Bayesian Structured Hazard Regression

Thomas Kneib

Department of Statistics
Ludwig-Maximilians-University Munich

Joint work with
Ludwig Fahrmeir & Andrea Hennerfeind

LUDWIG-
MAXIMILIANS-
UNIVERSITÄT
MÜNCHEN

16.11.2007

Outline

1. Geoadditive Modelling of Continuous Survival Times.

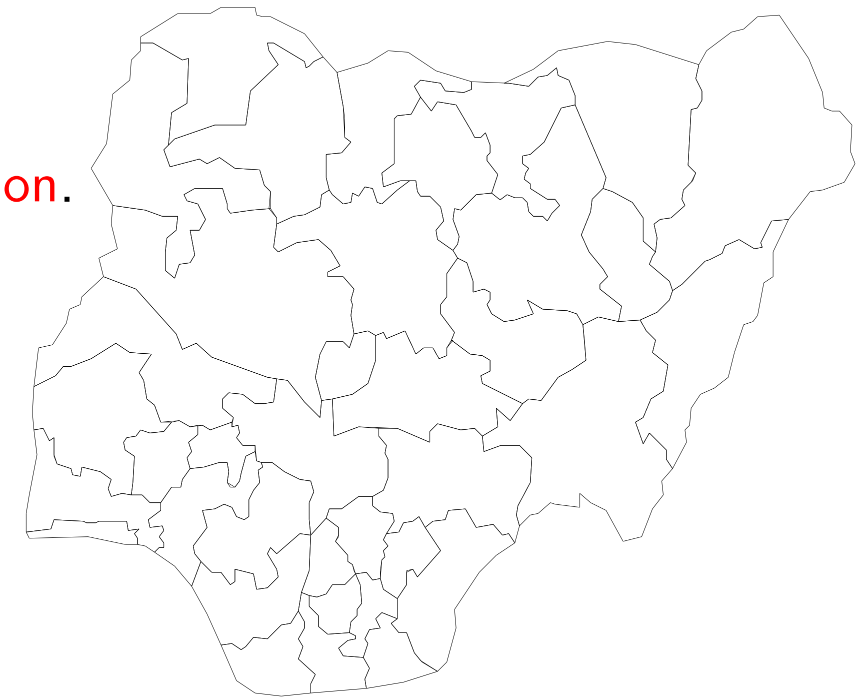
2. Inferential Concepts: Empirical Bayes vs. Full Bayes.

3. Continuous Time Multi-State Models.

Childhood mortality in Nigeria

- Data from the 2003 Demographic and Health Survey (DHS) in Nigeria.
- Retrospective questionnaire on the health status of women in reproductive age and their children.
- Survival time of $n = 5323$ children.
- Numerous covariates including spatial information.
- Analysis based on the Cox model:

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

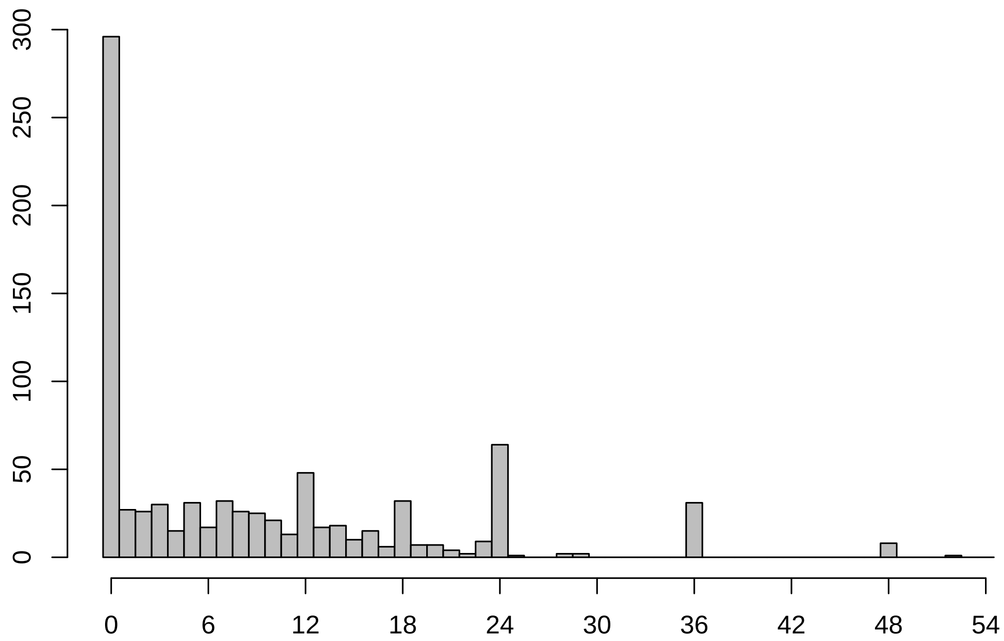


- Limitations of the classical Cox model:
  - Restricted to right censored observations.
  - Post-estimation of the baseline hazard.
  - Proportional hazards assumption.
  - Parametric form of the predictor.
  - No spatial correlations.

$\Longrightarrow$ Structured hazard regression.

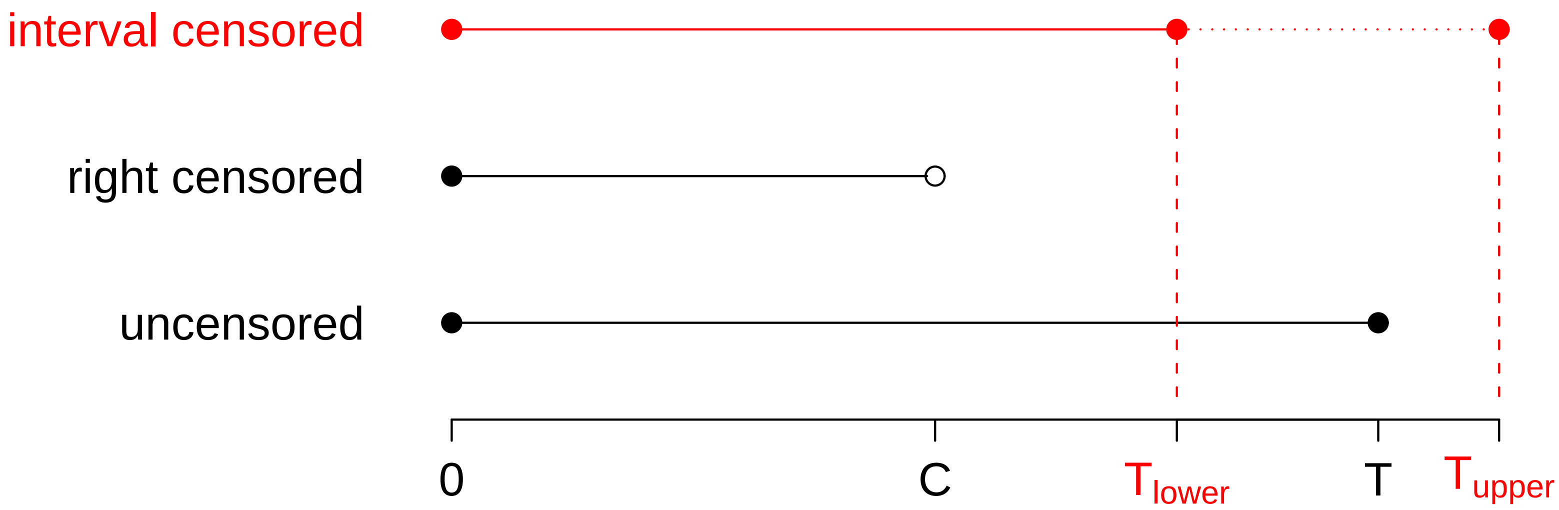
Interval censored survival times

- In theory, survival times should be available in days.
- Retrospective questionnaire ⇒ most uncensored survival times are rounded (Heaping).



- In contrast: censoring times are given in days.

⇒ Treat survival times as interval censored.


interval censored
right censored
uncensored
0
C
T_lower
T
T_upper

- Likelihood contributions:

$$P(T > C) = S(C) = \exp\left[-\int_0^C \lambda(t)dt\right].$$

$$P(T \in [T_{lower}, T_{upper}]) = S(T_{lower}) - S(T_{upper}) = \exp\left[-\int_0^{T_{lower}} \lambda(t)dt\right] - \exp\left[-\int_0^{T_{upper}} \lambda(t)dt\right].$$

- Derivatives of the log-likelihood become much more complicated for interval censored survival times.

- Numerical integration techniques have to be used in both cases.

- Piecewise constant time-varying covariates and left truncation can easily be included.

Structured hazard regression

- Introduce a more flexible, semiparametric hazard rate model

$$\lambda(t;\cdot) = \exp\left[g_0(t) + \sum_{j=1}^{q} g_j(t) z_j(t) + \sum_{k=1}^{p} f_k(x_k(t)) + f_{spat}(s) + u(t)'\gamma\right]$$

where

– $g_0(t) = \log(\lambda_0(t))$ is the log-baseline-hazard,

– g_j are time varying effects of covariates $z_j(t)$,

– f_k are nonparametric functions of continuous covariates $x_k(t)$,

– f_{spat} is a spatial function,

– $u(t)'\gamma$ are parametric effects.

