

# Bayesian Structured Hazard Regression

Andrea Hennerfeind, Ludwig Fahrmeir & Thomas Kneib  
Department of Statistics  
Ludwig-Maximilians-University Munich



29.05.2006

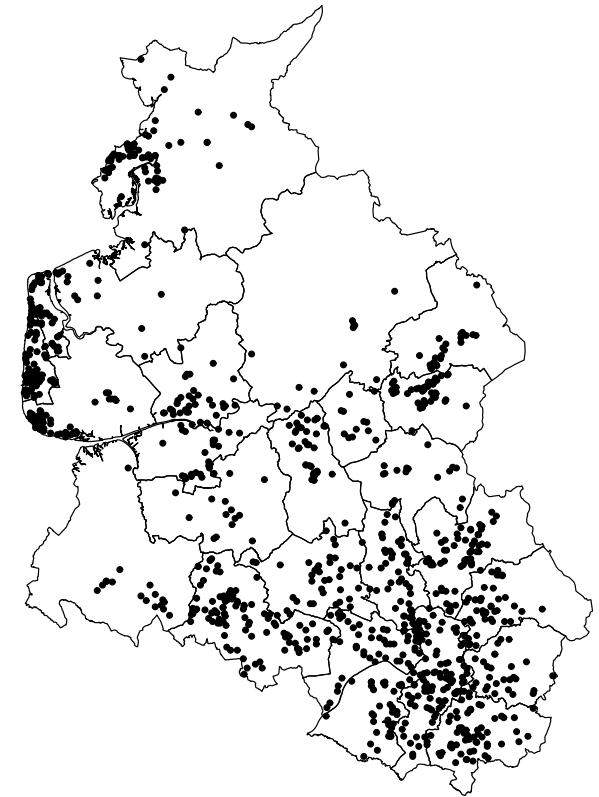


# Outline

- Leukemia survival data.
- Structured hazard regression for survival times.
- Bayesian inference in structured hazard regression.
  - Full Bayesian inference based on MCMC.
  - Empirical Bayes inference using mixed model methodology.
- Multi-state models for the analysis of human sleep.

## Leukemia Survival Data

- Survival time of adults after diagnosis of acute myeloid leukemia.
- 1,043 cases diagnosed between 1982 and 1998 in Northwest England.
- 16 % (right) censored.
- **Continuous** and **categorical** covariates:
  - age* age at diagnosis,
  - wbc* white blood cell count at diagnosis,
  - sex* sex of the patient,
  - tpi* Townsend deprivation index.
- **Spatial information** in different resolution.



- Classical Cox **proportional hazards model**:

$$\lambda(t; x) = \lambda_0(t) \exp(x' \gamma).$$

- **Baseline hazard**  $\lambda_0(t)$  is a nuisance parameter and **remains unspecified**.
- Estimate  $\gamma$  based on the partial likelihood.
- Questions / Limitations:
  - **Simultaneous estimation** of baseline hazard rate and covariate effects.
  - **Flexible** modelling of covariate effects (e.g. nonlinear effects, interactions).
  - **Spatially correlated** survival times.
  - **Non-proportional hazards** models / **time-varying effects**.

⇒ **Structured hazard regression models.**

- Replace usual parametric predictor with a **flexible semiparametric** predictor

$$\lambda(t; \cdot) = \lambda_0(t) \exp[f_1(\text{age}) + f_2(\text{wbc}) + f_3(\text{tpi}) + f_{\text{spat}}(s_i) + \beta_1 \text{sex}]$$

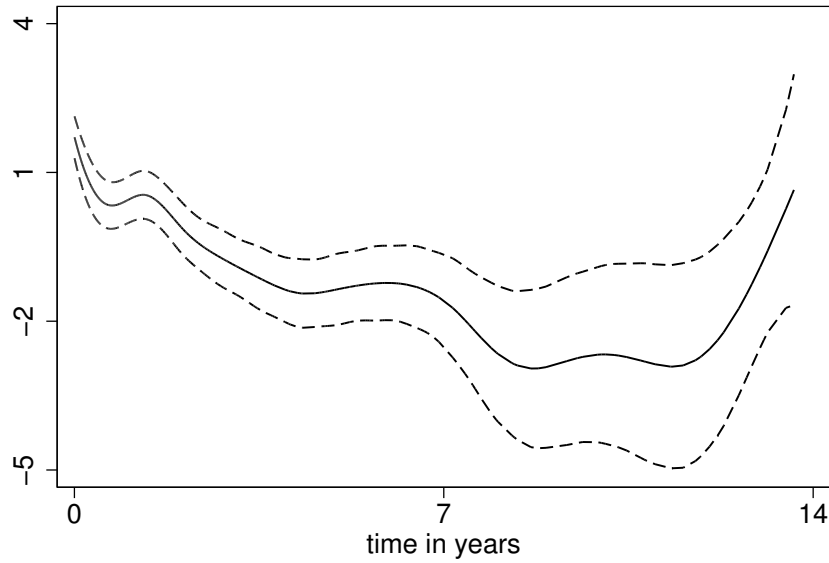
and **absorb the baseline**

$$\lambda(t; \cdot) = \exp[g_0(t) + f_1(\text{age}) + f_2(\text{wbc}) + f_3(\text{tpi}) + f_{\text{spat}}(s_i) + \beta_1 \text{sex}]$$

where

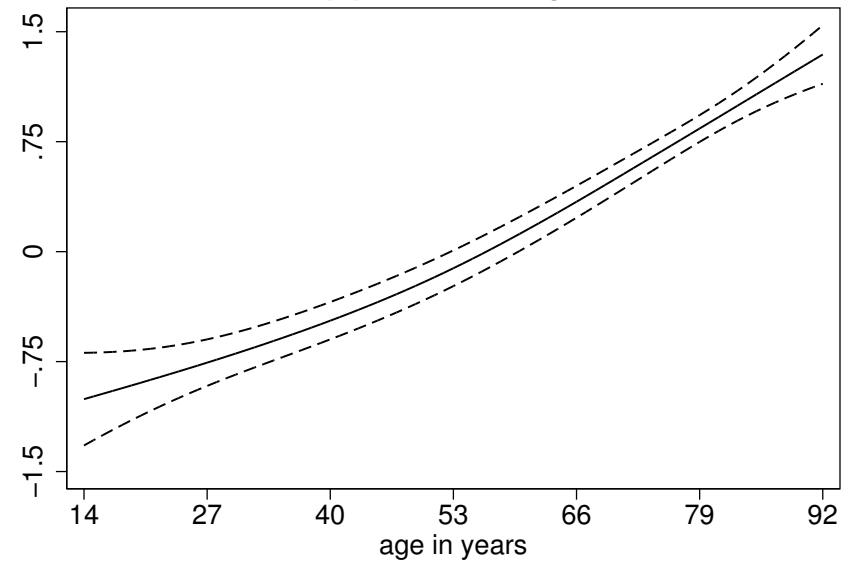
- $g_0(t) = \log(\lambda_0(t))$  is the **log-baseline hazard**,
- $f_1, f_2, f_3$  are **nonparametric** functions of age, white blood cell count and deprivation, and
- $f_{\text{spat}}$  is a **spatial** function.

