Statistical Modelling

http://smj.sagepub.com/

Bayesian Tobit quantile regression model for medical expenditure panel survey data

Yu Ryan Yue and Hyokyoung Grace Hong

Statistical Modelling 2012 12: 323

DOI: 10.1177/1471082X1201200402

Statistical Modelling 2012; **12**(**4**): 323–346

Bayesian Tobit quantile regression model for medical expenditure panel survey data

Yu Ryan Yue[1] **and Hyokyoung Grace Hong**[1]
[1]Zicklin School of Business, Baruch College, The City University of New York, New York.

Abstract: High expenditure on healthcare is an important segment of the U.S. economy, making healthcare cost modelling valuable in decision-making processes over a wide array of domains. In this paper, we analyze medical expenditure panel survey (MEPS) data. Tobit regression model has been popularly used for the medical expenditures. However, it is no longer sufficient for the MEPS data because: *(i)* the distribution of the expenditures shows skewness, heavy tails and heterogeneity; *(ii)* most predictors are categorical, including binary, nominal and ordinal variables; *(iii)* there are a few predictors which may be nonlinearly related to the response. We therefore propose a Bayesian Tobit quantile regression model to describe a complete distributional view on how the medical expenditures depend on the various predictors. Specifically, we assume an asymmetric Laplace error distribution to adapt the quantile regression to a Bayesian setting. Then, we propose a modified group Lasso for categorical factor selection, and a smoothing Gaussian prior for modelling the nonlinear effects. The estimates and their uncertainties are obtained using an efficient Monte Carlo Markov Chain sampling method. The effectiveness of our approach is demonstrated by modelling 2007 MEPS data.

Key words: asymmetric Laplace distribution; group lasso; MCMC; medical expenditure panel survey; nonlinear effect; Tobit quantile regression

Received November 2011; **revised** December 2011; **accepted** December 2011

1 Introduction

1.1 Medical expenditure panel survey data

The Medical Expenditure Panel Survey (MEPS), which began in 1996, is a set of large-scale surveys of families and individuals, their medical providers (doctors, hospitals, pharmacies, etc.), and employers across the United States. The MEPS collects data on the specific health services that Americans use, how frequently they use them, the cost of these services and how they are paid for, as well as data on the cost, scope and breadth of health insurance held by and available to U.S. workers. It intended to provide nationally representative estimates of health expenditure, utilization, payment

Address for correspondence: Yu Ryan Yue, Zicklin School of Business, Baruch College, The City University of New York, New York, NY 10010. E-mail: yu.yue@baruch.cuny.edu

 10.1177/1471082X1201200402

sources, income, employment, health status and health insurance coverage among the non-institutionalized, nonmilitary population of the United States. This series of government-produced datasets can be used to examine how individuals interact with the medical care system in the United States. The current publicly available MEPS component is the *Household Component*, consisting of six data files which describe the demographics and characteristics of the survey population and eight event-level files which capture all interactions with the U.S. medical system. The file of our interest is the *Full-Year Consolidated Data* file, which includes all demographic and medical characteristics, as well as patient-reported responses to the main survey questions. More information about MEPS can be found at its official website: http://www.meps.ahrq.gov/mepsweb/.

In this paper, we examine 2007 MEPS data using regression-based approach. The response variable is the total healthcare expenditure (the MEPS variable is totexp07), including insurance spending and annual out-of-pocket spending, measured in dollars. Summary statistics are reported in Table 1. For many econometric explorations, the following prominent features of these expenditure data are typically important to accommodate. First, the expenditures are, for most practical purposes, nonnegative. Second, a sizable fraction of observations (approximately 17% in the MEPS) are measured as zero. As a result, the distribution is a mixture of a point mass in zero and a continuous distribution truncated at zero. Third, the data exhibit a 'heavy' upper tail: in the MEPS data, almost 10% of the expenditures exceed \$10 000. Fourth, with a small probability, households face extremely large medical expenditure, resulting in a right-skewed distribution; note that skewness *per se* does not imply a heavy upper tail. These features are clearly shown in Figure 1. As for explanatory variables, there are socioeconomic factors such as the number of years of education, poverty level, region, etc., and personal characteristics such as self-rated health, general-risk-taking attitude, seat-belt use, etc. Since conventional survey items only allow a limited number of response options, most explanatory variables are categorical, including binary, nominal and ordinal variables.

