Scale-Dependent Priors for Variance Parameters in Structured Additive Distributional Regression[1]

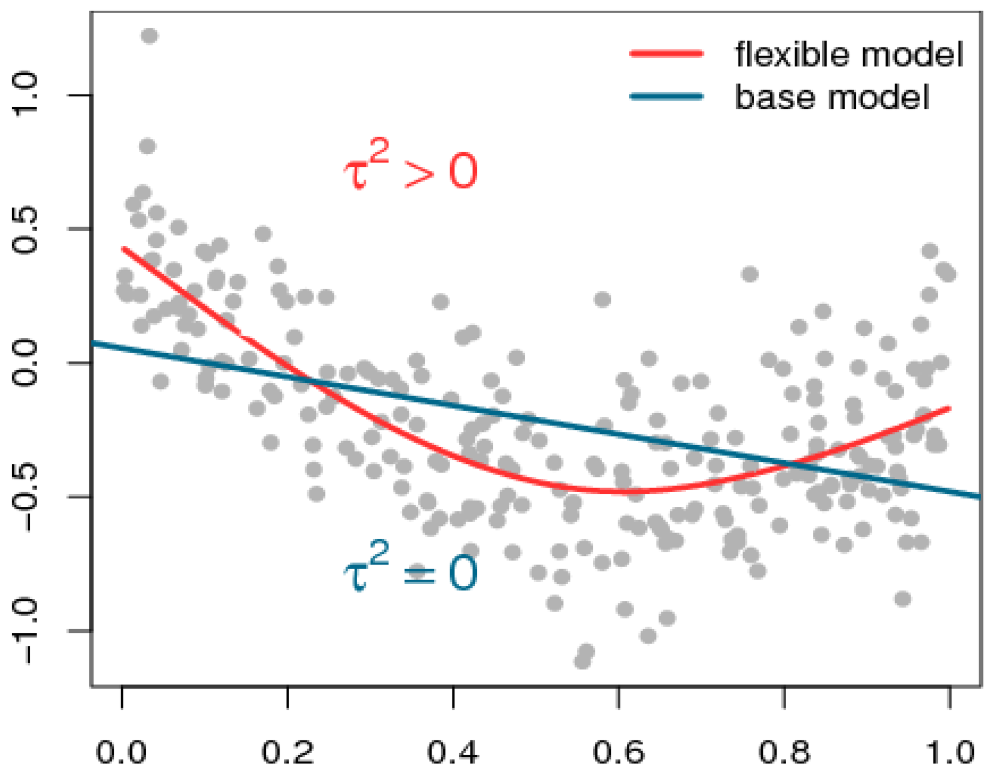
Nadja Klein, Thomas Kneib

Georg-August-University Göttingen

Introduction

Base Models and Increased Complexity

- Nested structure inherent in many model components.
- Hyperparameter τ^2 to determine the deviation of a flexible alternative to the base model.
- **Aims:** Construct principled prior distributions for variances of additive components in structured additive distributional regression (motivated from penalised complexity priors[2]).
  Develop Bayesian inference and elaborate similar approach for other priors from the literature.



- Alternative priors for τ^2 include half-normal (HN), half-Cauchy (HC) and uniform (U) priors for τ.
- HN priors for τ correspond to gamma, HC to generalised beta priors for τ^2.

Distributional Regression

M1: $\boldsymbol{y} \mid \boldsymbol{x} \sim F$ with parametric density $f(\boldsymbol{y}|\boldsymbol{\vartheta}_1, \dots, \boldsymbol{\vartheta}_K)$

M2: $\boldsymbol{\vartheta}_k(\boldsymbol{x}) = h_k(\boldsymbol{\eta}_k)$, $k = 1,\dots K$ with structured additive predictors

$$\eta_k = \sum_j g_j(\boldsymbol{x}) = \sum_j \boldsymbol{Z}_j \boldsymbol{\beta}_j$$

Standard Prior Specifications

- Regularisation or smoothness can be achieved via
  $$p(\boldsymbol{\beta}|\tau^2) \propto \exp(-1/(2\tau^2)\boldsymbol{\beta}'\boldsymbol{K}\boldsymbol{\beta})$$
  with prior precision $\boldsymbol{K}$ (specific to j,k, often rank deficient).
- For variances τ^2 inverse gamma (IG(a,b)) priors are used:

a,b=ε, small (flat on log-scale)	a=1,b=ε (flat for $1/\tau^2$)
a=-1,b=0 (flat for τ^2)	a=-0.5,b=0 (flat for τ).

