University of Innsbruck

Bayesian semiparametric additive quantile regression

Elisabeth Waldmann, Thomas Kneib, Yu Ryan Yu, Stefan Lang

Working Papers in Economics and Statistics

2012-06

Research Platform eeecon
Empirical and Experimental Economics

University of Innsbruck
http://eeecon.uibk.ac.at/

University of Innsbruck
Working Papers in Economics and Statistics

The series is jointly edited and published by

Department of Economics

Department of Public Finance

Department of Statistics

Contact Address:
University of Innsbruck
Department of Public Finance
Universitaetsstrasse 15
A-6020 Innsbruck
Austria
Tel: + 43 512 507 7171
Fax: + 43 512 507 2788
e-mail: eeecon@uibk.ac.at

The most recent version of all working papers can be downloaded at
http://eeecon.uibk.ac.at/wopec/

For a list of recent papers see the backpages of this paper.

Bayesian Semiparametric Additive Quantile Regression

Elisabeth Waldmann, Thomas Kneib
Department of Economics
Georg August University Göttingen

Yu Ryan Yue
Zicklin School of Business, Baruch College,
City University of New York

Stefan Lang
Department of Statistics
University of Innsbruck

Abstract

Quantile regression provides a convenient framework for analyzing the impact of covariates on the complete conditional distribution of a response variable instead of only the mean. While frequentist treatments of quantile regression are typically completely nonparametric, a Bayesian formulation relies on assuming the asymmetric Laplace distribution as auxiliary error distribution that yields posterior modes equivalent to frequentist estimates. In this paper, we utilize a location-scale-mixture of normals representation of the asymmetric Laplace distribution to transfer different flexible modeling concepts from Gaussian mean regression to Bayesian semiparametric quantile regression. In particular, we will consider high-dimensional geoadditive models comprising LASSO regularization priors and mixed models with potentially non-normal random effects distribution modeled via a Dirichlet process mixture. These extensions are illustrated using two large-scale applications on net rents in Munich and longitudinal measurements on obesity among children.

1 Introduction

Quantile regression allows to determine the influence of covariates on the conditional quantiles of the distribution of a dependent variable. Therefore, one of the main advantages over mean regression is that quantile regression permits to supply detailled information about the complete conditional distribution instead of only the

mean. In addition, outliers and extreme data are usually less influential in quantile regression due to the inherent robustness of quantiles.
For classical linear quantile regression as introduced by Koenker and Bassett (1978), estimation of the quantile-specific regression coefficients $\boldsymbol{\beta}_\tau$ relies on minimizing the sum of asymmetrically weighted absolute deviations (AWAD)

$$\min_{\boldsymbol{\beta}_\tau} \sum_{i=1}^{n} \rho_\tau(y_i - \boldsymbol{x}_i' \boldsymbol{\beta}_\tau)$$

where $(y_i, \boldsymbol{x}_i)$, $i = 1, \ldots, n$ are the observed response and covariate values for n observations, the check function

$$\rho_\tau(y_i - \boldsymbol{x}_i'\boldsymbol{\beta}_\tau) = \begin{cases} \tau|y_i - \boldsymbol{x}_i'\boldsymbol{\beta}_\tau| & \text{if } y_i \geq \boldsymbol{x}_i'\boldsymbol{\beta}_\tau \\ (1-\tau)|y_i - \boldsymbol{x}_i'\boldsymbol{\beta}_\tau| & \text{if } y_i < \boldsymbol{x}_i'\boldsymbol{\beta}_\tau \end{cases}$$

defines asymmetrically weighted absolute residuals and $\tau \in (0,1)$ is the quantile of interest. This approach is completely nonparametric and does not require the assumption of a specific response distribution. No closed form solution for the minimization problem exists and quantile regression estimates are typically obtained based on linear programming (see Koenker, 2005, for details).
Recent interest in quantile regression has focused on broadening the scope of supported model specifications. For example, additive quantile regression models have gained considerable attention. Yue and Rue (2011) and Oh et al. (2011) propose differentiable approximations to the AWAD criterion that allow to employ different types of smoothing approaches while Li et al. (2010) show ways to incorporate regularisation on fixed effects into linear Bayesian quantile regression. Fenske et al. (2011) propose boosting approaches for flexible, additive quantile regression models, where penalized least squares estimates are utilized as base-learners. Koenker et al. (1994) added an L_1-norm penalty to the AWAD criterion that allows to still use linear programming techniques in the context of quantile smoothing splines. Koenker and Mizera (2004) extend this approach to surface estimation based on triograms.
In this paper, we will introduce yet more flexible types of quantile regression models motivated by two large-scale applications on rents for flats in the city of Munich and on longitudinal childhood obesity measurements. The German tenancy law gives restrictions to the increase of rents and forces the landlords to keep the prize in a range defined by flats which are comparable in size, location and quality. To

make it easier for tenants and owners to assess if the rent is appropriate for a flat, so-called rental guides are derived based on large samples of flats. In the following, we will use data from the 2007 Munich rental guide with about 3,000 observations and 250 covariates. Kneib et al. (2011) suggested a high-dimensional geoadditive model

$$y_i = \boldsymbol{x}_i'\boldsymbol{\beta} + f_1(\text{size}_i) + f_2(\text{year}_i) + f_{\text{spat}}(s_i) + \varepsilon_i \tag{1}$$

for analyzing the *expectation* of the net rent per square meter y_i in terms of nonlinear effects f_1 and f_2 of the size of the flat in square meters and the year of construction, a spatial effect f_{spat} based on district information s_i and a high-dimensional vector of mostly categorical covariates $\boldsymbol{x}_i$ (such as for example presence of a fridge, attic, garden or balcony) with linear effects $\boldsymbol{\beta}$. While penalized splines and a Gaussian Markov random field have been employed for the nonlinear and spatial effects, respectively, LASSO and ridge penalization have been applied to the vector $\boldsymbol{\beta}$ to achieve regularization. It turned out that, for mean regression, a geoadditive model with LASSO regularization outperforms a model of moderate dimension resulting from expert knowledge and has slight advantages over a comparable model with ridge regularization. We therefore aim at extending the high-dimensional geoadditive model to quantile regression, to enable a more detailled view on the conditional distribution of the net rents. In particular, quantile regression allows to determine flexible bounds for the net rent based on, for example, the 5% and the 95% quantiles without imposing strong assumptions on the error distribution.

The second application deals with longitudinal measurements on obesity among children in the LISA (Influences of *L*ife-style factors on the development of the *I*mmune *S*ystem and *A*llergies in East and West Germany) study. The objective of the study is to analyze the variations in individual body mass index (BMI) patterns while simultaneously determining the impact of factors driving childhood obesity such as the breast-feeding behaviour or maternal BMI. Since there is considerable variation in the highly non-linear individual profiles, Heinzl et al. (2012) suggest a flexible additive mixed model

$$y_{it} = f(t) + \boldsymbol{x}_{it}'\boldsymbol{\beta} + \boldsymbol{z}_{it}'\boldsymbol{b}_i + \varepsilon_{it} \tag{2}$$

where $t = 1, \ldots, T_i$, denotes the time, $i = 1, \ldots, n$, the individual, $f(t)$ represents the overall trend in the BMI measurements, $\boldsymbol{x}_{it}'\boldsymbol{\beta}$ contains parametric, fixed effects common to all children and the random effects term $\boldsymbol{z}_{it}'\boldsymbol{b}_i$ contains individual-specific deviations from the overall trend. Since there appear to be different groups of children with specific patterns in their individual-specific deviations and to account for potential non-normality of the random effects distribution, Heinzl et al. (2012) utilized a Dirichlet process mixture (DPM) as random effects distribution which allows for very flexible random effects distributions and model-based clustering of the random effects.

In summary, our aim is to make flexible components in semiparametric regression models such as nonlinear effects, spatial effects, LASSO regularized coefficient blocks or non-normal random effects applicable in the context of quantile regression. These are typically difficult to combine with linear programming or other direct maximization approaches. Instead, we rely on a Bayesian formulation of quantile regression based on the asymmetric Laplace distribution as an auxiliary error distribution as suggested in Yu and Moyeed (2001). Therefore we consider the alternative representation of the quantile regression problem as

$$y_i = \eta_{i,\tau} + \varepsilon_{i,\tau}$$

where $\eta_{i,\tau}$ is the predictor of the τth quantile in the regression model and $\varepsilon_{i,\tau}$ is an appropriate error term. Instead of assuming zero mean for the errors as in mean regression, one then imposes the restriction that the τ-quantile of the error distribution is zero. In the Bayesian framework, we have to assume a specific distribution for the errors (or equivalently the responses) to be able to set up a likelihood. The asymmetric Laplace distribution $y_i \sim \mathrm{ALD}(\eta_{i,\tau}, \delta^2, \tau)$ with location parameter $\eta_{i,\tau}$, precision parameter δ^2, asymmetry τ and density

$$p(y_i|\eta_{i,\tau}, \delta^2, \tau) = \tau(1-\tau)\delta^2 \exp(-\delta^2 \rho_\tau(y_i - \eta_{i,\tau})) \tag{3}$$

is particularly useful since it yields posterior mode estimates that are equivalent to the minimizers of the AWAD criterion.