Model Components and Priors

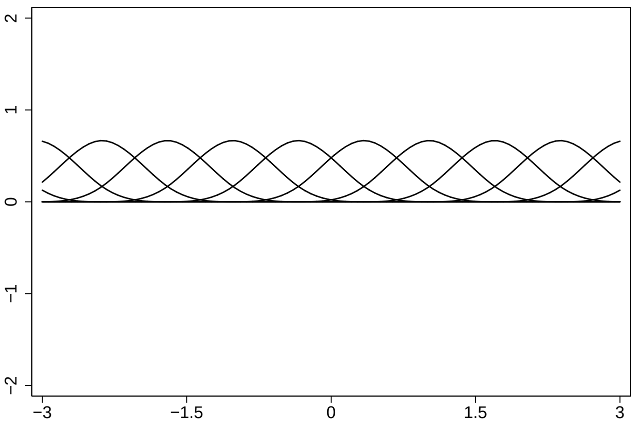
- Penalised splines for log-baseline, time-varying effects and nonparametric effects.
  - Approximate g_j (or f_k) by a weighted sum of B-spline basis functions

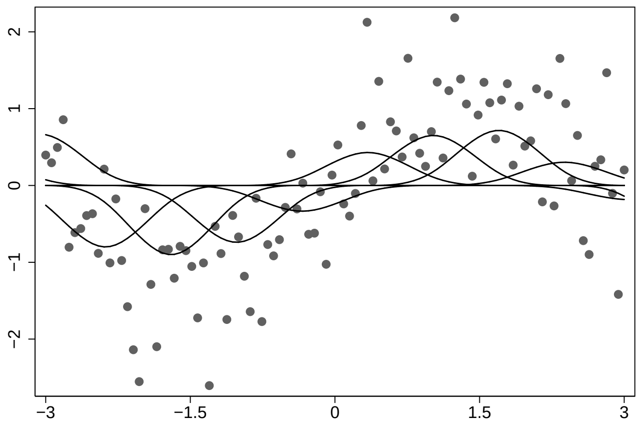
$$f(x) = \sum \xi_j B_j(x).$$

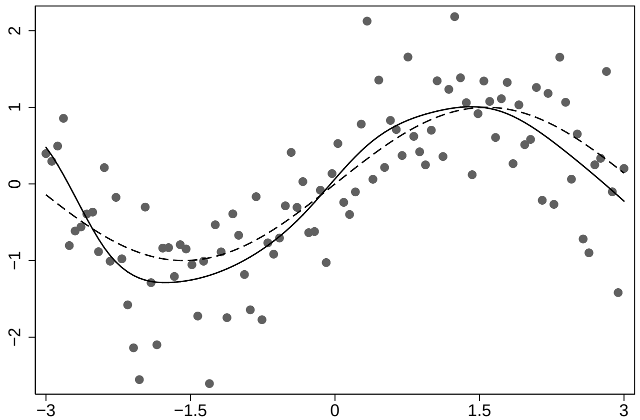
  - Employ a large number of basis functions to enable flexibility.
  - Penalise differences between parameters of adjacent basis functions to ensure smoothness

$$\frac{1}{2\tau^2} \sum (\xi_j - \xi_{j-1})^2 \qquad \text{(first order differences)}$$

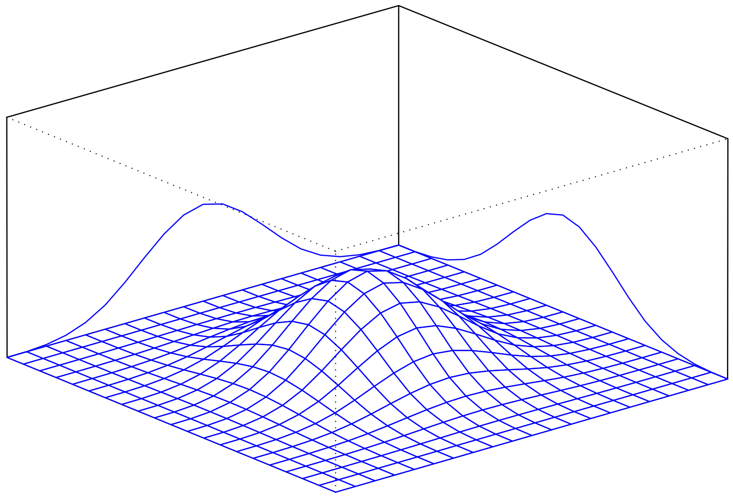
$$\frac{1}{2\tau^2} \sum (\xi_j - 2\xi_{j-1} + \xi_{j-2})^2 \qquad \text{(second order differences)}$$

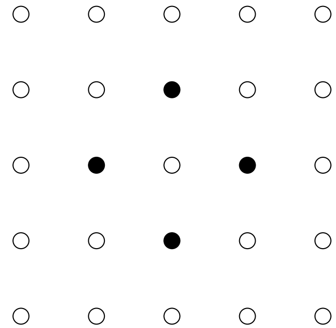

2
1
0
−1
−2
−3
−1.5
0
1.5
3


2
1
0
−1
−2
−3
−1.5
0
1.5
3


2
1
0
−1
−2
−3
−1.5
0
1.5
3

- Bivariate penalised splines.



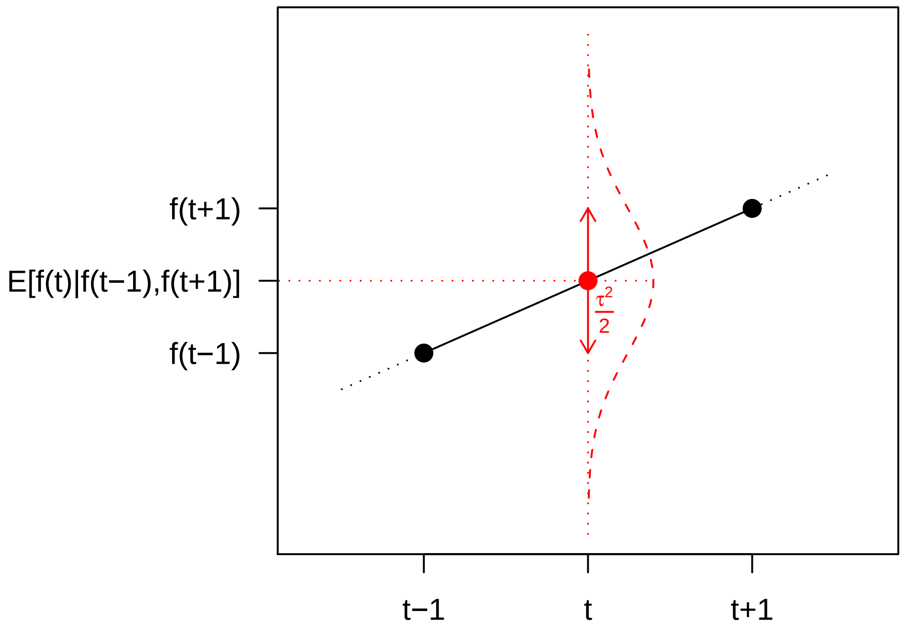


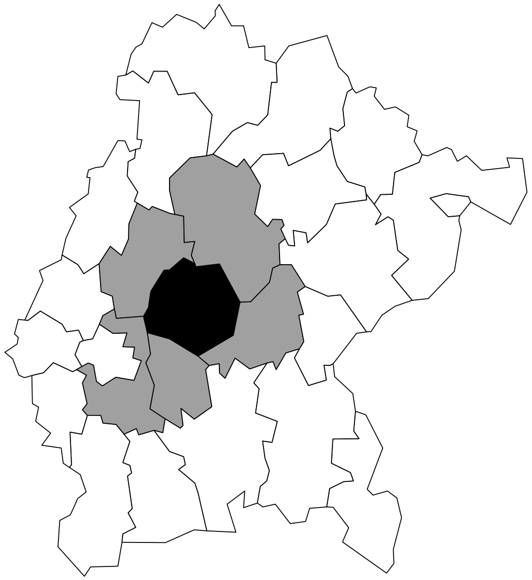
- Varying coefficient models.
  - Effect of covariate x varies smoothly over the domain of a second covariate z:

