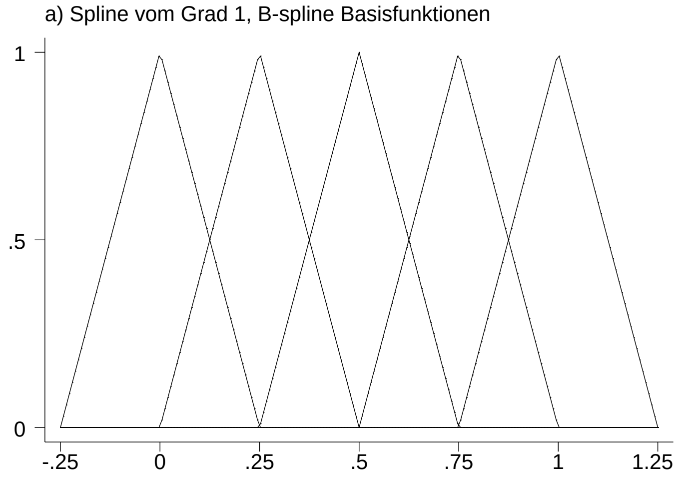
$$
\begin{aligned}
f(x) &= \text{Spline of degree } l \text{ with equally spaced inner knots} \\
&\quad\ \xi_1, ..., \xi_r \text{ between } x_{(min)} \text{ and } x_{(max)} \\
&= \beta_1 B_1(x) + \cdots + \beta_{r+l+1} B_{r+l+1}(x)
\end{aligned}
$$

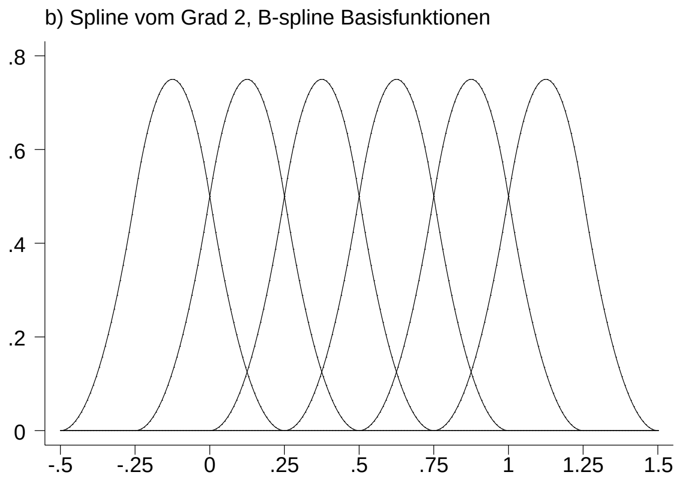
$B_1, ..., B_{r+l+1}$ B-spline Basis

X design matrix with elements $X(i,j) = B_j(x_i)$


a) Spline vom Grad 0
1.5
1
.5
0
.25
.5
.75
1
b) Spline vom Grad 1
.5
.45
.4
0
.25
.5
.75
1
c) Spline vom Grad 2
1
.8
.6
.4
0
.25
.5
.75
1


a) Spline vom Grad 0, B-spline Basisfunktion B^0_1
b) Spline vom Grad 0, B-spline Basisfunktion B^0_2
c) Spline vom Grad 0, B-spline Basisfunktion B^0_3
d) Spline vom Grad 0, B-spline Basisfunktion B^0_4
0
.5
1
0
.25
.5
.75
1


a) Spline vom Grad 1, B-spline Basisfunktionen
1
.5
0
-.25
0
.25
.5
.75
1
1.25


b) Spline vom Grad 2, B-spline Basisfunktionen
.8
.6
.4
.2
0
-.5
-.25
0
.25
.5
.75
1
1.25
1.5

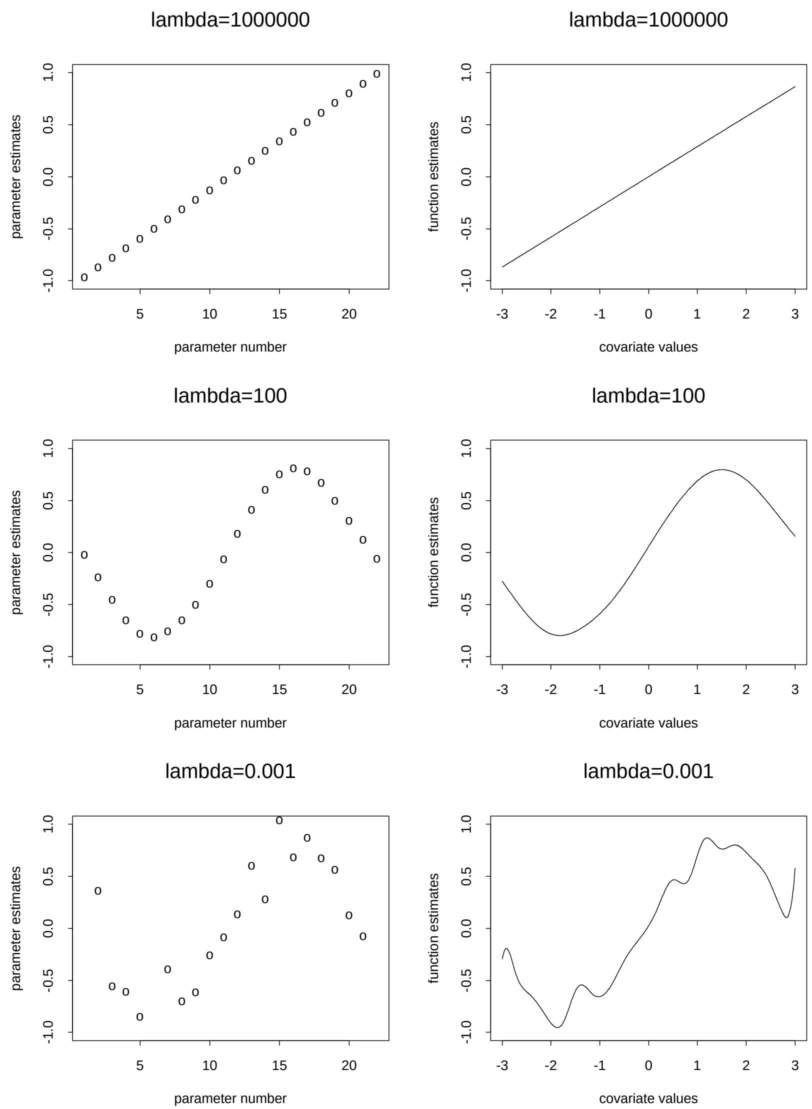
Example 1: P-splines, frequentist version