To make Bayesian inference feasible, Yue and Rue (2011) introduced a location-scale mixture representation of the asymmetric Laplace distribution that allows to rewrite Bayesian quantile regression as a conditionally Gaussian regression with offset and weights. As a consequence, Bayesian inferential schemes developed for Gaussian regression models can then (at least conceptually) be easily transferred to quantile regression. However, Yue and Rue (2011) observed severe mixing and convergence in their approach to sampling-based Bayesian quantile regression and therefore had to resort to an approximate solution based on integrated nested Laplace approximations. We propose a different updating scheme (based on basis coefficients of nonparametric effects instead of function evaluations) that overcomes the difficulties encountered by Yue and Rue (2011). Moreover, we imbed Bayesian semiparametric quantile regression in a generic framework that enables the flexible inclusion of hyperprior structures on the variance (and mean) parameters of the conditional Gaussian priors of the regression effects. Such hyperprior structured can be used to re-cast extensions of semiparametric regression such as LASSO regularization or Dirichlet process mixtures in the context of conditionally Gaussian Bayesian quantile regression.

The rest of this paper is organized as follows: In Section 2, we first introduce the location-scale mixture representation of the asymmetric Laplace distribution and present a generic Markov chain Monte Carlo simulation algorithm for Bayesian quantile regression with conditionally Gaussian priors. Afterwards we introduce different special cases and the corresponding hyperprior specifications. Section 3 presents the applications based on a high-dimensional geoadditive regression model in case of the Munich rental guide and a nonparametric random effects model for the longitudinal obesity measurements.

2 Bayesian Semiparametric Quantile Regression

2.1 Generic Bayesian Quantile Regression with Auxiliary Error Distribution

While the asymmetric Laplace distribution (3) provides a convenient way to express quantile regression in a Bayesian framework based on an auxiliary error distribution, it complicates inference based on Markov chain Monte Carlo simulations due to the inherent non-differentiability of the check function ρ_τ. We therefore follow Yue and Rue (2011) and utilize a scale mixture of Gaussians representation of the asymmetric Laplace distribution. Let $Z \sim \mathrm{N}(0,1)$ and $W \sim \mathrm{Exp}(\delta^2)$ be two independent random variables following a standard normal and exponential distribution with rate parameter δ^2, respectively. Then the distribution of

$$Y = \eta + \xi W + \sigma Z \sqrt{\frac{W}{\delta^2}}$$

with $\xi = \frac{1-2\tau}{\tau(1-\tau)}$ and $\sigma^2 = \frac{2}{\tau(1-\tau)}$ follows the ALD(η, δ^2, τ) distribution. As a consequence, the Bayesian quantile regression problem can be reformulated as a conditionally Gaussian regression with offsets ξW and weights $\sigma\sqrt{\frac{W}{\delta^2}}$ after imputing W as a part of the MCMC sampler.

To be more specific, we assume that n independent realisations $y_i \sim \text{ALD}(\eta_i, \delta^2, \tau)$ are given with generic semiparametric predictor

$$\eta_i = \sum_{j=1}^{J} f_j(\boldsymbol{v}_i).$$

The predictor comprises various functions f_j that are defined on the complete vector of covariates $\boldsymbol{v}_i$. For example, specific components may be given by *(i) linear functions* $f_j(\boldsymbol{v}_i) = \boldsymbol{x}_i'\boldsymbol{\beta}$ where $\boldsymbol{x}_i$ is a subvector of $\boldsymbol{v}_i$, *(ii) univariate nonlinear functions* $f_j(\boldsymbol{v}_i) = f(x_i)$ where x_i is a single continuous element of $\boldsymbol{v}_i$, *(iii) spatial effects* $f_j(\boldsymbol{v}_i) = f_{\text{spat}}(s_i)$ where s_i is a spatial location variable, or *(iv) random effects* $f_j(\boldsymbol{v}_i) = x_i\boldsymbol{b}_{c_i}$, where x_i is some covariate (potentially including a constant for random intercepts) and c_i is a cluster variable that groups the observations, see Fahrmeir et al. (2004) or Kneib et al. (2009) for similar generic model specifications. In matrix notation, we can always write the generic model as

$$\eta = \boldsymbol{Z}_1\boldsymbol{\gamma}_1 + \ldots + \boldsymbol{Z}_J\boldsymbol{\gamma}_J$$

where the design matrices $\boldsymbol{Z}_j$ are obtained by suitable basis expansions and $\boldsymbol{\gamma}_j$ contain the corresponding basis coefficients.

Our assumptions imply the observation model

$$\boldsymbol{y} \,|\, \boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_J, \boldsymbol{w}, \delta^2 \sim N(\boldsymbol{\eta} + \xi\boldsymbol{w}, \sigma^2/\delta^2\boldsymbol{D})$$

where $\boldsymbol{y} = (y_1, \ldots, y_n)'$ and $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_n)'$ are the vectors of response observations and predictors, respectively, $\boldsymbol{w} = (w_1, \ldots, w_n)'$ is the vector of i.i.d Exp(δ^2) distributed weights implied by the scale mixture, and $\boldsymbol{D} = \text{diag}(w_1, \ldots, w_n)$ is a corresponding diagonal matrix of the weights.

In order to enforce specific properties of the basis coefficients such as (for example spatial) smoothness, we assume conditionally Gaussian, possibly partially improper priors

$$\boldsymbol{\gamma}_j \,|\, \boldsymbol{m}_j, \boldsymbol{\theta}_j, \delta^2 \propto \exp\left(-\frac{1}{2}(\boldsymbol{\gamma}_j - \boldsymbol{m}_j)'\boldsymbol{K}_j(\boldsymbol{\theta}_j)(\boldsymbol{\gamma}_j - \boldsymbol{m}_j)\right) \tag{4}$$

where the prior precision $\boldsymbol{K}_j(\boldsymbol{\theta}_j)$ may depend on a vector of further hyperparameters $\boldsymbol{\theta}_j$ for which additional hyperpriors have to be defined depending on the specific type of effect. A term $\boldsymbol{f} = \boldsymbol{Z}\boldsymbol{\gamma}$ (e.g. P-spline, LASSO component etc.) is then specified by defining

- a design matrix $\boldsymbol{Z}$,
- a precision or penalty matrix $\boldsymbol{K}(\boldsymbol{\theta})$,
- a prior $p(\boldsymbol{m})$ for $\boldsymbol{m}$,
- a prior $p(\boldsymbol{\theta})$ for the hyperparameter(s) $\boldsymbol{\theta}$.

We will give specific examples below in Sections 2.2 – 2.4.
Since the location-scale mixture representation and the conditionally Gaussian prior structure yield a conjugate model hierarchy, the full conditionals for the regression coefficients are again Gaussian $\boldsymbol{\gamma}_j|\cdot \sim \mathrm{N}(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)$ with expectation and covariance matrix

$$\begin{aligned}\boldsymbol{\mu}_j &= \boldsymbol{\Sigma}_j^{-1}\left(\tfrac{\delta^2}{\sigma^2}\boldsymbol{Z}_j'\boldsymbol{D}^{-1}(\boldsymbol{y} - \xi\boldsymbol{w} - \boldsymbol{\eta}_{-j}) + \boldsymbol{K}_j(\boldsymbol{\theta}_j)\boldsymbol{m}_j\right)\\ \boldsymbol{\Sigma}_j &= \left(\boldsymbol{K}_j(\boldsymbol{\theta}_j) + \tfrac{\delta^2}{\sigma^2}\boldsymbol{Z}_j'\boldsymbol{D}^{-1}\boldsymbol{Z}_j\right)^{-1}\end{aligned} \tag{5}$$

where $\boldsymbol{\eta}_{-j} = \boldsymbol{\eta} - \boldsymbol{Z}_j\boldsymbol{\gamma}_j$ is the partial predictor without the j-th effect. When comparing the full conditionals with those arising from mean regression, where

$$\boldsymbol{\mu}_j = \delta^2\boldsymbol{\Sigma}_j^{-1}(\boldsymbol{Z}_j'(\boldsymbol{y} - \boldsymbol{\eta}_{-j}) + \boldsymbol{K}_j(\boldsymbol{\theta}_j)\boldsymbol{m}_j) \qquad \text{and} \qquad \boldsymbol{\Sigma}_j = (\boldsymbol{K}_j(\boldsymbol{\theta}_j) + \delta^2\boldsymbol{Z}_j'\boldsymbol{Z}_j)^{-1}$$

we find only minor differences corresponding basically to the imputed weights and the offset.
Resulting from the scale mixture representation of the asymmetric Laplace distribution, the weights w_i are a priori i.i.d exponentially distributed given the prior precision δ^2, i.e. $w_i \,|\, \delta^2 \sim \mathrm{Exp}(\delta^2)$. This implies that the full conditionals for imputing the inverse of the weights are inverse Gaussian:

$$w_i^{-1}|\cdot \sim \mathrm{InvGauss}\left(\sqrt{\frac{\xi^2 + 2\sigma^2}{(y_i - \eta_i)^2}}, \frac{\delta^2(\xi^2 + 2\sigma^2)}{\sigma^2}\right). \tag{6}$$

If the prior for the precision parameter δ^2 is chosen to be the conjugate gamma distribution $\mathrm{Ga}(a_0, b_0)$, the resulting full conditional is also gamma:

$$\delta^2|\cdot \sim \mathrm{Ga}\left(a_0 + \frac{3n}{2}, b_0 + \frac{1}{2\sigma^2}\sum_{i=1}^{n} w_i^{-1}(y_i - \eta_i - \xi w_i)^2 + \sum_{i=1}^{n} w_i\right). \quad (7)$$

Since the prior for the weights also depends on δ^2, these are also part of the updated gamma parameters in the full conditional, yielding a slight change compared to the corresponding full conditional in mean regression.
In summary, we obtain the prior structure shown in Figure 1 that induces the following algorithm for generic Bayesian quantile regression:

i. For $j = 1, \ldots, J$ sample $\boldsymbol{\gamma}_j$ from the Gaussian distribution with parameters (5),

ii. for $j = 1, \ldots, J$ sample $\boldsymbol{\theta}_j$ and $\boldsymbol{m}_j$ from the corresponding hyper full conditional (as detailled in the following sections),

iii. for $i = 1, \ldots, n$ sample w_i from the inverse Gaussian distribution (6),

iv. sample δ^2 from the gamma distribution (7).

