Geoadditive hazard regression for interval censored survival times

Thomas Kneib & Ludwig Fahrmeir
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

25th EMS OSLO 2005

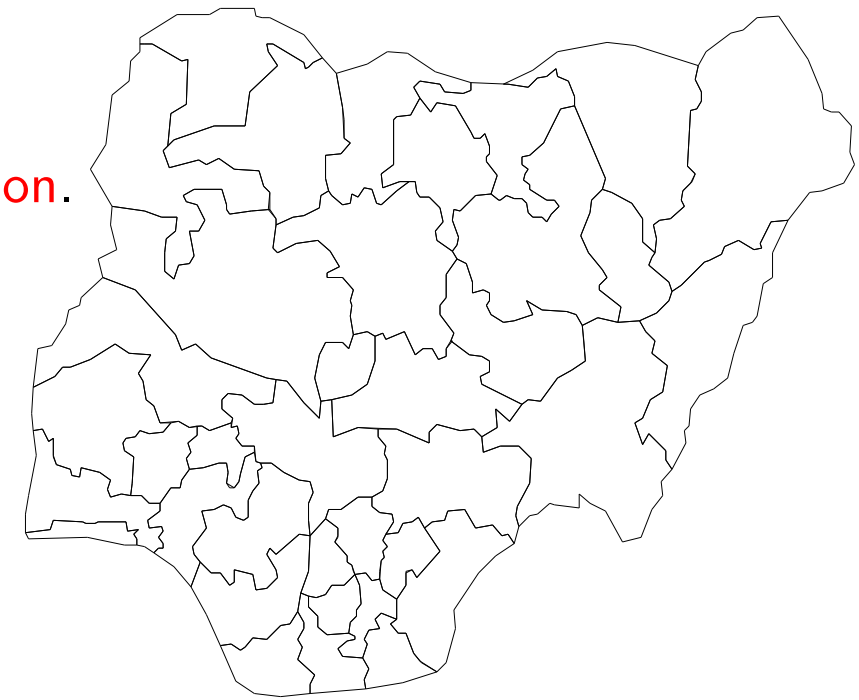
27.7.2005

SFB 386

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:

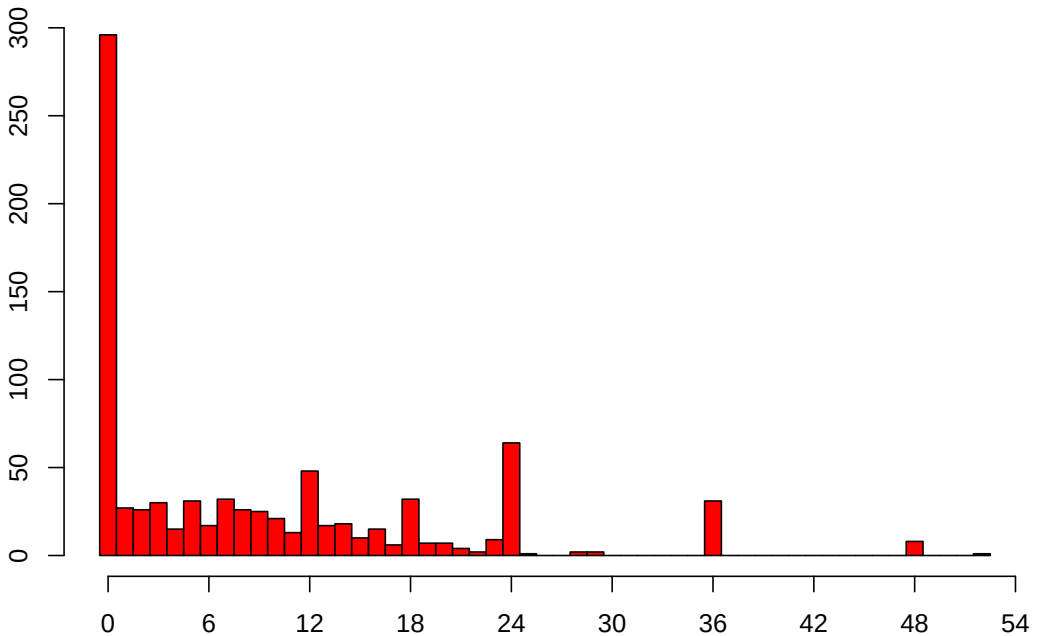
$$\lambda(t; u) = \lambda_0(t) \exp(u'\gamma).$$



- 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.