- relatively large number of inner knots
- difference penalty for $\beta_1, ..., \beta_{r+l+1}$ to penalize too rough functions f
- Leads to penalized likelihood estimation

$$L = l - \lambda \sum_{s=k+1}^{m} (\Delta^k \beta_s)^2$$

Δ^k denotes the difference operator of order k.

- Problem: Estimation of the smoothing parameter λ.


lambda=1000000
lambda=1000000
lambda=100
lambda=100
lambda=0.001
lambda=0.001
parameter estimates
function estimates
parameter number
covariate values

Example 1: P-splines, Bayesian approach

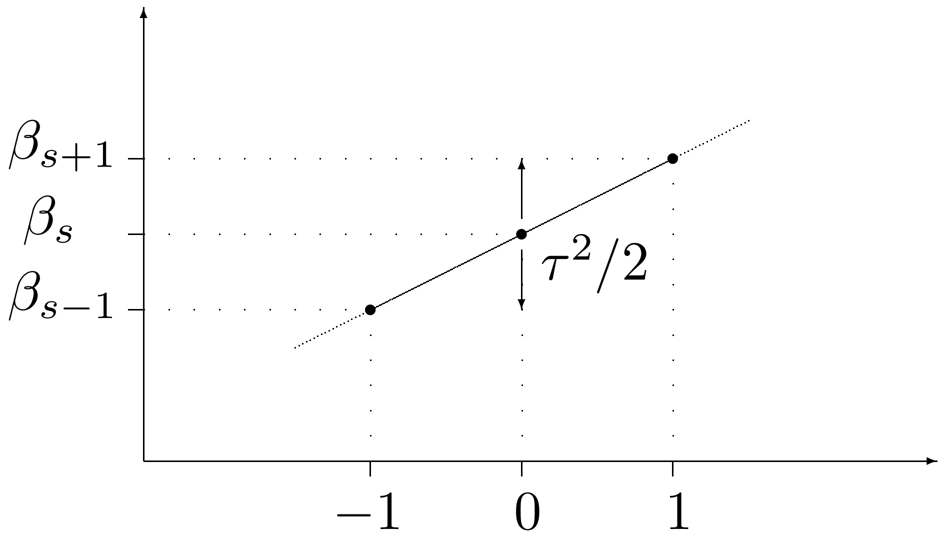
- replace difference penalties by their stochastic analogues
- smoothness prior for $\beta_1, ..., \beta_{r+l+1}$ to penalize too rough functions f
- use first or second order random walks as smoothness prior:

$\beta_t = \beta_{t-1} + u_t$ (RW1)

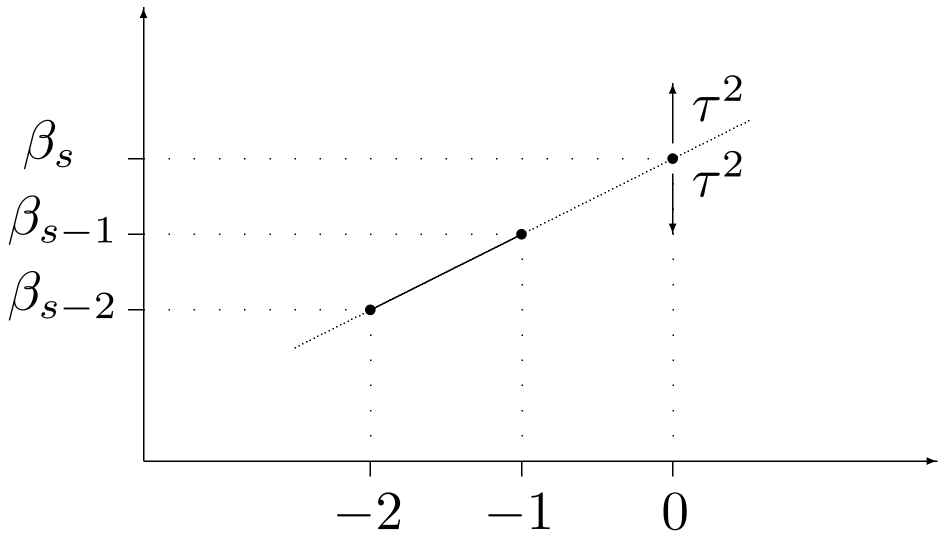
$\beta_t = 2\beta_{t-1} - \beta_{t-2} + u_t$ (RW2)