Note that our MCMC-sampler for geoadditive quantile regression differs from the one proposed in Yue and Rue (2011) in the update related to nonparametric effects. While Yue and Rue (2011) update the vector of function evaluations $\boldsymbol{f}_j$, our sampler is based on the corresponding basis coefficients $\boldsymbol{\gamma}_j$. This has two major advantages: On the one hand, the dimensionality of the parameter vector is considerably smaller, inducing a tremendous reduction in computing time. On the other hand, it avoids the severe mixing and convergence problems observed by Yue and Rue (2011) and therefore renders estimation og models with more than one nonparametric effect possible.
In the following sections, we present specific examples for modeling the functions f_j and updating the corresponding hyperparameters $\boldsymbol{\theta}_j$ and $\boldsymbol{m}_j$. For notational simplicity the index j will be suppressed, as there will be always a focus on one special class of effects.

2.2 Geoadditive Quantile Regression

Continuous covariates. For approximating potentially nonlinear effects, Bayesian P-splines can be used, see Eilers and Marx (1996) and Brezger and Lang (2006) for full details. Here the $n \times K$ design matrix $\boldsymbol{Z}$ is composed of B-spline basis functions

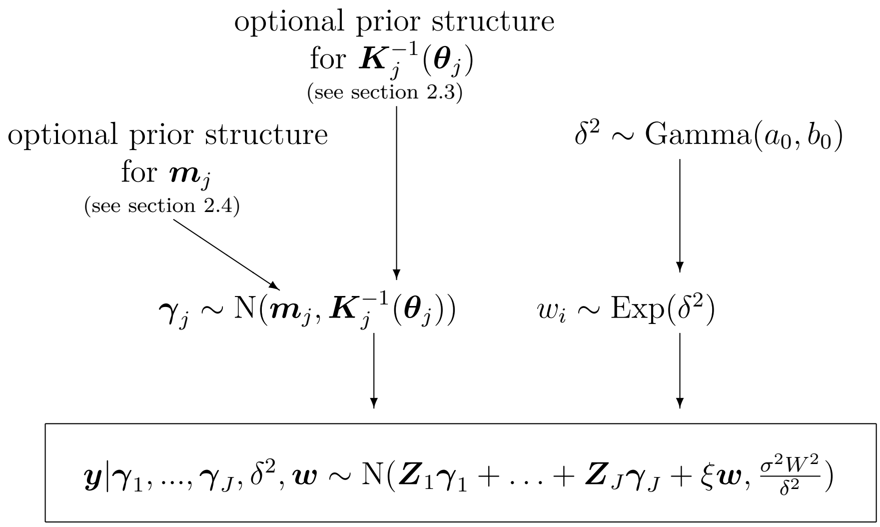

Figure 1: Structure of a simple quantile regression model

evaluated at the observations x_i. Assuming a first or second order random walk for $\boldsymbol{\gamma}$, i.e.

$$\gamma_k \,|\, \gamma_{k-1}, \theta^2 \sim N(\gamma_{k-1}, \frac{1}{\theta^2}), \qquad k = 2, \ldots, K$$

or

$$\gamma_k \,|\, \gamma_{k-1}, \gamma_{k-2}, \theta^2 \sim N(2\gamma_{k-1} - \gamma_{k-2}, \frac{1}{\theta^2}), \qquad k = 3, \ldots, K$$

as smoothness prior with diffuse priors for initial values yields the penalty matrix $\boldsymbol{K}(\boldsymbol{\theta}) = \theta^2 \boldsymbol{R}'\boldsymbol{R}$ where $\boldsymbol{R}$ is a first or second order difference matrix. The prior also implies $\boldsymbol{m} = \boldsymbol{0}$. The vector of additional parameters $\boldsymbol{\theta}$ collapses to a single precision parameter θ^2 that governs the trade off between fidelity to the data and smoothness. The standard prior is $\theta^2 \sim \text{Ga}(a, b)$ implying the full conditional

$$\theta^2 \,|\, \cdot \sim \text{Ga}(a + 0.5\text{rank}(\boldsymbol{K}(\boldsymbol{\theta})), b + 0.5\boldsymbol{\gamma}'\boldsymbol{K}(\boldsymbol{\theta})\boldsymbol{\gamma}). \tag{8}$$

Spatial effects. For data observed on a regular or irregular lattice as in our case study on rents in Munich a common approach for the spatial effect is based on Markov random fields, see Rue and Held (2005). Let $s_i \in \{1, \ldots, K\}$ denote the spatial index or region of the i-th observations. Then we assume $f_{\text{spat}}(s_i) = \gamma_{s_i}$, i.e. separate parameters $\gamma_1, \ldots, \gamma_K$ for each region are estimated. The $n \times K$ design matrix $\boldsymbol{Z}$ is an incidence matrix whose entry in the i-th row and k-th column is equal to one if observation i has been observed at location k and zero otherwise.

The most simple Markov random field prior for the regression coefficients γ_s is defined by

$$\gamma_s \mid \gamma_u, u \neq s, \theta \sim N\left(\sum_{u \in \partial_s} \frac{1}{N_s}\gamma_u, \frac{1}{N_s\theta^2}\right),$$

where N_s is the number of adjacent regions of s, and ∂_s denotes the regions which are neighbors of region s. This implies the penalty matrix $\boldsymbol{K}(\boldsymbol{\theta})$ with elements

$$\boldsymbol{K}(\boldsymbol{\theta})[k,u] = \frac{1}{\theta^2}\begin{cases} -1 & k \neq u, k \sim u, \\ 0 & k \neq u, k \nsim u, \\ |N(k)| & k = u \end{cases}$$

and $\boldsymbol{m} = \mathbf{0}$. Again the vector $\boldsymbol{\theta}$ of additional parameters collapses to a single precision parameter θ with gamma prior $\theta^2 \sim \mathrm{Ga}(a,b)$ and corresponding full conditional (8).

2.3 LASSO Regularization

In this section, we will show how the concept of the LASSO (Tibshirani, 1996) can be adapted to Bayesian quantile regression by specifying suitable hyperpriors for $\boldsymbol{\theta}$ following the ideas presented in Park and Casella (2008). Suppose that the regression coefficients in a K-dimensional vector $\boldsymbol{\gamma}$ shall be subject to LASSO-regularization. Then a prior structure that yields posterior mode estimates which can be interpreted as the Bayesian analogue to LASSO-regularized penalized maximum likelihood estimates is given by the Laplace prior

$$p(\boldsymbol{\gamma}) = \prod_{k=1}^{K} \lambda^2 \exp\left(-\lambda|\gamma_k|\right),$$

To enable the inclusion of Bayesian LASSO-regularization in the basic geoadditive quantile regression sampler outlined in the previous section, we again rely on a scale-mixture representation of the Laplace prior yielding $\boldsymbol{\gamma}|\boldsymbol{\theta} \sim N(\mathbf{0}, \boldsymbol{K}^{-1}(\boldsymbol{\theta}) = \mathrm{diag}(1/\theta_1^2, \dots, 1/\theta_K^2))$ and hyperpriors $\theta_k \sim \mathrm{Exp}(\lambda^2)$, $k = 1, \dots, K$ and $\lambda^2 \sim \mathrm{Ga}(a,b)$. The full conditionals for the LASSO-specific parameters are

$$\theta_k^{-2}|\cdot \sim \mathrm{InvGauss}\left(\frac{|\lambda|}{\gamma_k}, \lambda^2\right) \quad \text{and} \quad \lambda^2|\cdot \sim \mathrm{Ga}\left(a + K, b + 0.5\sum_{k=1}^{K} 1/\theta_k^2\right).$$

According to Park and Casella (2008) the LASSO prior specified so far may result in a multimodal posterior. We can avoid this problem by the scale-dependent prior $\boldsymbol{\gamma}|\boldsymbol{\theta} \sim N(\mathbf{0}, \frac{1}{\delta^2}\boldsymbol{K}^{-1}(\boldsymbol{\theta}) = \text{diag}(1/\theta_1^2, \ldots, 1/\theta_K^2))$ where the prior covariance is scaled by the inverse of δ^2. Then all full conditionals have to be slightly modified accordingly in analogy to Park and Casella (2008).