$$f(x, z) = x \cdot g(z)$$

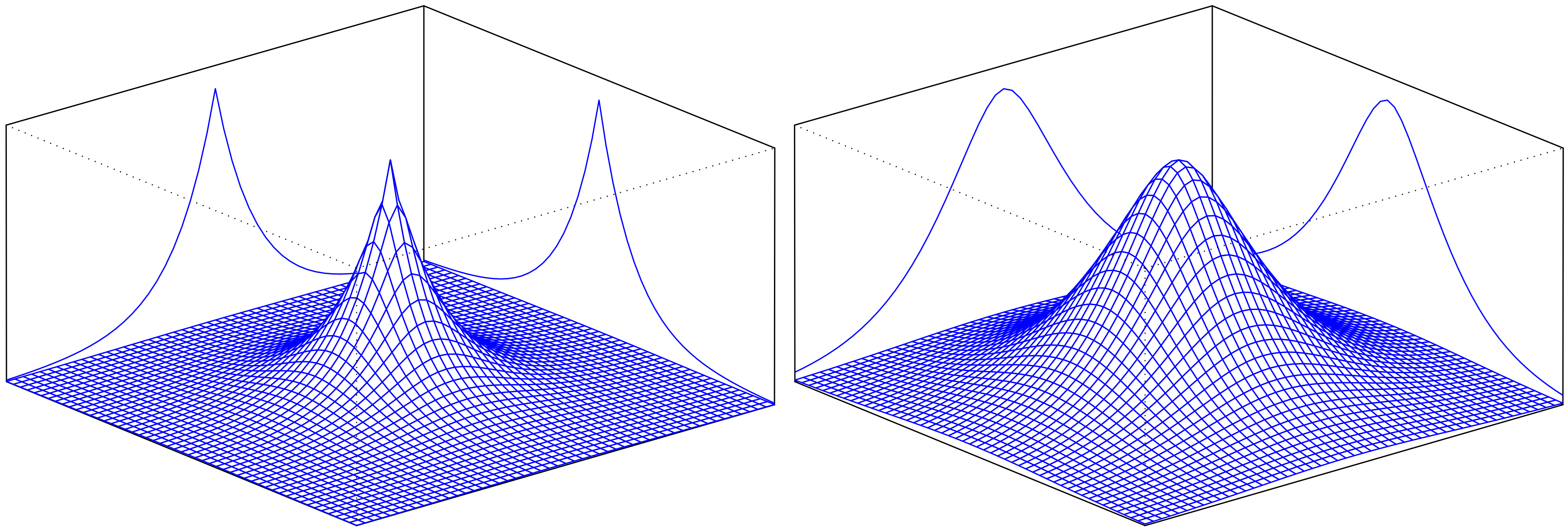
  - Survival time as effect modifier $\Rightarrow$ Time-varying effects $x \cdot g(t)$.

- Spatial effect for regional data: 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 $f_{spat}(s)$ is the average of the function evaluations of adjacent sites.





- Spatial effect 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.



- All effects can be cast into one general framework.

- All vectors of function evaluations f_j can be expressed as

$$f_j = Z_j \xi_j$$

with design matrix Z_j and regression coefficients ξ_j.

- Generic form of the prior for ξ_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, $\mathrm{rank}(K_j) = k_j \leq d_j = \dim(\xi_j)$.

- $\tau_j^2 \geq 0$ can be interpreted as a variance or (inverse) smoothness parameter.

Bayesian Inference

- Fully Bayesian inference:
  - All parameters (including the variance parameters τ^2) are assigned suitable prior distributions.
  - Typically, estimation is based on MCMC simulation techniques.
  - Usual estimates: Posterior expectation, posterior median (easily obtained from the samples).
- Empirical Bayes inference:
  - Differentiate between parameters of primary interest (regression coefficients) and hyperparameters (variances).
  - Assign priors only to the former.
  - Estimate the hyperparameters by maximising their marginal posterior.
  - Plugging these estimates into the joint posterior and maximising with respect to the parameters of primary interest yields posterior mode estimates.

- MCMC-based inference:
  - Assign inverse gamma prior to τ_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 τ_j^2,
	$b_j = 0,\ a_j = -\frac{1}{2}$	Flat prior for standard deviation τ_j.

  - Conditions for proper posteriors in structured additive regression are available.
  - Gibbs sampler for $\tau_j^2|\cdot$:

    Sample from an inverse Gamma distribution with parameters

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

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

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

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

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

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

- Empirical Bayes inference.
  - Consider the variances τ_j^2 as unknown constants to be estimated from their marginal posterior.
  - Consider the regression coefficients ξ_j as correlated random effects with multivariate Gaussian distribution

    $\Rightarrow$ Use mixed model methodology for estimation.

- Problem: In most cases partially improper random effects distribution.

- Mixed model representation: Decompose

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

where

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

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

- This yields a variance components model with pedictor

$$\eta = X\beta + Vb$$

where in turn

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

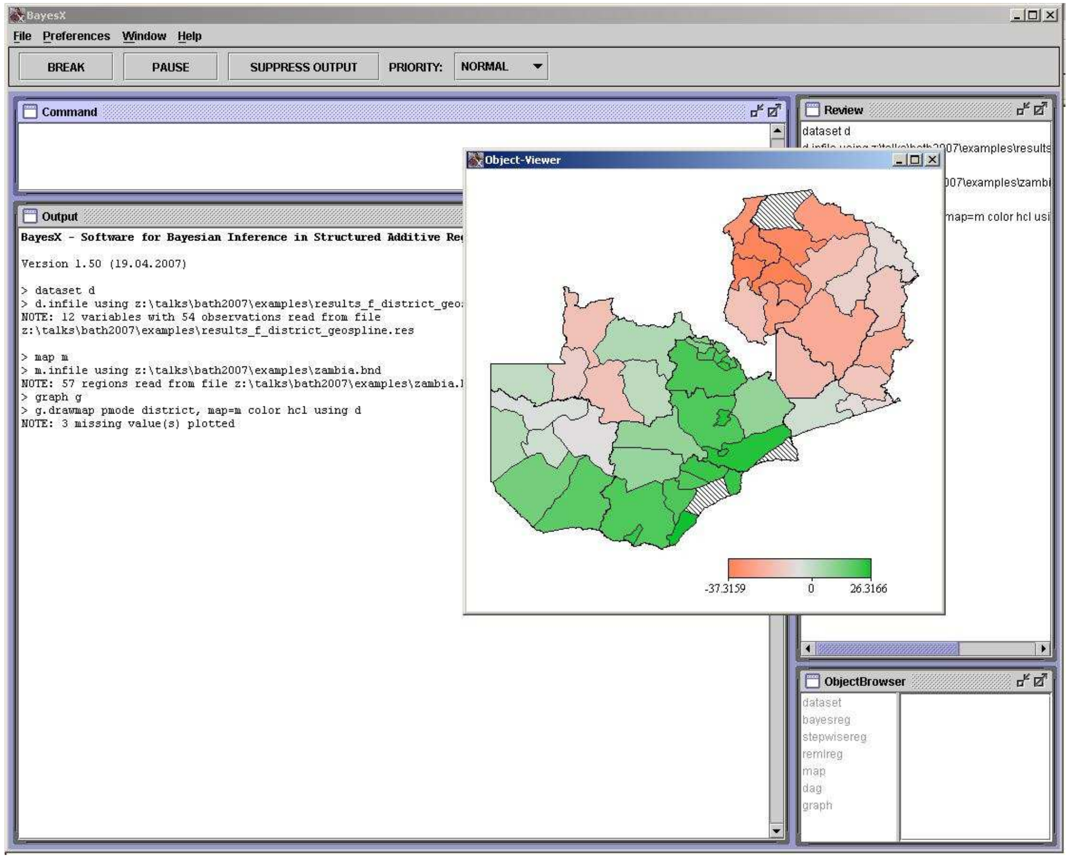
- Obtain empirical Bayes estimates / penalized likelihood estimates via iterating
  - Penalized maximum likelihood for the regression coefficients β and b.
  - Restricted Maximum / Marginal likelihood for the variance parameters in Q:

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

- Involves a Laplace approximation to the marginal likelihood (corresponding to REML estimation of variances in Gaussian mixed models).