(a) log(baseline)



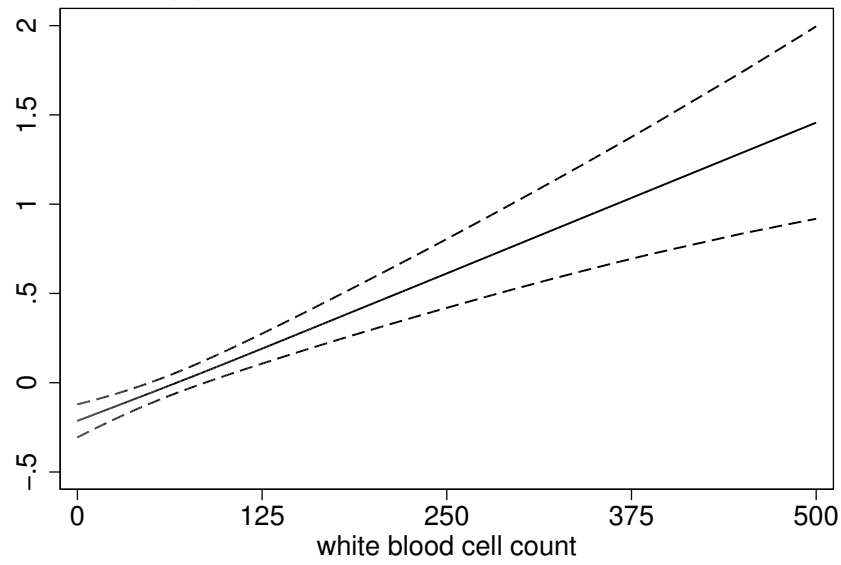
Log-baseline hazard.

(b) effect of age



Effect of age at diagnosis.

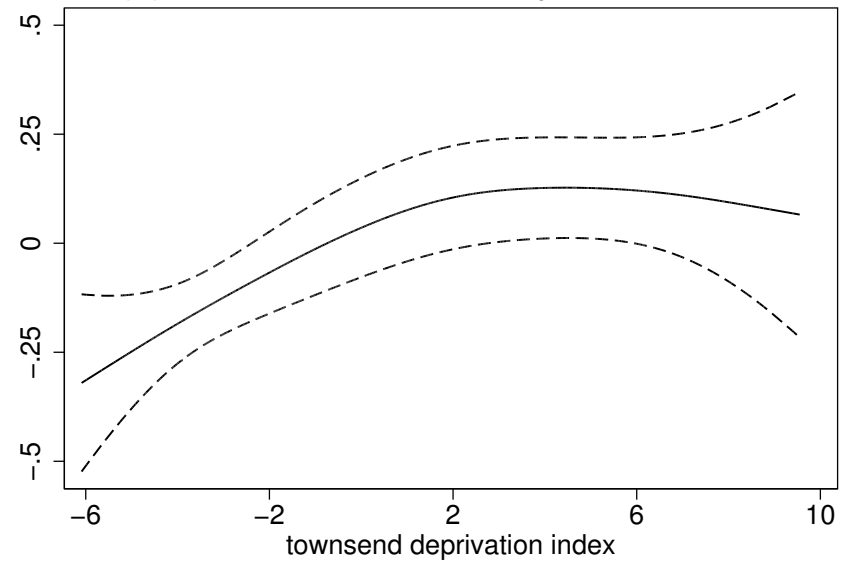
(c) effect of white blood cell count

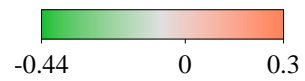
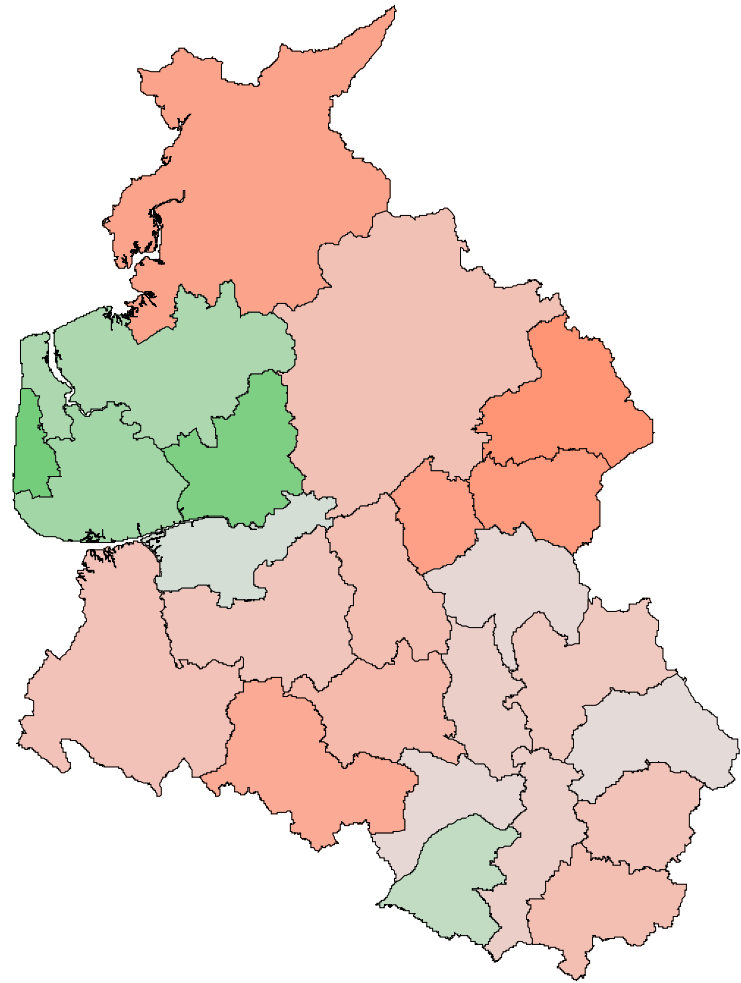


Effect of white blood cell count.