2.4 Dirichlet Process Mixtures for Random Effects

A further extension of the predictor for Bayesian quantile regression results when considering random effects with potentially non-normal random effects distribution specified via a Dirichlet process mixture prior (see Ghosh and Ramamoorthi, 2010, for a recent review). The latter allows to specify a hyperprior on the space of all random effects distributions while simultaneously enabling model-based clustering of the random effects to retrieve groups of observations with similar random effects profiles. In the following $\boldsymbol{\gamma}_i$ will denote the random effects vector for individual i, referred to as $\boldsymbol{b}_i$ in equation (2).
Our prior specification and implementation of DPMs is based on the stick-breaking representation introduced by Sethuraman (1994). Let G denote the random effects distribution and assume that a Dirichlet hyperprior is specified for G, i.e.

$$\boldsymbol{\gamma}_i \sim G, \quad G \sim DP(\nu_0, G_0)$$

where G_0 is a base distribution and $\nu_0 > 0$ specifies a concentration parameter that determines a priori expected deviations of G from the base distribution G_0. Then it follows from the stick-breaking representation of Dirichlet processes that

$$G(\cdot) = \sum_{k=1}^{\infty} \pi_k \delta_{\boldsymbol{\phi}_k}(\cdot),$$

where $\delta_{\boldsymbol{\phi}_k}(\cdot)$ are Dirac measures (i.e. point masses) located at cluster-specific parameter vectors $\boldsymbol{\phi}_k$ drawn from a base distribution G_0, i.e.

$$\boldsymbol{\phi}_k \overset{\text{i.i.d.}}{\sim} G_0,$$

independently from the (random) weights π_k. The weights are generated through the *stick-breaking process*

$$\pi_1 = v_1, \quad \pi_k = v_k \left(1 - \sum_{j=1}^{k-1}(1-\pi_j)\right) = v_k \prod_{j=1}^{k-1}(1-v_j), \quad k = 2, 3, \ldots,$$

with $v_k \overset{\text{i.i.d.}}{\sim} \text{Be}(1, \nu_0)$, where Be denotes the beta distribution. As a consequence, realisations of the Dirichlet process can be constructed as infinite mixtures of point masses at locations generated as i.i.d. draws from a base measure. The weights π_k are generated by first breaking the part $\pi_1 = v_1$ from a stick of the length $1 = \pi_1 + \pi_2 + \ldots$, then breaking off the part $\pi_2 = v_2(1 - \pi_1)$ from the remaining stick of length $1 - \pi_1$, and so on.

The stick-breaking representation of Dirichlet processes enables an intuitive interpretation since it may be considered an infinite extension of finite mixture models. However, it also reveals that realisations from a Dirichlet process are almost surely discrete with all probability mass concentrated on the locations $\boldsymbol{\phi}_k$, $k = 1, 2, \ldots$. To overcome this limitation, we do not specify the Dirichlet process directly for the random effects distribution but for the hyperparameters of this distribution, yielding Dirichlet process mixtures with the following prior hierarchy:

$$\begin{aligned} \boldsymbol{\gamma}_i &\overset{\text{ind.}}{\sim} p(\boldsymbol{\gamma}_i|\boldsymbol{\phi}_i), \\ \boldsymbol{\phi}_i &\overset{\text{i.i.d.}}{\sim} G, \\ G &\sim DP(\nu_0, G_0). \end{aligned}$$

Here, the random effects are assumed to be realized independently from distributions $p(\boldsymbol{\gamma}_i|\boldsymbol{\phi}_i)$ with individual-specific parameters $\boldsymbol{\phi}_i$. These are generated according to a probability measure obtained from a Dirichlet process with concentration parameter ν_0 and base distribution G_0. Since the realization of the Dirichlet process is almost surely discrete, ties among the individual-specific parameters $\boldsymbol{\phi}_i$ will arise and therefore there will be groups of individuals sharing the same random effects distribution.

Specific choices we make for the prior specification in case of random effects are

$$\begin{aligned} \boldsymbol{\gamma}_i &\overset{\text{ind.}}{\sim} N(\boldsymbol{m}_i, \boldsymbol{K}(\boldsymbol{\theta})^{-1}), \\ \boldsymbol{m}_i &\overset{\text{i.i.d.}}{\sim} G, \\ G &\sim DP(\nu_0, G_0), \end{aligned}$$

i.e. the random effects are independent Gaussian distributed with a common covariance matrix $\boldsymbol{K}(\boldsymbol{\theta})^{-1} = \boldsymbol{\Omega}$ but differing means $\boldsymbol{m}_i$ following a Dirichlet process prior.

We will choose G_0 to be Gaussian and assign ν_0 a Gamma distribution. The hyper-

$$\begin{tikzcd}
\nu_0 \sim \text{Gamma}(a,b) \arrow[dd] & \\
 & G_0 = N(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0) \arrow[dl] \\
G \sim DP(\nu_0, G_0), \arrow[d] & \\
\boldsymbol{m}_i \sim G \arrow[d] & \sigma_{\gamma_r} \sim \text{Gamma}(a_r, b_r), r = 1, \dots, q \arrow[d] \\
\boldsymbol{m}_i & \boldsymbol{K}^{-1}(\boldsymbol{\theta}) = \text{diag}(\sigma_{\gamma_1}, \dots, \sigma_{\gamma_q})
\end{tikzcd}$$

Figure 2: Hyper-prior structure for the mean in the DPM context.

prior structures of $\boldsymbol{\mu}_{\gamma_i}$ and $\boldsymbol{K}(\boldsymbol{\theta})$ are visualized in Figure 2.
For the random effects $\boldsymbol{\gamma}_i$ we need to take into account the data distribution and the prior with the parameters generated via the DPM. The fact, that we are using DPMs does not change the equation in comparison to what we would get in a normal mixed model approach. The full conditional of $\boldsymbol{m}$ is of the same structure as (5).
As the stick breaking process can obviously not be conducted to infinity, the standard approach is to truncate the process in a certain N – see Ishwaran and James (2001) – and only take into account the first N terms of the sum. In our example, N is chosen to be 100.
A common way to implement the DPM – also Ishwaran and James (2001) – is to introduce a vector of latent classification variables $\boldsymbol{c}$ of the length n, which consists in each iteration of the Gibbs sampler of m different values of $1, \dots, n$. The subvector $\boldsymbol{c}^*$ denotes the vector comprising only the distinct values corresponding to non-empty clusters. The auxiliary variables ϕ_k ($k \in 1, \dots, N$) are drawn from two different types of distributions depending on the fact if $k \in \boldsymbol{c}$ or not.
For the k which are not one of the different values in the set of c_i, ϕ_k is drawn from the base distribution:

$$\phi_k | \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0 \sim N(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0).$$

If $k \in \boldsymbol{c}^*$, ϕ_k is drawn as follows:

$$\phi_k|\sigma^2_{\gamma_r}, \mu_{0_r}, \sigma^2_{0_r}, \boldsymbol{\gamma}, \boldsymbol{c} \sim N(\mu^*_{0_r}, \sigma^{2*}_{0_r}),$$

for the different k

$$\mu^*_{0_r} = \left(\frac{n_h}{\sigma^2_{\gamma_r}} + \frac{1}{\sigma^2_{0_r}}\right)^{-1} \left(\frac{n_h}{\sigma^2_{\gamma_r}} \bar{b}_{r,h} + \frac{\mu_{0_r}}{\sigma^2_{0r}}\right) \quad \text{and} \quad \sigma^{2*}_{0_r} = \left(\frac{n_h}{\sigma^2_{\gamma_r}} + \frac{1}{\sigma^2_{0_r}}\right)^{-1}.$$

The $\sigma^2_{0_r}, r = 1, \ldots, q$ are the diagonal elements of $\boldsymbol{\Sigma}_0$ and their priors are, again, gamma distributions. The latent classification variables are drawn from a mixture of the likelihood for the different ϕ, weighted with different π_k, which are drawn via the stickbreaking representation of the Dirichlet distribution:

$$c_i|\boldsymbol{\pi}, \boldsymbol{\phi}, \boldsymbol{\gamma}_i, \boldsymbol{\Sigma}_\gamma \sim \sum_{h=1}^{N} \pi_k p(\boldsymbol{\gamma}_i|\phi_k, \boldsymbol{\Sigma}_\gamma)\delta_k(\cdot)$$

with $\delta_k(\cdot)$ being the Dirac measure in k.
As a nex step, the π_k are constructed in a stick-breaking step using an auxiliary variable v_k, which is beta distributed, as allready described. The π_k are then obtained through the above mentioned product

$$\pi_k = v_k \prod_{l<k} (1 - v_l).$$

Finally the precision, which was assigned a gamma distribution as prior, for the DP is drawn from a gamma distribution

$$\nu_0|\boldsymbol{\pi} \sim \text{Gamma}\left(N - 1 + a_\alpha, b_\alpha - \sum_{h=0}^{N-1} \log(1 - V_k)\right).$$

The $\boldsymbol{m}_i$ themselves are estimated by using $\boldsymbol{m}_i = \phi_{c_i}$. As $\boldsymbol{c}$ contains only m different values (with $m \leq n$) the clustering mechanism follows directly from the construction. It is obvious that, if this algorithm is repeated many times in an MCMC simulation, due to the randomness of each step we get different values for each individual and furthermore different clusters. Therefore, we will use a nearest neighbor approach to get a clustering which is valid over all iterations.

3 Applications

3.1 High-Dimensional Geoadditive Regression for the Munich Rental Guide

As a first application of semiparametric Bayesian quantile regression, we consider the high-dimensional geoadditive regression model (1) but extended to conditional quantile specifications. We chose to model the 5%, 20%, 50%, 80% and 95% quantiles to give a detailled summary on both the central part of the distribution and the boundaries. This reveales information not only about the expected rent for a flat but also about the span, the rent is supposed to be in.

Due to the high dimensionality of the vector of parametric covariates (238 covariates in total), we will consider Bayesian LASSO regularization for all components (except the intercept). For the nonlinear effects of the size of the flat and the year of construction, we consider cubic P-splines with second order random walk prior and 20 equidistant knots. The spatial effect was calculated with a Markov random field where two subquarters are treated as neighbors if they share a common boundary. For all gamma type priors, we chose hyperparameters $a = b = 0.001$. The number of iterations for the MCMC sampler was fixed at 35,000 with a burn in period of 5,000 iterations and a thinning parameter of 30 (yielding 1,000 samples for determining posterior means). A graphical analysis of mixing and convergence showed no inadequacies.