$u_t \sim N(0, \tau^2)$

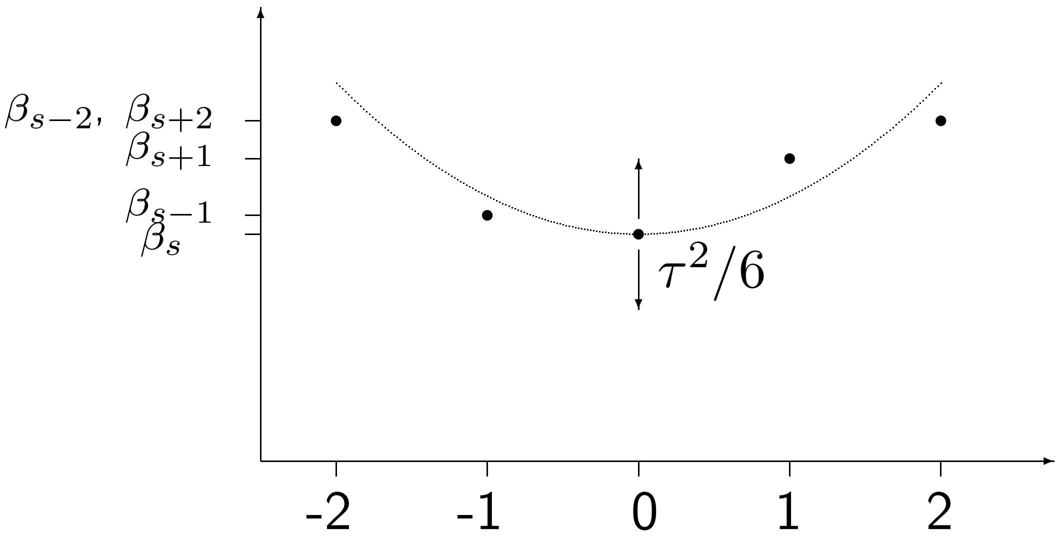
RW1: $P(\beta_s|\beta_{s-1}, \beta_{s+1})$



RW2: $P(\beta_s|\beta_{s-1}, \beta_{s-2})$



RW2: $P(\beta_s|\beta_{s-1}, \beta_{s-2}, \beta_{s+1}, \beta_{s+2})$



Example 2: Markov random fields

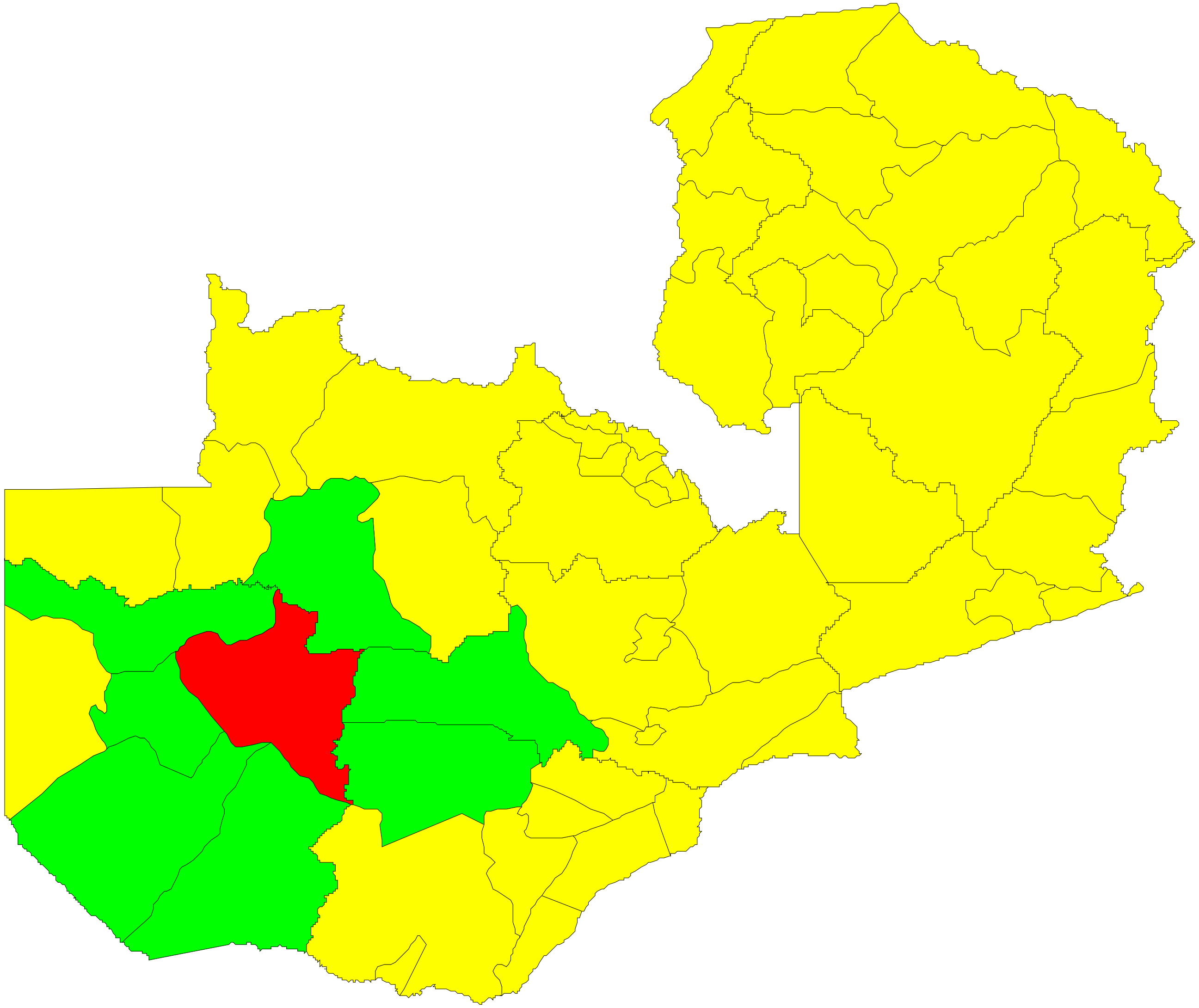
- Markov random fields (Besag, York, Mollie 1991), e.g.

$$\beta_s | \beta_{-s}, \tau^2 \sim N\left(\sum_{j \in \partial_s} \frac{1}{N_s} \beta_j, \frac{1}{N_s} \tau^2 \right)$$

∂_s denoting the sites, that are neighbors of site s

N_s number of neighbors

- X 0/1 design matrix



Example 3: 2-dimensional surfaces

$$\eta = \cdots + f_1(x_1) + f_2(x_2) + f_{1,2}(x_1, x_2) + \cdots$$

$$\begin{aligned} f_{1,2} &= \text{tensor product of one dimensional B-splines} \\ &= \sum_{\rho=1}^{m} \sum_{\nu=1}^{m} \beta_{\rho,\nu} B_{1,\rho}(x_1) B_{2,\nu}(x_2). \end{aligned}$$

spatial smoothness prior fo coefficients $\beta_{\rho,\nu}$, e.g.