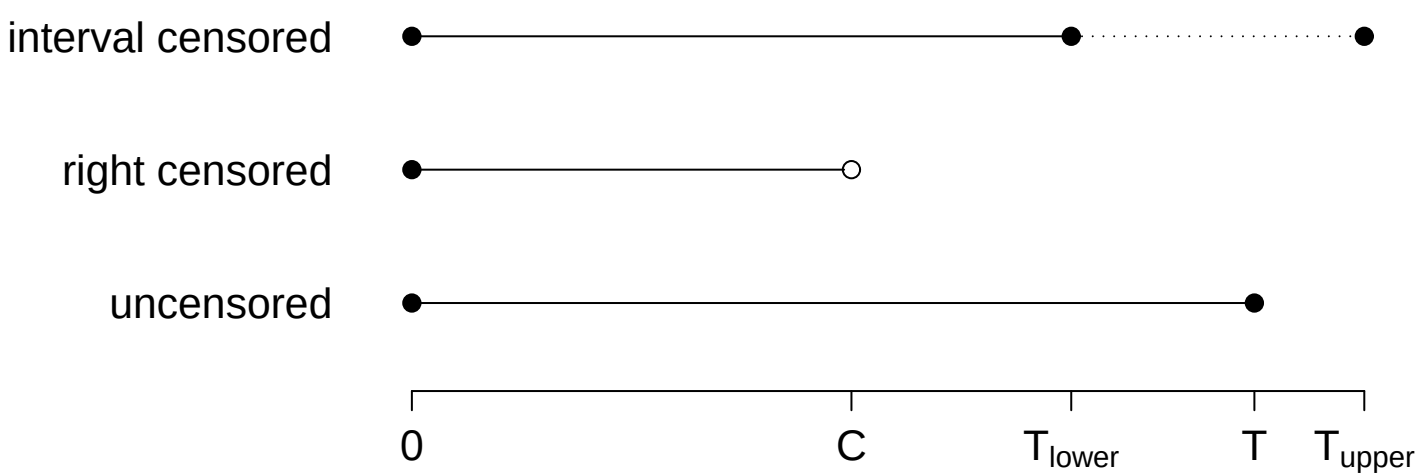
Interval censored survival times

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



- In contrast: censoring times are given in days.

$\Rightarrow$ Treat survival times as interval censored.



- Likelihood contributions:

$$P(T \in [T_{lower}, T_{upper}]) = S(T_{lower}) - S(T_{upper}).$$

- Derivatives of the log-likelihood become much more complicated.

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

Structured hazard regression

- Introduce a more flexible, semiparametric hazard rate model

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

where

- $g_0(t) = \log(\lambda_0(t))$ is the log-baseline-hazard,
- g_j are time varying effects of covariates z_j,
- f_k are nonparametric functions of continuous covariates x_k,
- f_{spat} is a spatial function,
- $u(t)$ are possibly time-varying covariates.

- Log-baseline, time-varying effects and nonparametric effects can be estimated based on penalized splines.

- Spatial effects depend on data structure:
  - Region data: Markov random fields.
  - Exact locations: stationary Gaussian random fields (Kriging).

- 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 improper multivariate Gaussian prior / random effects distribution:

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

- K_j is a penalty matrix, τ_j^2 a smoothing parameter.

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

- Obtain empirical Bayes estimates via iterating
  - Penalized maximum likelihood for regression coefficients.
  - Restricted Maximum / Marginal likelihood for variance parameters.
- Requires numerical integration techniques.

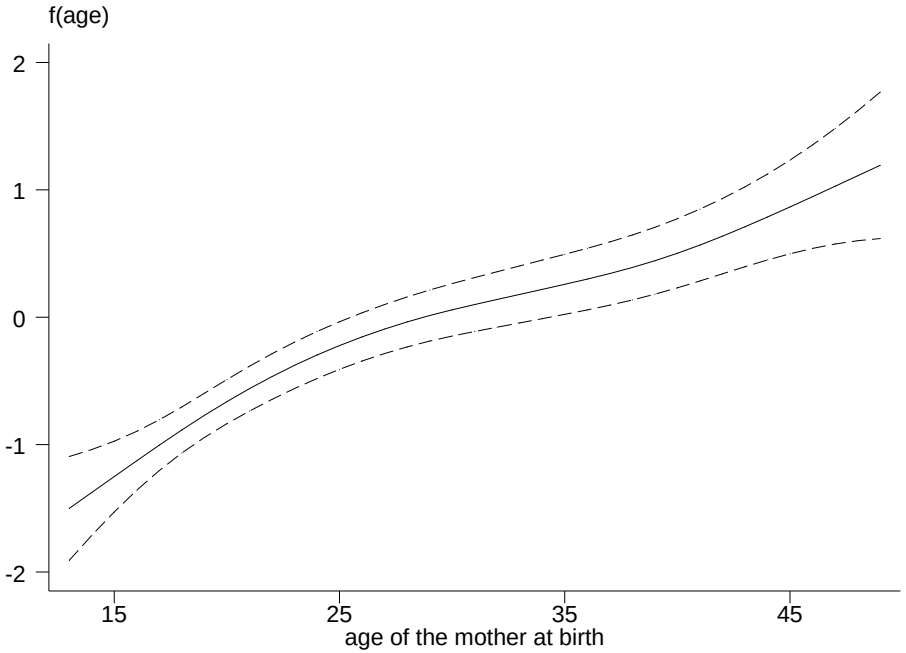
Software

- Implemented in the software package BayesX.

