Bayesian inference for structured additive quantile regression models

Yu Yue *
Baruch College, City University of New York
and
Håvard Rue
Norwegian University for Science and Technology

July 5, 2009

Abstract

Most quantile regression problems in practice require flexible semiparametric forms of the predictor for modeling the dependence of responses on covariates. Furthermore, it is often necessary to add random effects accounting for overdispersion caused by unobserved heterogeneity or for correlation in longitudinal data. We present a unified approach for Bayesian quantile inference via Markov chain Monte Carlo (MCMC) simulation and approximate inference using integrated nested Laplace approximations (INLA) in additive and semiparametric mixed models. Different types of covariates are all treated within the same general framework by assigning appropriate Gaussian Markov random field (GMRF) priors with different forms and degrees of smoothness. We applied the approach in several simulation and case studies, showing that the methods are also computationally efficient in problems with many covariates and large datasets.

Keywords: Nonparametric quantile regression; Asymmetric Laplace distribution; Gaussian Markov random fields; Integrated nested Laplace approximations; Gibbs sampling; Additive and semiparametric mixed models; Spatial effect.

*Address for correspondence: Yu (Ryan) Yue, Department of Statistics and CIS, Baruch College, City University of New York, New York, NY 10010; *Email: yu.yue@baruch.cuny.edu*

1 Introduction

This paper concerns the models of the following type

$$
\begin{aligned}
y_j &= \eta_j + \varepsilon_j, \quad \varepsilon_j \sim F(\varepsilon_j|\boldsymbol{\theta}), \quad j \in J \\
\eta_i &= \boldsymbol{x}_i^T\boldsymbol{\beta} + \sum_{k=1}^{n_f} f_k(u_{ki}) + b_i, \quad i = 1, \dots, n
\end{aligned} \tag{1}
$$

where y_j denotes jth observation and η_j is its predictor; J is a subsect of $\{1, \dots, n\}$, that is, not necessarily all latent η_j are observed through the data y_j; F is an unknown distribution function of error terms ε_i and may depend on some additional parameters $\boldsymbol{\theta}$; $\boldsymbol{x}_i$ is a vector of n_β covariates with a linear effect and $\boldsymbol{\beta}$ is the corresponding unknown vector; b_i are unstructured random effects; $f_k(u_{ki})$ is the effect of a generic covariate k which assumes value u_{ki} for ith observation. The functions f_k, $k = 0, \dots, n_f$, comprise usual nonlinear effect of continuous covariates, time trends and seasonal effects, as well as two dimensional surfaces, interactions, and spatial effects.

Assuming y_j belong to an exponential family and its mean is linked to η_i through certain link function, we would have structured additive regression (STAR) models. Thanks to the various forms that the functions $\{f_k(\cdot)\}$ can take, the STAR models have a wealth of applications (see, e.g., Fahrmeir and Tutz, 2001, for a detailed account). However, it might be the case that the upper or lower quantiles of y_j may depend on the covariates quite differently from the center. Therefore, inference on conditional quantiles can provide a more complete description of functional changes than focusing solely on the mean. The STAR models, unfortunately, only estimate conditional mean. To overcome this limitation, we extend a subclass of the STAR models, named *latent Gaussian models* (Rue et al., 2009), to quantile regression context.

Quantile regression, a completely distribution free approach, has emerged as a useful supplement to ordinary mean regression. Its value has been demonstrated by rapidly expanding literatures in econometrics, social sciences, and biomedical studies; see Koenker (2005) for a comprehensive review. Given a fixed and known quantile $\tau \in (0, 1)$, we assume that the τth quantile of the error distribution in (1) is zero, i.e., $F^{-1}(\tau|\boldsymbol{\theta}) = 0$. Then the corresponding quantile function of the continuous response variable Y_i is

$$
\mathrm{Q}_{Y_i}(\tau|\boldsymbol{x}_i, \boldsymbol{u}_i) = \eta_{\tau i} = \boldsymbol{x}_i^T\boldsymbol{\beta}_\tau + \sum_{k=1}^{n_f} f_{\tau k}(u_{ki}) + b_{\tau i}, \tag{2}
$$

where predictor $\eta_{\tau i}$ is composed of a linear effect $\boldsymbol{\beta}_\tau$, a sum of nonlinear effects $f_{\tau k}$ for both univariate and bivariate covariates, and an unstructured random effect $\boldsymbol{b}_\tau$. We here term the proposed model (2) as structured additive quantile regression (STAQ) model.

An alternative representation of the STAQ quantile regression model can be achieved via the following minimization problem:

$$\arg\min_{\boldsymbol{\eta}_\tau} \sum_{i=1}^{n} \rho_\tau(y_i - \eta_{\tau i}), \quad \text{where} \quad \rho_\tau(u) = \begin{cases} u\tau & u \geq 0 \\ u(\tau - 1) & u < 0 \end{cases}, \tag{3}$$

is the so called 'check function' of Koenker and Bassett (1978), and $\eta_{\tau i}$ has the same structure as in (2). Thus, the check function is the appropriate loss function for quantile regression problems regarded from a decision theoretical point of view. To estimate STAQ model using criterion (3), we develop a fully Bayesian inferential framework. The method is based on assuming Y_i have iid (identically independent distributed) asymmetric Laplace distributions and taking appropriate Gaussian type priors on the additive components in the predictor. Two different tools are proposed for such Bayesian quantile inference. The first approach is carried by Markov chain Monte Carlo (MCMC) sampling while the second one is a relatively new technique using integrated nested Laplace approximation (INLA) introduced in Rue et al. (2009). Both approaches are fairly efficient and provide reliable quantile inference as we shall demonstrate via extensive simulation studies together with data analysis.

From the frequentist point of view, there is extensive literature on additive quantile regression, a special case of our STAQ model. To review a few, Koenker et al. (1994) and Koenker and Mizera (2004) used total variation regularization approach to estimate possible univariate and bivariate nonparametric terms; De Gooijer and Zerom (2003), Yu and Lu (2004), Horowitz and Lee (2005), and Cai and Xu (2008) based their estimation on local polynomial fitting; Takeuchi et al. (2006) and Li et al. (2007) explored reproducing kernel Hilbert space (RKHS) norm for nonparametric quantile estimation; Fenske et al. (2009) used boosting algorithm to allow for data-driven determination of the amount of smoothness required for the nonlinear effects and combines model selection with an automatic variable selection property.

Compared to its frequentist counterparts, STAQ offers advantages in the following aspects. First, it is much more flexible. Besides ordinary linear and nonlinear effects, STAQ can easily account for time trends, for two dimensional surfaces, interactions and spatial effect, for overdispersion caused by unobserved heterogeneity, or for correlation in longitudinal data. Second, with suitable hyperpriors there are no tuning parameters to estimate. This contrasts with virtually all other frequentist methods, which require some kind of *ad hoc* bandwidth or smoothing parameter selectors (e.g., AIC and SIC). Third, the posterior captures all the uncertainty in an estimate, including the uncertainty that would come from selecting a smoothing parameter. Finally, the computation is fairly efficient for large datasets and a standard R package is available for easy implementation.

The remainder of this paper is organized as follows. Section 2 provides necessary information about asymmetric Laplace distribution which would be used as likelihood in our Bayesian quantile regression models. Section 3 specifies all the priors assigned for the latent parameters in model (1). The MCMC and INLA inference approaches are both introduced in Section 4. The simulated and real data examples are provided in Section 5, followed by a general discussion in Section 6.

2 Asymmetric Laplace distribution

Bayesian inference requires likelihood. We need an assumption on data distribution for Bayesian quantile inference because the classical quantile regression has no such restriction. A possible parametric link between the minimization problem (3) and the maximum likelihood theory is the asymmetric Laplace density. This skewed distribution appeared in Koenker et al. (1999) and Yu and Moyeed (2001), among others. We say that a random variable Y has asymmetric Laplace distribution with parameters η, δ_0 and τ, and write it as $Y \sim \text{ALD}(\eta, \delta_0, \tau)$, if the corresponding probability density function is given by

$$[y \mid \eta, \delta_0, \tau] = \tau(1-\tau)\delta_0 \exp\left\{-\delta_0 \rho_\tau\left(y-\eta\right)\right\}, \tag{4}$$

where $\eta \in \mathbb{R}$ is the location parameter, $\delta_0 \in \mathbb{R}^+$ is the scale parameter, and $0 < \tau < 1$ is the skewness parameter. Note that the check function ρ_τ assigns weight τ or $1-\tau$ to the observations greater or less than η respectively, and $P(Y \leq \eta) = \tau$. Therefore, the distribution splits along the scale parameter into two parts, one with probability τ to the left, and one with probability $1-\tau$ to the right. See Yu and Zhang (2005) for further properties of this distribution.