BayesX

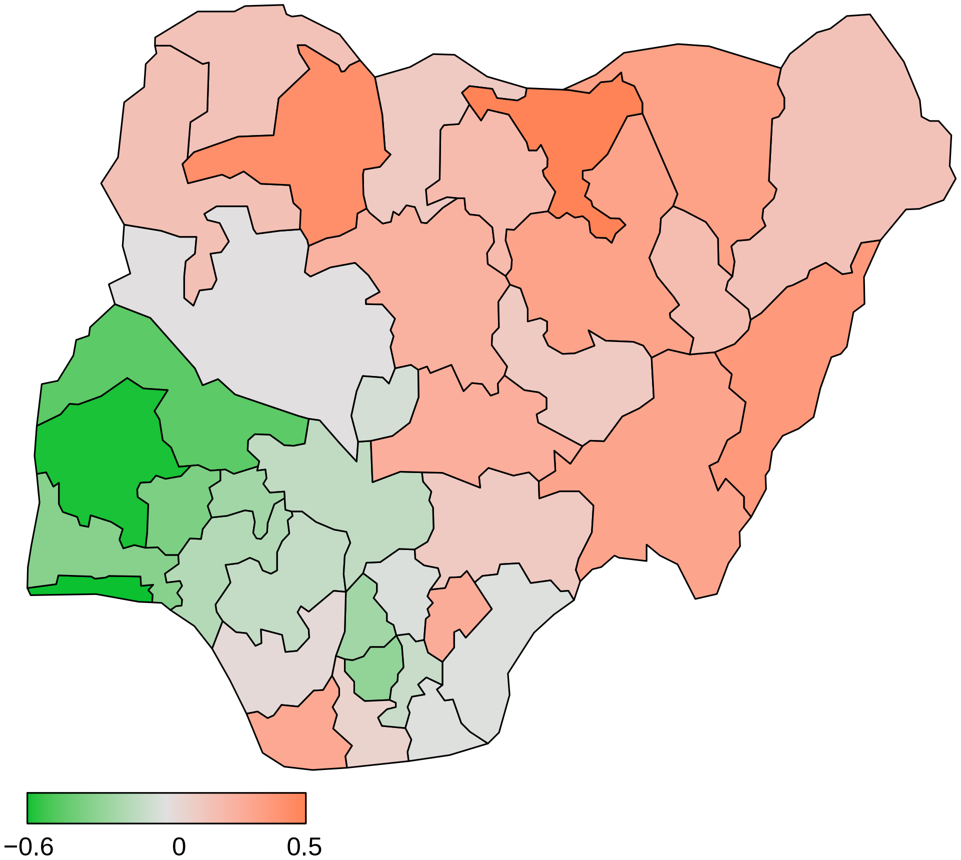
- BayesX is a software tool for estimating structured additive regression models.



- Available from

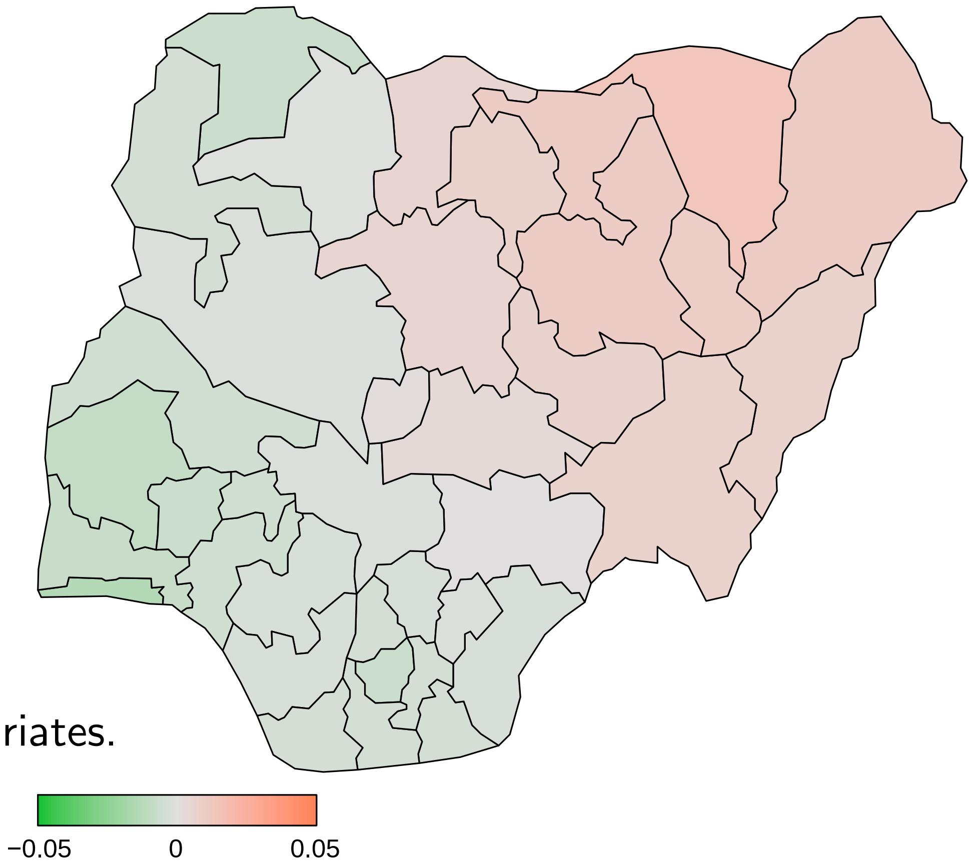
`http://www.stat.uni-muenchen.de/~bayesx`

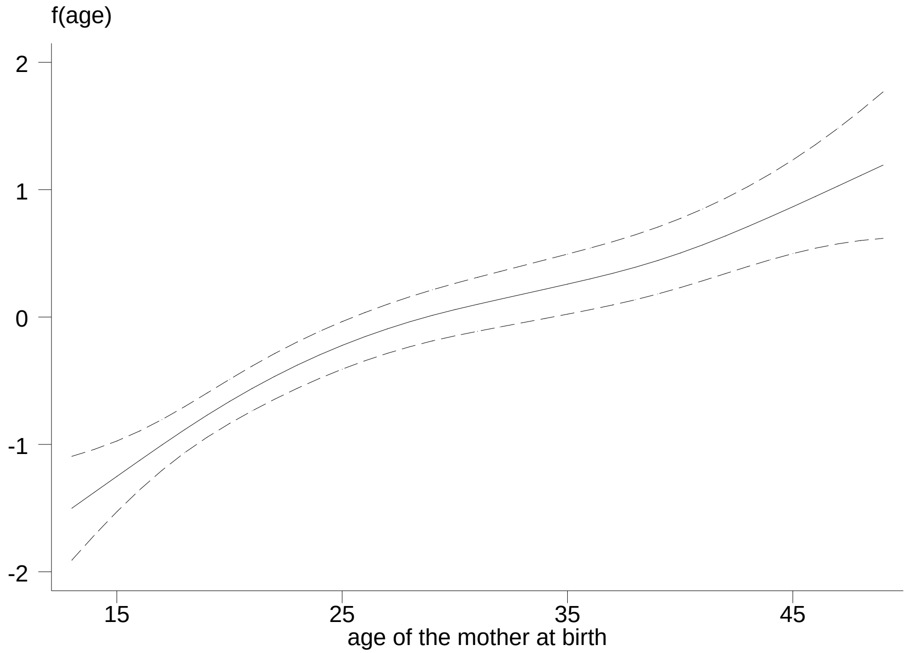
Childhood mortality in Nigeria II



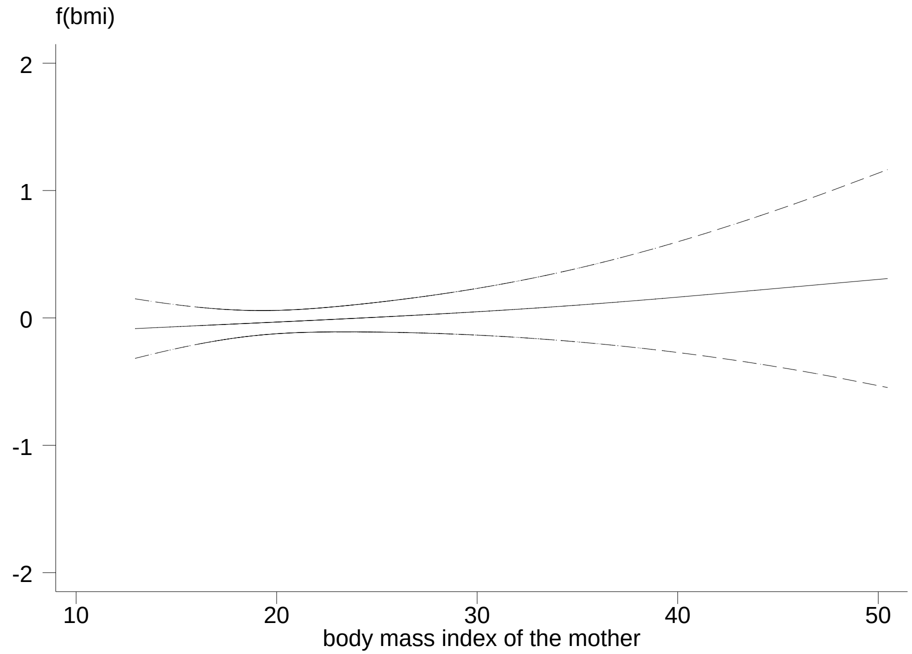
Spatial effect without covariates.

Spatial effect including covariates.

