A mixed model approach for structured hazard regression with interval censored survival times

Thomas Kneib
Department of Statistics
Ludwig-Maximilians-University Munich

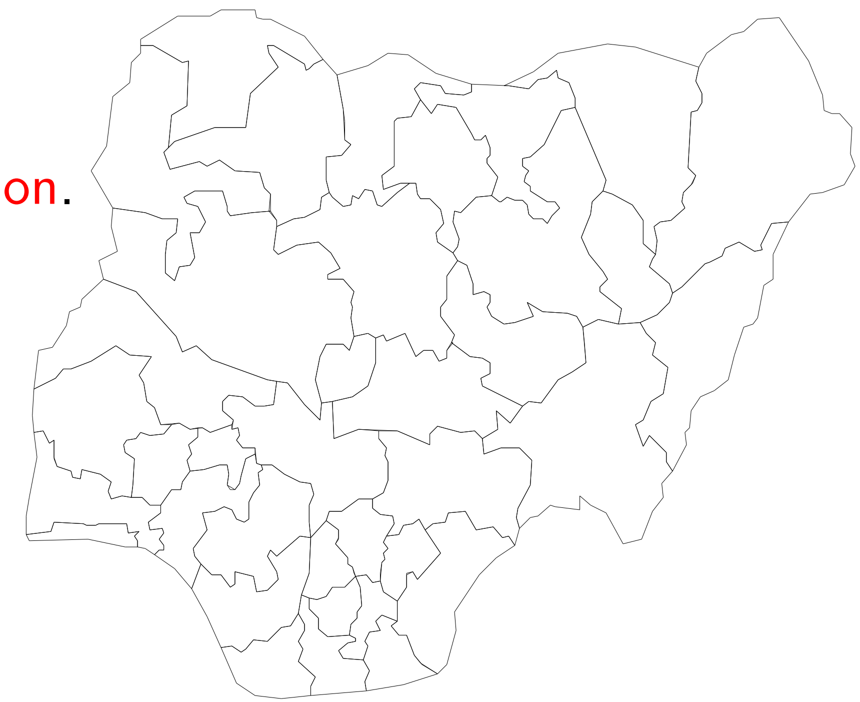
1. Childhood mortality in Nigeria
2. Interval censored survival times
3. Structured hazard regression
4. Mixed model based inference
5. Software
6. Childhood mortality in Nigeria II

30.10.2005

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:

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

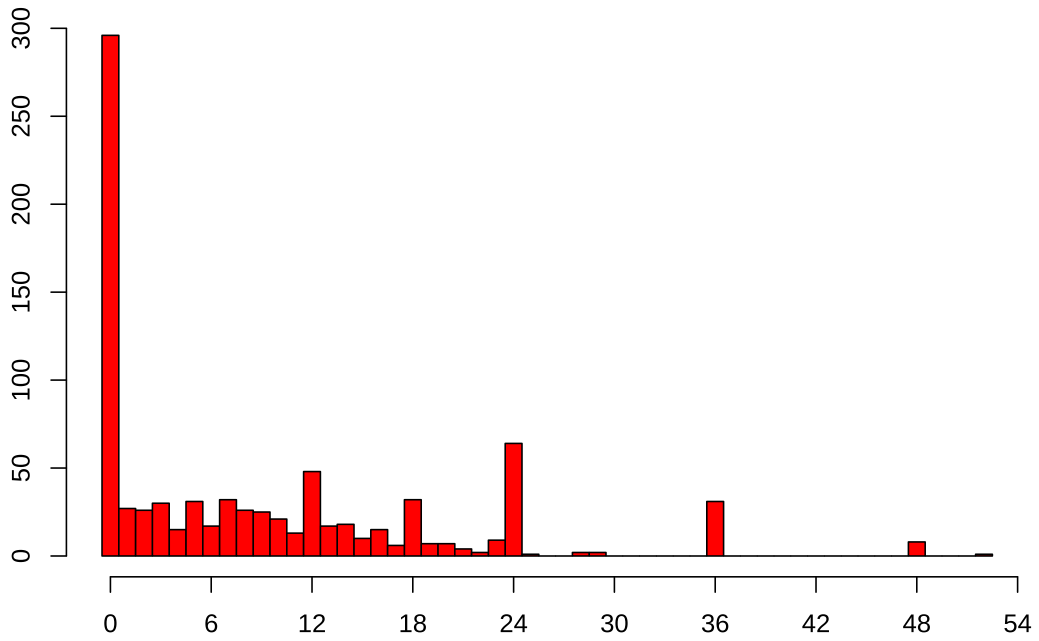


- Limitations of the classical Cox model:
  - Restricted to right censored observations.
  - Post-estimation of the baseline hazard.
  - Proportional hazards assumption.
  - Parametric form of the predictor.
  - No spatial correlations.