2-dimensional random walks

Further examples

- random intercepts and slopes
- varying coefficient models
- time varying seasonal effects

Bayesian Inference via MCMC

- Draw random numbers from the posterior.
- Estimate characteristics of the posterior by their empirical analogue.
- Efficiency guaranteed by matrix operations for sparse matrices.

Details in
Fahrmeir, Lang (2001a,b)
Lang and Brezger (2002)

Example: Gaussian response

$$y_i = f_1(x_{i1}) + f_2(x_{i2}) + \epsilon_i \qquad y = X_1\beta_1 + X_2\beta_2 + \epsilon$$

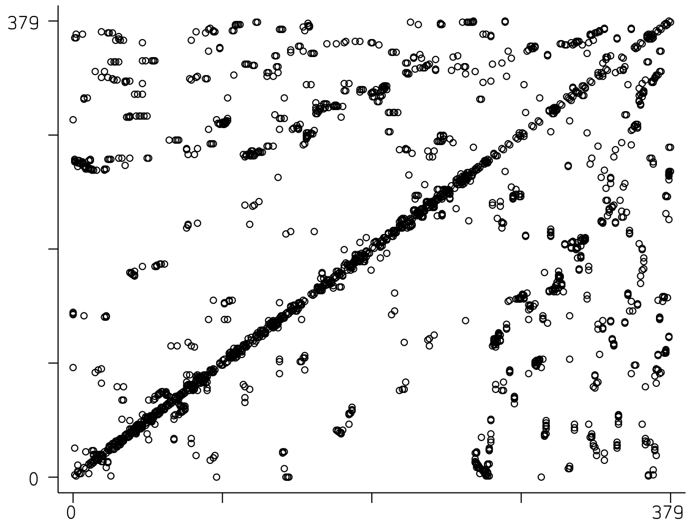
- Draw random number from $\beta_1|\cdot$ which is Gaussian with

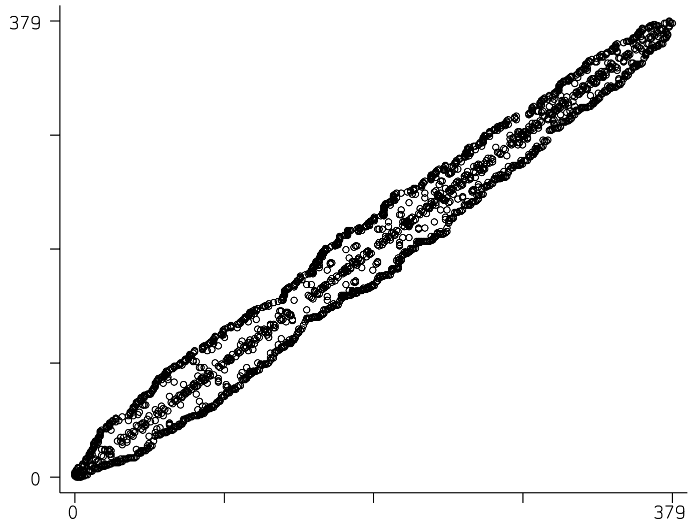
$$\Sigma_1 = \left(\frac{X_1'X_1}{\sigma^2} + \frac{K_1}{\tau_1^2}\right)^{-1} \qquad \mu_1 = \frac{1}{\sigma^2}\Sigma_1 X_1'(y - X_2\beta_2)$$

- Draw random number from $\beta_2|\cdot$ which is Gaussian with

$$\Sigma_2 = \left(\frac{X_2'X_2}{\sigma^2} + \frac{K_2}{\tau_2^2}\right)^{-1} \qquad \mu_2 = \frac{1}{\sigma^2}\Sigma_2 X_2'(y - X_1\beta_1)$$

- Draw random numbers from full conditionals of variance parameters $\sigma^2|\cdot, \quad \tau_1^2|\cdot, \quad \tau_2^2|\cdot.$


379
0
0
379


379
0
0
379

Application : "rental guides" for flats

Response variable

$R =$ monthly rent per square meter in German Marks

Covariates

F	floor space in square meters
A	age of building, that is year of construction
L	location in subquarters of the building in Munich
z	vector of categorical covariates characterizing the flat