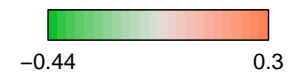
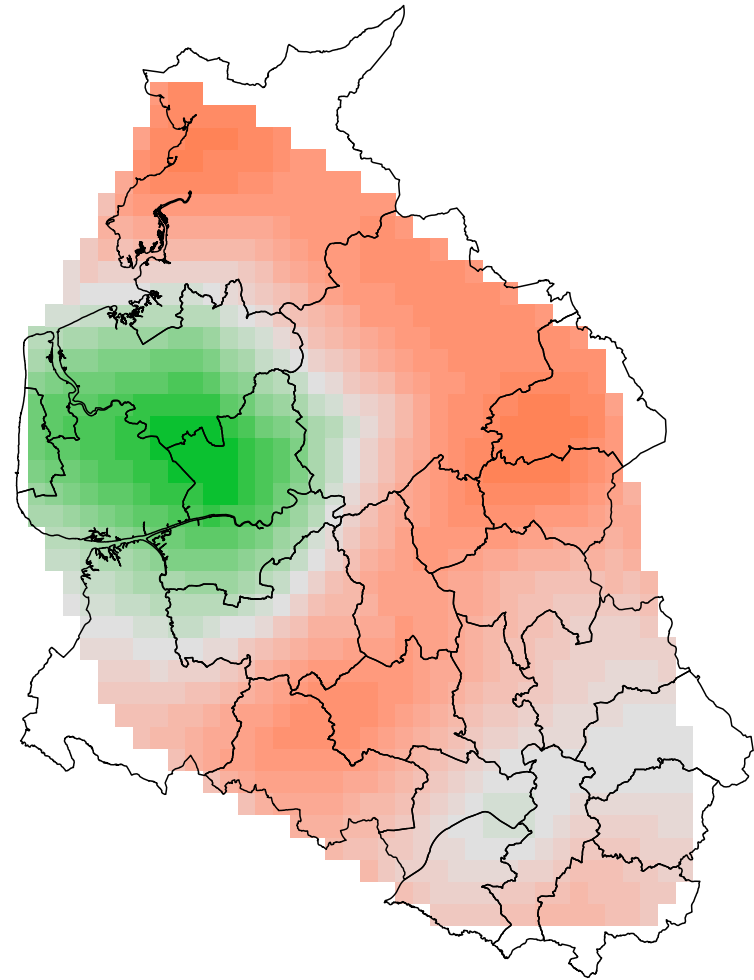
Effect of deprivation.

(d) effect of townsend deprivation index





District-level analysis



Individual-level analysis

# Structured Hazard Regression

- A general structured hazard regression model consists of an arbitrary combination of the following model terms:
  - Log baseline hazard  $g_0(t) = \log(\lambda_0(t))$ .
  - Time-varying effects  $g_l(t)u_l$  of covariates  $u_l$ .
  - Nonparametric effects  $f_j(x_j)$  of continuous covariates  $x_j$ .
  - Spatial effects  $f_{spat}(s)$  of a spatial location variable  $s$ .
  - Interaction surfaces  $f_{j,k}(x_j, x_k)$  of two continuous covariates.
  - Varying coefficient interactions  $u_j f_k(x_k)$  or  $u_j f_{spat}(s)$ .
  - Frailty terms  $b_g$  (random intercept) or  $x_j b_g$  (random slopes).

- **Penalised splines** for the baseline effect, time-varying effects, and nonparametric effects:
  - Approximate  $f(x)$  (or  $g(t)$ ) by a weighted sum of **B-spline basis** functions

$$f(x) = \sum \xi_j B_j(x).$$

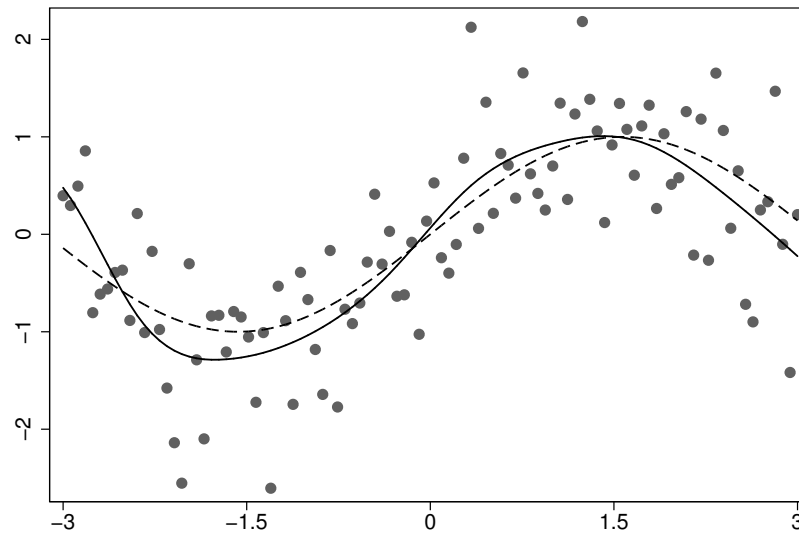
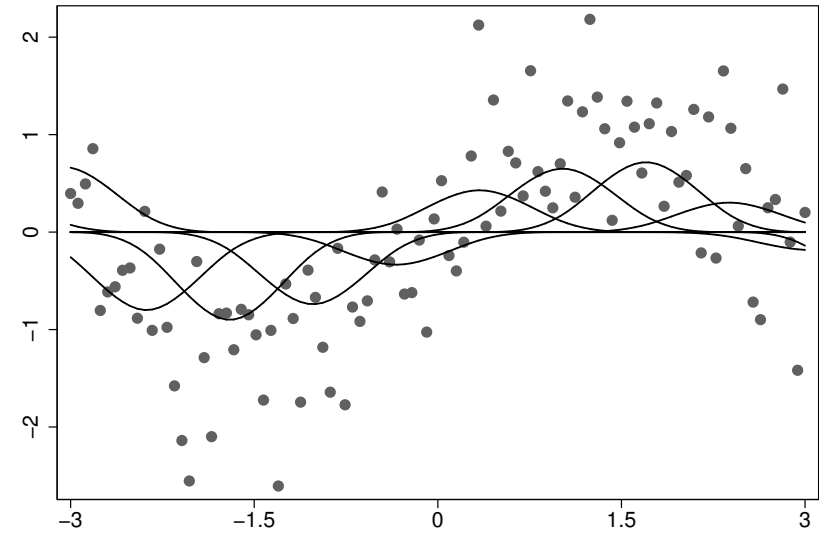
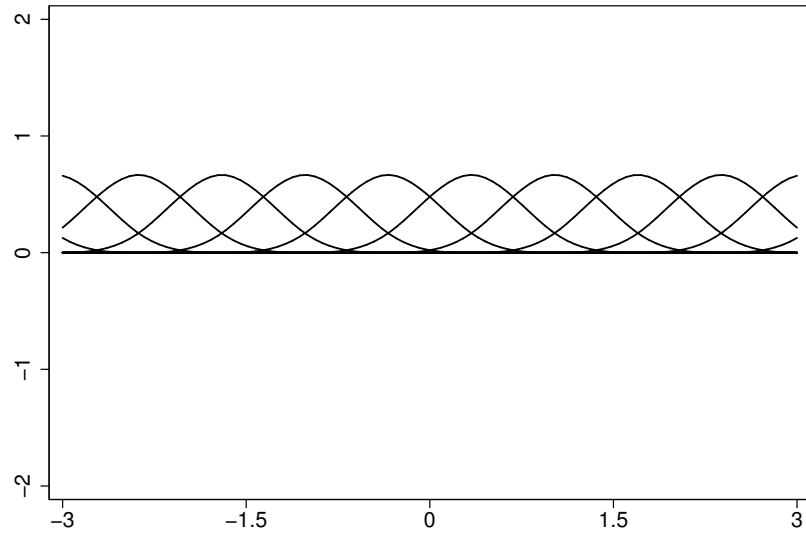
- Employ a large number of basis functions to enable flexibility.
- **Penalise differences** between parameters of adjacent basis functions to ensure smoothness:

$$Pen(\xi|\tau^2) = \frac{1}{2\tau^2} \sum (\Delta_k \xi_j)^2.$$

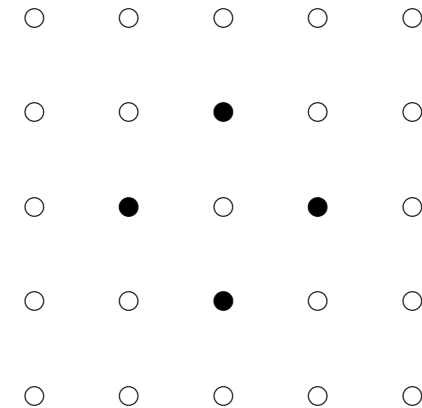
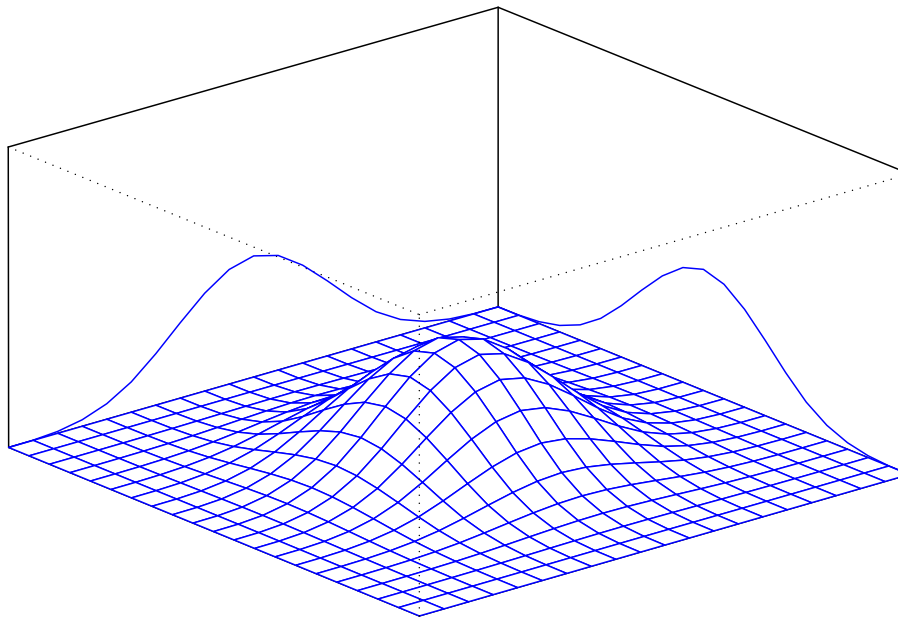
- Bayesian interpretation: Assume a  $k$ -th order **random walk prior** for  $\xi_j$ , e.g.

$$\xi_j = \xi_{j-1} + u_j, \quad u_j \sim N(0, \tau^2) \quad (\text{RW1}).$$

$$\xi_j = 2\xi_{j-1} - \xi_{j-2} + u_j, \quad u_j \sim N(0, \tau^2) \quad (\text{RW2}).$$

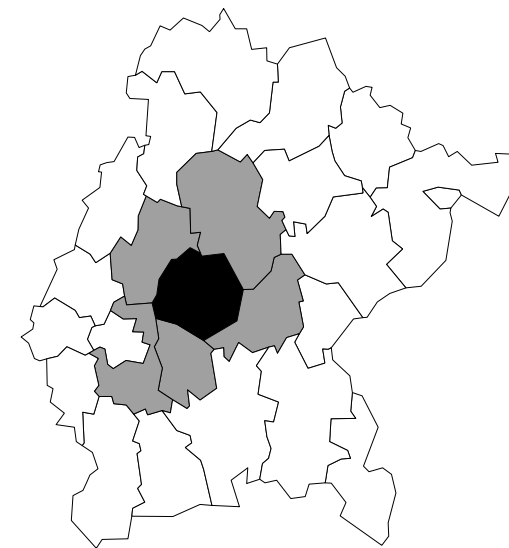
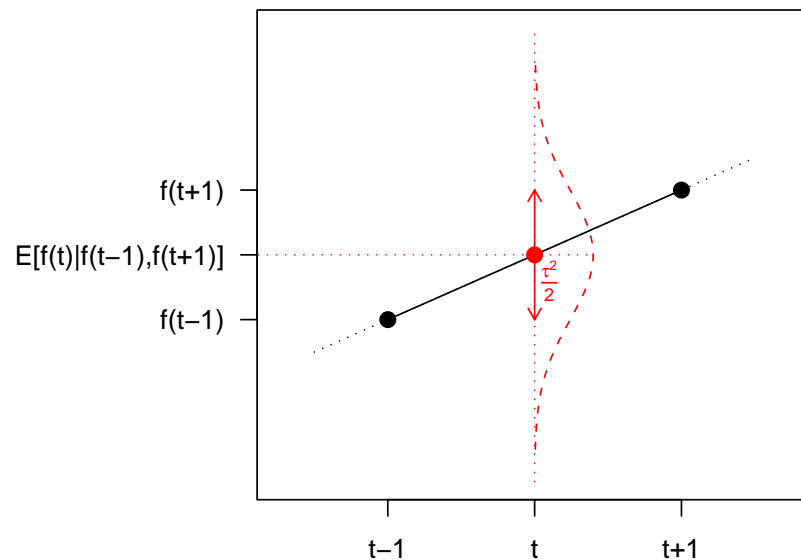


- **Bivariate** Tensor product P-splines for interaction surfaces:
  - Define bivariate basis functions (Tensor products of univariate basis functions).
  - Extend random walks on the line to random walks on a regular grid.