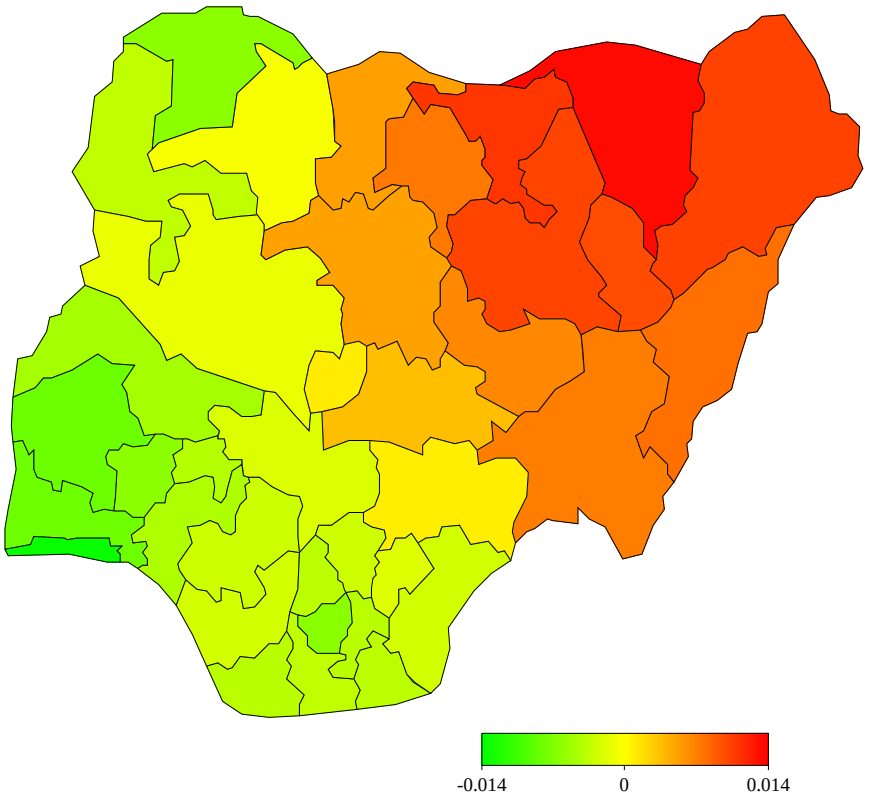
- Available from

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

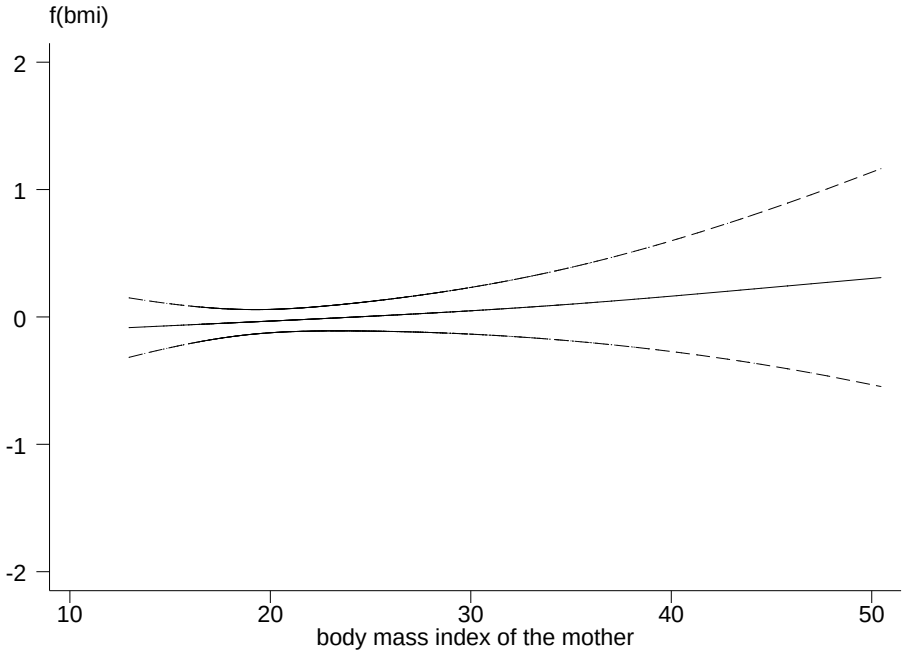
Childhood mortality in Nigeria II



Age of the mother at birth.