Assuming that $y_i \overset{iid}{\sim} \text{ALD}(\eta_i, \delta_0, \tau)$, then the likelihood for n independent observations is proportional to

$$L(\boldsymbol{\eta}, \delta_0; \boldsymbol{y}, \tau) \propto \delta_0^{n/2} \exp\Big\{ -\delta_0 \sum_{i=1}^{n} \rho_\tau(y_i - \eta_i) \Big\}. \tag{5}$$

It is easy to see that the maximization of the likelihood in (5) with respect to $\boldsymbol{\eta}$ given δ_0 is equivalent to the minimization of the loss function in (3). Thus, the asymmetric Laplace distribution is useful in terms of unifying the likelihood and the inference for quantile regression estimation. Following the same error assumption, Yu and Moyeed (2001), among others, implemented successful Bayesian inference for (parametric) quantile regression, and Geraci and Bottai (2007) and Yuan and Yin (2009) studied it for longitudinal data.

Another noteworthy feature about asymmetric Laplace distribution is that it can be represented as a scale mixture of normals (Tsionas, 2003; Rue and Held, 2005):

$$Y \overset{\mathcal{D}}{=} \eta + \xi W + \sigma Z \sqrt{\delta_0^{-1} W}, \quad \text{where} \quad \xi = \frac{1-2\tau}{\tau(1-\tau)}, \quad \sigma^2 = \frac{2}{\tau(1-\tau)} \tag{6}$$

are two scalars depending on τ. The random variables $W > 0$ and Z are independent and have exponential and normal distributions respectively,

$$\pi(w \mid \delta_0) = \delta_0 \exp(\delta_0 w) \quad \text{and} \quad z \sim N(0, 1).$$

This mixture expression will bring huge convenience to MCMC inference in Section 4.1.

There exist some other approaches to specify error distributions for Bayesian quantile regression. To avoid the restrictive parametric assumption, Kottas and Krnjajic (2009) and Taddy and Kottas (2009) constructed general classes of semiparametric and nonparametric distributions for likelihood using Dirichlet process mixture models. Reich et al. (2008) assume a flexible infinite mixture of normals for errors and use a stick-breaking construction for specifying priors. In addition, Dunson and Taylor (2005) used an approximate method based on the substitution likelihood for quantiles. The reasons we chose asymmetric Laplace distribution in this work are two-fold. First, it facilitates Bayesian inference and computation in the complicated additive models we concern there. Second, it is flexible enough to accommodate different kinds of error distributions (e.g., symmetric, heavy-tailed, or skewed) as we shall show in Section 5.1.

3 Prior distributions

3.1 Smoothness priors for functions

To facilitate description of our method, we will suppress the subscription τ of regression effects in the following. Priors for unknown functions $f_k(\cdot), k = 1, \ldots, n_f$, belong to the class of Gaussian Markov random fields (GMRF), whose specific forms depend on covariate types and on prior beliefs about the smoothness of f_k. Although only GMRF is used in this work, there exist some other options of smoothness priors for functions, e.g., Bayesian P-splines (Lang and Brezger, 2004).

Let $\boldsymbol{f} = (f(u_1), \ldots, f(u_n))^T$, a random vector of the response at u_i, $i = 1, \ldots, n$. We say $\boldsymbol{f}$ is a GMRF with mean $\boldsymbol{\mu}$ and precision (the inverse covariance) matrix $\delta\boldsymbol{Q}$ if and only if it has density of form

$$\pi(\boldsymbol{f} \mid \delta) \propto \delta^{(n-m)/2} \exp\bigg(-\frac{\delta}{2}(\boldsymbol{f} - \boldsymbol{\mu})'\boldsymbol{Q}(\boldsymbol{f} - \boldsymbol{\mu})\bigg), \tag{7}$$

where $\boldsymbol{Q}$ is a semidefinite matrix of constants with rank $n - m\,(m \geq 0)$. The properties of a particular GMRF are all reflected through matrix $\boldsymbol{Q}$. For instance, the Markov properties of GMRFs totally depend on the various sparse structures that the matrix $\boldsymbol{Q}$ may have. In this paper we use two kinds of GMRFs: continuous random walk (CRW) models (see, e.g. Wecker and Ansley, 1983) for metrical convariates and intrinsic autoregressive models (see, e.g. Besag and Kooperberg, 1995) for spatial covariates. Those two GMRFs share the form (7) but with different structures of $\boldsymbol{Q}$. The GMRFs have been sucessfully used in many applications; see Rue and Held (2005) for a comprehensive review.

3.1.1 Metrical covariates

Let $u_1 < u_2 < \cdots < u_n$ be the set of continuous locations and $z_i = f(u_i)$ be the function evalutions at u_i, for $i = 1, \ldots, n$. The construction of CRW model is based on a discretely observed continuous time process $z(u)$ that is a realization of an $m - 1$ fold integrated Wiener process given by

$$z(u) = \int_0^u \frac{(u-h)^{m-1}}{(m-1)!} dW(h), \tag{8}$$

where $W(h)$ is a standard Wiener process. One can show that density of $z(u)$ is Gaussian with zero mean and a completely dense matrix $\boldsymbol{Q}$. Since factorizing an $n \times n$ dense matrix is at cost $\mathcal{O}(n^3)$, such matrix $\boldsymbol{Q}$ would make Bayesian computation cumbersome with a huge dataset. Fortunately, the solution of (8) does have a Markov property on an augmented space

$$\tilde{\boldsymbol{z}} = \left[z_1, z_1^{(1)}, \ldots, z_1^{(m-1)}, \ldots, z_n, z_n^{(1)}, \ldots, z_n^{(m-1)}\right]^T,$$

where also the derivatives (up to mth order) at u_i are included. The mn dimensional vector $\tilde{\boldsymbol{z}}$ is also (singular) Gaussian with mean zero but a banded precision $\boldsymbol{Q}$ of bandwidth $2m-1$ (see Rue and Held, 2005, chapter 3.5 for details). For inference using MCMC, we will first use efficient algorithms for sparse matrices (Rue and Held, 2005) to simulate the full vector $\tilde{\boldsymbol{z}}$ at cost $\mathcal{O}(n)$, and then simply ignore the derivatives in the analysis. The CRW models have been successfully used as priors for smoothing splines in Wahba (1990) and for varying-coefficient models in Eubank et al. (2004), who instead used recursions based on the Kalman filter.

We now present the important case $m = 2$. Define $\kappa_i = u_{i+1} - u_i$ and the matrices

$$\boldsymbol{A}_i = \begin{pmatrix} 12/\kappa_i^3 & 6/\kappa_i^2 \\ 6/\kappa_i^2 & 4/\kappa_i \end{pmatrix} \quad \boldsymbol{B}_i = \begin{pmatrix} -12/\kappa_i^3 & 6/\kappa_i^2 \\ -6/\kappa_i^2 & 2/\kappa_i \end{pmatrix} \quad \boldsymbol{C}_i = \begin{pmatrix} 12/\kappa_i^3 & -6/\kappa_i^2 \\ -6/\kappa_i^2 & 4/\kappa_i \end{pmatrix}$$

for $i = 1, \ldots, n-1$. Then the vector $\tilde{\boldsymbol{z}} = [z_1, z_1^{(1)}, \ldots, z_n, z_n^{(1)}]^T$ has

$$\boldsymbol{Q} = \begin{pmatrix} \boldsymbol{A}_1 & \boldsymbol{B}_1 & & & & \\ \boldsymbol{B}_1' & \boldsymbol{A}_2 + \boldsymbol{C}_1 & \boldsymbol{B}_2 & & & \\ & \boldsymbol{B}_2' & \boldsymbol{A}_3 + \boldsymbol{C}_2 & \boldsymbol{B}_3 & & \\ & & \ddots & \ddots & \ddots & \\ & & & & & \boldsymbol{B}_{n-1} \\ & & & & \boldsymbol{B}_{n-1}' & \boldsymbol{C}_{n-1} \end{pmatrix}. \tag{9}$$

Here we have used diffuse priors for z_1 and $z_1^{(1)}$. The null space of $\boldsymbol{Q}$ is spanned by two vectors $(u_1, 1, u_2, 1, \ldots, u_n, 1)'$ and $(1, 0, 1, 0, \ldots, 1, 0)'$. When we go to higher-order models, the matrix $\boldsymbol{Q}$ has the same structure as (9), but the matrices $\boldsymbol{A}_i$, $\boldsymbol{B}_i$, and $\boldsymbol{C}_i$ will differ (see Rue and Held, 2005, for more information).