Model 1:

$$\eta = \gamma_0 + f_1(F) + f_2(A) + f_{12}(F, A) + b_L^{unstr} + b_L^{str} + \mathbf{z}'\gamma$$

f_1, f_2 P-spline of degree 3 with second order random walk penalty

f_{12} tensor product P-spline with 2-dim. first order random walk penalty

b_L^{unstr} uncorrelated random effect

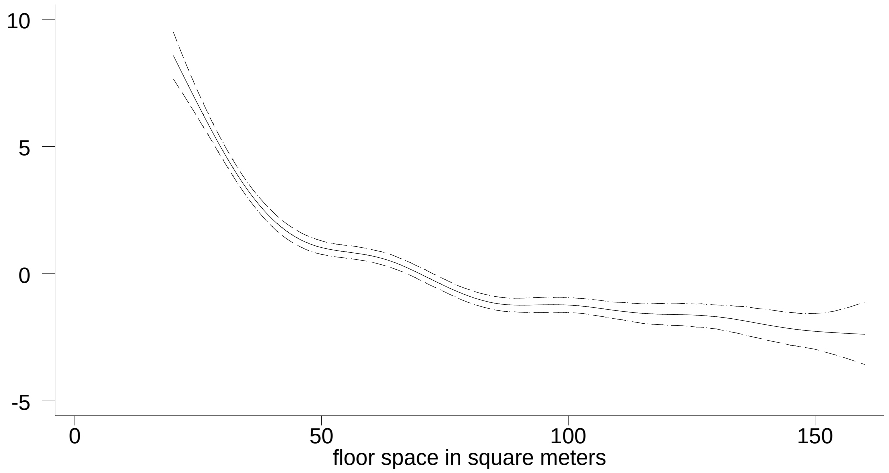
b_L^{str} spatially correlated random effect, MRF

Model 2:

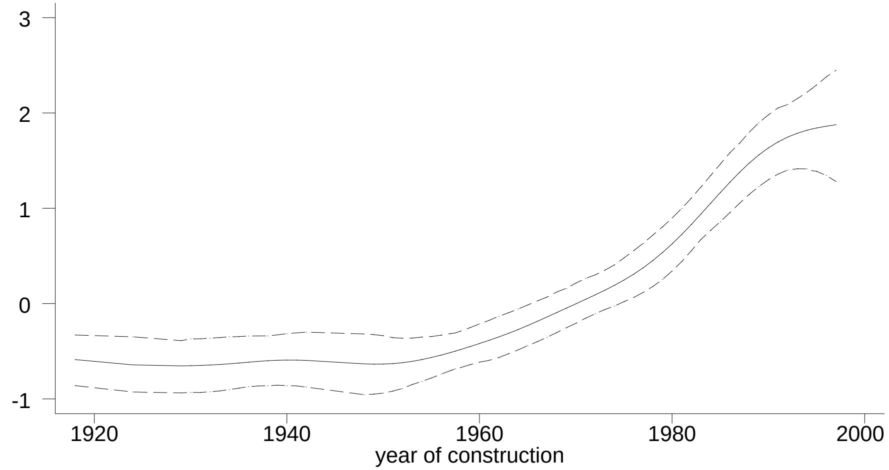
$$\eta = \gamma_0 + f_1(F) + f_2(A) + f_{12}(F, A) + b_L^{unstr} + b_L^{str} + \gamma_1\, gL + \gamma_2\, tL + \mathbf{z}'\gamma$$

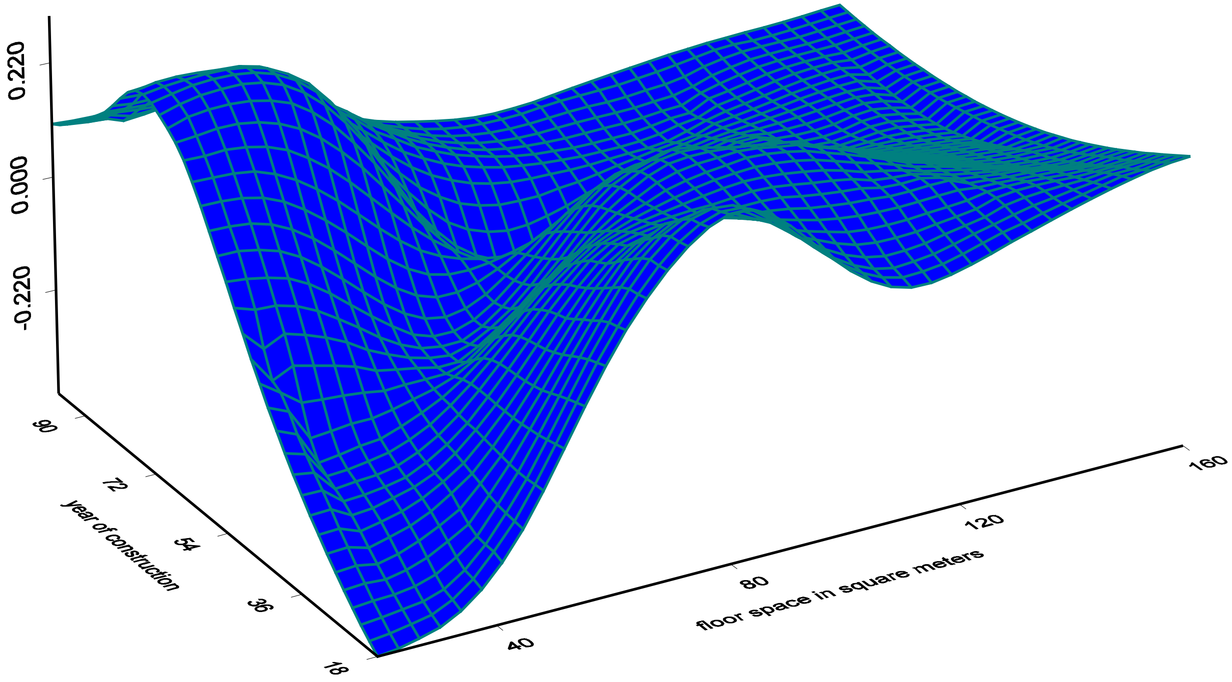
where gL and tL are $0/1$ indicators for good location and top location assessed by experts.