- Extensions usually deal with single issues but do not allow for a simultaneous treatment of all problems.

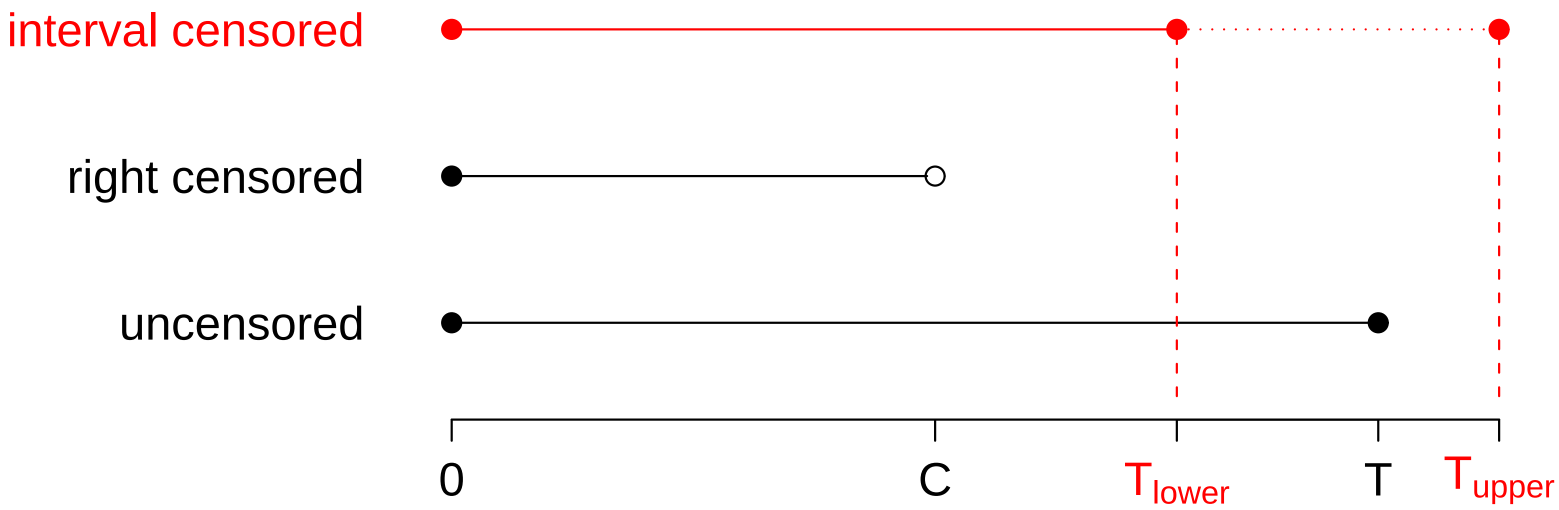
Interval censored survival times

- 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:

$$
\begin{aligned}
P(T > C) &= S(C) \\
&= \exp\left[-\int_0^C \lambda(t)dt\right].
\end{aligned}
$$

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

- 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.

- Log-baseline, time-varying effects and nonparametric effects can be estimated based on penalized splines.
  - Approximate g_j (or f_k) by a weighted sum of B-spline basis functions.
  - Employ a large number of basis functions to enable flexibility.
  - Penalize differences between adjacent parameters of adjacent basis functions to ensure smoothness.

- Spatial effect for regional data: Markov random fields.
  - Define appropriate neighborhoods for the regions.
  - Assume that the expected value of $f_{spat}(s)$ is the average of the function evaluations of adjacent sites.
  - Can be considered a bivariate extension of a first order random walk on the real line.

- Spatial effect for point-referenced data: Stationary Gaussian random fields.
  - Spatial effect follows a zero mean stationary Gaussian stochastic process.
  - Correlation of two arbitrary sites is defined by an intrinsic correlation function.
  - Well-known as Kriging in the geostatistics literature.
- Extensions:
  - Interaction surfaces (2d P-splines).
  - Varying coefficient terms (continuous and spatial effect modifiers).
  - Frailties (i.i.d. random effects).
- All effects can be cast into one general framework.

Mixed model based inference

- Each term in the predictor is associated with a vector of regression coefficients with multivariate Gaussian prior / random effects distribution:

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

- K_j is a penalty matrix, τ_j^2 a smoothing parameter.
- In most cases K_j is rank-deficient.