If we look at the results, we see that in fact most of the effects differ over the different quantiles (see for example the nonlinear effects in Figures 3 and 4). The picture shows the centered curves for the noncentral quantiles in comparison to the posterior interval of the median regression. The grey stripes in the background indicate the concentration of the data for the corresponding values. For reasons of lucidity and because the figures have the aim to compare the effects, the posterior intervals are only plotted for the median regression. The first thing which has to be mentioned is, that the posterior interval is broader in the parts with less observations. Another obvious fact is, that smoothing works better for the central quantiles than for the extreme ones. This might be caused by the sparsity of data in these areas. As for the differences of the impact on the different quantiles we see a tendency of the functions to tilt over for both effects. The effect of the size of the flat seems to be less expressed for the lower quantiles, expecially in the lowprice segment the size of the flat does not really have any influence at all and increasing variability is observed as τ grows. It is the other way round for the year of construction: there is no pronounced effet of the age of the building on the highest quantile. Another effect which shows in the latter is that the positive impact of old buildings does not exist for lower quantiles.

In the spatial effect, the most noticeable fact is the higher number of significant effects for the outer quantiles (see Figure 5). While for the 95% quantile, 93 subquarters are selected to be significant, there are only 54 subquarters, which have significant impact on the median. Furthermore, we see a tendency towards more subquarters which have positive effect on the price in the higher quantiles (50 subquarters with positive and 43 with negative influence in the 95% quantile) and the opposed effect for lower quantiles (52 negative to 29 negative in the 5% quantile). The second fact we can see in the graphics is that, just as in the nonlinear effects, the estimation for the regression on the outer quantiles is less smooth, than for the ones in the middle, indicating more variability in extreme quantiles.

Selection of the LASSO-regularized covariates with parametric effects via 95% posterior credible intervals led to a whole of 103 variables aggregated over all five models. The quantile for which the least effects where recognized as significant was the median with 47, while the highest number was 58 for $\tau = 0.95$. Analysing posteriors of these parameters, different comportments can be detected. There are covariates, which show very similar effects over all five quantiles, while others are only selected for parts of them. In most of the cases, we can see the same direction of effect, while in some the sign changes over the different quantiles (see Figure 6 for some exemplary effects).

In some cases, the reason for covariates not to be selected is obviously the lack of data (for example the fact that *sauna*, just as high class facilities, is only selected to have an influence on the 95% quantile, while *no modernisiations at all* only appears in the lowest quantile), while others just seem to be selected due to inner correlation. The latter is not surprising as the LASSO is known for only selecting a few or even one representative for highly correlated covariates.

To compare the results of the high-dimensional geoadditive model with those of an expert model with a restricted set of covariates and with quantiles calculated from a Gaussian mean regression model results, we performed a ten-fold cross validation comparing the empirical risk on the test data. The quantiles for each flat where estimated as the quantiles of the distribution:

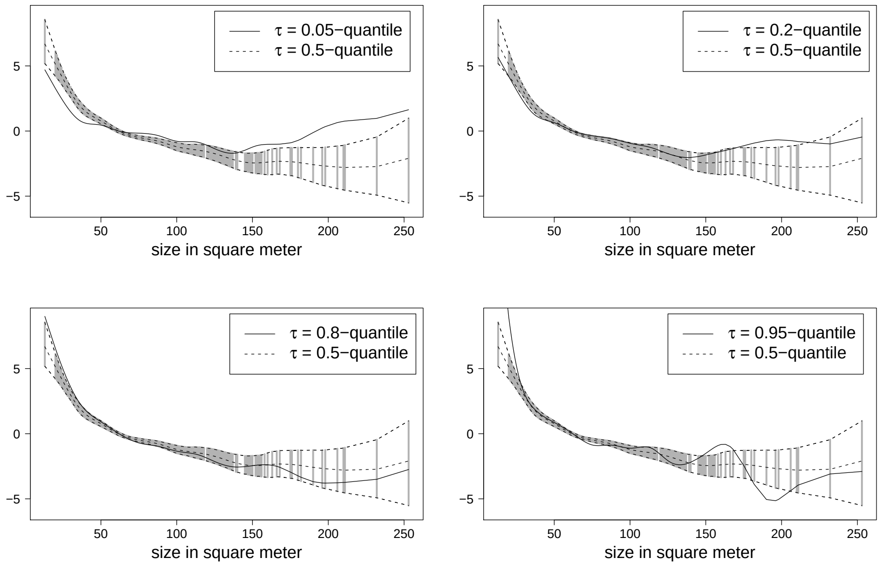

Figure 3: Munich rent index: nonlinear effect of size of the flat. Solid line: effect on the non central quantiles, dashed lines: effect on the median and 95%-posterior interval, lightgrey lines in background: concentration of data.

$$N(\hat{\eta}_i, \hat{\boldsymbol{\varepsilon}}^T \hat{\boldsymbol{\varepsilon}})$$

where $\hat{\eta}_i$ denotes the prediction for the ith flat and $\hat{\boldsymbol{\varepsilon}}$ is the vector of the residuals. Thus the quantiles are obtained by the underlying Gaussian distribution, using the prediction as mean and the squared residuals as variance. The empirical risk is obtained by evaluating the quantile loss function $\rho(y_i - \eta_{i,\tau})$ at the posterior mean estimates in the test set. It turns out that the empirical risk is lower for the LASSO as compared to the expert model for nearly all quantiles, see Table 3.1 for average risks over the folds. In comparison to the mean regression model both quantile regression models, expert and LASSO, perform better. These results are also illustrated in Figure 7 in terms of a parallel coordinate plot.

Geoadditive quantile regression including LASSO regularizaton is implemented in BayesX (Lang et al., 2005) and will be published in the next version of the software.

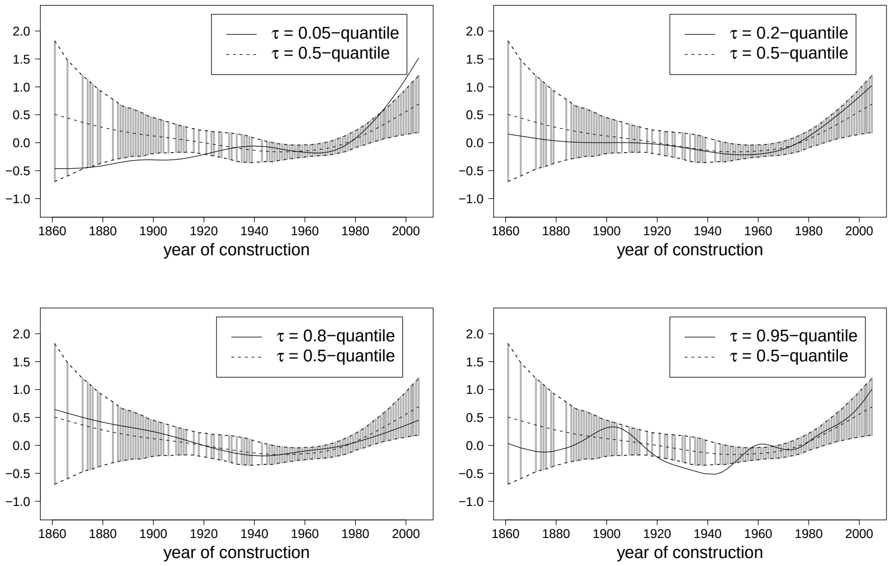

Figure 4: Munich rent index: nonlinear effect of year of contruction. Solid line: effect on the non central quantiles, dashed lines: effect on the median and 95%-posterior interval, lightgrey lines in background: concentration of data.

Quantile	LASSO Model	Expert Model	Mean Regression
$\tau = 0.05$	0.2285	0.2304	0.2146
$\tau = 0.20$	0.5043	0.5476	0.5516
$\tau = 0.50$	0.7100	0.7317	0.7339
$\tau = 0.80$	0.5005	0.5089	0.5168
$\tau = 0.95$	0.2301	0.2224	0.2039

Table 1: Munich rent index: mean risk for the high-dimensional geoadditive model, the expert model and the mean regression model, averaged over the ten cross validation folds.

3.2 Nonparametric Random Effects for Longitudinal Obesity Measurements in Children

As an application involving the specification of individual-specific random effects, we consider the childhood obesity data described in the introduction. These longitudinal data were collected for 3,097 healthy neonates over 60 months at nine mandatory medical examinations (2 weeks, 3, 6, 12, 24, 48 and 60 months). We restricted attention to complete cases, yielding a final sample size of 2,043 indi-