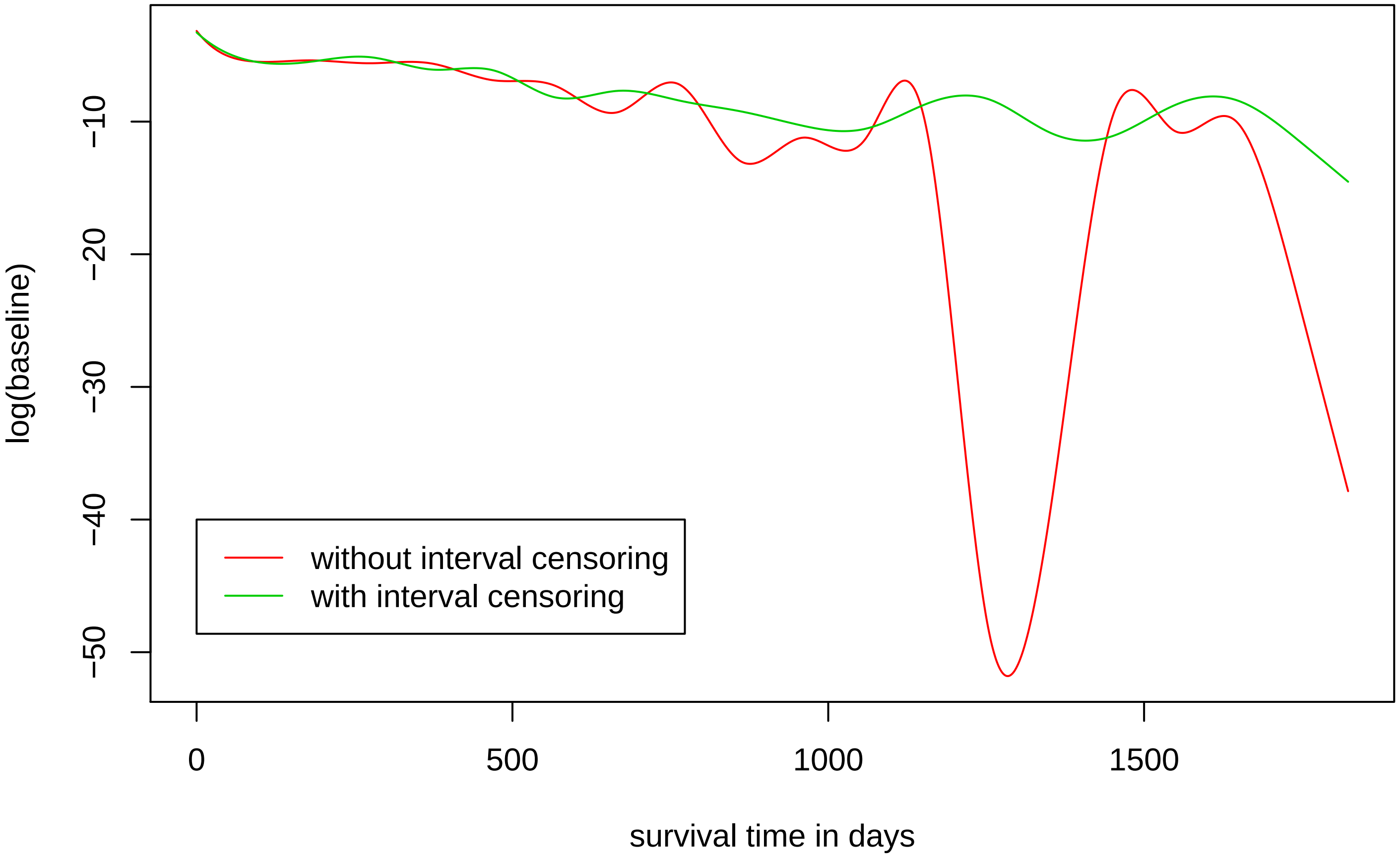




Age of the mother at birth.



Body mass index of the mother


log(baseline)
−10
−20
−30
−40
−50
0
500
1000
1500
survival time in days
without interval censoring
with interval censoring

Multi-State Models

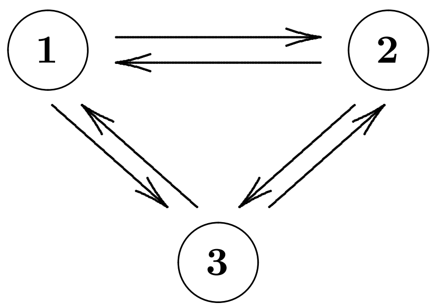
- Multi-state models form a general class for the description of the evolution of discrete phenomena in continuous time (i.e. event history analysis).

- We observe paths of a process

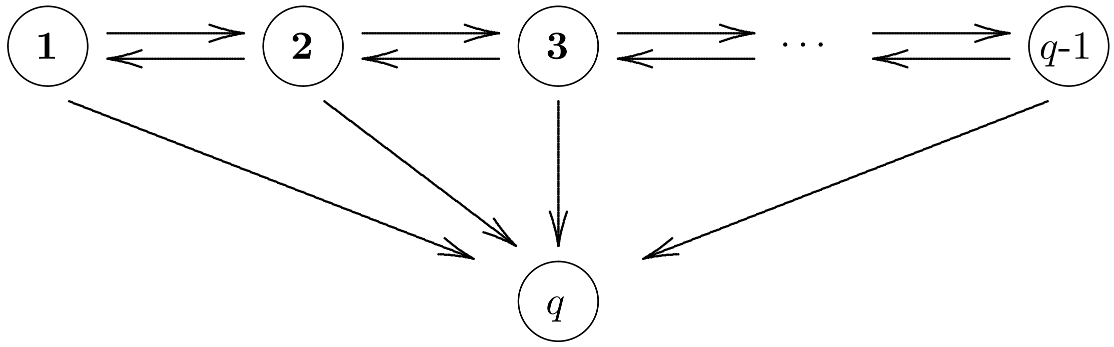
$$X = \{X(t), t \geq 0\} \quad \text{with} \quad X(t) \in \{1, \dots, q\}.$$

- Yields a similar data structure as for Markov processes.

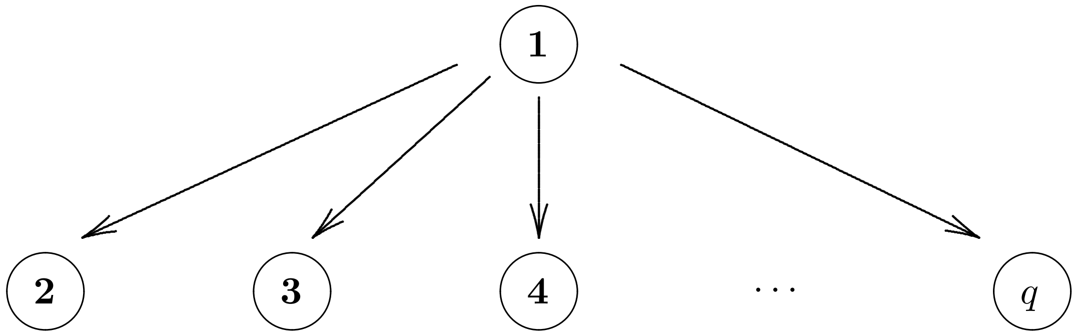
- Examples:
  - Recurrent events:



– Disease progression:



– Competing risks:



- (Homogenous) Markov processes can be compactly described in terms of the transition intensities

$$\lambda_{ij} = \lim_{\Delta t \to 0} \frac{P(X(t+\Delta t) = j | X(t) = i)}{\Delta t}$$