Methodology

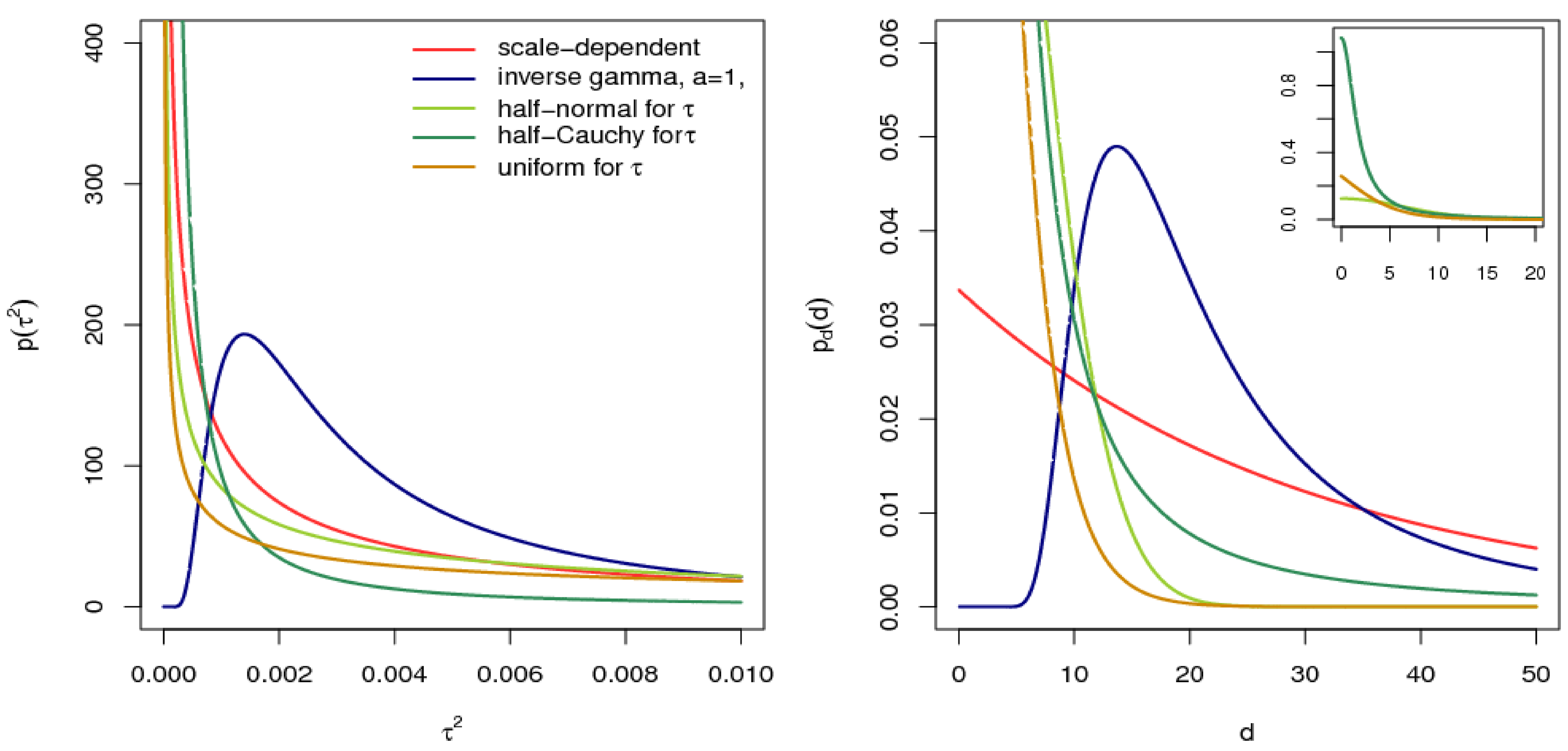
Scale-Dependent (SD) Priors

SD prior for τ^2 has density of a Weibull distribution with shape parameter a=0.5 and scale parameter θ resulting from three construction principles[2]:

Occam's razor: Prior should invoke principle of parsimony.

Measure of complexity: Increased complexity between base model p_b and flexible alternative p is measured by $d(p||p_b)=\sqrt{2\mathrm{KLD}(p||p_b)}$.