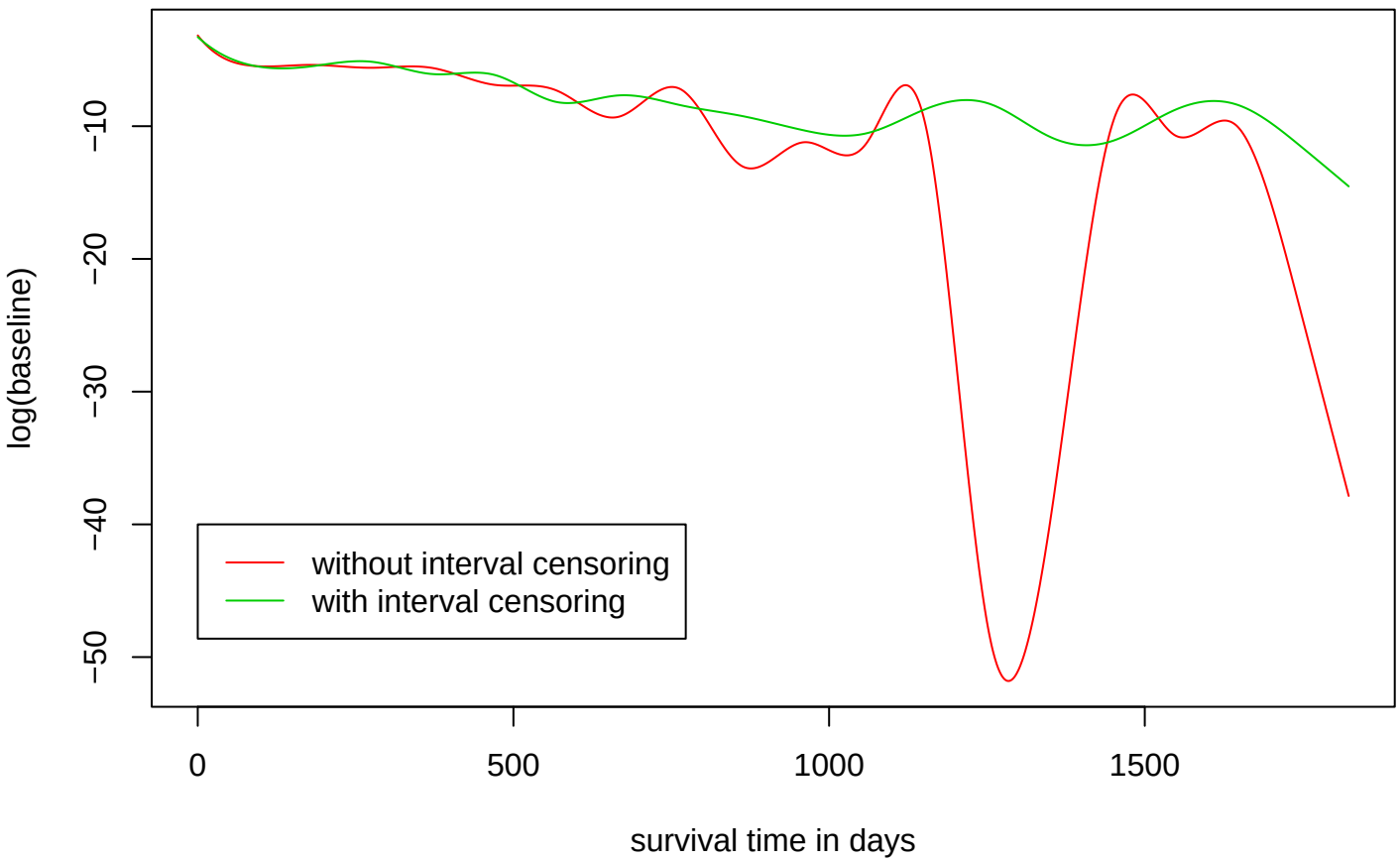


Spatial effect.



Body mass index of the mother.

variable	estimate	p-value
Intercept	-8.27	$<$0.0001
Breastfeeding (time-varying)	-4.27	$<$0.0001
Place of delivery	-0.54	0.0001
Long birth	0.28	0.0047
Christian	-0.52	0.0001
Other	-0.27	0.2879
Muslim (ref. category)		


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

Discussion

- 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`