$\Rightarrow$ Reparametrize the model to obtain a mixed model with proper distributions.

- Decompose

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

where

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

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

- This yields the variance components model

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

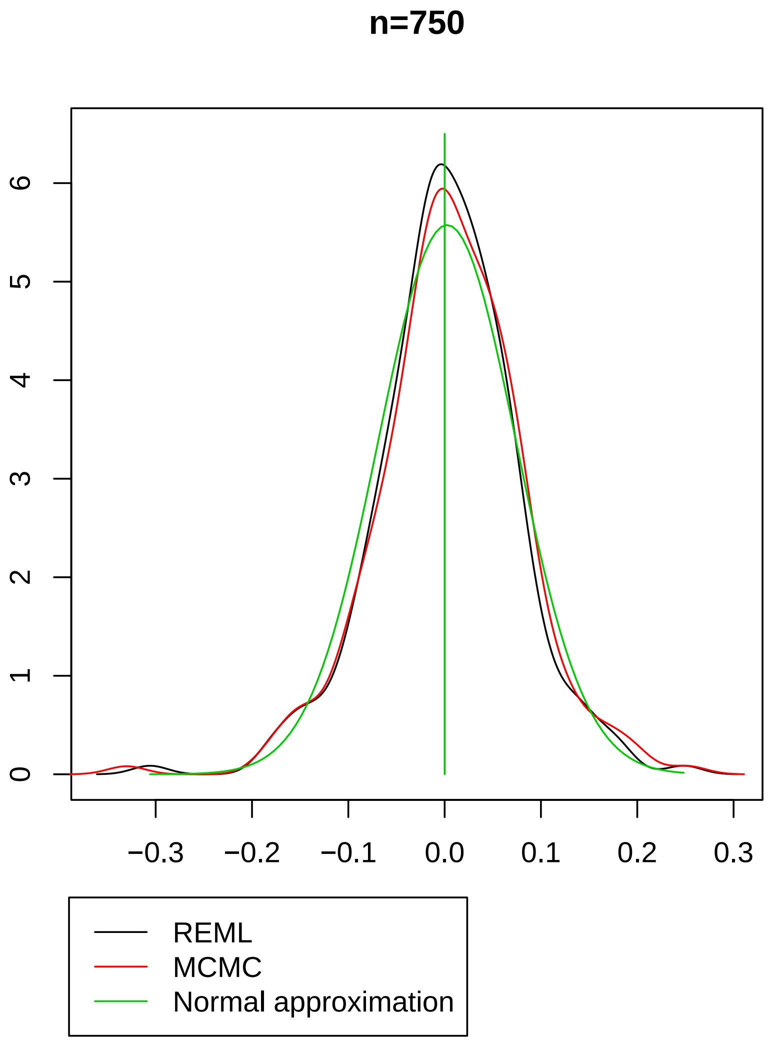
where in turn

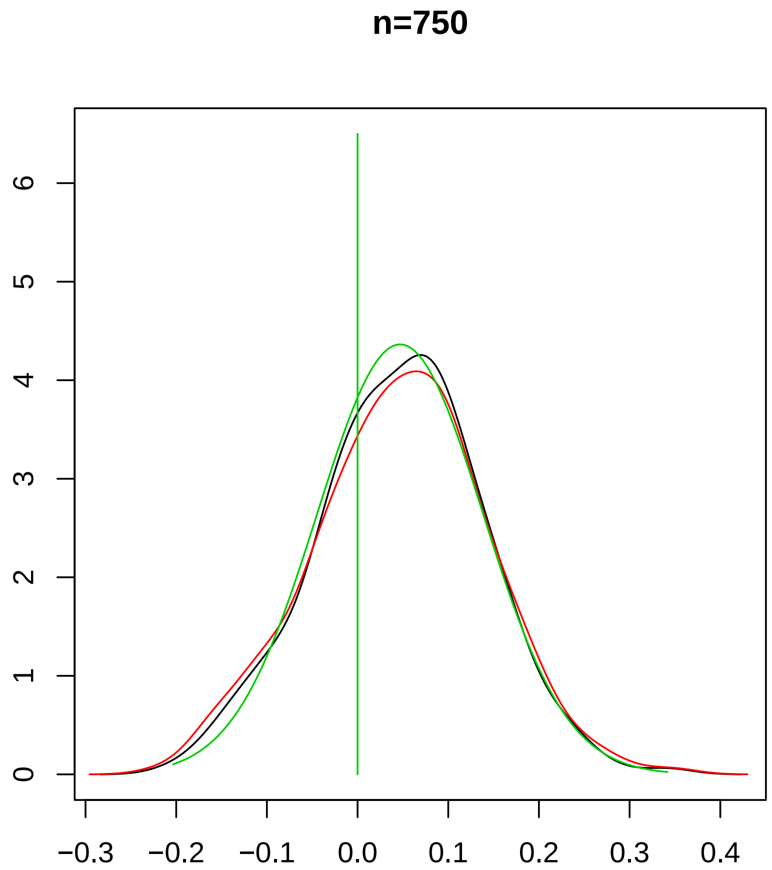
$$p(\beta) \propto const \qquad \text{and} \qquad b \sim N(0, Q).$$