3.1.2 Spatial covariates

Let us now turn our attention to a spatial covariate u, where the values of u represent the location or site in connected geographical regions. A common way to deal with spatial covariates is based on a set of predefined neighbors for each u_i. For geographical data as considered in this paper we usually assume that two sites u_i and u_j are neighbors if they share a common boundary. Letting n_i denote the number of neighbors of site u_i, we assume the following spatial smoothness prior for the function evaluations $f(u_i)$, $i = 1, \ldots, n$:

$$f(u_i) \mid \{f(u_j)\, j \neq i\}, \delta \sim N\left(\frac{1}{n_i} \sum_{j:j\sim i} f(u_j), \frac{1}{n_i \delta}\right), \tag{10}$$

where $j \sim i$ denotes that site u_i and u_j are neighbors. Thus the conditional mean of $f(u_i)$ is an unweighted average of evaluations of neighboring sites. The joint prior density of $\boldsymbol{f}$ in (10) has the same expression as in (7) with mean zero and matrix $\boldsymbol{Q}$ such that

$$Q_{ij} = \begin{cases} n_i & i = j, \\ -1 & i \sim j, \\ 0 & \text{otherwise.} \end{cases} \tag{11}$$

It is easy to see that the sparse matrix $\boldsymbol{Q}$ has rank $n - 1$.

A more general prior including expression (10) as a special case is given by

$$f(u_i) \mid \{f(u_j)\, j \neq i\}, \delta \sim N\left(\frac{\sum_{j:j\sim i} w_{ij} f(u_j)}{\sum_{j:j\sim i} w_{ij}}, \frac{1}{\delta \sum_{j:j\sim i} w_{ij}}\right),$$

where w_{ij} are known weights, e.g., the inverse Euclidian distance between the sites of u_i and u_j. The corresponding matrix $\boldsymbol{Q}$ has the same structure as in (11) but different entries (see Rue and Held, 2005, page 103). Other options for modeling spatial effects like conditional autoregressive (CAR) models and stationary Gaussian random field (Kriging) models are proper, but with the same general form, see Rue and Held (2005).

Let $\boldsymbol{f}_k$, $k = 1, \ldots, n_f$, be a vector of kth function component in (1). Since they are of the same form, we assume the priors on functions $\boldsymbol{f}_k$ are

$$\pi(\boldsymbol{f}_k \mid \delta_k) \propto \delta_k^{(n-m_k)/2} \exp\left(-\frac{\delta_k}{2} \boldsymbol{f}_k' \boldsymbol{Q}_k \boldsymbol{f}_k\right),$$

where $\boldsymbol{Q}_k$ has rank $n - m_k$ and its structure depends on the type of covariates.

3.2 Further prior assumptions

For a fully Bayesian analysis, hyperpriors for all precisions $(\delta_0, \delta_1, \ldots, \delta_{n_f})$ are introduced in a further stage of the hierarchy. Common choices are Gamma priors with parameters a and c

so that the mean is a/c and the variances is a/c^2. We make Gamma priors proper and highly dispersed, avoiding problems arising with improper priors.

For the fixed effect parameters $\{\beta_j\}$s, we shall assume independent diffuse priors $\pi(\beta_j) \propto$ constant or a weakly informative Gaussian prior $\boldsymbol{\beta} \sim N(\mathbf{0}, \phi^{-1}\boldsymbol{I})$ with small precision ϕ. If $\boldsymbol{\beta}$ is a high-dimensional vector, one may consider using Bayesian regularization priors developed in Kneib et al. (2009), where conditionally Gaussian priors are assigned with suitable hyperprior assumptions on the variances inducing the desired shrinkage and sparseness on coefficient estimates. Those priors can be easily utilized in the models proposed here with small effort on modifying algorithms.

For unstructured random effects, we make the usual assumption that the b_i are iid Gaussian with precision ν,

$$b_i \mid \nu \sim N(0, \nu^{-1}), \quad i = 1, \ldots, n,$$

and define again a highly dispersed hyperprior for ν.

4 Posterior inference

4.1 Markov chain Monte Carlo approaches

Recall that we assume likelihood to be asymmetric Laplace distribution, which has a representation as given by (6). Letting $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_{n_f})'$, the posterior distribution can be obtained through

$$[\boldsymbol{\eta} \mid \boldsymbol{y}] \propto \int p(y \mid \boldsymbol{\eta}, \delta_0, \boldsymbol{w})\, \pi(\boldsymbol{w} \mid \delta_0)\, \pi(\boldsymbol{\eta} \mid \boldsymbol{\delta}, \boldsymbol{\nu})\, \pi(\boldsymbol{\nu})\, \pi(\boldsymbol{\delta})\, \pi(\delta_0)\, d\boldsymbol{w}\, d\boldsymbol{\nu}\, d\boldsymbol{\delta}\, d\delta_0, \tag{12}$$

where $y \mid \boldsymbol{\eta}, \delta_0, \boldsymbol{w} \sim N(\boldsymbol{\eta} + \xi \boldsymbol{w}, \sigma^2 \boldsymbol{w} \delta_0^{-1})$ with the priors and hyperpriors defined in Section 3. The expression (12) yields a traceable and efficient Gibbs sampler that works as follows:

1. Simulate $w_i^{-1} \mid \cdot \sim$ Inverse Gaussian(μ', λ'), $i = 1, \ldots, n$, where

   $$\mu' = \sqrt{\frac{\xi^2 + 2\sigma^2}{(y_i - \eta_i)^2}} \quad \text{and} \quad \lambda' = \frac{\delta_0(\xi^2 + 2\sigma^2)}{\sigma^2},$$

   in the parameterization of the inverse Gaussian density given by

   $$f(x) = \sqrt{\frac{\lambda'}{2\pi}} x^{-3/2} \exp\left\{ -\frac{\lambda'(x - \mu')^2}{2(\mu')^2 x} \right\}, \quad x > 0;$$

   see, e.g., Chhikara and Folks (1989).

2. Simulate fixed effect parameters $\boldsymbol{\beta} \mid \cdot \sim N(\boldsymbol{\mu_\beta}, \sigma^2\delta_0^{-1}\boldsymbol{\Sigma_\beta})$ where

$$\boldsymbol{\mu_\beta} = \boldsymbol{\Sigma_\beta D}_w^{-1}(\boldsymbol{y} - \boldsymbol{\eta}_{-\boldsymbol{\beta}} - \xi\boldsymbol{w}), \quad \boldsymbol{\Sigma_\beta} = \left(\boldsymbol{X}'\boldsymbol{D}_w^{-1}\boldsymbol{X}\right)^{-1},$$

and $\boldsymbol{D}_w = \mathrm{diag}(w_1, \ldots, w_n)$.

3. Simulate function vectors $\boldsymbol{f}_k \mid \cdot \sim N(\boldsymbol{\mu}_k, \sigma^2\delta_0^{-1}\boldsymbol{\Sigma}_k)$, $k = 1, \ldots, n_f$, where

$$\boldsymbol{\mu}_k = \boldsymbol{\Sigma}_k\boldsymbol{D}_{\boldsymbol{w}}^{-1}(\boldsymbol{y} - \boldsymbol{\eta}_{-k} - \xi\boldsymbol{w}), \quad \boldsymbol{\Sigma}_k = (\boldsymbol{D}_{\boldsymbol{w}}^{-1} + \sigma^2\lambda_k\boldsymbol{Q}_k)^{-1},$$

and $\lambda_k = \delta_k/\delta_0$. For identifiability, we add sum-to-zero constraint to each $\boldsymbol{f}_k$ by computing $\boldsymbol{f}_k^* = \boldsymbol{f}_k - \boldsymbol{\Sigma}_k\mathbf{1}(\mathbf{1}'\boldsymbol{\Sigma}_k\mathbf{1})^{-1}\mathbf{1}'\boldsymbol{f}_k$ (see Rue and Held, 2005, section 2.3.3).