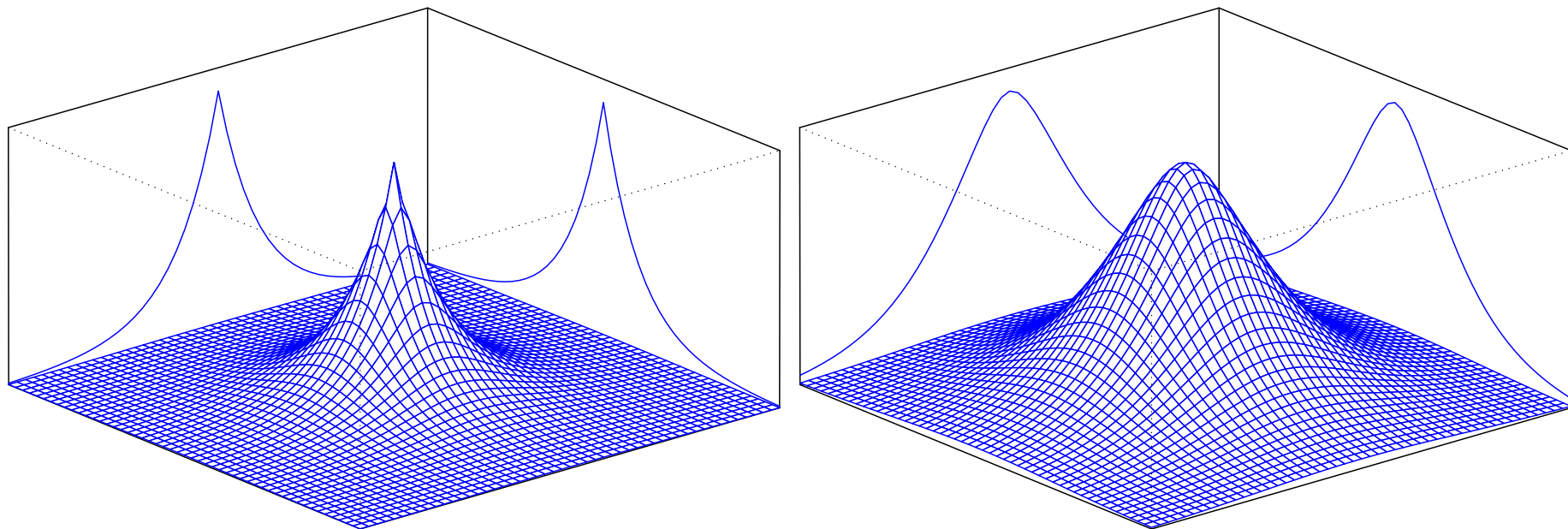


- Spatial effects for regional data  $s \in \{1, \dots, S\}$ : **Markov random fields**.
  - Bivariate extension of a first order random walk on the real line.
  - Define appropriate **neighbourhoods** for the regions.
  - Assume that the expected value of  $f_{spat}(s) = \xi_s$  is the **average of the function evaluations of adjacent sites**:

$$\xi_s | \xi_{s'}, s' \neq s, \tau^2 \sim N \left( \frac{1}{N_s} \sum_{s' \in \partial_s} \xi_{s'}, \frac{\tau^2}{N_s} \right).$$



- Spatial effects for point-referenced data: **Stationary Gaussian random fields**.
  - Well-known as **Kriging** in the geostatistics literature.
  - Spatial effect follows a zero mean stationary Gaussian stochastic process.
  - Correlation of two arbitrary sites is defined by an **intrinsic correlation function**.
  - Can be interpreted as a basis function approach with **radial basis functions**.



- Cluster-specific **frailty terms**:
  - Account for unobserved heterogeneity.
  - Easiest case: i.i.d Gaussian frailty.
- All covariates in the discussed model terms are allowed to be **piecewise constant time-varying**.

# Bayesian Inference

- **Generic representation** of structured hazard regression models:

$$\lambda(t) = \exp [x(t)' \gamma + f_1(z_1(t)) + \dots + f_p(z_p(t))]$$

- For example:

$$f(z(t)) = g(t)$$

$$z(t) = t$$

log-baseline effect,

$$f(z(t)) = u(t)g(t)$$

$$z(t) = (u, t)$$

time-varying effect of  $u(t)$ ,

$$f(z(t)) = f(x(t))$$

$$z(t) = x(t)$$

smooth function of a continuous covariate  $x(t)$ ,

$$f(z(t)) = f_{spat}(s)$$

$$z(t) = s$$

spatial effect,

$$f(z(t)) = f(x_1(t), x_2(t))$$

$$z(t) = (x_1(t), x_2(t))$$

interaction surface,

$$f(z(t)) = b_g$$

$$z(t) = g$$

i.i.d. frailty  $b_g$ ,  $g$  is a grouping index.

- The generic representation facilitates description of inferential details.

- All vectors of function evaluations  $f_j$  can be expressed as

$$f_j = Z_j \xi_j$$

with design matrix  $Z_j$ , constructed from  $z_j(t)$ , and regression coefficients  $\xi_j$ .

- **Generic form** of the prior for  $\xi_j$ :

$$p(\xi_j | \tau_j^2) \propto (\tau_j^2)^{-\frac{k_j}{2}} \exp\left(-\frac{1}{2\tau_j^2} \xi_j' K_j \xi_j\right)$$

- $K_j \geq 0$  acts as a **penalty matrix**,  $\text{rank}(K_j) = k_j \leq d_j = \text{dim}(\xi_j)$ .
- $\tau_j^2 \geq 0$  can be interpreted as a **variance** or (inverse) **smoothness parameter**.
- Relation to **penalized likelihood**: Penalty terms

$$P_{\lambda_j}(\xi_j) = \log[p(\xi_j | \tau_j^2)] = -\frac{1}{2} \lambda_j \xi_j' K_j \xi_j, \quad \lambda_j = \frac{1}{\tau_j^2}.$$

- Likelihood for right censored survival times under the assumption of noninformative censoring:

$$\prod_{i=1}^n \lambda_i(T_i)^{\delta_i} \exp \left( - \int_0^{T_i} \lambda_i(t) dt \right),$$

where  $\delta_i$  is the censoring indicator.

- In general, **numerical integration** has to be used to evaluate the cumulative hazard rate (e.g. the trapezoidal rule).

## Fully Bayesian inference based on MCMC

- Assign **inverse gamma prior** to  $\tau_j^2$ :

$$p(\tau_j^2) \propto \frac{1}{(\tau_j^2)^{a_j+1}} \exp\left(-\frac{b_j}{\tau_j^2}\right).$$

Proper for	$a_j > 0, b_j > 0$	Common choice: $a_j = b_j = \varepsilon$ small.
Improper for	$b_j = 0, a_j = -1$	Flat prior for variance $\tau_j^2$ ,
	$b_j = 0, a_j = -\frac{1}{2}$	Flat prior for standard deviation $\tau_j$ .

- **Conditions for proper posteriors** in structured hazard regression: Enough uncensored observations and either
  - proper priors for the variances or
  - $a_j < b_j = 0$  and rank deficiency in the prior for  $\xi_j$  not too large.

- MCMC sampling scheme:

- **Metropolis-Hastings** update for  $\xi_j | \cdot$ :

Propose new state from a multivariate Gaussian distribution with precision matrix and mean

$$P_j = Z_j' W Z_j + \frac{1}{\tau_j^2} K_j \quad \text{and} \quad m_j = P_j^{-1} Z_j' W (\tilde{y} - \eta_{-j}).$$

**IWLS-Proposal** with appropriately defined working weights  $W$  and working observations  $\tilde{y}$ .

- **Gibbs sampler** for  $\tau_j^2 | \cdot$ :

Sample from an inverse Gamma distribution with parameters

$$a'_j = a_j + \frac{1}{2} \text{rank}(K_j) \quad \text{and} \quad b'_j = b_j + \frac{1}{2} \xi_j' K_j \xi_j.$$

- Efficient algorithms make use of the sparse matrix structure of  $P_j$  and  $K_j$ .