The MEPS data have been extensively used for the econometrical analysis of the healthcare expenditures (e.g., Cameron and Trivedi, 2010; Clements and Hendry, 2011). The high expenditures on healthcare have been an important segment of the U.S. economy, accounting for about 16% of GDP in 2007, highest among all the developed countries. The effective healthcare cost modelling has fundamentally or

Table 1 Summary statistics of total healthcare expenditure in 2007

Observations	15 890
0.25 Percentile	164
0.50 Percentile	1152
0.75 Percentile	3997
0.95 Percentile	18 808
Mean	4498
Standard deviation	12 728

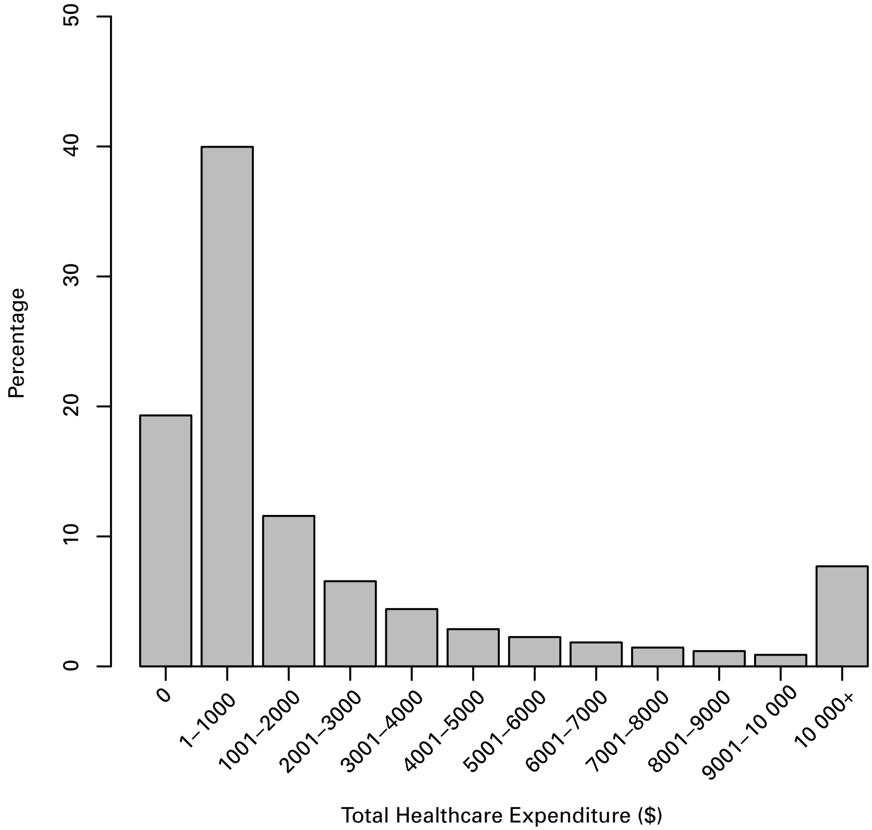

Figure 1 Histogram of the medical expenditures in 2007 MEPS data

peripherally informed decision making over a wide array of domains: risk-adjusted provider payments; provider utilization review/profiling; cost-of-illness assessment; cost aspects of evaluation studies and future projections of disease-specific healthcare cost burdens (Mullahy, 2009).

1.2 Statistical modelling of medical expenditure data

The medical expenditure variable is a so-called *limited dependent* variable whose distribution is mostly continuous but has a point mass at one or more specific values, such as zero. There are a multitude of statistical approaches to modelling of a limited dependent variable, e.g., the two-part model, the Tobit model, the sample selection model (SSM), hurdle models and finite mixture models. For an excellent comprehensive survey of this literature, see Jones (2000). Here, we only briefly review the Tobit model because it is closely related to the method that we are going to propose in this article.