4. Simulate random effect parameters $\boldsymbol{b} \mid \cdot \sim N(\boldsymbol{\mu_b}, \sigma^2\delta_0^{-1}\boldsymbol{\Sigma_b})$ where

$$\boldsymbol{\mu_b} = \boldsymbol{\Sigma_b D}_w^{-1}(\boldsymbol{y} - \boldsymbol{\eta}_{-\boldsymbol{b}} - \xi\boldsymbol{w}), \quad \boldsymbol{\Sigma_b} = \left(\boldsymbol{D}_w^{-1} + \nu\boldsymbol{I}_n\right)^{-1}.$$

5. Simulate error precision

$$\delta_0 \mid \cdot \sim \mathrm{Gamma}\left(a_0 + \frac{3n}{2}, c_0 + \frac{1}{2\sigma^2}\sum_{i=1}^{n} w_i^{-1}(y_i - \eta_i - \xi w_i)^2 + \sum_{i=1}^{n} w_i\right).$$

6. Simulate smoothing precision δ_k, $k = 1, \ldots, n_f$

$$\delta_k \mid \cdot \sim \mathrm{Gamma}\left(a_k + \frac{n - m_k}{2}, c_k + \frac{1}{2}\boldsymbol{f}_k'\boldsymbol{Q}_k\boldsymbol{f}_k\right).$$

7. Simulate random effect precision

$$\nu \mid \cdot \sim \mathrm{Gamma}\left(a_\nu + \frac{n}{2}, b_\nu + \frac{1}{2}\boldsymbol{b}'\boldsymbol{b}\right).$$

Drawing samples iteratively from those regular full conditional distributions are straightforward, without any further need for tuning. Still, some of the updating steps require additional comments. For the high-dimensional vector $\boldsymbol{\beta}$, it will typically be necessary to apply a block-update since a complete update is extremely expensive due to the high-dimensional matrix computation. For the smooth and spatial effects, one can take advantage of sparse matrix structures possessed by the precision matrices of the full conditionals. Employing the sparse matrix Cholesky decomposition described in Rue (2001) and Rue and Held (2005), the simulation of the high-dimensional

Gaussian effects becomes feasible as demonstrated by the examples in this paper. The notorious correlation between GMRF and its precision parameter tends to make MCMC methods exhibit poor performance. One may use the one-block approach (see Rue and Held, 2005, chapter 4) to improve Markov chain's mixing, or employ a conceptually different inference tool as described in next section to completely avoid using MCMC.

4.2 Inference by integrated nested Laplace approximations

Integrated nested Laplace approximations (INLA) is a new approach to statistical inference for latent Gaussian models introduced by Rue and Martino (2007) and Rue et al. (2009). It provides a fast, deterministic alternative to MCMC which is the standard tool for inference in such models. The main advantage of the INLA approach over MCMC is that it is much faster to compute; it gives answers in minutes and seconds where MCMC requires hours and days. In addition, the INLA inference does not suffer from the slow convergence and poor mixing issues that the MCMC methods often have for the models concerned here.

In short, the INLA approach provides a recipe for fast Bayesian inference using accurate approximations to the marginal posterior density for the hyperparameters and the posterior marginal densities for the latent variables. The approximate posterior marginals can then be used to compute summary statistics of interest, such as posterior means, variances or quantiles. The theory behind INLA is thoroughly described in Rue et al. (2009) and will not be repeated here.

Unfortunately, we cannot directly implement the INLA method to our models since the asymmetric Laplace distribution is assumed to be the likelihood and the second order derivative of its log likelihood is zero. However, the log likelihood has an "overall curvature". We therefore chose to approximate the Laplace distribution using

$$\tilde{\rho}_{\tau,\gamma}(x) = \begin{cases} \gamma^{-1}\log(\cosh(\tau\gamma|x|)) & \text{if } x \geq 0 \\ \gamma^{-1}\log(\cosh((1-\tau)\gamma|x|)) & \text{if } x < 0 \end{cases},$$

which has second order derivatives everywhere. Green (1990) used a similar type of approximation in Bayesian image analysis. The parameter $\gamma > 0$ is fixed and the approximation approaches $|x|$ as $\gamma \to \infty$; see Figure 1. In practice one may tune parameter γ to obtain an adequate approximation of the Laplace distribution for likelihood. In our experience, γ between 2 and 5 is sufficient for all applications.

As a result, using INLA inference for quantile regression has two sources of errors: one is from approximating the asymmetric Laplace distribution and another is from approximating posterior marginals. However, those approximation errors are negligible in practice; see Section 5.1.

5 Examples

5.1 Simulation studies

There are three purposes in this simulation section. First, verify the adequacy of using asymmetric Laplace distribution as likelihood by considering several differently distributed errors. Second, compare the MCMC and INLA, two different inference tools for STAQ, regarding both accuracy and computational efficiency. Third, compare the proposed method with an alternative approach: boosting quantile regression models in Fenske et al. (2009), which have been shown to have at least equally fine performance compared to the state-of-art quantile smoothing splines via total variation regularization in Koenker et al. (1994) and Koenker and Mizera (2004).

The datasets are simulated by the following two additive models

$$\begin{aligned} \text{Model 1}: y_i &= 0.4u_i + 0.5\sin(2.7u_i) + 1.1/(1+u_i^2) + \varepsilon_i, \quad u_i \in [-3,3], \\ \text{Model 2}: y_i &= 2 + \sin(2u_i/3) + 0.5(1+(x-3)^2)\varepsilon_i, \quad u_i \in [0,6], \end{aligned} \tag{13}$$

where u_i are equally-spaced drawn from the specified ranges, $\varepsilon_i \overset{iid}{\sim} F$, and F is a distribution function. The resulting quantile functions have nonlinear predictor structures given by

$$\begin{aligned} \text{Model 1}: Q_1(\tau \mid u_i) &= 0.4u_i + 0.5\sin(2.7u_i) + 1.1/(1+u_i^2) + F^{-1}(\tau), \\ \text{Model 2}: Q_2(\tau \mid u_i) &= 2 + \sin(2u_i/3) + 0.5(1+(x-3)^2)F^{-1}(\tau). \end{aligned}$$

The number of observations is fixed to be $n = 400$ for all setups. We consider three different error distributions: a standard normal distribution, a t-distribution with 2 degrees of freedom and a gamma distribution with shape 4 and scale 1. Figure 2 displays data examples from both models with all kinds of error terms. Note that Model 2 has a heteroscedastic data structure where the quantile curves are no longer parallel shifted as in Model 1. Additionally, the heavy-tailed t distribution leads to some extreme outliers in both models. Similar examples have been analyzed in Taddy and Kottas (2009), Kottas and Krnjajic (2009), and Fenske et al. (2009).

For each of the generated datasets, we estimated nonparametric functions for a fixed quantile grid on $\tau \in \{0.25, 0.50, 0.75\}$ by our STAQ models using MCMC and INLA, and by boosting algorithm (function `gamboost()` from package **mboost**). Following Fenske et al. (2009), we used cubic penalized spline base-learners with second order difference penalty, 20 inner knots and three degrees of freedom. In addition, an extra test dataset was created to determine the number of boosting iterations for each of the simulated datasets. For both MCMC and INLA, the CRW2 model is used as the function prior and diffuse Gamma priors ($a = 1, c = .001$) are assumed for all precision parameters. The values of tuning parameter γ for INLA are chosen to be 3 for all setups. The results from MCMC methods are based on 40,000 total iterations with a 15,000 burn-in period and every fifth drawing. The Markov chains converge fairly fast and mix quite well.