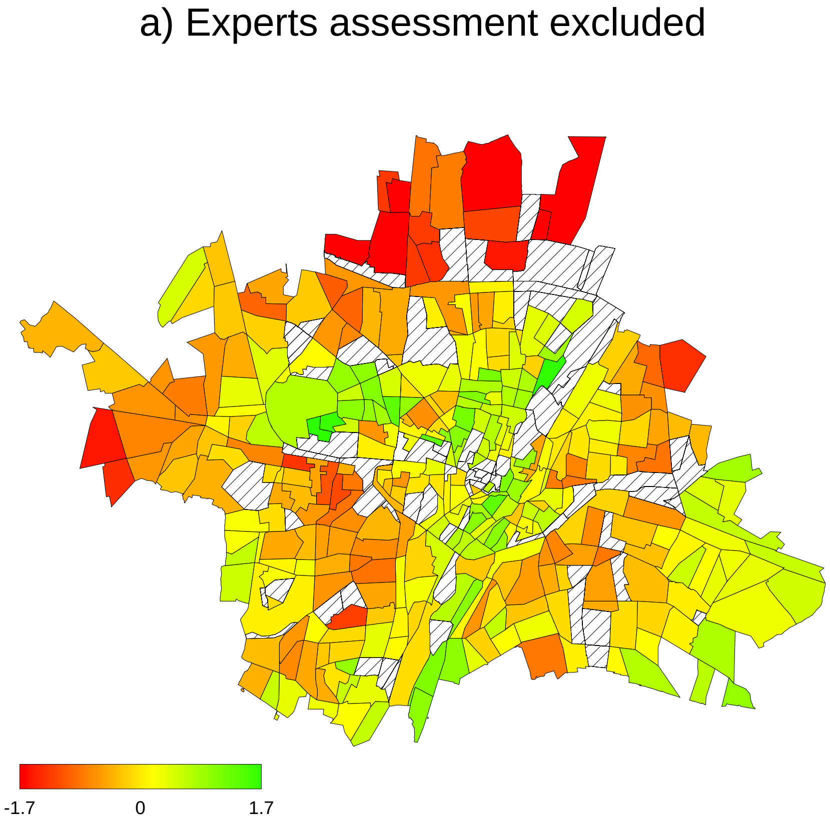
a) Main effect of floor space

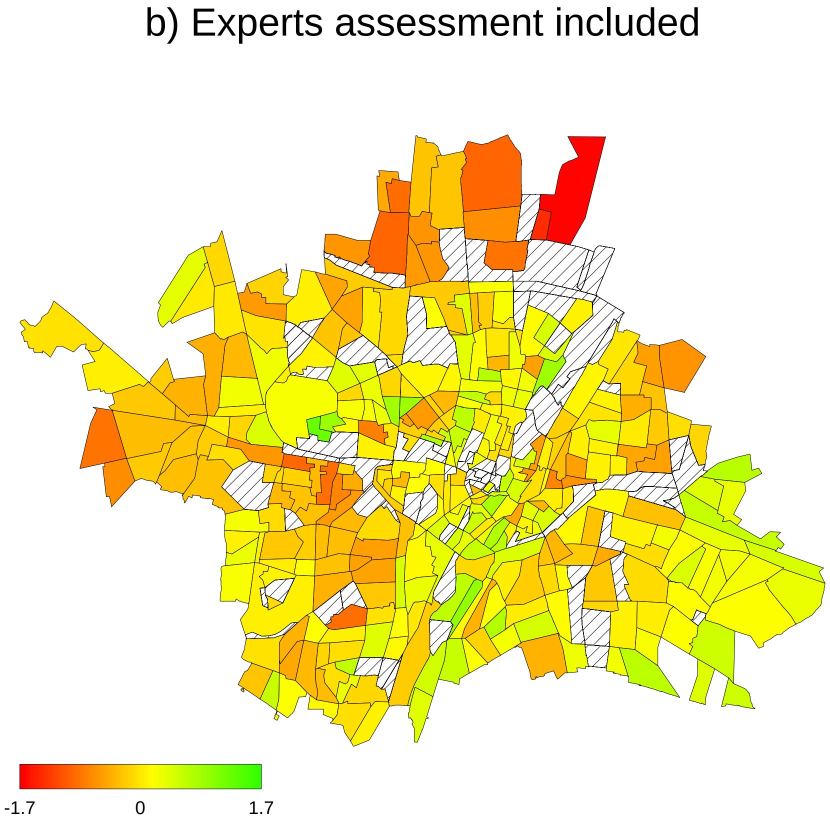


b) Main effect of year of construction




0.220
0.000
-0.220
90
72
54
36
18
year of construction
40
80
120
160
floor space in square meters


a) Experts assessment excluded
-1.7
0
1.7


b) Experts assessment included
-1.7
0
1.7

Example: Forest health data

Variable of primary interest

y_{it}, the degree of defoliation of beech i in year t, measured in three catogories:

$y_{it} = 1$	no defoliation
$y_{it} = 2$	defoliation 25% or less
$y_{it} = 3$	defoliation above 25%