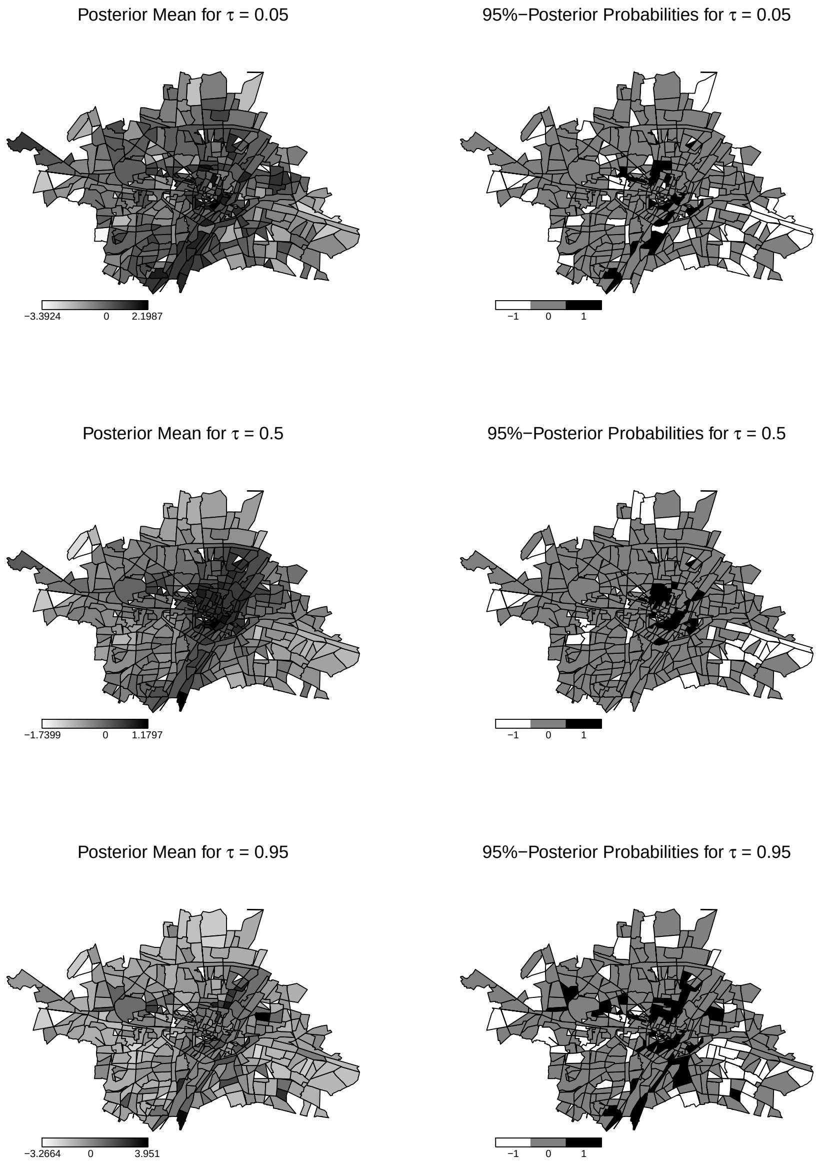

Figure 5: Munich rent index: spatial effects on the median (in the middle) and the outer quantiles ($\tau = 0.05$ on the top and $\tau = 0.95$ at the bottom). On the left: centralized effects, on the right: quarters with significant negative effect in white, quarters with significant positive effect in black and quarters with nonsignificant effects in grey.

viduals. Collected covariates are *sex* (gender), *breast* (nutrition until the age of 4 months, 0 = bottle-fed or mixture of bottle-fed and breast-fed, 1 = breast-fed only), *mSmoke* (maternal smoking during pregnancy, 0 = no, 1 = yes), *area* (0 = rural,

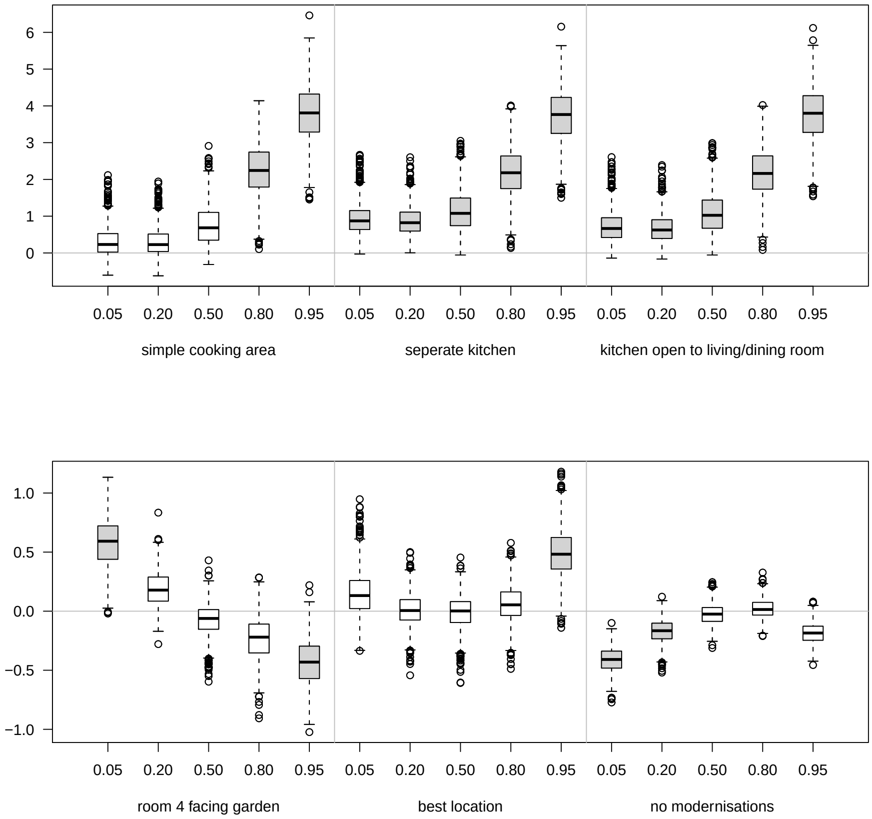

Figure 6: Munich rent index: examples for regularized linear effects with different behaviour in different quantiles; the grey boxes represent significant effects and the white ones non significant effects.

1= urban), *ageY* (age in years), *mBMI* (maternal BMI at pregnancy begin) and *mDiffBMI* (maternal BMI gain during pregnancy). The latter two variables were used in centered form.

We consider a quantile-specific version of model equation (2), where the temporal trend is specified as a cubic P-spline with twenty inner knots and second order random walk prior. The random effects comprise subject-specific intercepts, a random slope for a linear time trend and an additional random slope for the nonlinear time transformation $\log(t+1)/(t+1)^2$. The prior for the precision of the DPM was a gamma distribution with $a = 0.5$ and $b = 10$, the rest of the priors was handeled in the same way as in the analysis of the Munich rental guide. The analysis was performed for the same five quantiles (5%, 20%, 50%, 80% and 95%) as in the Mu-

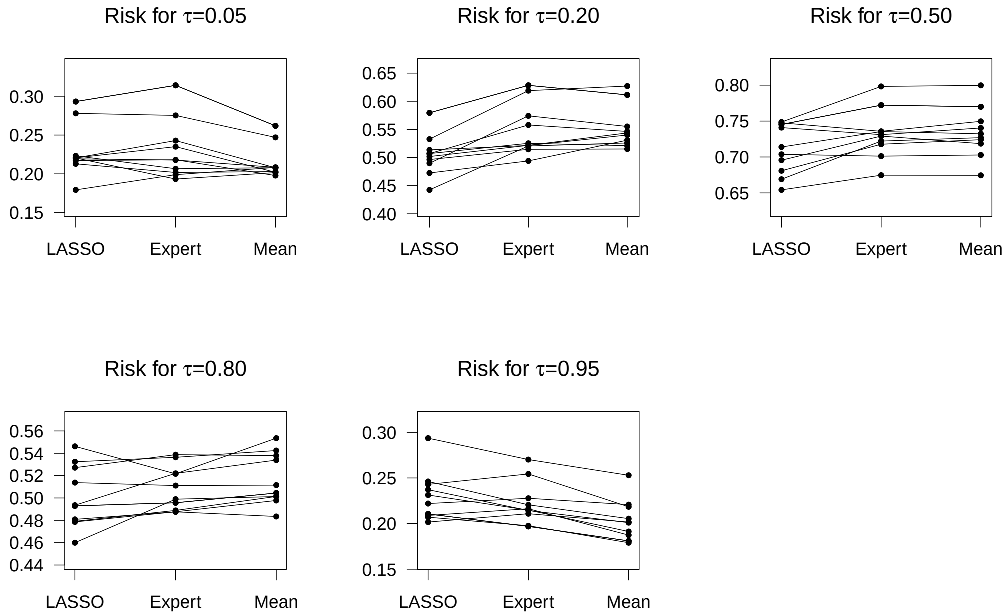

Figure 7: Munich rent index: parallel coordinate plots of risk functions for LASSO, expert and mean regression model.

nich rental guide while the number of MCMC iterations had to be higher to achieve satisfactory results for mixing and convergence. Due to high autocorrelations especially in the estimation for the nonlinear trend function, we used a burn in period of 100,000 iterations, a thinning parameter of 200 and did a total of 300,000 iterations in order to obtain posteriors of the size of 1,000 samples. Note that the mixing performance of all parameters (in particular the random effects) except the time trend was satisfactory already after a much smaller number of iterations.
To account for correlations arising from the longitudinal arrangement of the data while allowing for potential non-normality of the random effects distribution, we included random effects using the Dirichlet process mixtures as explained in Section 2.4. The resulting number of clusters varies between eight and ten clusters for the different quantiles. The clusters are not build analogously for the quantiles, that means that for different parts of the distribution children may belong to different groups.
The results for the median regression are quite similar to those of the mean regression obtained by Heinzl et al. (2012), with the obvious difference in robustness. Figure 8 shows the BMI conditional on time for four different individuals for the median

	0.05	0.2	0.5	0.8	0.95
sex	0.2260	0.2254	0.1915	0.1977	0.2177
breast	0.0874	0.0752	0.0557	0.0526	0.0696
mSmoke	0.0367	0.0097	-0.0123	0.0166	0.0163
area	-0.0120	-0.0341	-0.0757	-0.0752	-0.0868
mBMI	0.0468	0.0454	0.0453	0.0485	0.0491
mDiffBMI	0.0712	0.0755	0.0828	0.0871	0.0895

Table 2: Obesity study: Linear effects for the five different quantiles.