To evaluate the estimation performance, we compute mean absolute deviation errors (MADE) as given by

$$\text{MADE} = \frac{1}{n}\sum_{i=1}^{n}\left|Q_k(\tau \mid u) - \hat{Q}_k(\tau \mid u)\right|, \quad k = 1, 2,$$

for 200 replications, then obtain the quartiles of those 200 MADE. Table 1 and 2 show the described performance criteria of all competing methods for Model 1 and 2 respectively. We first concern the comparison between INLA and MCMC approaches. From the tables, two algorithms yield close performances: the MADE results are mostly located in the same range. We also present the estimated quantile curves for Model 2 of normally and t distributed errors in Figure 4, from which we can hardly tell any differences. Furthermore, the Laplace approximation in INLA provides marginal posterior distributions that are close to the samples of a long MCMC run as shown in Figure 4 (bottom). Compared to the boosting algorithm, our method performed equally well in Model 1 and slightly outperformed in Model 2 where the heterogenous error terms were considered. We therefore conclude that: (i) assuming asymmetric Laplace distribution on likelihood is appropriate for Bayesian nonparametric quantile inference; (ii) the performance of our STAQ model is competitive; (iii) INLA provides quite good approximations to the posterior distributions for valid inference.

To demonstrate computational efficiency of our method, we experimented with both MCMC (10,000 iterations) and INLA on the simulated datasets of five different sample sizes $n = 400$, 800, 1500, 3000 and 6000. The MCMC computations were performed on quad-core Intel Pentium 2.84GHz CPUs using the Fortran program (code available upon request), while the INLA computations were performed on a single-processor 2.13-GHz laptop using R-interface to the 'inla' program (Rue and Martino, 2009). The computational time is recorded in Table 3. Note that we only post the results for estimating medians of datasets simulated from Model 1 with normal error terms. Similar performances were found in other situations. As we can see, the MCMC computation is really fast and increases with sample size at order n. The INLA approach is even much faster: it took less than 20 seconds for $n = 6,000$. (The INLA results were scaled for CPU difference.)

5.2 Munich rental guide

In the following, we will analyze the 2003 Munich rental data (see Rue and Held, 2005, Sec. 4.2.1). The response variable Y is the rent (per square meter in Euros) for a flat and the covariates are the spatial location ($\boldsymbol{u}_i$), floor space (size$_i$), year of construction (year$_i$), and various indicator variables such as an indicator for a flat with no central heating, no warm water, more refined kitchen equipment and so on (see Table 4 for details). The dataset consists of $n = 2035$ observations. There are 380 districts in Munich. The floor size varies from 17 to 185 square meters and the year of construction goes from 1918 to 2001, both of which are well known to have nonlinear impact on the rent from previous analyses (Fahrmeir et al., 2004; Rue and Held, 2005). The fact that the response variable is of right skewness and heterogeneity indicates that quantile regression fitting is quite suitable for the analysis of this dataset.

We build the following *geoadditive* model for τth quantile regression

$$\eta_{\tau i} = \boldsymbol{x}_i^T \boldsymbol{\beta}_\tau + f_{\tau 1}(\text{size}_i) + f_{\tau 2}(\text{year}_i) + f_{\tau 3}(\boldsymbol{u}_i), \tag{14}$$

where $\boldsymbol{x}_i$ represents a 13-dimensional vector of categorical covariates with effects summarized in the vector $\boldsymbol{\beta}_\tau$ that is assigned a diffuse normal prior. The functions $f_{\tau 1}$ and $f_{\tau 2}$ are nonlinear effects of floor size and year of construction. Both are modeled as the CRW2 that we described in Section 3.1.1. The function $f_{\tau 3}$ represents the spatial effect modeled as the intrinsic GMRF in equation (10) based on the discrete spatial information on the districts of Munich. The estimation of three quartiles ($\tau = .25, .50, .75$) was implemented by INLA (see a code example in appendix), with tuning parameter $\gamma = 2$ and hyperparameters $a = 1, b = .001$. As we shall see, the quantile inference can offer some useful insights that could not be detected via mean regression.

The effects of year of construction and floor size in the model (14) for rents are shown in Figure 5 (top and middle). Generally speaking, the rents decreased for buildings constructed during the 1920s to 1930s, and then started to increase for newly constructed buildings until 1960s; after being relatively stable for about 10 years, the rents have kept increasing since 1970s. A closer look shows that the quartiles of rents have slightly different changing patterns over years. Compared to the expensive flats ($\tau = .75$), the rents of cheap flats ($\tau = .25$) seem to decrease more quickly in the first 10 years but also increase more quickly in the following years. It is the evidence of increasing variance from 1920 to 1930, followed by decreasing variance. Additionally, during the stable period (1960-1970) the effect of year slightly increased for lower quartile, was flat for median, and decreased a little bit for upper quartile. It can be explained by a boom in construction building in those years, with flats having comparably homogenous quality. For all quartiles, the effect of floor size is shown as expected with high rents per square meter for smaller flats and only very little variation for flats of size larger than about 50 square meters (except for the size larger than 130 whose odd behavior can be explained by the lack of data information).

Figure 5 (bottom) also presents the spatial effect of location in Munich. As what we know from expert assessments, high rents for flats are in the center of Munich and some popular districts along the river Isar and near to parks. In contrast, significantly negative effects are found for some districts at the border of Munich. Furthermore, the upper quartile appears to have lighter colors in the central regions than the lower quartile while all quartiles have similar colors in the other districts. It provides interesting evidence of increased variance in the central districts of Munich, whereas some of the suburban districts are more homogeneous.

The estimated linear effects as well as their posterior standard deviations are listed in Table 5. In general, all the estimated coefficients are significant and explain the response variable appropriately. It is also interesting to see that a particular variable may contribute to different quantiles in different ways. For example, as τ increases variables Keine.Zh and Kein.Badkach have significantly increasing effects while Keine.Wwv has decreasing effect. It means the absence of warm water supply would yield a larger rent reduction in upper quartile than in lower quartile, while the absences of central heating system and flagging in bathroom have reverse effects on quartiles. It can be explained by the fact that the warm water supply is a basic feature with which only a few flats are not equipped with and those flats tend to be cheap; however, the central heating and flagging are less common and the flats without them could be expensive.

6 Final remarks

In this paper, we proposed a unified approach for Bayesian quantile inference via MCMC and INLA methods in the additive and semiparametric mixed models. Different types of covariates are all treated within the same general framework by assigning appropriate Gaussian type priors with different forms and degrees of smoothness. Extensive simulation studies demonstrated our approach in terms of both accuracy and computational efficiency. The Munich rental example showed how effective of our model is to account for nonlinear and spatial effects from a large dataset. Although we did not present any example here, the proposed model is ready to make quantile inference in the varying coefficient models and semiparametric mixed models for longitudinal data

It is possible to extend our model to estimating quantile functions of count or binary data. It is well known that the transformed quantile regression equals to the quantile regression of the transformed variable, i.e., $Q(t(y)) = t(Q(y))$ where $t(\cdot)$ denotes a transformation and $Q(\cdot)$ denotes the quantile. Hence we may find a transformation to create an appropriate continuous response variable, which has the same quantiles as the original discrete data, and then apply the model developed here to the transformed data.

Acknowledgements

Yu Yue's research has been supported by PSC-CUNY research award #60147-39 40. The authors thank Professor Thomas Kneib for his assistance in simulation studies and Munich rental data example.

Appendix

In order to make the results of our data analysis reproducible, we provide the following R code for estimating the median function in Munich rental example. Other quantiles can be obtained via changing values of alpha in function inla().

```
> library(INLA)
> data(Munich)
> g = system.file("demodata/munich.graph", package="INLA")
> formula = rent ~ f(location,model="besag",graph.file=g,param=c(1,.001)) +
                   f(year,model="crw2",values=seq(1918,2001),param=c(1,.001)) +
                   f(floor.size,model="crw2",param=c(1,.001)) +
                   Gute.Wohnlage + Beste.Wohnlage + Keine.Wwv + Keine.Zh +
                   Kein.Badkach + Besond.Bad + Gehobene.Kueche +
                   zim1 + zim2 + zim3 + zim4 + zim5 + zim6 -1
> mod = inla(formula,data = Munich, verbose = T, family = "laplace",
             control.data=list(alpha=.5, gamma=2, epsilon=.01),
             control.predictor = list(initial = 12), control.inla = list(h=1e-4))
```

References

Besag, J. and Kooperberg, C. (1995). On conditional and intrinsic autoregressions. *Biometrika* **82**, 733–746.