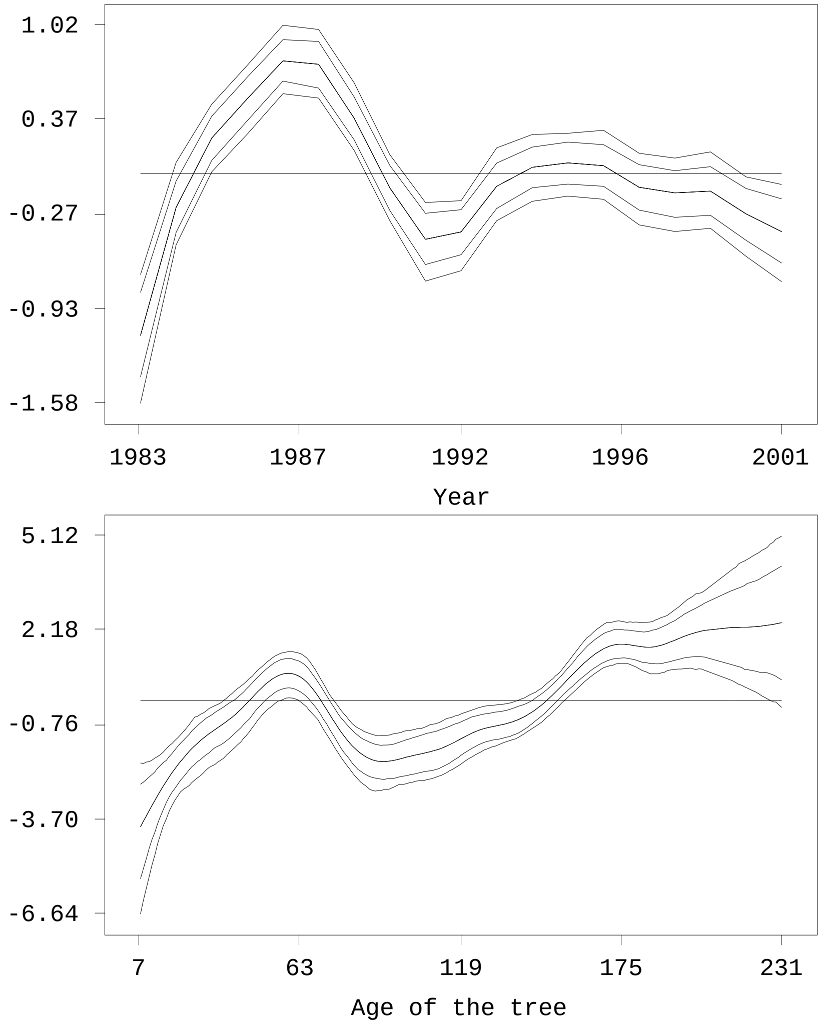
Covariates

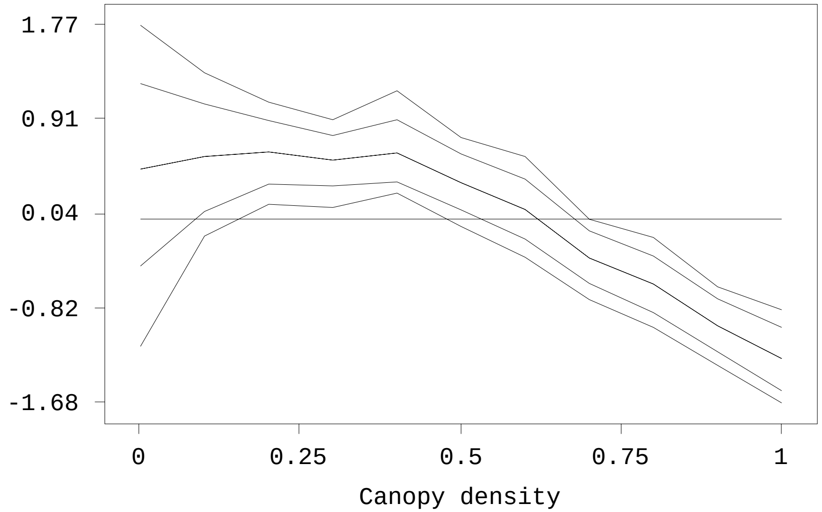
T	Calendar time in years (1983-2001)
S	site of the beech (84 sites)
A	age of the tree in years (7-231)
C	Canopy density at the stand measured in percentages 0%,10%,...,90%,100%

Model

Cumulative probit model, i.e.

$$U_{it} = f_1(t) + f_2(A_{it}) + f_3(C_{it}) + f_{spat}(s_i) + z_{it}'\gamma + \epsilon_{it}$$


1.02
0.37
-0.27
-0.93
-1.58
1983
1987
1992
1996
2001
Year
5.12
2.18
-0.76
-3.70
-6.64
7
63
119
175
231
Age of the tree


1.77
0.91
0.04
-0.82
-1.68
0
0.25
0.5
0.75
1
Canopy density

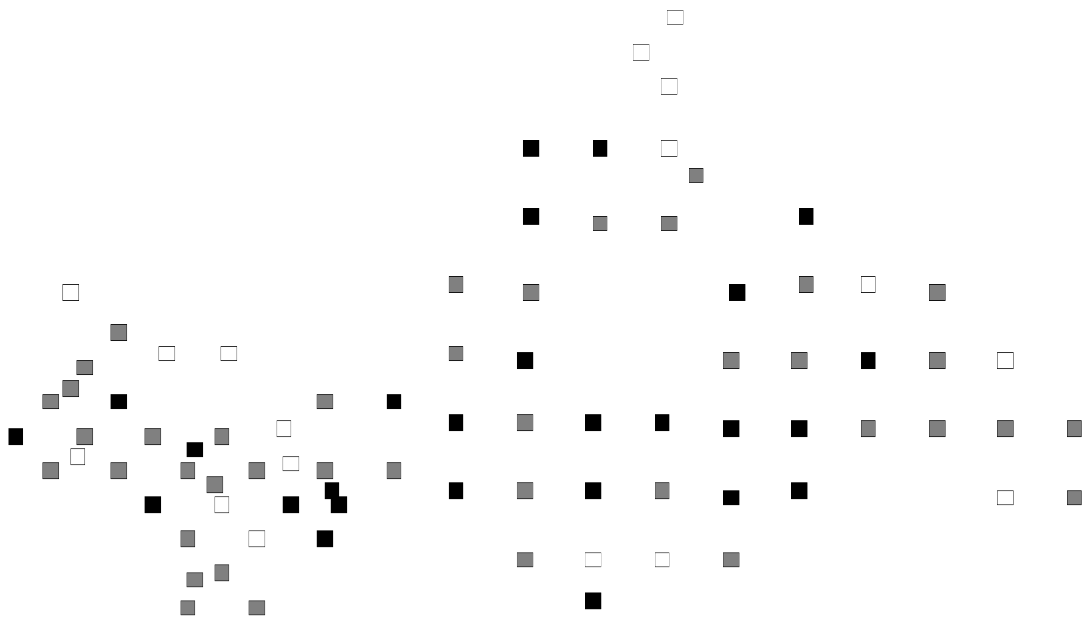


Figure 1: *Posterior probabilities of the spatial effect for a nominal level of 80%.*