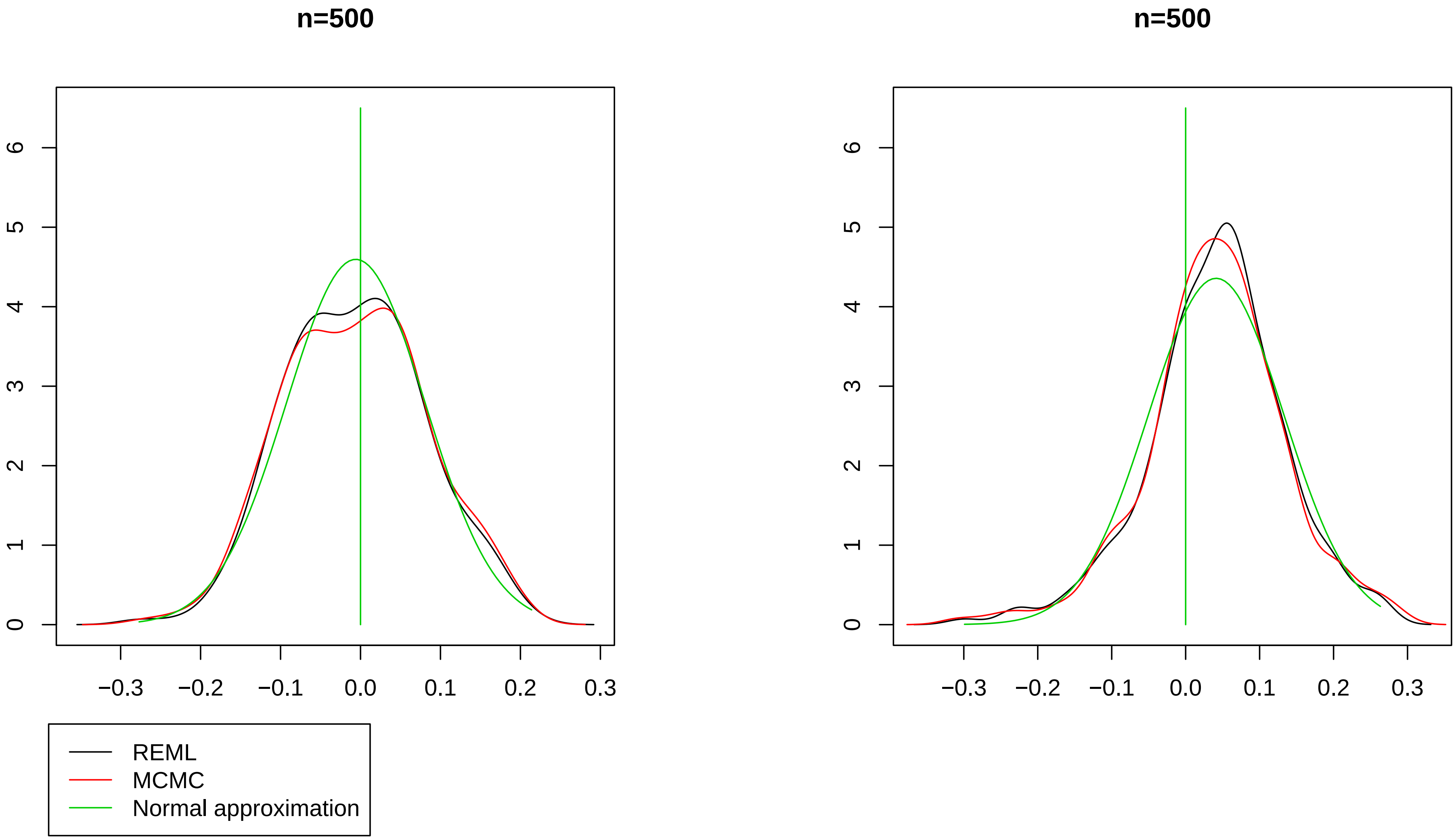
- Obtain empirical Bayes estimates / penalized likelihood estimates via iterating
  - Penalized maximum likelihood for the regression coefficients β and b.
  - Restricted Maximum / Marginal likelihood for the variance parameters in Q:

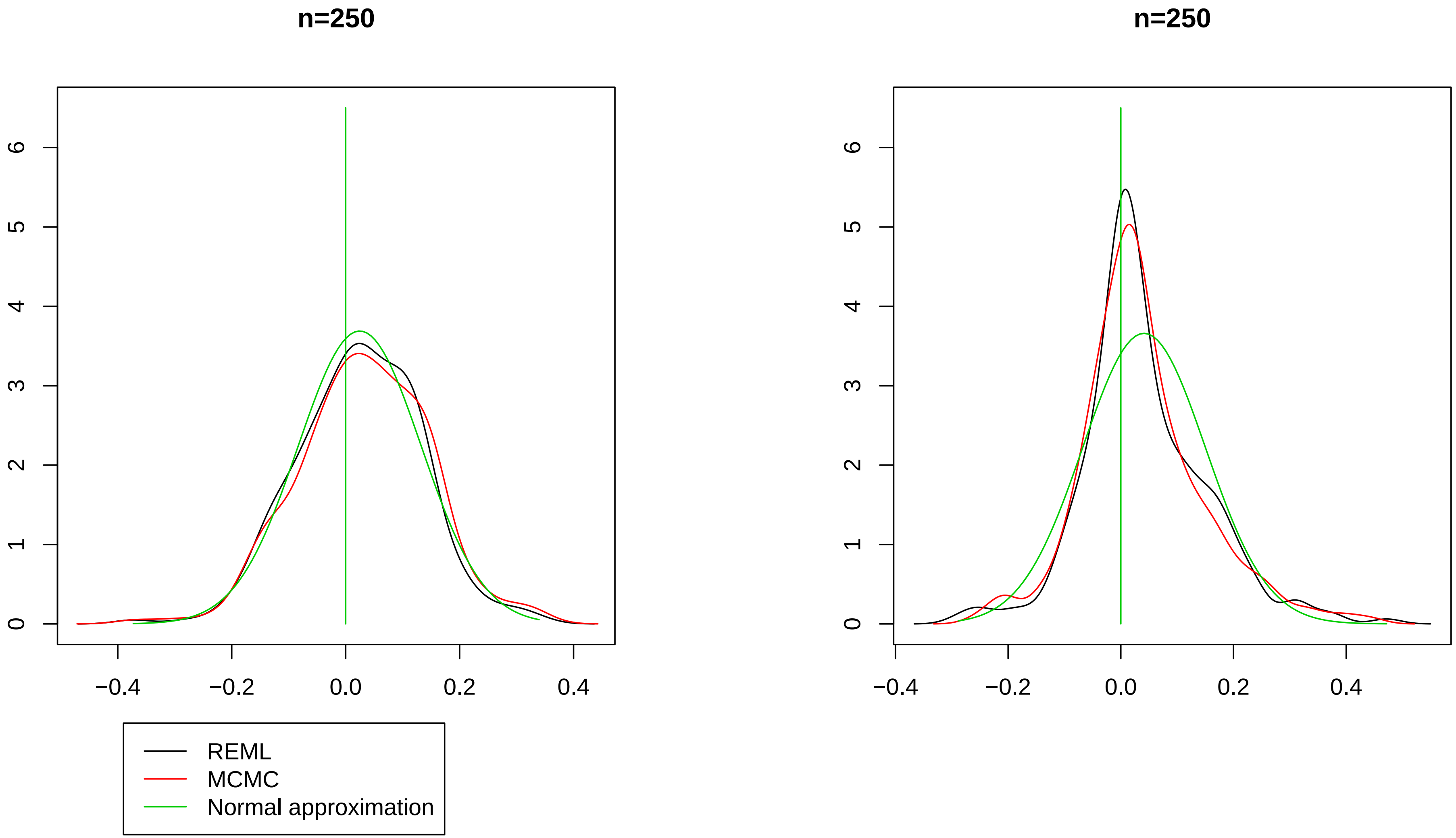
$$L(Q) = \int L(\beta, b, Q)p(b)d\beta db \to \max_{Q}.$$

- Involves Laplace approximation to the marginal likelihood.
- These approximations have proven to be quite accurate in simulation studies.