# Empirical Bayes inference based on mixed model methodology

- Consider the variances  $\tau_j^2$  as **unknown constants** to be estimated.
- Idea: Consider  $\xi_j$  a **correlated random effect** with multivariate Gaussian distribution and use mixed model methodology.
- Problem: In most cases **partially improper random effects distribution**.
- Mixed model representation: Decompose

$$\xi_j = X_j\beta_j + Z_j b_j,$$

where

$$p(\beta_j) \propto \text{const} \quad \text{and} \quad b_j \sim N(0, \tau_j^2 I_{k_j}).$$

$\Rightarrow \beta_j$  is a **fixed effect** and  $b_j$  is an **i.i.d. random effect**.

- This yields the **variance components model**

$$\lambda(t; \cdot) = \exp [x'\beta + z'b],$$

where in turn

$$p(\beta) \propto \text{const} \quad \text{and} \quad b \sim N(0, Q).$$

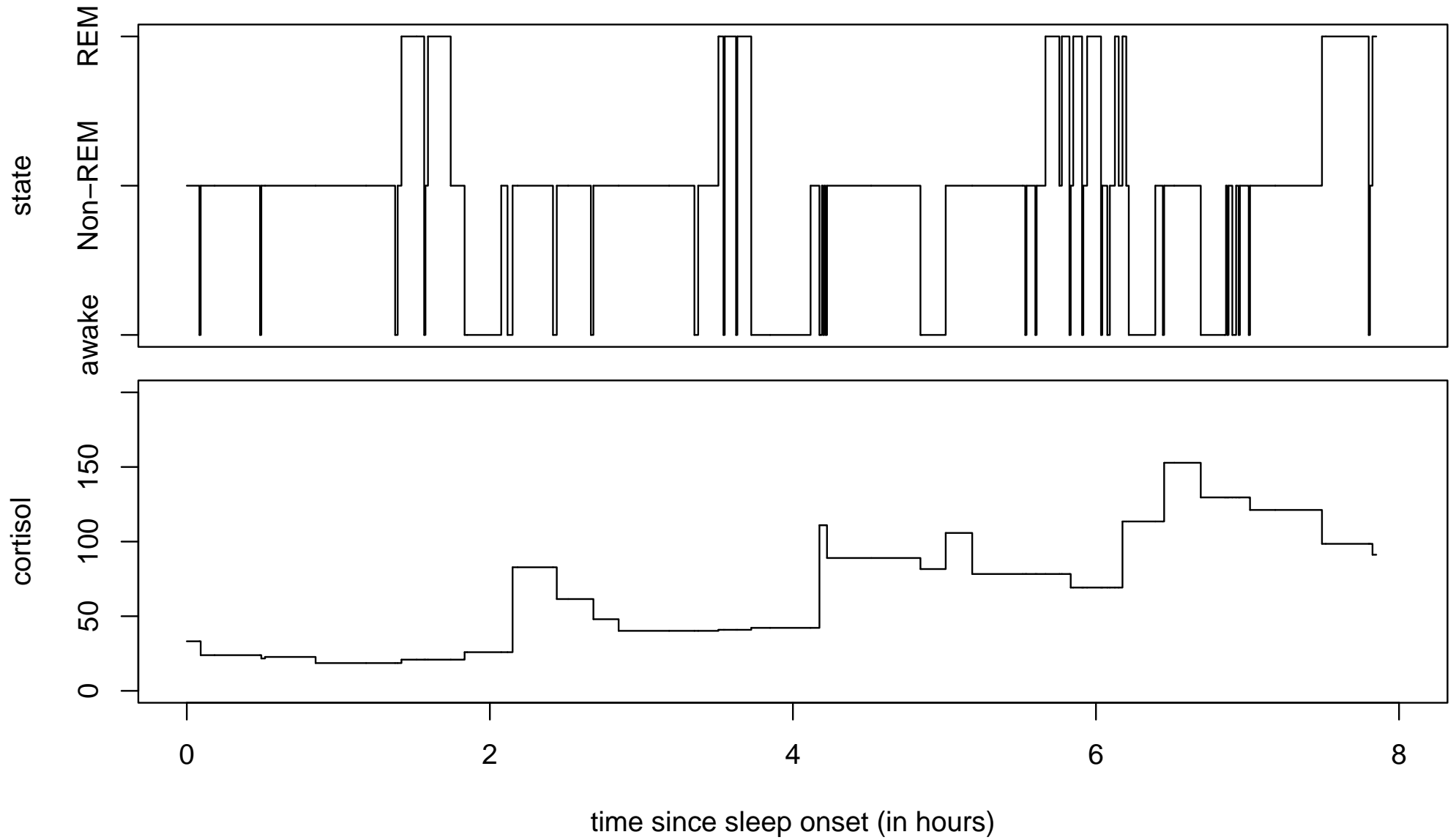
- Obtain **empirical Bayes estimates** / **penalized likelihood estimates** via iterating
  - Penalized maximum likelihood for the regression coefficients  $\beta$  and  $b$ .
  - Restricted Maximum / Marginal likelihood for the variance parameters in  $Q$ :