The standard 'Tobit' (Tobin, 1958) regression model can be easily described using the concept of a latent desired level of expenditure, denoted by y_i^*. The classic linear regression model is then used for the latent variable:

$$y_i^* = \boldsymbol{x}_i'\boldsymbol{\beta} + \varepsilon_i, \quad \varepsilon_i \overset{iid}{\sim} N(0, \sigma^2),$$

where $\boldsymbol{x}_i$ contains predictors of interest and error ε_i follows an identically independent normal distribution with mean zero and variance σ^2. The observed expenditure is assumed to be related to the latent value by the following:

$$y_i = \begin{cases} y_i^*, & \text{if } y_i^* > 0, \\ 0, & \text{otherwise.} \end{cases}$$

Econometrically speaking, the Tobit model assumes a so-called *single decision-making* process. The individual chooses the level of medical expenditure that maximizes his or her welfare. Positive expenditures correspond to desired expenditures. Zero expenditure represents a corner solution, in which income and/or preferences for health are so low that spending nothing on healthcare is best for the individual (O'Donnell *et al.*, 2008). The regression coefficients $\boldsymbol{\beta}$ in the Tobit model are ususally estimated by maximum likelihood approach and the resulting estimates are consistent.

Unfortunately, the classic Tobit model is not appropriate for the 2007 MEPS data. First, the distribution of the expenditures is highly skewed to the right with a heavy upper tail, making the conditional mean not appropriate to summarize the relationship between the expenditures and the predictors. Using logarithmic transformations

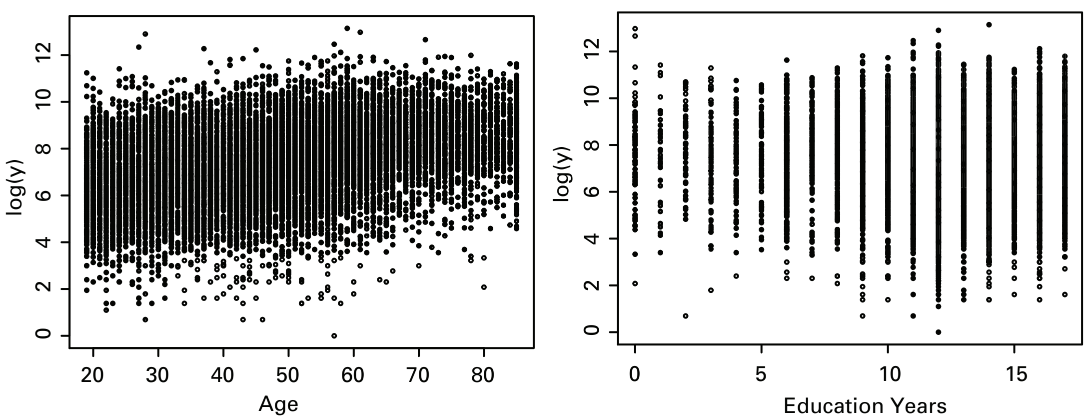

Figure 2 Violation of homoscedastic assumption: plots of logarithm of healthcare expenditure against *Age* (left) and *Education* (right)

may relieve the skewness, but this raises the problem of retransforming to the original scale (e.g., dollars rather than log-dollars), in order to make inferences that are relevant for policy (Duan, 1983). Second, the assumption of homoscedasticity is violated since the variability of errors changes across the subjects (see Figure 2). As a result, different parts of the distribution may depend on the predictors in different ways. Third, most predictors in the MEPS data are categorical variables, including both ordinal and nominal variables. It is well known that the ordinary regression models are not efficient especially when lots of such variables are used as predictors. Fourth, previous empirical research showed that the healthcare costs strongly depend on some predictors, e.g., *age*, in a nonlinear way (see Alemayehu and Warner, 2004; Jung and Tran, 2010; Bell *et al.*, 2011).