Cai, Z. and Xu, X. (2008). Nonparametric quantile estimations for dynamic smooth coefficient models. *Journal of the American Statistical Association* **103**, 1595–1608.

Chhikara, R. S. and Folks, L. (1989). *The Inverse Gaussian Distribution: Theory, Methodology, and Applications*. New York: Marcel Dekker.

De Gooijer, J. and Zerom, D. (2003). On additive conditional quantiles with high-dimensional covariates. *Journal of the American Statistical Association* **98**, 135–146.

Dunson, D. and Taylor, J. (2005). Approximate bayesian inference for quantiles. *Journal of Nonparametric Statistics* **17**, 385–400.

Eubank, R. L., Huang, C., Maldonado, Y. M. n., Wang, N., Wang, S. and Buchanan, R. J. (2004). Smoothing spline estimation in varying-coefficient models. *Journal of the Royal Statistical Society, Series B: Statistical Methodology* **66**, 653–667.

Fahrmeir, L., Kneib, T. and Lang, S. (2004). Penalized structured additive regression for space-time data: a Bayesian perspective. *Statistica Sinica* **14**, 731–761.

Fahrmeir, L. and Tutz, G. (2001). *Multivariate Statistical Modeling based on Generalized Linear Models*. Berlin: Springer.

Fenske, N., Kneib, T. and Hothorn, T. (2009). Identifying risk factors for severe childhood malnutrition by boosting additive quantile regression. Technical Report 52, Department of Statistics, University of Munich.

Geraci, M. and Bottai, M. (2007). Quantile regression for longitudinal data using the asymmetric laplace distribution. *Biostatistics* **8**, 140–154.

Green, P. J. (1990). Bayesian reconstructions from emission tomography data using a modified em algorithm. *IEEE Transactions on Medical Imaging* **9**, 84–93.

Horowitz, J. L. and Lee, S. (2005). Nonparametric estimation of an additive quantile regression model. *Journal of the American Statistical Association* **100**, 1238–1249.

Kneib, T., Konrath, S. and Fahrmeir, L. (2009). High-dimensional structured additive regression models: Bayesian regularisation, smoothing and predictive performance. Technical report, Department of Statistics, University of Munich.

Koenker, R. (2005). *Quantile Regression (Econometric Society Monographs)*. Cambridge University Press.

Koenker, R. and Bassett, J., Gilbert (1978). Regression quantiles. *Econometrica* **46**, 33–50.

Koenker, R., Machado, J. A. F. and Machado, J. A. F. (1999). Goodness of fit and related inference processes for quantile regression. *Journal of the American Statistical Association* **94**, 1296–1310.

Koenker, R. and Mizera, I. (2004). Penalized triograms: Total variation regularization for bivariate smoothing. *Journal of the Royal Statistical Society, Series B: Statistical Methodology* **66**, 145–163.

Koenker, R., Ng, P. and Portnoy, S. (1994). Quantile smoothing splines. *Biometrika* **81**, 673–680.

Kottas, A. and Krnjajic, M. (2009). Bayesian semiparametric modeling in quantile regression. *Scandinavian Journal of Statistics* (to appear).

Lang, S. and Brezger, A. (2004). Bayesian P-splines. *Journal of Computational and Graphical Statistics* **13**, 183–212.

Li, Y., Liu, Y. and Zhu, J. (2007). Quantile regression in reproducing kernel hilbert spaces. *Journal of the American Statistical Association* **102**, 255–268.

Reich, B., Bondell, H. and Wang, H. (2008). Flexible Bayesian quantile regression for independent and clustered data. Technical report (submitted), North Carolina State University.

Rue, H. (2001). Fast sampling of Gaussian Markov random fields. *Journal of the Royal Statistical Society, Series B* **63**, 325–338.

Rue, H. and Held, L. (2005). *Gaussian Markov Random Fields: Theory and Applications*, volume 104 of *Monographs on Statistics and Applied Probability*. Chapman & Hall, London.

Rue, H. and Martino, S. (2007). Approximate Bayesian inference for hierarchical Gaussian Markov random fields. *Journal of Statistical Planning and Inference* **137**, 3177–3192.

Rue, H. and Martino, S. (2009). Implementing approximate Bayesian inference for latent Gaussian models by using integrated nested laplace approximations: a manual for the inla-program. Technical report, Department of Mathematical Sciences, Norwegian University of Science and Technology, Trondheim.

Rue, H., Martino, S. and Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models using integrated nested Laplace approximations (with discussion). *Journal of the Royal Statistical Society, Series B* **71**, 319–392.

Taddy, M. and Kottas, A. (2009). A Bayesian nonparametric approach to inference for quantile regression. *Journal of Business and Economic Statistics* (to appear).

Takeuchi, I., Le, Q., Sears, T. and Smola, A. (2006). Nonparametric quantile estimatio. *Journal of Machine Learning Research* **7**, 1231–1264.

Tsionas, E. G. (2003). Bayesian quantile inference. *Journal of Statistical Computation and Simulation* **73**, 659–674.

Wahba, G. (1990). *Spline Models for Observational Data*. SIAM [Society for Industrial and Applied Mathematics].

Wecker, W. E. and Ansley, C. F. (1983). The signal extraction approach to nonlinear regression and spline smoothing. *Journal of the American Statistical Association* **78**, 81–89.

Yu, K. and Lu, Z. (2004). Local linear additive quantile regression. *Scandinavian Journal of Statistics* **31**, 333–346.

Yu, K. and Moyeed, R. A. (2001). Bayesian quantile regression. *Statistics & Probability Letters* **54**, 437–447.

Yu, K. and Zhang, J. (2005). A three-parameter asymmetric laplace distribution and its extension. *Communications in Statistics: Theory and Methods* **34**, 1867–1879.

Yuan, Y. and Yin, G. (2009). Bayesian quantile regression for longitudinal studies with nonignorable missing data. *Biometrics* (to appear).

Table 1: Median with lower and upper quartiles of estimated MADE from 200 replications of Model 1 simulation setups with three types of errors.

Standard normal errors			
τ	MCMC	INLA	boost
0.25	.164 (.139, .192)	.156 (.130, .187)	.160 (.133, .184)
0.50	.145 (.125, .164)	.139 (.113, .162)	.143 (.120, .170)
0.75	.165 (.140, .190)	.153 (.128, .182)	.159 (.131, .181)
Student t distributed errors			
τ	MCMC	INLA	boost
0.25	.215 (.178, .256)	.217 (.174, .250)	.216 (.188, .252)
0.50	.166 (.137, .195)	.162 (.136, .185)	.166 (.139, .204)
0.75	.214 (.178, .253)	.218 (.173, .252)	.220 (.185, .255)
Gamma distributed errors			
τ	MCMC	INLA	boost
0.25	.229 (.195, .270)	.225 (.198, .284)	.233 (.193, .275)
0.50	.280 (.225, .321)	.271 (.222, .311)	.268 (.226, .332)
0.75	.357 (.296, .426)	.355 (.307, .420)	.345 (.293, .417)

Table 2: Median with lower and upper quartiles of estimated MADE from 200 replications of Model 2 simulation setups with three types of errors.

Standard normal errors			
τ	MCMC	INLA	boost
0.25	.261 (.196, .335)	.249 (.181, .312)	.317 (.250, .368)
0.50	.206 (.144, .287)	.212 (.138, .298)	.294 (.230, .351)
0.75	.233 (.159, .304)	.235 (.183, .308)	.316 (.265, .376)
Student t distributed errors			
τ	MCMC	INLA	boost
0.25	.322 (.249, .401)	.321 (.248, .415)	.413 (.339, .521)
0.50	.217 (.160, .300)	.221 (.159, .311)	.328 (.278, .403)
0.75	.322 (.209, .405)	.311 (.236, .416)	.412 (.340, .502)
Gamma distributed errors			
τ	MCMC	INLA	boost
0.25	.387 (.298, .486)	.388 (.283, .547)	.451 (.365, .521)
0.50	.493 (.382, .610)	.422 (.323, .514)	.528 (.412, .622)
0.75	.706 (.559, .839)	.707 (.520, .842)	.682 (.517, .826)

Table 3: Computational time (seconds) versus sample size

Sample size	400	800	1500	3000	6000
MCMC	5.96	12.32	23.13	45.34	89.21
INLA	0.94	1.83	3.57	7.66	16.80

Table 4: Variables in the Munich rental dataset.