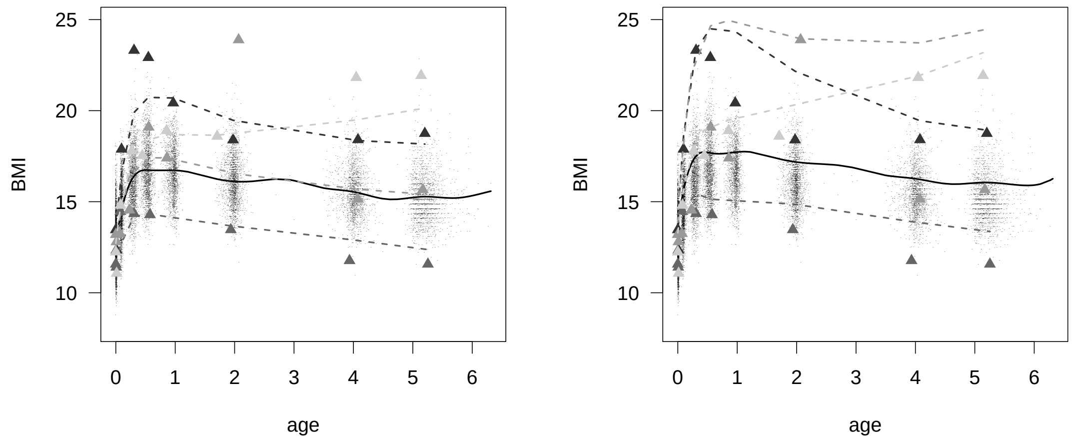

Figure 8: Obesity study: effect of the age on the BMI in 50% – quantile (on the left) and 95% – quantile regression (on the right).

regression (on the left side) and the 95%-quantile regression on the right. While the outlier for the children shown in grey and green are ignored in the median regression, they are obviously taken into account in the curve on the right side. The same effect arises for the clustered version.

The linear effects are presented in Table 2 and Figure 9. In the latter, significant effects, i.e. the ones in which did not contain zero in their 95% posterior credible interval, are marked in grey. The most interesting result is the observation, that breastfeeding does only have a significant effect on the lowest quantile. This is a fact, which was already stated in the work of Beyerlein et al. (2008). The negative effect on the higher quantiles, which was seen as even more important in this work, could not be confirmed by our analysis.

The calculated models were compared to two simpler models via the DIC. The DIC

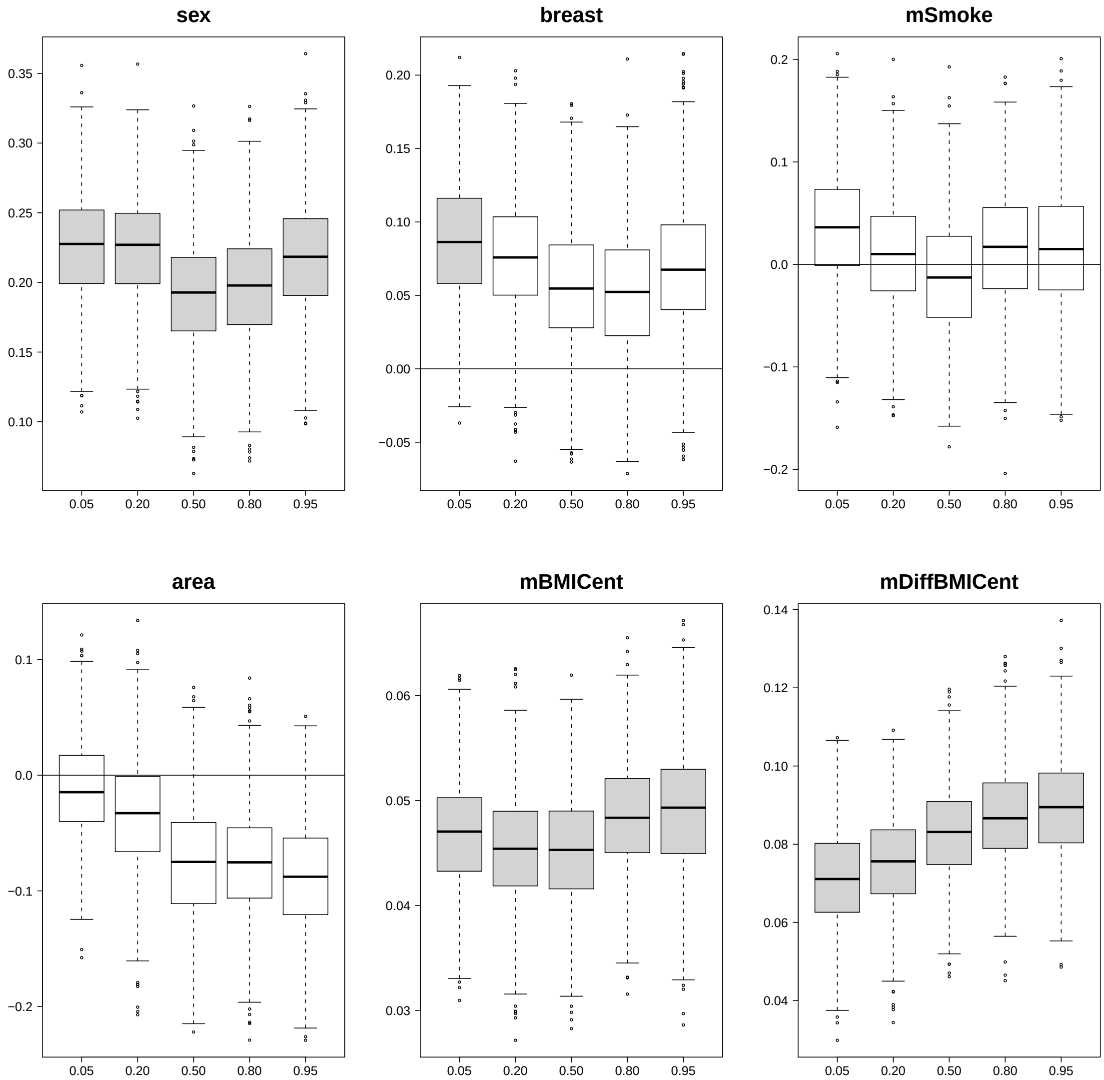

Figure 9: Obesity Study: boxplots of the different effects on the different quantiles; the grey boxes represent significant effects and the white ones non significant effects.

was calculated as

$$\mathrm{DIC} = 2\sum_{b=1}^{B} \mathrm{dev}(\boldsymbol{\theta}_b) - \mathrm{dev}(\bar{\boldsymbol{\theta}})$$

i.e. as the difference of two times the mean over the deviance in each sample and the deviance which is calculated with the sample mean of the parameters where b indexes the MCMC iterations. The DPM-Model outperformed a model without random effects with high difference in all quantiles (the mean difference was 6447.81). Using a normal Gaussian prior for the random effects performed worse than the

τ	DPM	Gaussian random effects	no random effects
0.05	15296.03	16586.32	23031.73
0.20	32957.19	34246.35	38377.08
0.50	34665.83	35774.21	39813.83
0.80	34379.04	35301.70	40103.93
0.95	16636.18	17732.05	24846.75

Table 3: Obesity study: DIC for different models.

more complex DPM-Model over all quantiles, too (mean difference: 1141.272). For an overview over the different DICs see table 3. These results can, however, only be accepted with reservation, since the DIC calculations are based on the auxiliary assumption of the asymmetric Laplace distribution for the error terms.
The MCMC algorithm for the DPM random effects model was implemented in C++ and R. Just like the scheme itself the program was based on work by Heinzl et al. (2012).

4 Conclusions and Discussion

The presented possibilities of modeling quantiles are useful extensions for the regression toolbox. Obviously the mixture representation of ALD allows for considerable flexibility and, in particular, other Bayesian approaches from mean regression can easily be incorporated. Especially the combination of different effects is rendered possible and seems to be nearly unlimited when using latent Gaussian model formulations. In future work, this concept could be transferred to variational approximations Ormerod and Wand (2010), which have the advantage to avoid simulation-based inference.
One principle difficulty with Bayesian quantile regression is that the ALD is only a working model that yields a model misspecification. Therefore significance and uncertainty statements or quantities derived from the samples, like the DIC, have to be interpreted with care. A possible alternative is to include the estimation of the error density in the MCMC algorithm, for example via mixtures – see for example Kottas and Krnjajic (2009), Dunson and Taylor (2005) and Taddy and Kottas (2010). Apart from this drawback, the presented method is still a progress in calculating highly demanding models for conditional quantiles. The possibility to regularize linear effects individually for each quantile can lead to interesting findings,

concerning different influential structures, while using Dirichlet process mixtures for the random effects gives rise to a model which accounts for dependencies in clusters.
Acknowledgement: Financial support from the German Science Foundation (DFG), grant KN 922/4-1 is gratefully acknowledged. We also thank Felix Heinzl for providing his code on DPMs in mean regression.

References

Beyerlein, A., A. M. Toschke, and R. von Kries (2008). Breastfeeding and childhood obesity: Shift of the entire BMI distribution or only the upper parts? *Obesity 16*, 2730–2733.

Brezger, A. and S. Lang (2006). Generalized structured additive regression based on Bayesian P-splines. *Computational Statistics and Data Analysis 50*, 967–991.

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

Eilers, P. H. C. and B. D. Marx (1996). Flexible smoothing using B-splines and penalized likelihood. *Statistical Science 11*, 89–121.

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

Fenske, N., T. Kneib, and T. Hothorn (2011). Identifying risk factors for severe childhood malnutrition by boosting additive quantile regression. *Journal of the American Statistical Association 106*, 494–510.