1.3 Tobit quantile regression

We adopt the idea of Tobit quantile regression, which turns out to be an ideal tool to analyse MEPS data. Again, we describe the model in the latent variable framework, where the observed response variable can then be written as $y_i = \max\{0, y_i^*\}$. Given a sample of independent observations $\boldsymbol{y} = (y_1, \ldots, y_n)$ and associated m covariates $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m)$, the latent variable y_i^* is modelled as follows:

$$y_i^* = \eta_\tau(\boldsymbol{x}_i \mid \boldsymbol{\theta}) + \varepsilon_{\tau i}, \quad \varepsilon_{\tau i} \sim F_{\tau i} \text{ subject to } F_{\tau i}(0|\boldsymbol{x}_i) = \tau, \tag{1.1}$$

where $\eta_\tau(\cdot \mid \boldsymbol{\theta})$ is a function with the parameters $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, and the random error $\varepsilon_{\tau i}$ follows a cumulative distribution function $F_{\tau i}$ whose τth quantile conditional on $\boldsymbol{x}_i$ equals zero. It is easy to see that model (1.1) defines η_τ to be the τth conditional quantile function of y_i^* given $\boldsymbol{x}_i$. The error distribution F_τ is often left unspecified in the classical literature. Assuming linear model $\eta_\tau(\boldsymbol{x}_i \mid \boldsymbol{\theta}) = \boldsymbol{x}_i'\boldsymbol{\beta}_\tau$ ($\boldsymbol{\beta}_\tau \in \mathbb{R}^p$), an intuitive estimator for the Tobit model under the above quantile restriction solves

$$\arg\min_{\boldsymbol{\beta}_\tau} \sum_{i=1}^{n} \rho_\tau(y_i - \max\{0, \boldsymbol{x}_i'\boldsymbol{\beta}_\tau\}), \quad \text{where} \quad \rho_\tau(u) = \begin{cases} u\tau, & u \geq 0, \\ u(\tau - 1), & u < 0, \end{cases} \tag{1.2}$$

is the so-called 'check function' of Koenker and Bassett (1978).

Initiated by Chib's (1992) work in standard Tobit model, Yu and Stander (2007) pioneered the Bayesian approach of Tobit quantile regression. Their method is based on assuming an asymmetric Laplace (AL) distribution for the error term $\varepsilon_{\tau i}$ in model (1.1) and using Monte Carlo Markov Chain (MCMC) technique to simulate samples from the model's posterior distribution. Estimates and their uncertainties can be easily calculated using the posterior samples. However, Yu and Stander used Metropolis-Hastings method in their MCMC algorithm rather than took advantage of the mixture representation of AL density to create a more efficient Gibbs sampler. Their approach is thus limited to the simple linear quantile functions only. Taddy and Kottas (2010) proposed a nonparametric method of Tobit quantile regression using Dirichlet process mixture models for the joint distribution of the response and the covariates.

In this paper, we extend Yu and Stander's (2007) work by developing a general Bayesian framework for flexible Tobit quantile regression models. The proposed method offers following advantages. First, it describes a distributional view of the medical expenditures dependent on the predictors by examining various conditional quantiles. Second, it takes into account for the point mass at zero of the distribution without lack of convexity problem. Third, appropriate regularization priors are taken on the categorical predictors, yielding accurate estimation and efficient variable selection. Fourth, it allows us to consider a possible nonlinear relationship between the medical expenditure and certain predictors. Fifth, it easily provides the Bayesian credible intervals for taking into account the uncertainty of estimation. This, however, would be a much harder task for the frequentists who might consider our model setting. Finally, it has an efficient MCMC algorithm to implement the Bayesian inference.

The remainder of the paper is organized as follows. In Section 2, we present the proposed method, introducing observation model and different kinds of priors. The MCMC simulation method is shown in Section 3. Results are summarized in Section 4, followed by a conclusion in Section 5.