Variable	Explanation
rent	net rent per square meter
floor.size	size of the flat in square meters
year	year of construction of the building in which the flat is located
location	location index (in terms of subquarters)
Gute.Wohnlage	dummy variable for good locations/good neighborhoods
Beste.Wohnlage	dummy variable for very good locations/very good neighborhoods
Keine.Wwv	dummy variable for absence of warm water supply
Keine.Zh	dummy variable for absence of central heating system
Kein.Badkach	dummy variable for absence of flagging in the bathroom
Besond.Bad	dummy variable for special features of the bathroom
Gehobene.Kueche	dummy variable for more refined kitchen equipment
zim1	dummy variable for flats with 1 room
zim2	dummy variable for flats with 2 room
zim3	dummy variable for flats with 3 room
zim4	dummy variable for flats with 4 room
zim5	dummy variable for flats with 5 room
zim6	dummy variable for flats with 6 room

Table 5: Munich rental guide: estimated linear effects and their posterior standard deviations.

Variable	$\tau = .25$	$\tau = .50$	$\tau = .75$
Gute.Wohnlage	0.578 (0.127)	0.712 (0.110)	0.674 (0.108)
Beste.Wohnlage	1.536 (0.306)	1.582 (0.317)	1.892 (0.323)
Keine.Wwv	-1.526 (0.293)	-1.972 (0.277)	-2.360 (0.298)
Keine.Zh	-1.622 (0.211)	-1.486 (0.202)	-1.123 (0.214)
Kein.Badkach	-0.374 (0.281)	-0.592 (0.108)	-0.597 (0.113)
Besond.Bad	0.479 (0.166)	0.358 (0.153)	0.442 (0.151)
Gehobene.Kueche	1.026 (0.210)	1.152 (0.166)	1.146 (0.172)
zim1	7.062 (0.248)	8.129 (0.242)	9.104 (0.246)
zim2	7.015 (0.211)	8.261 (0.181)	9.633 (0.181)
zim3	6.861 (0.197)	8.118 (0.164)	9.485 (0.169)
zim4	6.497 (0.174)	7.697 (0.182)	9.006 (0.186)
zim5	6.284 (0.349)	7.622 (0.337)	9.473 (0.349)
zim6	6.381 (0.574)	7.513 (0.527)	8.930 (0.566)

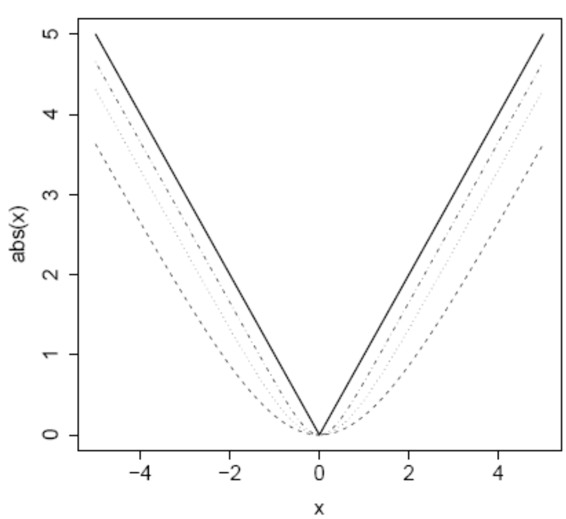

Figure 1: The function $|x|$ and the approximation $\log(\cosh(\gamma x))/\gamma$ for $\gamma = 0.5$, 1, and 2. The approximation improves for increasing γ.

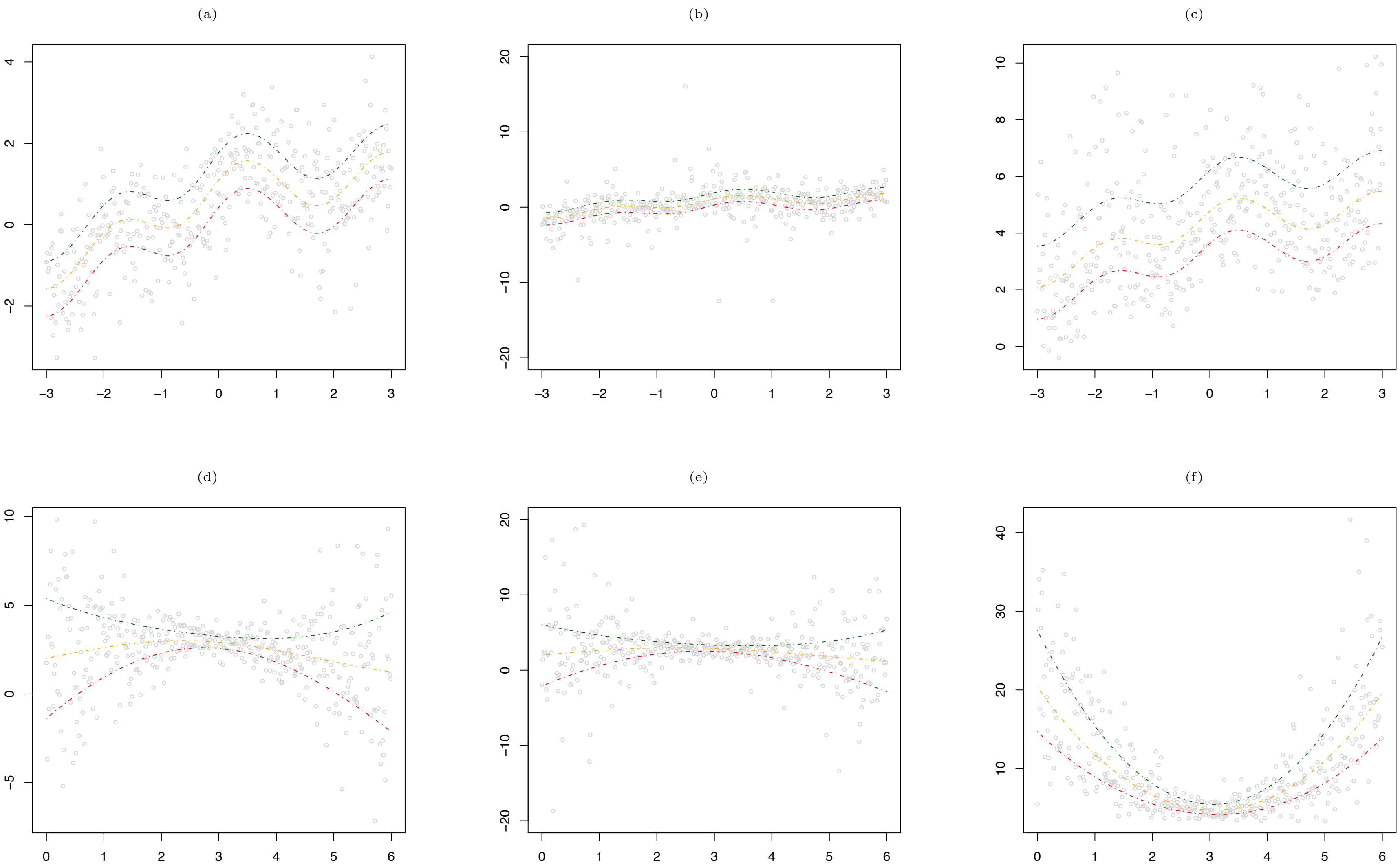

Figure 2: Data examples for nonparametric function setups in Model 1 (top) or Model 2 (bottom) with standard normal (left) or Student t (middle) or gamma (right) distributed error terms. Lines represent true underlying quantile curves for $\tau \in \{0.25, 0.50, 0.75\}$.