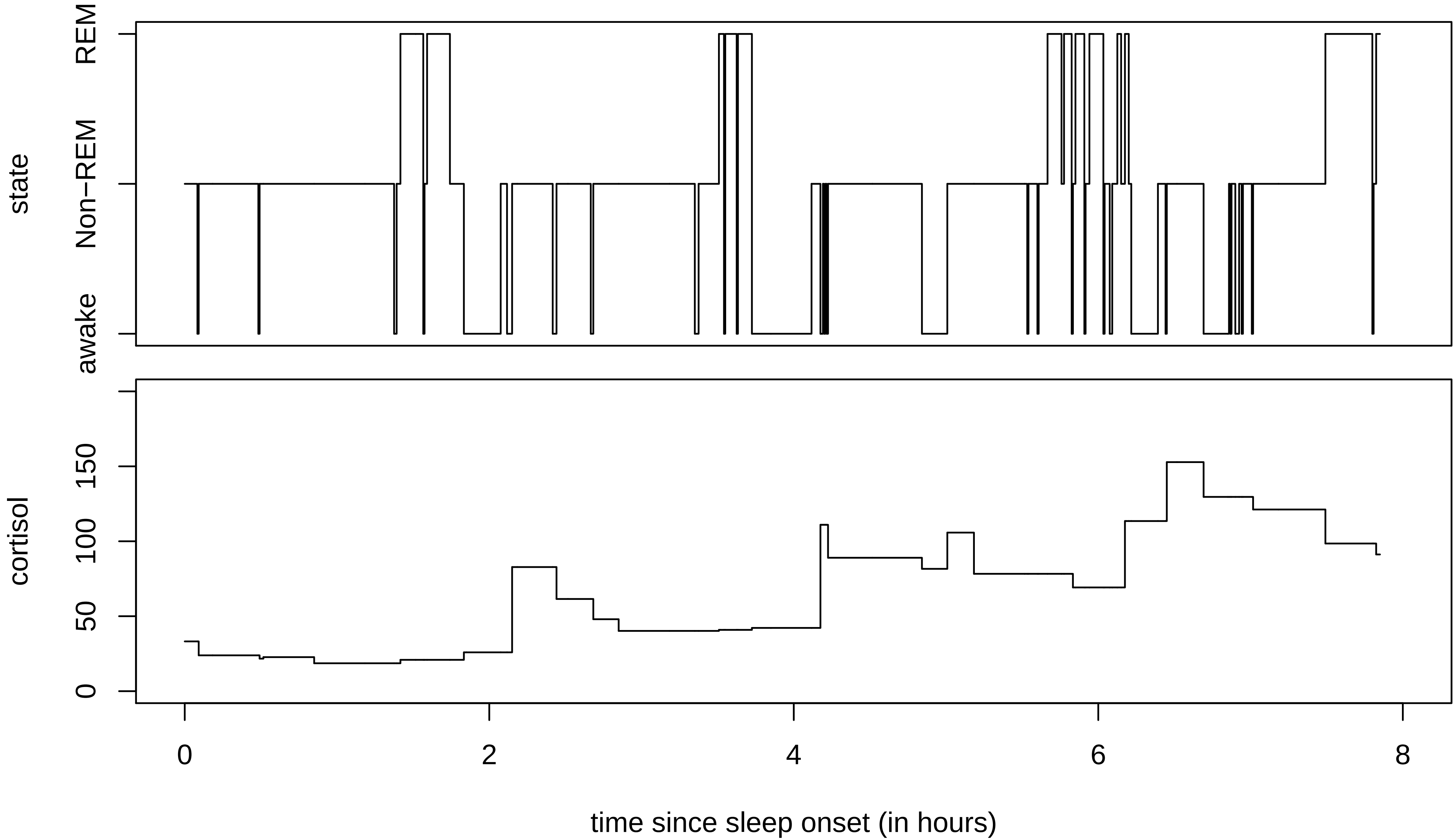
Human Sleep Data

- Human sleep can be considered an example of a recurrent event type multi-state model.

- State Space:

Awake	Phases of wakefulness
REM	Rapid eye movement phase (dream phase)
Non-REM	Non-REM phases (may be further differentiated)

- Aims of sleep research:
  - Describe the dynamics underlying the human sleep process.
  - Analyse associations between the sleep process and nocturnal hormonal secretion.
  - (Compare the sleep process of healthy and diseased persons.)


state
REM
Non−REM
awake
cortisol
0
50
100
150
0
2
4
6
8
time since sleep onset (in hours)

- Data generation:
  - Sleep recording based on electroencephalographic (EEG) measures every 30 seconds (afterwards classified into the three sleep stages).
  - Measurement of hormonal secretion based on blood samples taken every 10 minutes.
  - A training night familiarizes the participants of the study with the experimental environment.

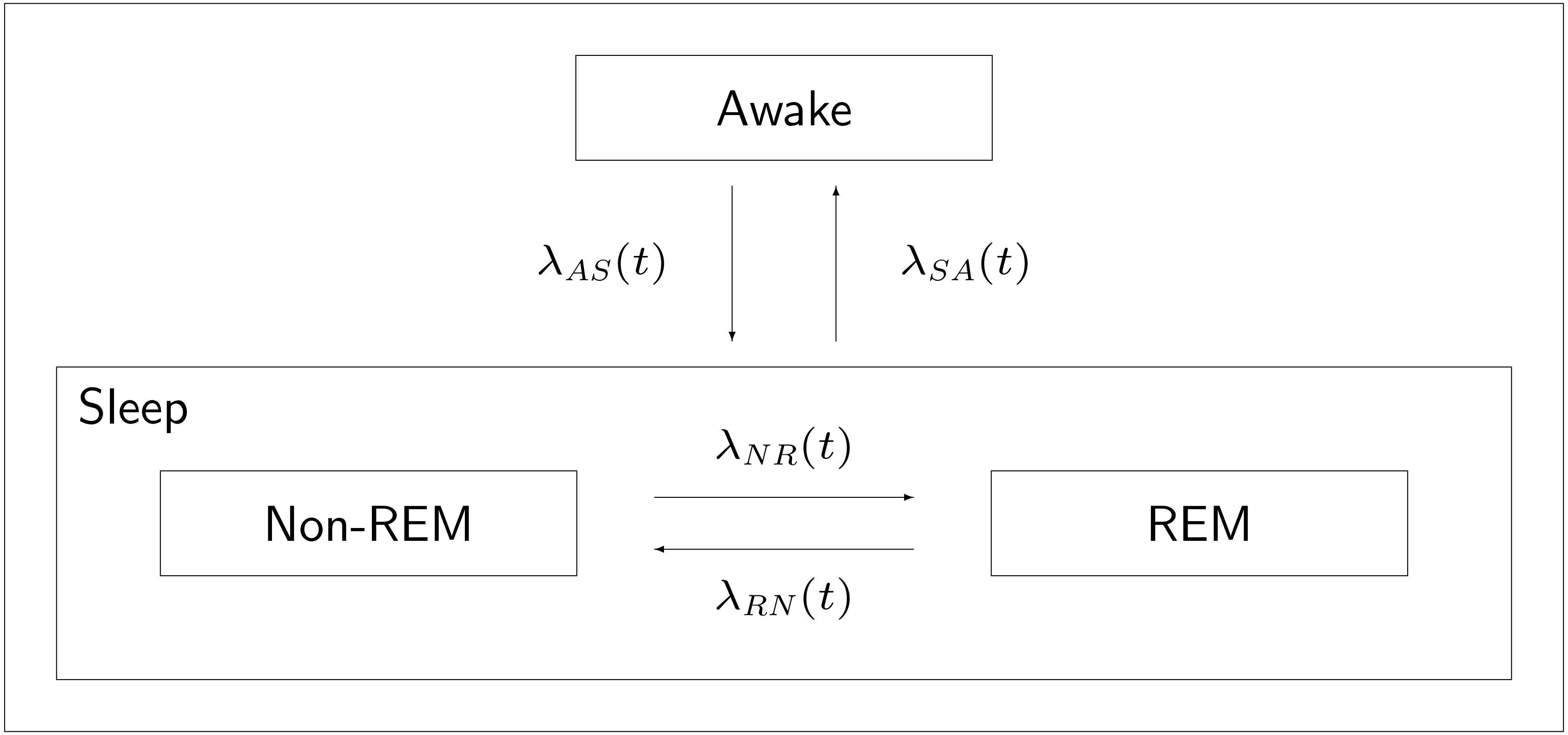
  $\Rightarrow$ Sleep processes of 70 participants.

- Simple parametric approaches are not appropriate in this application due to
  - Changing dynamics of human sleep over night.
  - The time-varying influence of the hormonal concentration on the transition intensities.
  - Unobserved heterogeneity.

$\Rightarrow$ Model transition intensities nonparametrically.

Specification of Transition Intensities

- To reduce complexity, we consider a simplified transition space:



- Model specification:

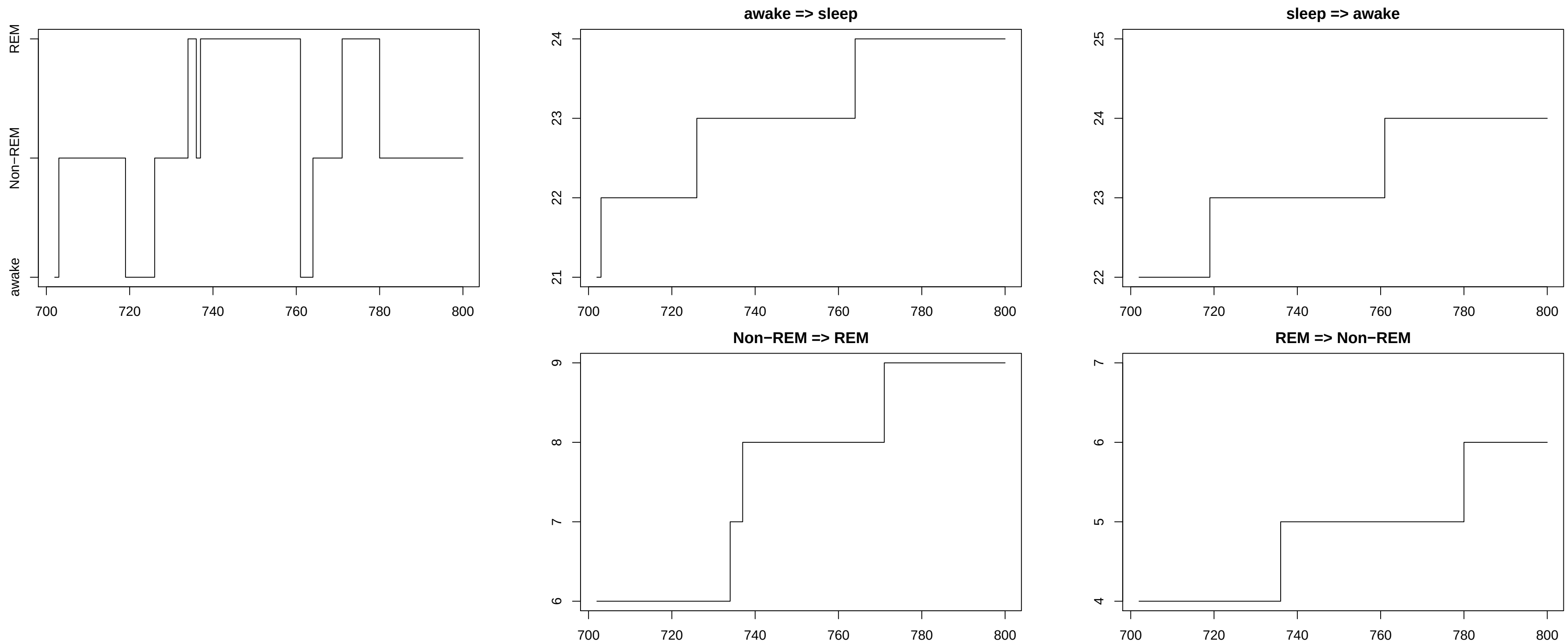
$$
\begin{aligned}
\lambda_{AS,i}(t) &= \exp\left[\gamma_0^{(AS)}(t) + b_i^{(AS)}\right] \\
\lambda_{SA,i}(t) &= \exp\left[\gamma_0^{(SA)}(t) + b_i^{(SA)}\right] \\
\lambda_{NR,i}(t) &= \exp\left[\gamma_0^{(NR)}(t) + c_i(t)\gamma_1^{(NR)}(t) + b_i^{(NR)}\right] \\
\lambda_{RN,i}(t) &= \exp\left[\gamma_0^{(RN)}(t)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