black=positive effect,white=negative effect

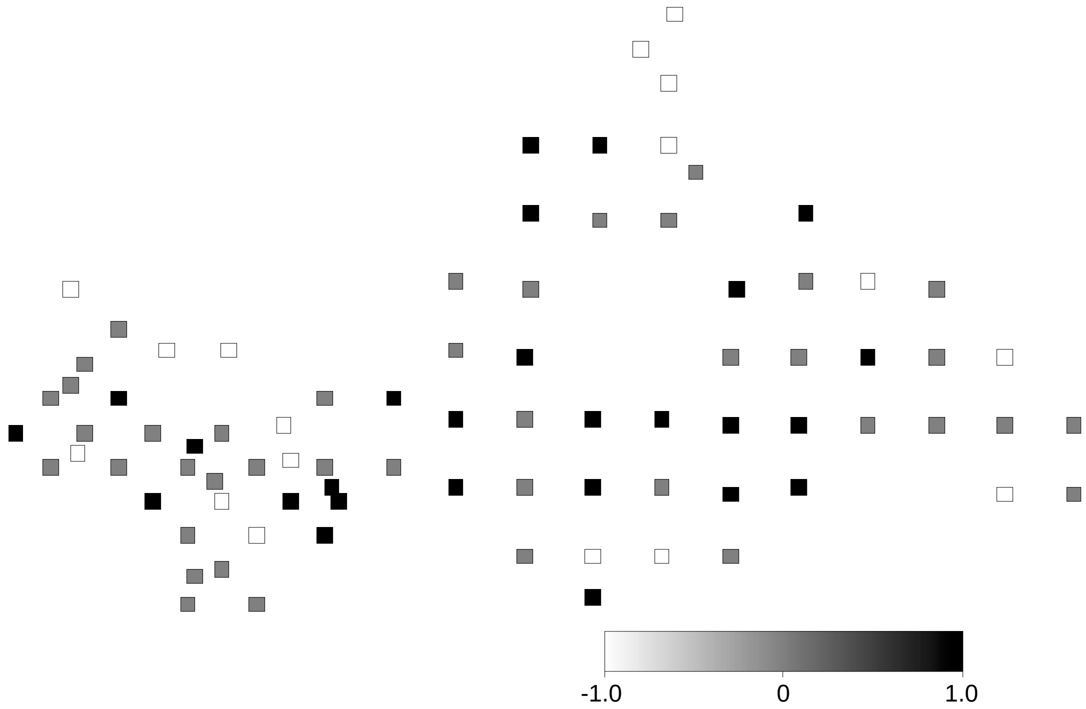


Figure 2: *Posterior probabilities of the spatial effect for a nominal level of 95%.*

black=positive effect,white=negative effect

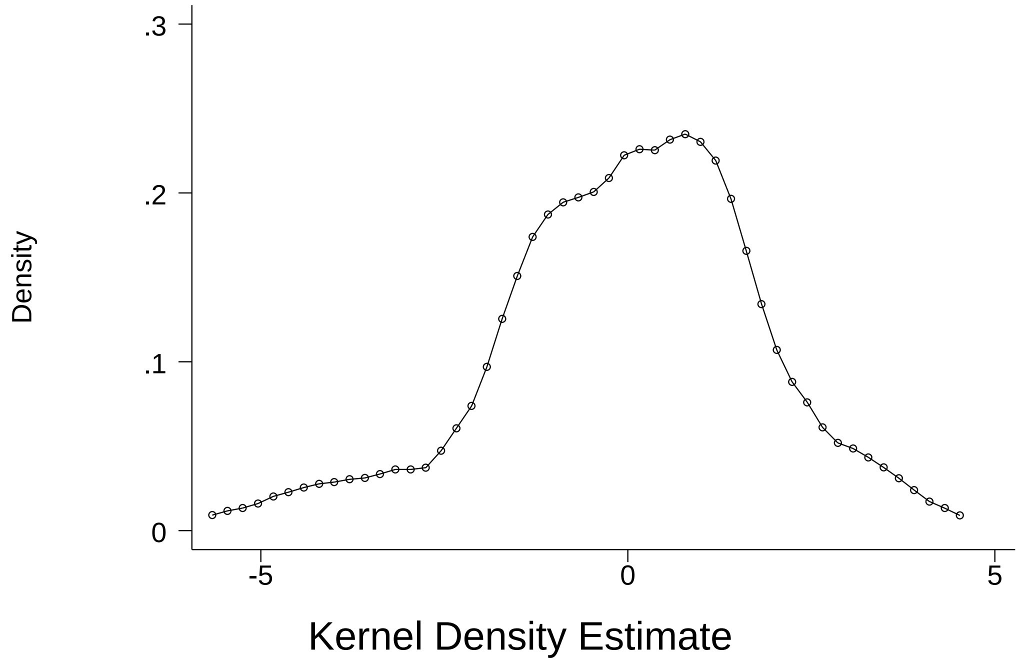


Figure 3: *Kernel density estimator of the posterior mean of the spatial effect*

Classification table

y_{it}	$\hat{y_{it}}$		
	1	2	3
1	896	74	1
2	116	366	11
3	0	40	45

Missclassified: 15.5%

Classification table if spatial effect is not considered

y_{it}	$\hat{y_{it}}$		
	1	2	3
1	860	111	0
2	233	256	4
3	12	69	4

Missclassified: 27.6%