2 Bayesian Tobit quantile regression

2.1 Observation model

In order to make Bayesian quantile inference, we need to specify a distribution on latent variable y_i^*. Following Yu and Stander (2007), we use AL distribution, denoted by AL(η, δ_0, τ), whose probability density function is given by

$$p(y^* \mid \eta, \delta_0, \tau) = \tau(1-\tau)\delta_0 \exp\left\{-\delta_0 \rho_\tau\left(y^* - \eta\right)\right\}, \tag{2.1}$$

where $\eta \in \mathbb{R}$ is a location parameter, $\delta_0 > 0$ is a scale parameter and $0 < \tau < 1$ is a skewness parameter. Since the check function ρ_τ assigns weight τ or $1-\tau$ to the observations greater than or less than η, respectively, the τth quantile of y^* is η despite of the value of δ_0. Another attractive feature about this skewed distribution is that it can be represented as a scale mixture of normals (e.g., Kotz *et al.*, 2001):

$$Y^* \stackrel{\mathcal{D}}{=} \eta + \xi W + \sigma Z \sqrt{\delta_0^{-1} W}, \quad \text{where} \quad \xi = \frac{1-2\tau}{\tau(1-\tau)}, \quad \sigma^2 = \frac{2}{\tau(1-\tau)} \tag{2.2}$$

are two scalars depending on τ. The independent random variables $W > 0$ and Z follow exponential distribution with mean δ_0^{-1} and standard normal distribution, respectively. This mixture representation makes it easy to sample from AL distribution, leading to its extensive use in Bayesian quantile regression; see, e.g., Yu and Moyeed (2001), Tsionas (2003), Kozumi and Kobayashi (2011) and Yue and Rue (2011).

We now define quantile function η_τ in model (1.1). For the general representation of categorical predictors x_j, we use dummy coding. That means, with $K_j + 1$ denoting the number of factor levels of x_j, for each x_j we have dummy variables $x_{j0}, \ldots, x_{jK_j}$, i.e., $x_{jk} = 1$ when $x_j = k$ and $x_{jk} = 0$ otherwise. We model the continuous variables, which appear to have nonlinear relationships with medical expenditures, as unknown smooth functions. As a result, a semiparametric additive model is assumed as follows:

$$\eta_\tau = \alpha + \sum_{j=1}^{p} \sum_{k=0}^{K_j} \beta_{jk} x_{jk} + \sum_{\ell=1}^{q} f_\ell(t_\ell). \tag{2.3}$$

Note that the α, β_{jk} and f_ℓ depend on the τth quantile. Since the quantiles are estimated separately in our approach, we will not show τ in those notations. For identifiability, we specify reference category $k = 0$, so that $\beta_{j0} = 0$ for all j. We also add sum-to-zero constraint to f_ℓ for all ℓ to make them identifiable from α. In matrix notation, $\boldsymbol{y}^* = (y_1^*, \ldots, y_n^*)'$ denotes the vector of latent response values; $\boldsymbol{X}_j$ is the design matrix containing observed (non-redundant) dummy variables $x_{j1}, \ldots, x_{jK_j}$; $\boldsymbol{f}_\ell$ denotes the vector of the function values and $\boldsymbol{P}_\ell$ is the corresponding incidence matrix. Letting $\boldsymbol{\beta}_j = (\beta_{j1}, \ldots, \beta_{jK_j})'$, the model has the following matrix form:

$$\boldsymbol{y}^* = \alpha \mathbf{1} + \sum_{j=1}^{p} \boldsymbol{X}_j \boldsymbol{\beta}_j + \sum_{\ell=1}^{q} \boldsymbol{P}_\ell \boldsymbol{f}_\ell + \boldsymbol{\varepsilon}_\tau, \tag{2.4}$$

where $\mathbf{1} = (1, \ldots, 1)'$, $\boldsymbol{\varepsilon}_\tau = (\varepsilon_{\tau 1}, \ldots, \varepsilon_{\tau n})'$ and $\varepsilon_{\tau i} \overset{iid}{\sim} \mathrm{AL}(0, \delta_0, \tau)$. To implement Bayesian inference, we need to specify prior distributions on scale parameter δ_0, coefficients $\boldsymbol{\beta}$ and unknown functions $\boldsymbol{f}_\ell$. Following Park and Casella (2008), we take the non-informative scale-invariant prior $p(\delta_0) \propto 1/\delta_0$ on δ_0. Moreover, a vague normal prior $N(0, \delta_\alpha^{-1})$ with small precision δ_α is assigned for α. The priors taken on $\boldsymbol{\beta}_j$ and $\boldsymbol{f}_\ell$ are described in the following sections.