n=750
0
1
2
3
4
5
6
−0.3
−0.2
−0.1
0.0
0.1
0.2
0.3
REML
MCMC
Normal approximation


n=750
0
1
2
3
4
5
6
−0.3
−0.2
−0.1
0.0
0.1
0.2
0.3
0.4


n=500
n=500
REML
MCMC
Normal approximation


n=250
0
1
2
3
4
5
6
−0.4
−0.2
0.0
0.2
0.4
REML
MCMC
Normal approximation
n=250
0
1
2
3
4
5
6
−0.4
−0.2
0.0
0.2
0.4

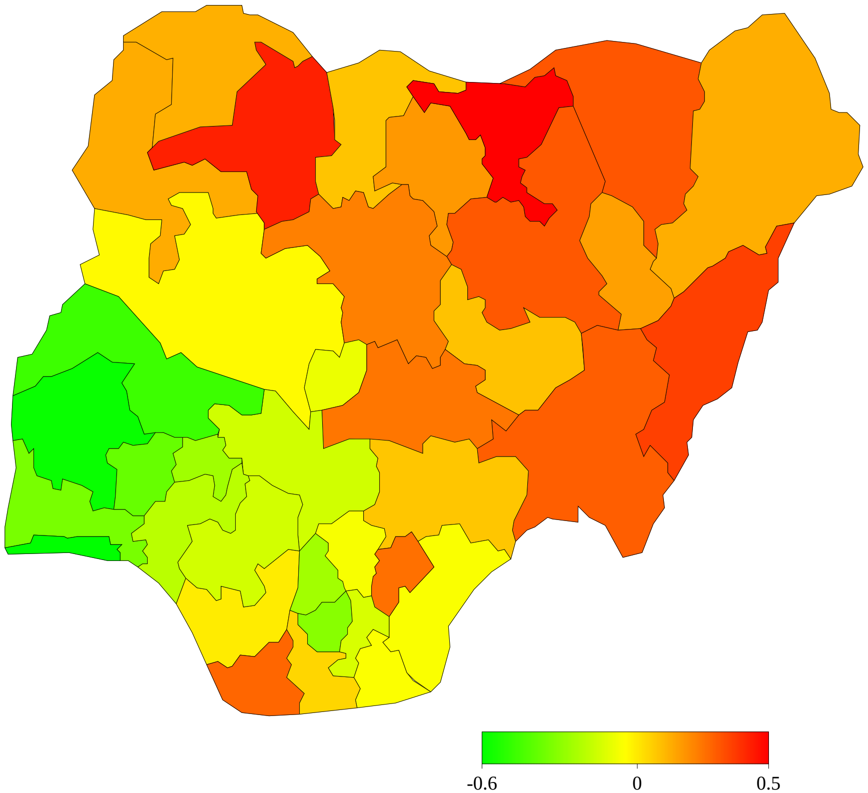
Software

- Implemented in the software package BayesX.