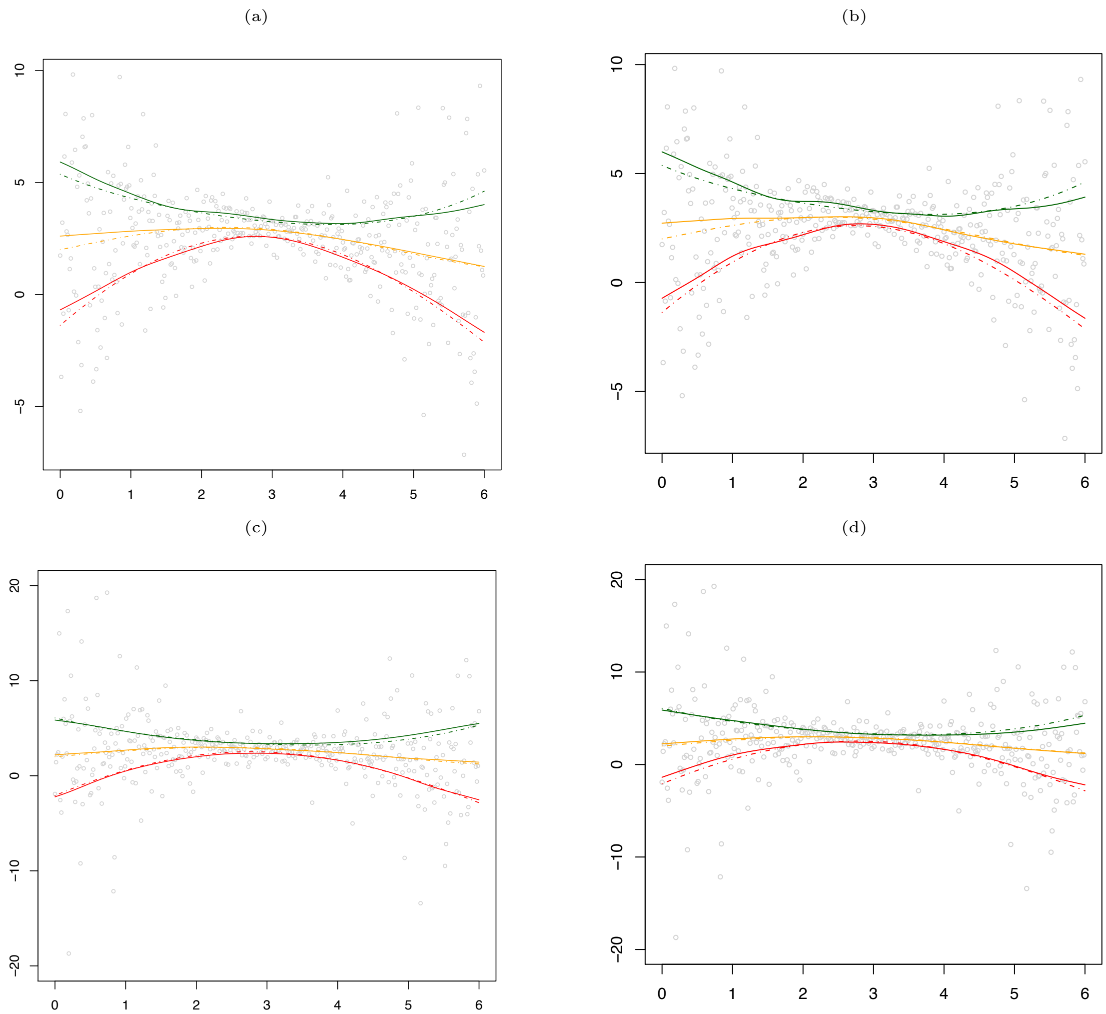

Figure 3: Example for estimated quantile curves for Model 2 with normal (top) and t (bottom) distributed error terms using INLA (left) and MCMC (right).

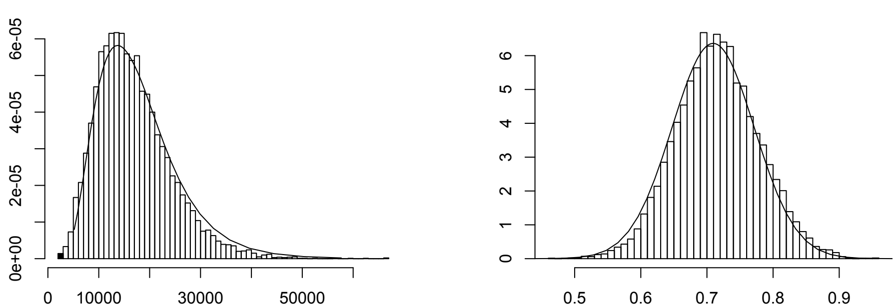

Figure 4: Comparison between marginal posterior distributions from INLA (line) and MCMC samples (histogram).

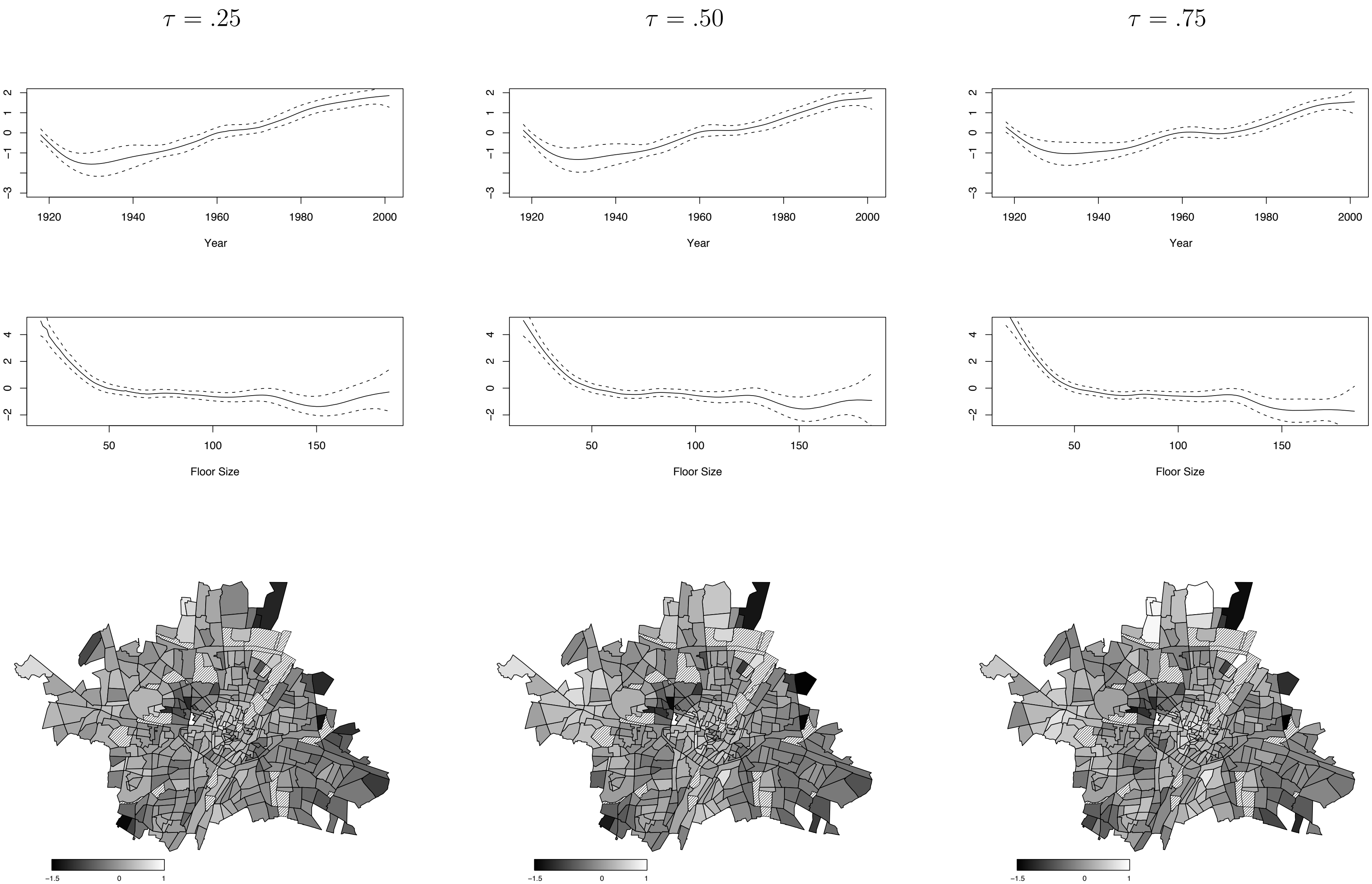

Figure 5: Munich rental guide: Estimated nonparametric effects with 95% pointwise credible bands (top and middle rows) and estimated spatial effects (bottom row) for the quartiles; the shaded areas are districts with no houses, such as parks or fields.