2.2 Group Lasso prior for categorical covariates

In the MEPS data, most variables are categorical, including ordinal, nominal and binary factors. For example, the poverty status is given as an ordinal predictor with five levels, and census region as a nominal with four values. Usually, such data are analysed via standard linear regression modelling, with dummy coded categorial explanatory variables. In the present situation, such modelling is possible, since the number of observations ($n = 15\,890$) is quite high compared to the number of dummy variables ($\sum_{j=1}^{p} K_j = 28$). Nevertheless, from the viewpoint of interpretation, model selection is often desired with the focus on reducing model complexity.

The group Lasso (Yuan and Lin, 2006) is a modification of the original Lasso (Tibshirani, 1996) which is designed for the selection of grouped variables, as dummy coded factors. It elegantly combines penalization within groups of variables and groupwise selection by using a Lasso penalty at the factor level, and a ridge-type penalty within groups of (e.g., dummy) coefficients. For demonstration purposes only, we here consider the regularized quantile regression model (without nonlinear terms) as in Li *et al.* (2010):

$$\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} \rho_\tau(y_i^* - \boldsymbol{x}_i'\boldsymbol{\beta}) + \lambda J(\boldsymbol{\beta}), \tag{2.5}$$

with smoothing parameter λ and penalty

$$J(\boldsymbol{\beta}) = \sum_{j=1}^{p} \sqrt{\boldsymbol{\beta}_j' \boldsymbol{\Omega}_j \boldsymbol{\beta}_j}, \tag{2.6}$$

where $\boldsymbol{\Omega}_j$ is some positive definite matrix. Via the L_1-norm penalty imposed by the square root, the group Lasso encourages sparsity at the factor level. Typically, a (scaled) identity matrix is used for the penalty matrices $\boldsymbol{\Omega}_j$; see Yuan and Lin (2006) for details.

The identity matrix, which has been used for the group Lasso to date, is applicable to categorical predictors in general. Ordinal covariates, however, provide more information than nominal covariates since the labels' ordering is meaningful. In Gertheiss and Tutz (2009) and Gertheiss *et al.* (2011), a difference penalty for ordinal predictors is proposed, where the differences between coefficients of adjacent levels of predictor x_j are penalized. They showed that this penalty led to a distinct improvement in accuracy of parameter estimation and prediction over simple ridge estimation, pure dummy coding or linear regression on the group labels. However, the response is forced to change slowly between two adjacent categories of x_j. In other words, they tried to avoid high jumps and prefer smoother coefficient subvectors $\boldsymbol{\beta}_j$. Such smoothness restriction may not be desirable in practice. We thus propose an extended version of this penalty as follows: let $J(\boldsymbol{\beta}) = \sum_{j=1}^{p} J_j(\boldsymbol{\beta}_j)$ with

$$J_j(\boldsymbol{\beta}_j) = \left\{ K_j \sum_{k=1}^{K_j} v_{jk}(\beta_{jk} - \beta_{j,k-1})^2 \right\}^{1/2}, \tag{2.7}$$

with $\beta_{j0} = 0$ for all j. That means for $\boldsymbol{\Omega}_j$ in equation (2.6) we use $\boldsymbol{\Omega}_j = K_j(\boldsymbol{U}_j'\boldsymbol{V}_j\boldsymbol{U}_j)$ with

$$\boldsymbol{U}_j = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ -1 & 1 & \cdots & 0 \\ 0 & \ddots & \ddots & 0 \\ 0 & \cdots & -1 & 1 \end{pmatrix} \quad \text{and} \quad \boldsymbol{V}_j = \begin{pmatrix} v_{j1} & 0 & \cdots & 0 \\ 0 & v_{j2} & \cdots & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & \cdots & 0 & v_{jK_j} \end{pmatrix}$$

are both $K_j \times K_j$ matrices. The parameters v_{jk} allow us to locally smooth the differences between coefficients of adjacent levels. The amount of smoothing for each difference may vary within or between factors according to the data. As a result, the proposed penalty has more flexibility on smoothing levels within $\boldsymbol{\beta}_j$, e.g., the jumps are allowed if necessary. For nominal predictors, we simply let $\boldsymbol{\Omega}_j = K_j\boldsymbol{V}_j$ since no ordering information needs to be taken into account. Using local smoothing parameters, v_{jk} allows our group Lasso penalty to not only select the categorical predictors but also distinguish the levels within one predictor.