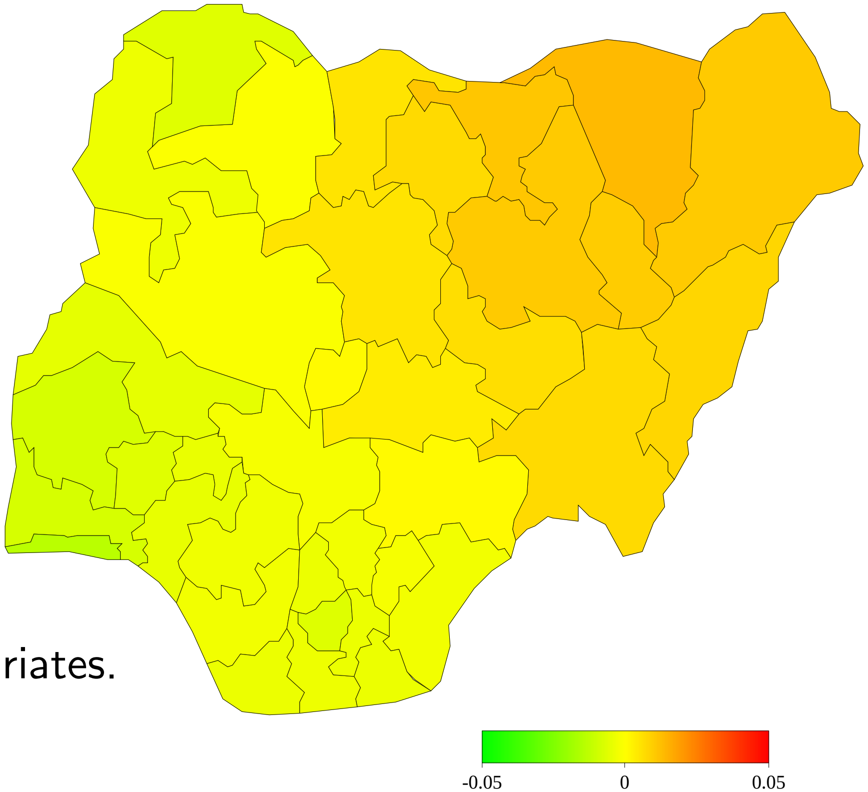
- Available from

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

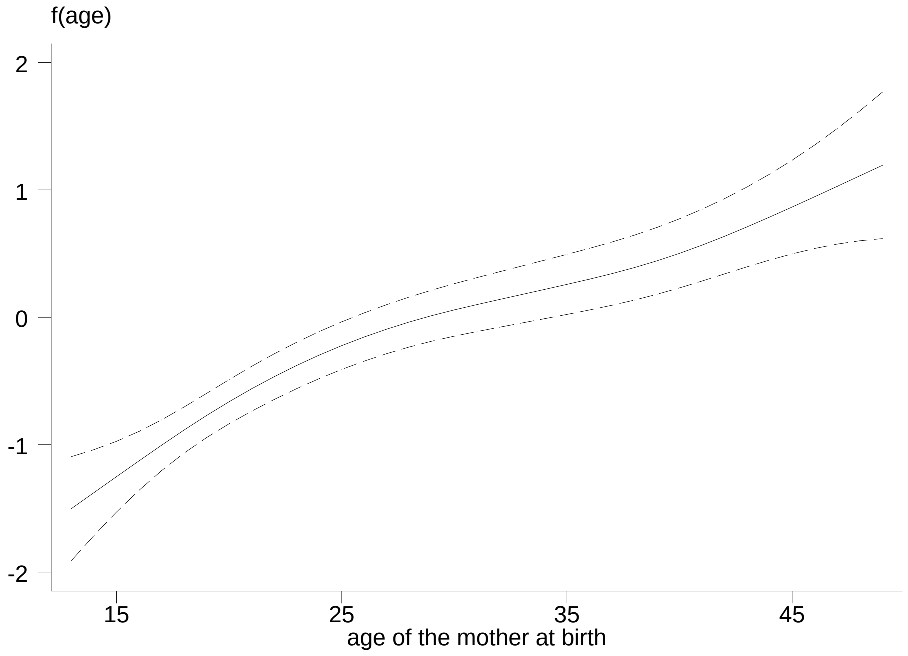
Childhood mortality in Nigeria II



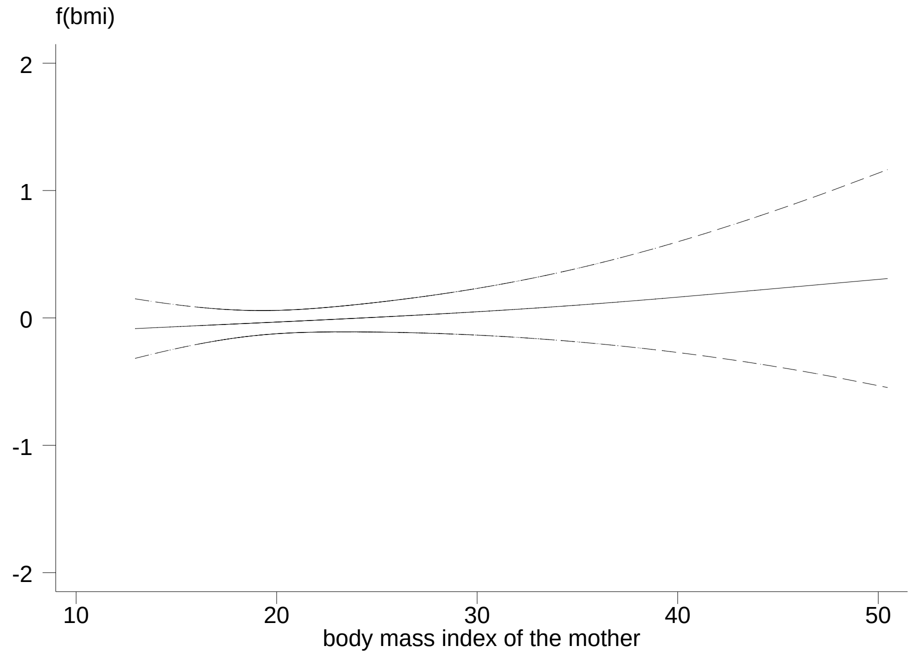
Spatial effect without covariates.