$$L(Q) = \int L(\beta, b, Q)p(b)d\beta db \rightarrow \max_Q.$$

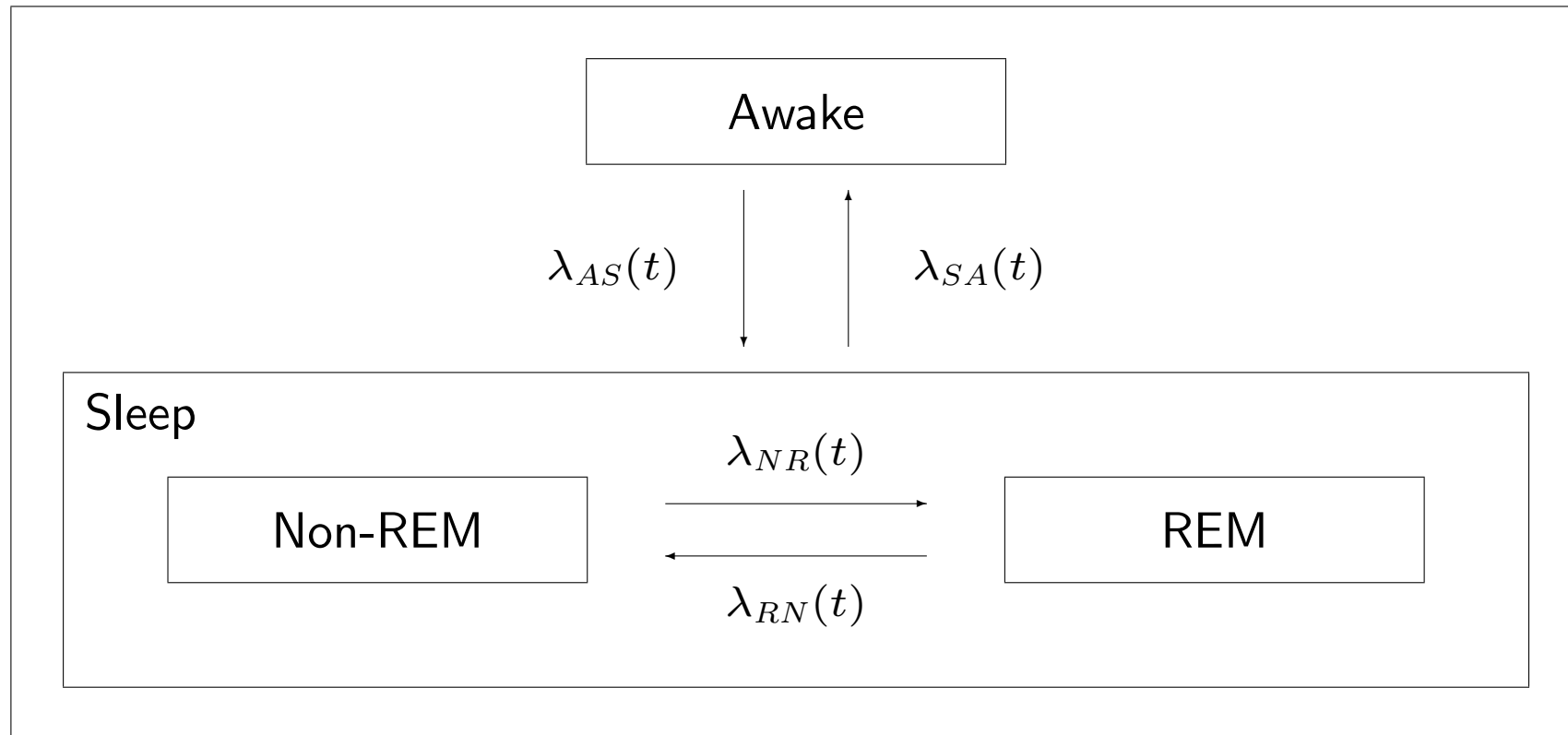
- Involves Laplace approximation to the marginal likelihood (similar as in the previous talk by Håvard Rue).
- Corresponds to REML estimation of variances in Gaussian mixed models.

# Human Sleep Data

- Consider individual human sleep data as independent realisations of **time-continuous** stochastic processes with **discrete state space** {awake, REM, non-REM}.
  - Compact description of such a process in terms of **transition intensities** between these states.
  - Simple approaches: Markov or Semi-Markov processes.
  - Limitations / Questions:
    - **Changing dynamics** of human sleep over night.
    - **Individual sleeping habits** to be described by covariates.
    - Only a small number of covariates is available (**unobserved heterogeneity**).
- ⇒ Model the transition intensities in analogy to survival models.



- Simplified structure for the transitions:



- Model for the transitions:

$$\begin{aligned}\lambda_{AS,i}(t) &= \exp \left[ \gamma_0^{(AS)}(t) + s_i \beta^{(AS)} + b_i^{(AS)} \right] \\ \lambda_{SA,i}(t) &= \exp \left[ \gamma_0^{(SA)}(t) + s_i \beta^{(SA)} + b_i^{(SA)} \right] \\ \lambda_{NR,i}(t) &= \exp \left[ \gamma_0^{(NR)}(t) + c_i(t) \gamma_1^{(NR)}(t) + s_i \beta^{(NR)} + b_i^{(NR)} \right] \\ \lambda_{RN,i}(t) &= \exp \left[ \gamma_0^{(RN)}(t) + c_i(t) \gamma_1^{(RN)}(t) + s_i \beta^{(RN)} + b_i^{(RN)} \right]\end{aligned}$$

where

$$c_i(t) = \begin{cases} 1 & \text{cortisol} > 60 \text{ n mol/l at time } t \\ 0 & \text{cortisol} \leq 60 \text{ n mol/l at time } t, \end{cases}$$

$$s_i = \begin{cases} 1 & \text{male} \\ 0 & \text{female,} \end{cases}$$

$$b_i^{(j)} = \text{transition- and individual-specific frailty.}$$

- Use penalized splines for the baseline and time-varying effects.
- I.i.d. Gaussian priors for the frailty terms (with transition-specific variances).
- The likelihood contribution for individual  $i$  can be constructed based on a **counting process formulation** of the model:

$$\begin{aligned}
 l_i &= \sum_{h=1}^k \left[ \int_0^{T_i} \log(\lambda_{hi}(t)) dN_{hi}(t) - \int_0^{T_i} \lambda_{hi}(t) Y_{hi}(t) dt \right] \\
 &= \sum_{j=1}^{n_i} \sum_{h=1}^k \left[ \delta_{hi}(t_{ij}) \log(\lambda_{hi}(t_{ij})) - Y_{hi}(t_{ij}) \int_{t_{i,j-1}}^{t_{ij}} \lambda_{hi}(t) dt \right].
 \end{aligned}$$

$N_{hi}(t)$  counting process for type  $h$  event.

$Y_{hi}(t)$  at risk indicator for type  $h$  event.

$t_{ij}$  event times of individual  $i$ .