where

$$
\begin{aligned}
c_i(t) &= \begin{cases} 1 & \text{cortisol} > 60 \text{ n mol/l at time } t \\ 0 & \text{cortisol} \leq 60 \text{ n mol/l at time } t, \end{cases} \\
b_i^{(j)} \sim N(0, \tau_j^2) &= \text{transition- and individual-specific frailty terms.}
\end{aligned}
$$

Counting Process Representation

- A multi-state model with k different types of transitions can be equivalently expressed in terms of k counting processes $N_h(t)$, $h = 1, \ldots, k$ counting these transitions.



- From the counting process representation we can derive the likelihood contributions.

- The counting process representation also provides a possibility for model validation based on martingale residuals.

- Every counting process is a submartingale and can therefore be (Doob-Meyer-) decomposed as

$$\begin{aligned} N_{hi}(t) &= A_{hi}(t) + M_{hi}(t) \\ &= \int_0^t \lambda_{hi}(t) Y_{hi}(t) du + M_{hi}(t), \end{aligned}$$

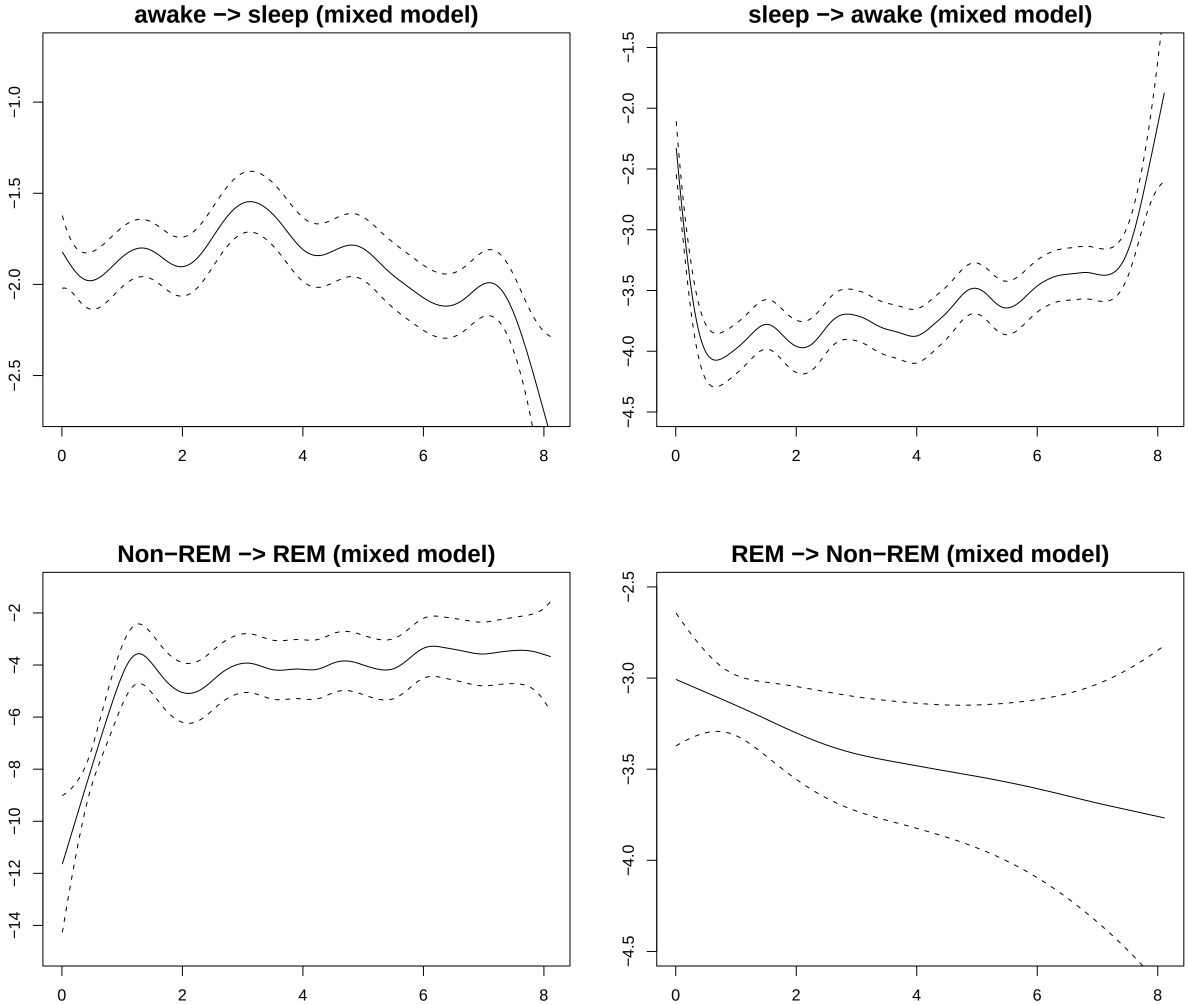
  where $M_{hi}(t)$ is a martingale and $A_{hi}(t)$ is the (predictable) compensator process of $N_{hi}(t)$.

- The martingales $M_{hi}(t)$ can be interpreted as continuous-time residuals.

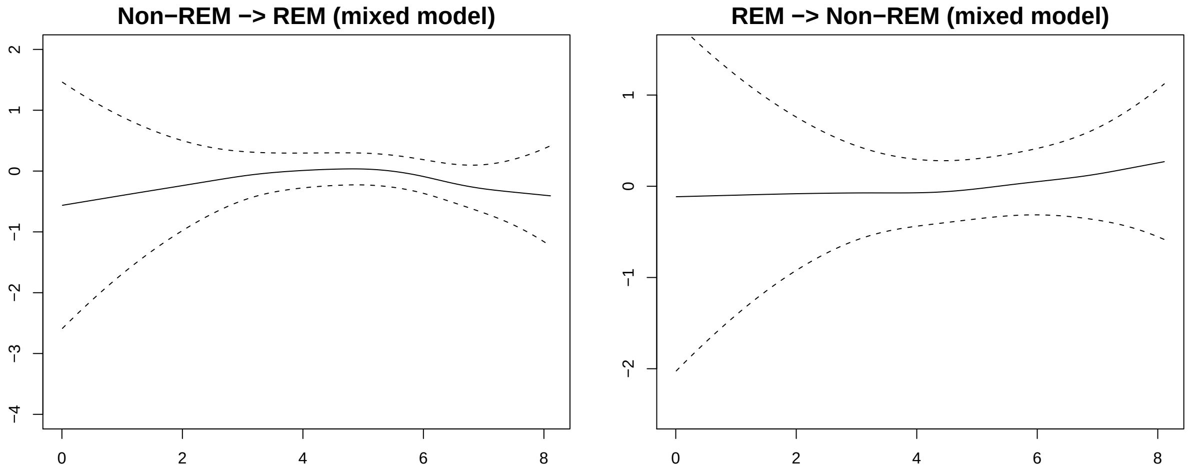
- Plots of $M_{hi}(t)$ against t can be used to compare models, evaluate the model fit, etc.