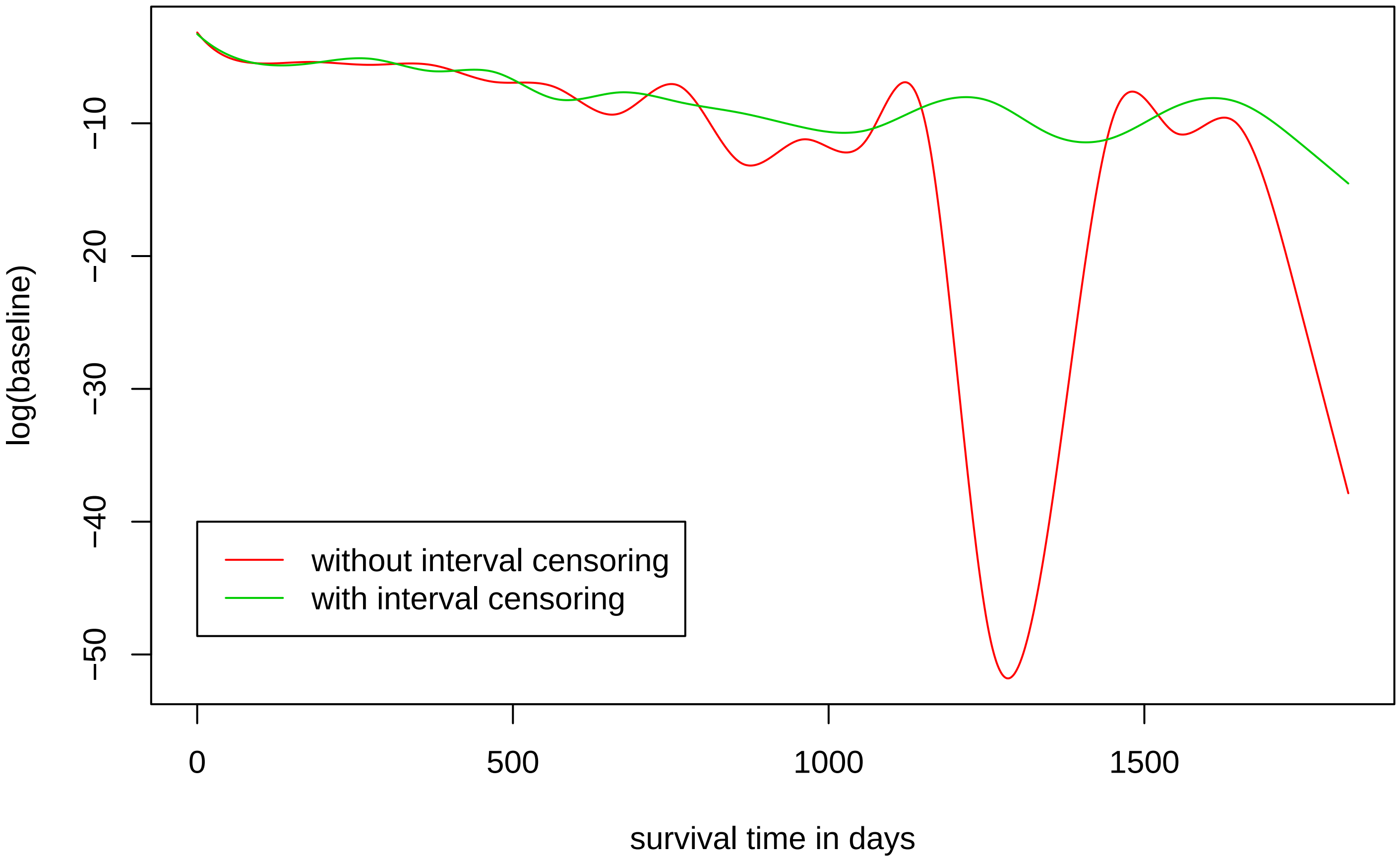
Spatial effect including covariates.



Age of the mother at birth.



Body mass index of the mother


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

Discussion

- Empirical Bayesian treatment of complex hazard regression models:
  - Combines geoadditive predictor with general censoring schemes.
  - Does not rely on MCMC simulation techniques.

    $\Rightarrow$ No questions on convergence and mixing of Markov chains, no hyperpriors.
  - Closely related to penalized likelihood estimation in a frequentist setting.
- Future work:
  - Multi state models.
  - Competing risks models.
  - Inclusion of interval censoring in these more general frameworks.

References

- Kneib, T. and Fahrmeir, L. (2004): A mixed model approach for structured hazard regression. SFB 386 Discussion Paper 400, University of Munich.

- Kneib, T. (2005): Geoadditive hazard regression for interval censored survival times. SFB 386 Discussion Paper 447, University of Munich.

- Available from

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