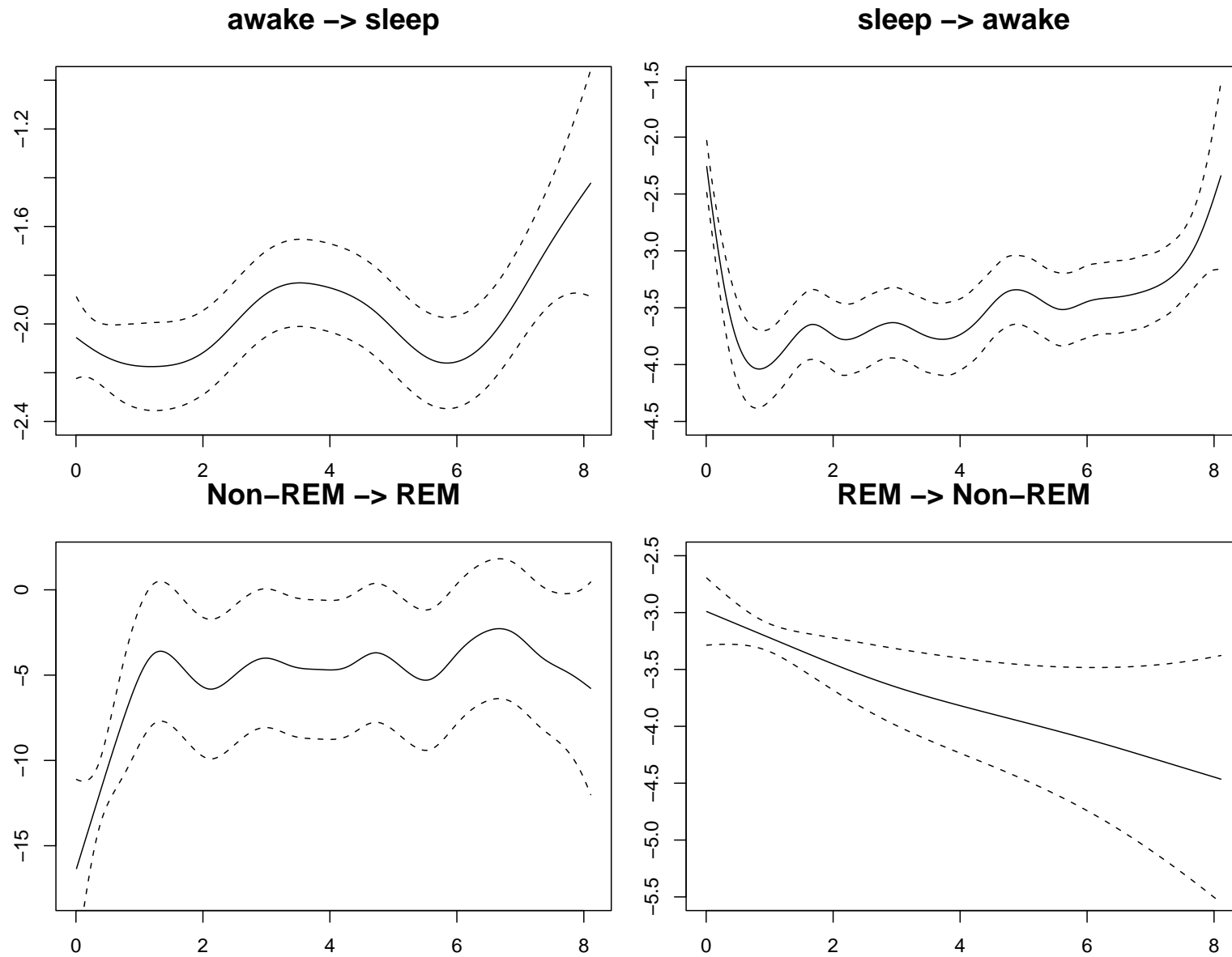
$n_i$  number of events for individual  $i$ .

$\delta_{hi}(t_{ij})$  transition indicator for type  $h$  transition.

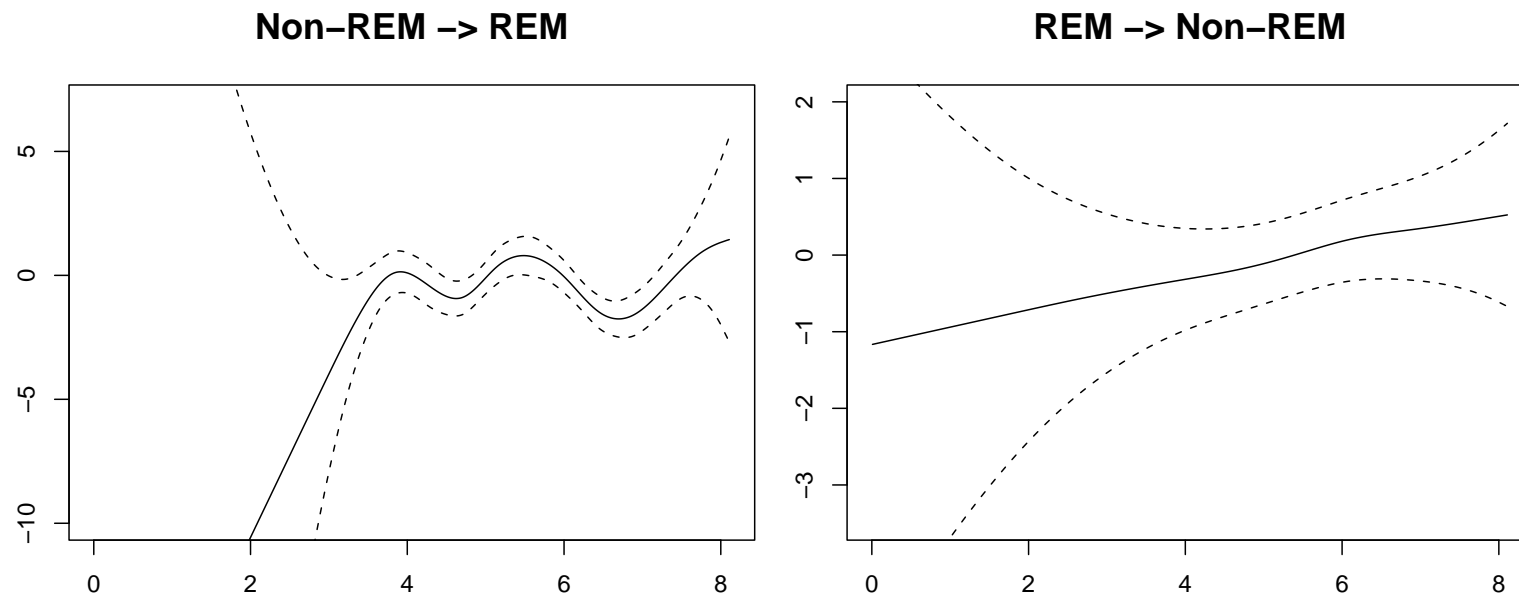
⇒ Concepts from survival analysis can be adapted.

- In particular:
  - Fully Bayesian inference based on MCMC and
  - Mixed model based empirical Bayes inference.

- Baseline effects:



- Time-varying effects for a high level of cortisol:



## Software

- Estimation was carried out using BayesX.
- Public domain software package for Bayesian inference in geosadditive and related models.



- Available from

<http://www.stat.uni-muenchen.de/~bayesx>

# Conclusions

- **Unified framework** for general regression models describing the hazard rate of survival models.
- Bayesian inference based on MCMC or mixed model methodology.
- Extendable to models for transition intensities in **multi state models**.
- Future work:
  - More general censoring mechanisms.
  - Conditions for propriety of posteriors.
  - Joint modelling of covariates and duration times.

## References

- BREZGER, KNEIB & LANG (2005). BayesX: Analyzing Bayesian structured additive regression models. *Journal of Statistical Software*, **14** (11).
- FAHRMEIR, KNEIB & LANG (2004): Penalized structured additive regression for space-time data: a Bayesian perspective. *Statistica Sinica* **14**, 731–761.
- HENNERFEIND, BREZGER, AND FAHRMEIR (2006): Geoadditive Survival Models. *Journal of the American Statistical Association*, to appear.
- KNEIB & FAHRMEIR (2006): A mixed model approach for geoadditive hazard regression. *Scandinavian Journal of Statistics*, to appear.
- KNEIB (2006): Geoadditive hazard regression for interval censored survival times. *Computational Statistics and Data Analysis*, conditionally accepted.