Constant rate penalisation: Implies exponential prior on distance scale d, i.e. constant rate of decay in the distance prior from the base model to stronger deviations.



User-Defined Scaling

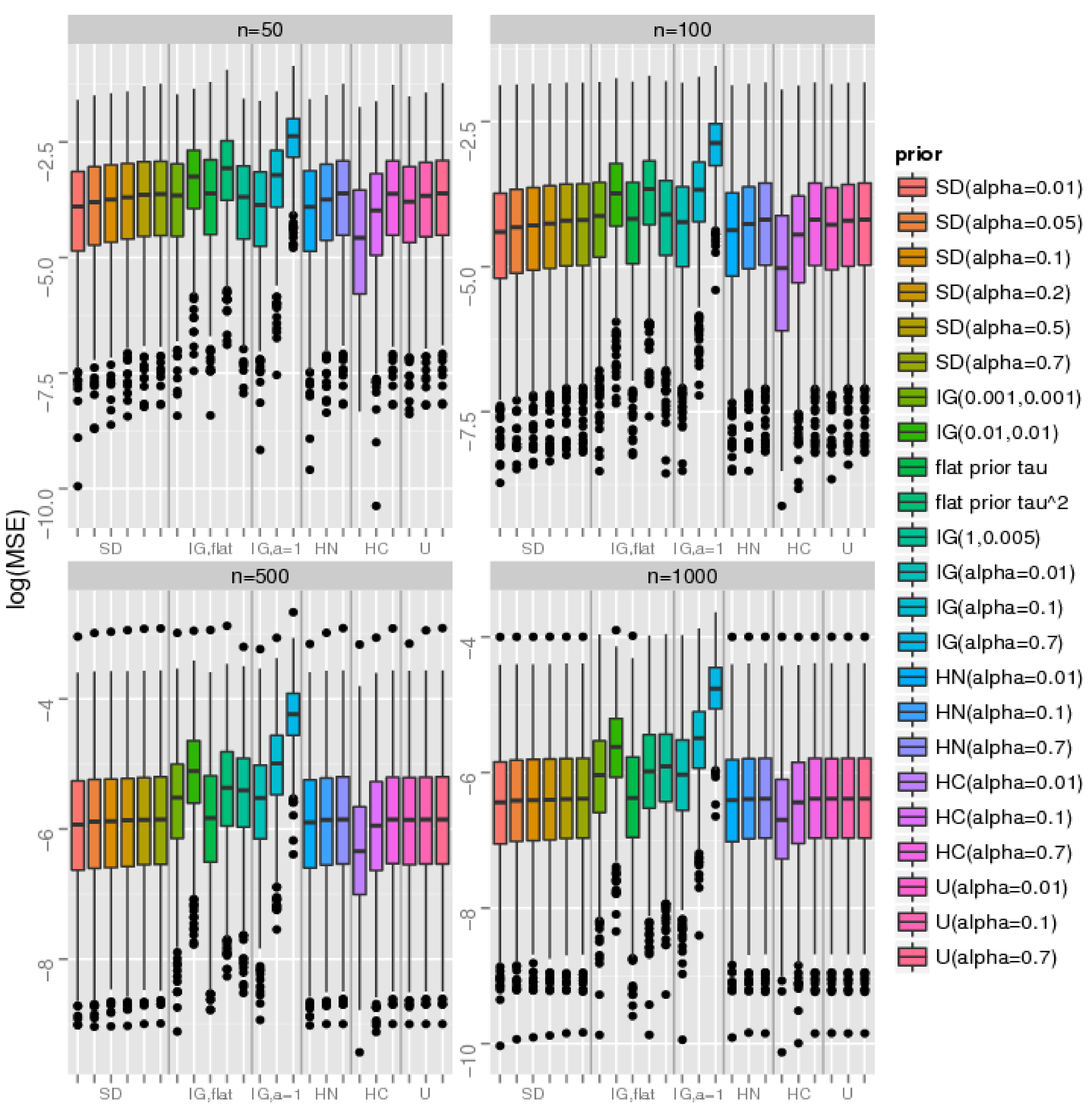
- Decay rate $\exp(-\lambda)$ in the exponential prior is controlled implicitly through scale parameter θ of the prior for τ^2 by $P(q(\tau^2 \leq c))$=1−α.
- Prior knowledge on effects g=$\boldsymbol{Z\beta}$ allows to specify interval with high marginal probability, $P(|g(\boldsymbol{x})| \leq c\ \forall \boldsymbol{x} \in D) \geq 1-\alpha$. Resulting integral can be approximated numerically.
- User-defined scaling has been transferred to HN, HC and U priors for τ (and IG priors for τ^2 when a=1).
- Optimising θ for given $\boldsymbol{Z}$ and $\boldsymbol{K}$ is implemented in the **R**-package `sdPrior` on CRAN.