We now consider a Bayesian interpretation of model (2.5). Li *et al.* (2010) showed that the group Lasso quantile estimates can be interpreted as posterior mode estimates when the regression parameters have independent and identical Laplace priors. Motivated by this connection, we consider a fully Bayesian analysis using a conditional Laplace prior specification of the form

$$p(\boldsymbol{\beta}_j \mid \delta_0, \lambda) = C_j\sqrt{|\boldsymbol{\Omega}_j|}(\delta_0\lambda)^{K_j}\exp\left(-\delta_0\lambda\sqrt{\boldsymbol{\beta}_j'\boldsymbol{\Omega}_j\boldsymbol{\beta}_j}\right), \tag{2.8}$$

where C_j is the normalizing constant depending on K_j. Following the equality in Andrews and Mallows (1974), the prior in (2.8) can be written as

$$p(\boldsymbol{\beta}_j \mid \delta_0, \lambda) = \int_0^\infty \sqrt{\frac{\delta_0|\boldsymbol{\Omega}_j|}{2\pi s_j}}\exp\left(-\frac{\delta_0}{2s_j}\boldsymbol{\beta}_j'\boldsymbol{\Omega}_j\boldsymbol{\beta}_j\right) \times \frac{(\lambda^2/2)^{(K_j+1)/2}}{\Gamma\left(\frac{K_j+1}{2}\right)} s_j^{(K_j-1)/2}\exp\left(-\frac{\lambda^2}{2}s_j\right)ds_j. \tag{2.9}$$

Consequently, our group Lasso prior is a scale mixture of normals:

$$\boldsymbol{\beta}_j \mid s_j, v_{jk}, \delta_0 \sim N\left(\mathbf{0}, s_j\delta_0^{-1}\boldsymbol{\Omega}_j^{-1}\right), \quad s_j \mid \lambda^2 \sim \text{Gamma}\left(\frac{K_j+1}{2}, \frac{\lambda^2}{2}\right). \tag{2.10}$$

This result allows us to efficiently implement group Lasso prior in our Tobit quantile regression model as shown in Section 3.

The λ^2 is the so-called Lasso parameter, which can be chosen by, e.g., cross-validation, from a frequentist point of view (Tibshirani, 1996). For fully Bayesian inference, we need to take a prior on λ^2. The improper scale-invariant prior $1/\lambda^2$ is tempting, but it leads to an improper posterior (Park and Casella, 2008). We thus use a conjugate gamma prior on λ^2, i.e., $\lambda^2 \sim \text{Gamma}(a_\lambda, b_\lambda)$. The prior density should approach 0 sufficiently fast as $\lambda^2 \to \infty$ (to avoid mixing problems) but should be relatively flat as well. For the MEPS data, we let $a_\lambda = b_\lambda = 1$, yielding prior mean of λ^2 to be 1. Since the data information dominates in our case, the results are fairly robust to the choice of the prior for λ^2. Regarding the adaptive smoothing parameters, we take $v_{jk} \sim \text{Gamma}(0.5, 0.5)$, which is a common prior used in dynamic modelling; see, e.g., Carter and Kohn (1996).

2.3 Smoothness priors for nonlinear terms

We model timescale covariates *age* and *edu* nonparametrically, assuming their relationships with medical expenditure can be explained by some smooth functions. Prior taken on the function space is the second-order random walk (RW2) model, which is much used in basic tasks, such as smoothing data and modelling response functions, where semiparametric regression, smoothing and penalized likelihood are methods used (Green and Silverman, 1994; Fahrmeir and Lang, 2001; Fahrmeir and Tutz, 2001).