Human Sleep Data II

- Baseline effects:

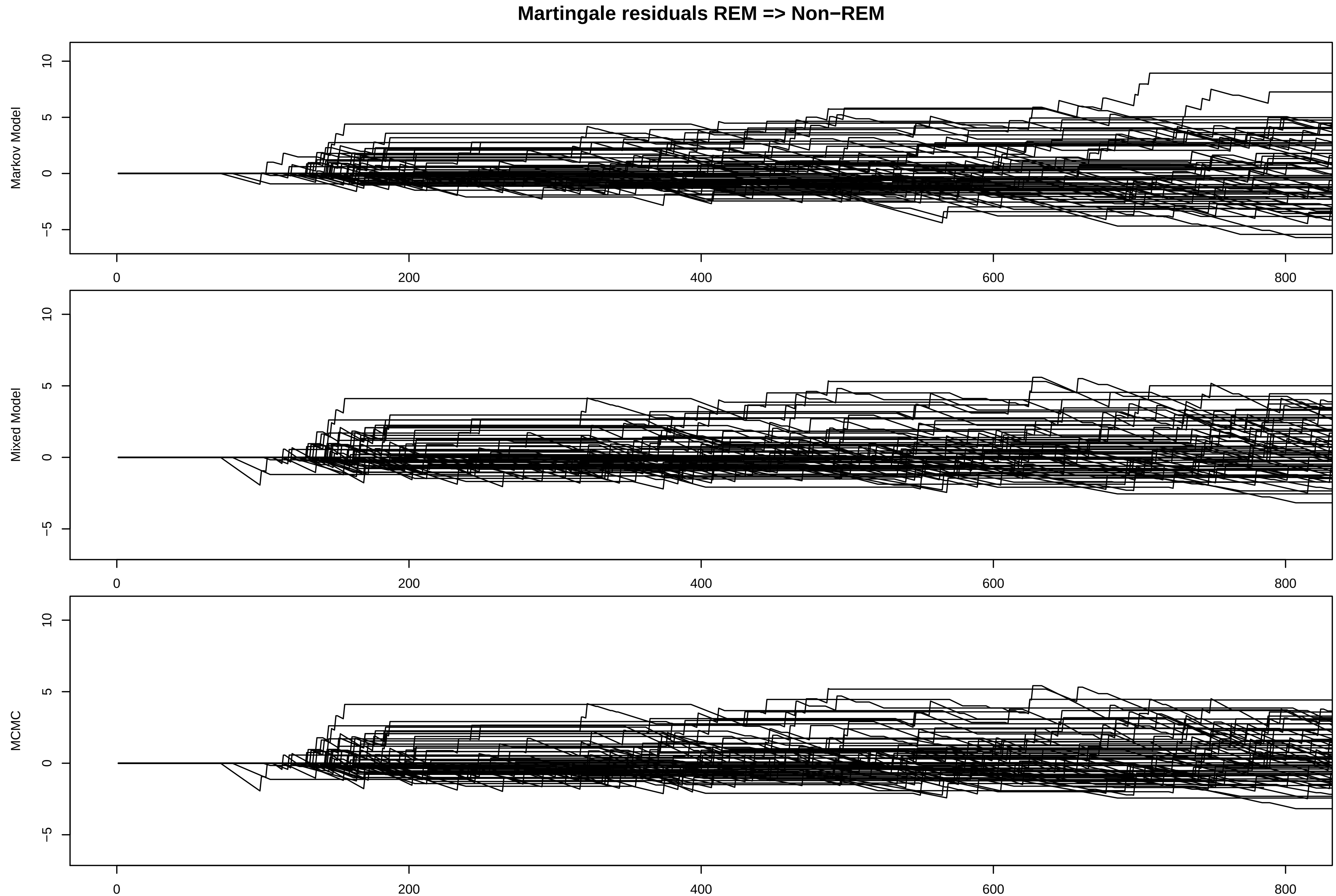


- Time-varying effects for a high level of cortisol:

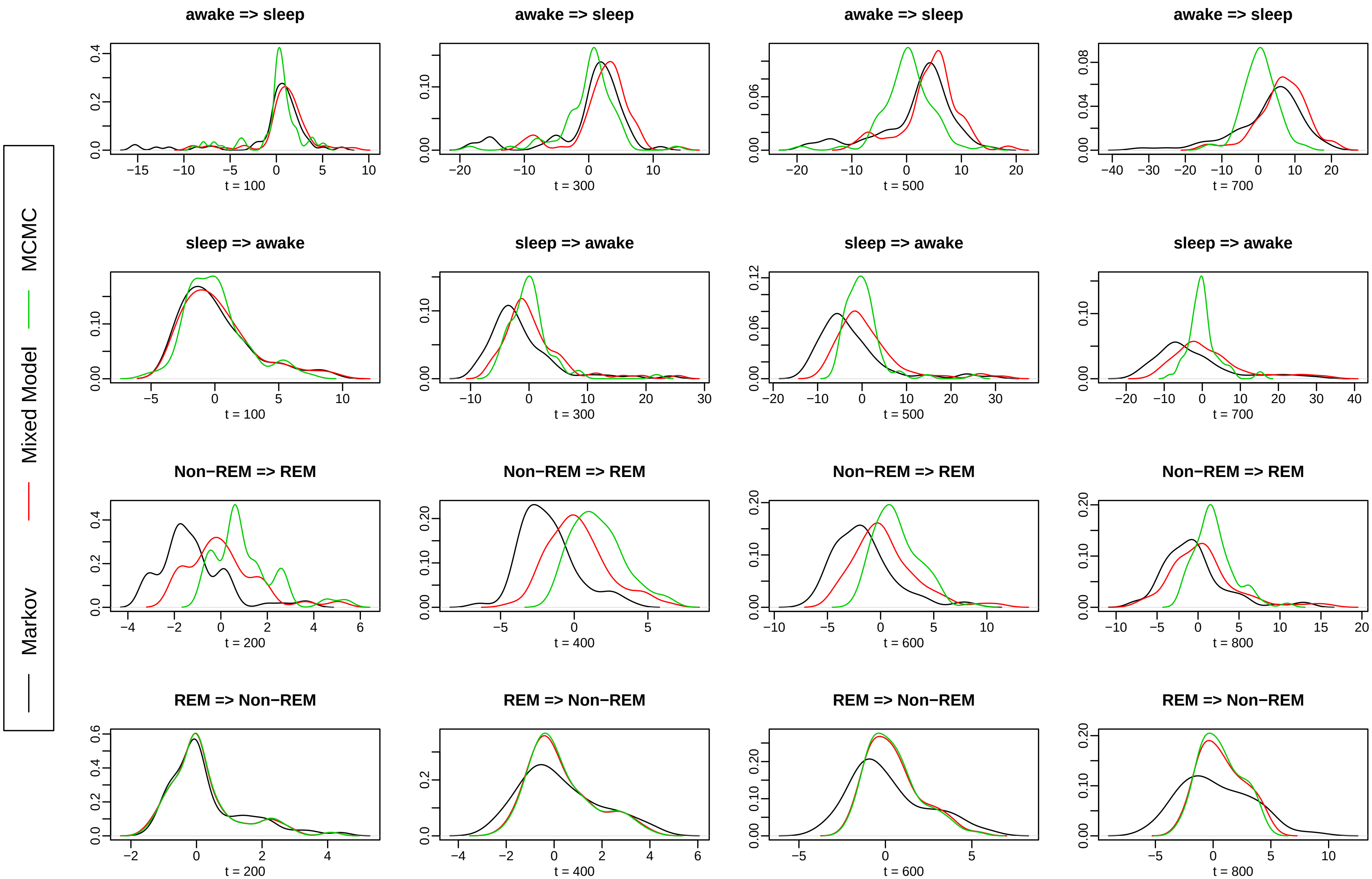


- The fully Bayesian approach detects individual-specific variation for all transitions.

- The empirical Bayes approach only detects individual-specific variation for the transition between REM and Non-REM.



Martingale residuals REM => Non−REM
Markov Model
Mixed Model
MCMC
10
5
0
−5
0
200
400
600
800


Markov
Mixed Model
MCMC
awake => sleep
t = 100
t = 300
t = 500
t = 700
sleep => awake
t = 100
t = 300
t = 500
t = 700
Non−REM => REM
t = 200
t = 400
t = 600
t = 800
REM => Non−REM
t = 200
t = 400
t = 600
t = 800

Summary and Outlook

- Computationally feasible semiparametric models for hazard rates / transition intensities.
- Fully Bayesian and empirical Bayes inference.
- General censoring mechanisms for analysing survival times.
- Model validation of multi-state models via martingale residuals.
- Future work:
  - Interval censored multi-state models (MCMC-based imputation of unobserved path information).
  - Correction for measurement in continuous covariates modelled semiparametrically.
  - Regularisation priors for high-dimensional covariate vectors.

References

- Hennerfeind, Brezger & Fahrmeir (2006): Geoadditive survival models. *Journal of the American Statistical Association*, **101**, 1065-1075.

- Kneib (2006): Geoadditive hazard regression for interval censored survival times. *Computational Statistics and Data Analysis*, **51**, 777-792.

- Kneib & Fahrmeir (2007): A mixed model approach for geoadditive hazard regression. *Scandinavian Journal of Statistics*, **34**, 207-228.

- Kneib & Hennerfeind (2008): Bayesian semiparametric multi-state models. *Statistical Modelling*, to appear.

- A place called home:

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