MCMC Inference

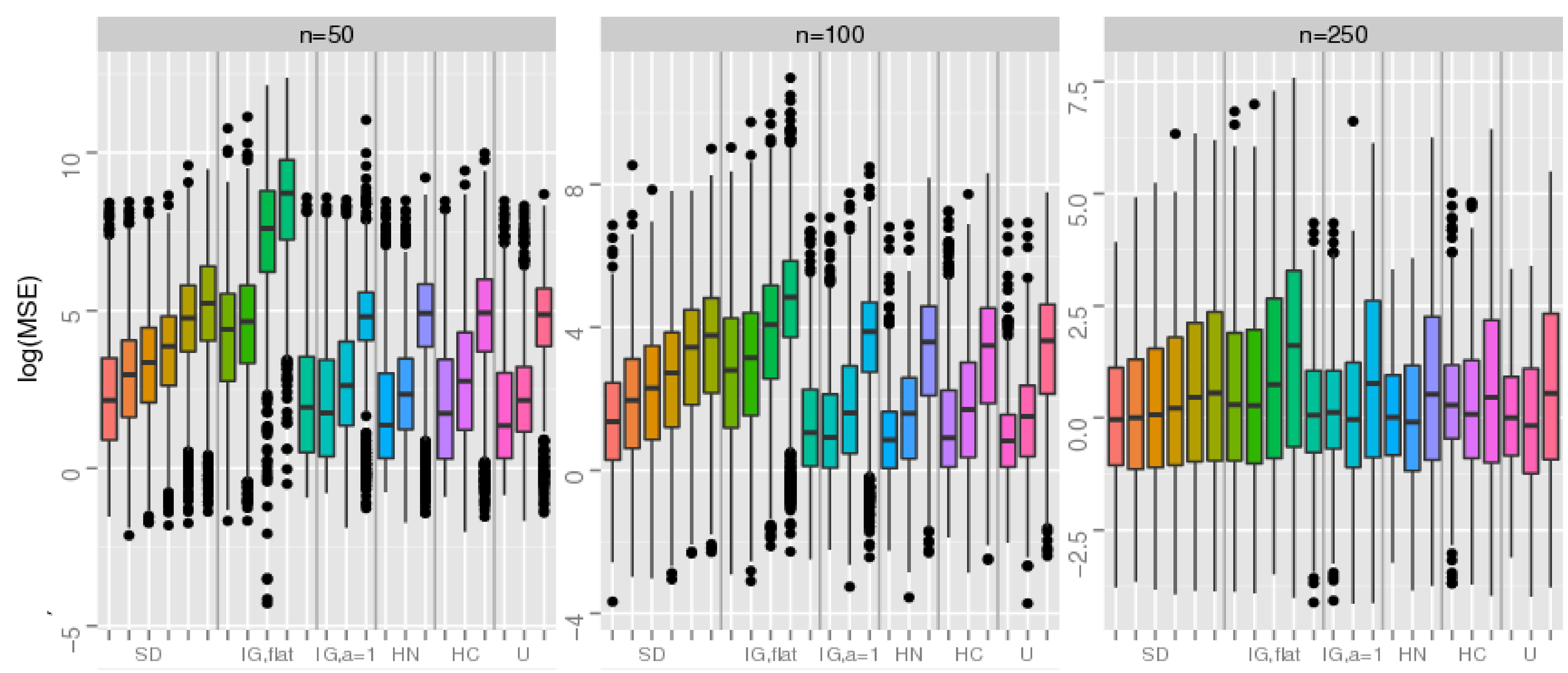
- Full conditional $p(\tau^2|\cdot)$ has no closed form when $p(\tau^2)$ is SD prior or priors derived from HN, HC and U for τ.
- Still feasible with Metropolis-Hastings steps, similar as for $\boldsymbol{\beta}$.
- Construction of proposal densities based on approximations to $\log(p(\log(\tau^2)\ |\cdot))$.
- Empirically, MCMC chains have good convergence behaviour.
- Conditions for propriety of joint posterior with SD prior are derived.
- Implementation in open software BayesX (`www.bayesx.org`).