Ghosh, J. K. and R. V. Ramamoorthi (2010). *Bayesian Nonparametrics*. Springer.

Heinzl, F., T. Kneib, and L. Fahrmeir (2012). Additive mixed models with dirichlet process mixture and p-spline priors. *Advances in Statistical Analysis 96*, 47–68.

Ishwaran, H. and L. James (2001). Gibbs sampling methods for stick-breaking priors. *Journal of Computational and Graphical Statistics 11*, 508–532.

Kneib, T., T. Hothorn, and G. Tutz (2009). Variable selection and model choice in geoadditive regression models. *Biometrics 65*, 626–634.

Kneib, T., S. Konrath, and L. Fahrmeir (2011). High-dimensional structured additive regression models: Bayesian regularisation, smoothing and predictive performance. *Applied Statistics 60*, 51–70.

Koenker, R. (2005). *Quantile Regression.* Number 9780521608275 in Cambridge Books. Cambridge University Press.

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

Koenker, R. and I. Mizera (2004). Penalized triograms: Total variation regularization for bivariate smoothing. *Journal of the Royal Statistical Society: Series B 66*(1), 145–163.

Koenker, R., P. Ng, and S. Portnoy (1994). Quantile smoothing splines. *Biometrika 81*(4), 673–680.

Kottas, A. and M. Krnjajic (2009). Bayesian nonparamteric modeling in quantile regression. *Scandinavian Journal of Statistics 36*, 297–319.

Lang, S., T. Kneib, and A. Brezger (2005). Bayesx: Analyzing Bayesian structural additive regression models. *Journal of Statistical Software 14*(i11).

Li, Q., R. Xi, and N. Lin (2010). Bayesian regularized quantile regression. *Bayesian Analysis 5*, 533–556.

Oh, H.-S., T. C. M. Lee, and D. W. Nychka (2011). Fast nonparametric quantile regression with arbitrary smoothing methods. *Journal of Computational and Graphical Statistics 20*, 510–526.

Ormerod, J. and M. Wand (2010). Explaining variational approximations. *The American Statistician 64*(2), 140–153.

Park, T. and G. Casella (2008). The Bayesian lasso. *Journal of the American Statistical Association 103*, 681–686.

Rue, H. and L. Held (2005). *Gaussian Markov Random Fields.* Boca Raton, FL: Chapman & Hall / CRC.

Sethuraman, J. (1994). A constructive definition of dirichlet priors. *Statistica Sinica 4*, 639–650.

Taddy, M. and A. Kottas (2010). A Bayesian nonparamteric approach to inference for quantile regression. *Journal of Business and Economic Statistics*, 357–369.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society (Series B) 58*, 267–288.

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

Yue, Y. and H. Rue (2011). Bayesian inference for additive mixed quantile regression models. *Computational Statistics and Data Analysis 55*, 84–96.

University of Innsbruck - Working Papers in Economics and Statistics
Recent Papers can be accessed on the following webpage:

http://eeecon.uibk.ac.at/wopec/

2012-06 **Elisabeth Waldmann, Thomas Kneib, Yu Ryan Yu, Stefan Lang:** Bayesian semiparametric additive quantile regression

2012-05 **Eric Mayer, Sebastian Rueth, Johann Scharler:** Government debt, inflation dynamics and the transmission of fiscal policy shocks

2012-04 **Markus Leibrecht, Johann Scharler:** Government size and business cycle volatility; How important are credit constraints?

2012-03 **Uwe Dulleck, David Johnston, Rudolf Kerschbamer, Matthias Sutter:** The good, the bad and the naive: Do fair prices signal good types or do they induce good behaviour?

2012-02 **Martin G. Kocher, Wolfgang J. Luhan, Matthias Sutter:** Testing a forgotten aspect of Akerlof's gift exchange hypothesis: Relational contracts with individual and uniform wages

2012-01 **Loukas Balafoutas, Florian Lindner, Matthias Sutter:** Sabotage in tournaments: Evidence from a natural experiment

2011-29 **Glenn Dutcher:** How Does the Social Distance Between an Employee and a Manager affect Employee Competition for a Reward?

2011-28 **Ronald Peeters, Marc Vorsatz, Markus Walzl:** Truth, trust, and sanctions: On institutional selection in sender-receiver games *forthcoming in Scandinavian Journal of Economics*

2011-27 **Haoran He, Peter Martinsson, Matthias Sutter:** Group Decision Making Under Risk: An Experiment with Student Couples *forthcoming in Economics Letters*

2011-26 **Andreas Exenberger, Andreas Pondorfer:** Rain, temperature and agricultural production: The impact of climate change in Sub-Sahara Africa, 1961-2009

2011-25 **Nikolaus Umlauf, Georg Mayr, Jakob Messner, Achim Zeileis:** Why Does It Always Rain on Me? A Spatio-Temporal Analysis of Precipitation in Austria *forthcoming in Austrian Journal of Statistics*

2011-24 **Matthias Bank, Alexander Kupfer, Rupert Sendlhofer:** Performance-sensitive government bonds - A new proposal for sustainable sovereign debt management

2011-23 **Gerhard Reitschuler, Rupert Sendlhofer:** Fiscal policy, trigger points and interest rates: Additional evidence from the U.S.

2011-22 **Bettina Grün, Ioannis Kosmidis, Achim Zeileis:** Extended beta regression in R: Shaken, stirred, mixed, and partitioned

2011-21 **Hannah Frick, Carolin Strobl, Friedrich Leisch, Achim Zeileis:** Flexible Rasch mixture models with package psychomix

2011-20 **Thomas Grubinger, Achim Zeileis, Karl-Peter Pfeiffer:** evtree: Evolutionary learning of globally optimal classification and regression trees in R

2011-19 **Wolfgang Rinnergschwentner, Gottfried Tappeiner, Janette Walde:** Multivariate stochastic volatility via wishart processes - A continuation

2011-18 **Jan Verbesselt, Achim Zeileis, Martin Herold:** Near Real-Time Disturbance Detection in Terrestrial Ecosystems Using Satellite Image Time Series: Drought Detection in Somalia *forthcoming in Remote Sensing and Environment*

2011-17 **Stefan Borsky, Andrea Leiter, Michael Pfaffermayr:** Does going green pay off? The effect of an international environmental agreement on tropical timber trade

2011-16 **Pavlo Blavatskyy:** Stronger Utility

2011-15 **Anita Gantner, Wolfgang Höchtl, Rupert Sausgruber:** The pivotal mechanism revisited: Some evidence on group manipulation

2011-14 **David J. Cooper, Matthias Sutter:** Role selection and team performance

2011-13 **Wolfgang Höchtl, Rupert Sausgruber, Jean-Robert Tyran:** Inequality aversion and voting on redistribution

2011-12 **Thomas Windberger, Achim Zeileis:** Structural breaks in inflation dynamics within the European Monetary Union

2011-11 **Loukas Balafoutas, Adrian Beck, Rudolf Kerschbamer, Matthias Sutter:** What drives taxi drivers? A field experiment on fraud in a market for credence goods

2011-10 **Stefan Borsky, Paul A. Raschky:** A spatial econometric analysis of compliance with an international environmental agreement on open access resources

2011-09 **Edgar C. Merkle, Achim Zeileis:** Generalized measurement invariance tests with application to factor analysis

2011-08 **Michael Kirchler, Jürgen Huber, Thomas Stöckl:** Thar she bursts - reducing confusion reduces bubbles *modified version forthcoming in American Economic Review*

2011-07 **Ernst Fehr, Daniela Rützler, Matthias Sutter:** The development of egalitarianism, altruism, spite and parochialism in childhood and adolescence

2011-06 **Octavio Fernández-Amador, Martin Gächter, Martin Larch, Georg Peter:** Monetary policy and its impact on stock market liquidity: Evidence from the euro zone

2011-05 **Martin Gächter, Peter Schwazer, Engelbert Theurl:** Entry and exit of physicians in a two-tiered public/private health care system

2011-04 **Loukas Balafoutas, Rudolf Kerschbamer, Matthias Sutter:** Distributional preferences and competitive behavior *forthcoming in Journal of Economic Behavior and Organization*

2011-03 **Francesco Feri, Alessandro Innocenti, Paolo Pin:** Psychological pressure in competitive environments: Evidence from a randomized natural experiment: Comment

2011-02 **Christian Kleiber, Achim Zeileis:** Reproducible Econometric Simulations

2011-01 **Carolin Strobl, Julia Kopf, Achim Zeileis:** A new method for detecting differential item functioning in the Rasch model

University of Innsbruck

Working Papers in Economics and Statistics

2012-06

Elisabeth Waldmann, Thomas Kneib, Yu Ryan Yu, Stefan Lang

Bayesian semiparametric additive quantile regression

Abstract
Quantile regression provides a convenient framework for analyzing the impact of covariates on the complete conditional distribution of a response variable instead of only the mean. While frequentist treatments of quantile regression are typically completely nonparametric, a Bayesian formulation relies on assuming the asymmetric Laplace distribution as auxiliary error distribution that yields posterior modes equivalent to frequentist estimates. In this paper, we utilize a location-scale-mixture of normals representation of the asymmetric Laplace distribution to transfer different flexible modeling concepts from Gaussian mean regression to Bayesian semiparametric quantile regression. In particular, we will consider high-dimensional geoadditive models comprising LASSO regularization priors and mixed models with potentially non-normal random effects distribution modeled via a Dirichlet process mixture. These extensions are illustrated using two large-scale applications on net rents in Munich and longitudinal measurements on obesity among children.

ISSN 1993-4378 (Print)
ISSN 1993-6885 (Online)