Empirical Results

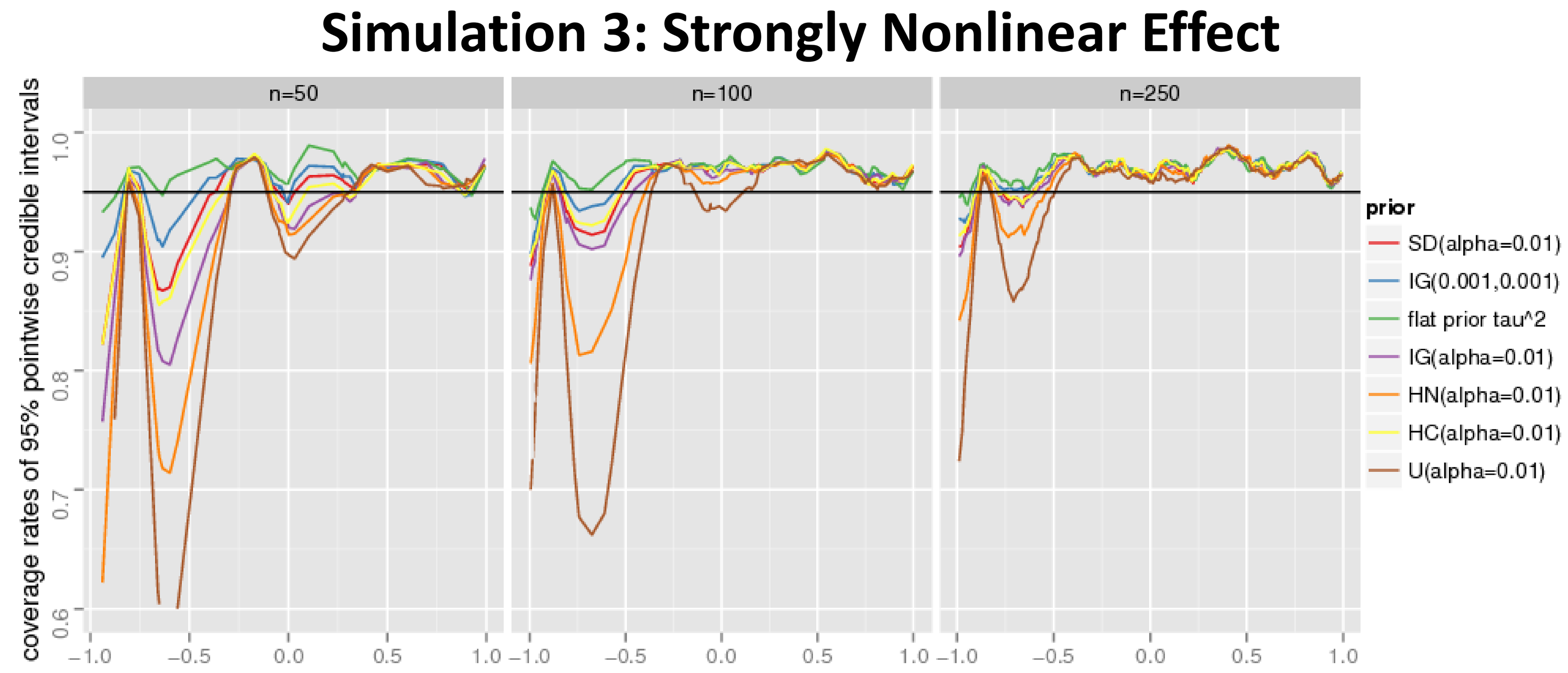
Simulation 1: Close to the Base Model



Simulation 2: Flat Likelihood



Simulation 3: Strongly Nonlinear Effect



Overall: SD priors are fairly robust in all considered scenarios.

1 Klein, N. and Kneib, T. (2015). Scale-Dependent Priors for Variance Parameters in Structured Additive Distributional Regression. Working paper under revision for *Bayesian Analysis*. Available on request.

2 Simpson, D., Martins, T., Riebler, A., Fuglstad, G., Rue, H. and Sørbye, S. (2015). Penalising model component complexity: A principled, practical approach to constructing priors. Working paper, arXiv:1403.4630.