Let us consider a smooth function $f(\cdot)$, which is observed on a sequence of equally spaced locations $t_1 < t_2 < \cdots < t_m$. Denoting $f_k = f(t_k)$ for $k = 3, \ldots, m$, the RW2 model has the density

$$p(\boldsymbol{f} \mid \delta) \propto \exp\left(-\frac{\delta}{2}(f_{k-1} - 2f_k + f_{k+1})^2\right), \tag{2.11}$$

where $\boldsymbol{f} = (f_1, \ldots, f_m)'$ and δ is the precision parameter. The density is invariant under addition of $a + bk$ to x_k for any constants a and b, and is therefore improper with rank $m - 2$. The term $f_{k-1} - 2f_k + f_{k+1}$ can be interpreted as an estimate of the second-order derivative of a continuous function $f(t)$ at $t = k$. Hence, the RW2 model is appropriate for representing 'smooth curves' with small squared second derivative. Furthermore, Yue *et al.* (2011) showed that the RW2 model can actually be derived by discretizing a cubic smoothing spline estimator (Wahba, 1990).

The RW2 model (2.11) can be written in matrix notation as

$$p(\boldsymbol{f} \mid \delta) \propto \delta^{(m-2)/2} \exp\left(-\frac{\delta}{2}\boldsymbol{f}' \boldsymbol{Q} \boldsymbol{f}\right), \tag{2.12}$$

where $\boldsymbol{Q} = \boldsymbol{R}'\boldsymbol{R}$ is a semi-definite matrix and

$$\boldsymbol{R} = \begin{pmatrix} 1 & -2 & 1 & 0 & \cdots & \cdots & 0 \\ 0 & 1 & -2 & 1 & 0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & 0 & 1 & -2 & 1 & 0 \\ 0 & \cdots & \cdots & 0 & 1 & -2 & 1 \end{pmatrix}_{(m-2)\times m}.$$

The sparse structure of matrix $\boldsymbol{R}$ indicates the Markov property, allowing for fast calculations of the related full conditionals in MCMC algorithms. Since it is the density of a singular normal distribution, we write (2.12) as $N(\boldsymbol{0}, \delta^{-1}\boldsymbol{Q}^{-})$, where $\boldsymbol{Q}^{-}$ denotes the generalized inverse of $\boldsymbol{Q}$. Note that the RW2 model can be easily extended to the irregularly spaced observations (Lindgren and Rue, 2008). To estimate parameter δ, we take a diffuse but proper gamma prior for δ, i.e., $\delta \sim \text{Gamma}(a_\delta, b_\delta)$, where $a_\delta = 1$ and $b_\delta = 0.001$, e.g. It is a common prior used for the smoothing parameter in nonlinear regression models (Fahrmeir and Lang, 2001; Yue and Rue, 2011; Yue *et al.*, 2011).

The RW2 model is also a Gaussian Markov random field (GMRF). The GMRF is a quite flexible class that can be used to model, for instance, nonlinear effects, time trends, seasonal effects, interactions and spatial effects (Rue and Held, 2005). The various GMRFs share the same form as in (2.12), but with different $\boldsymbol{Q}$. As a result, a variety of effects can be taken into account by the GMRFs in the proposed Tobit quantile regression model without changing the estimation procedure as described in Section 3.

3 Posterior inference

Using identity (2.2) with model (2.4) and the priors specified in Section 2, the hierarchical structure of our Tobit quantile regression model is given by

$$y = \begin{cases} y^*, & \text{if } y^* > 0, \\ 0, & \text{if } y^* \le 0, \end{cases}$$

$$\boldsymbol{y}^* \mid \boldsymbol{\eta}_\tau, \boldsymbol{w}, \delta_0 \sim N(\boldsymbol{\eta}_\tau + \xi\boldsymbol{w}, \sigma^2\delta_0^{-1}\boldsymbol{D}_w),$$

$$w_i \mid \delta_0 \sim \text{Exp}(\delta_0), \quad \delta_0 \sim 1/\delta_0,$$

$$\boldsymbol{\eta}_\tau = \alpha\boldsymbol{1} + \sum_{j=1}^{p} \boldsymbol{X}_j\boldsymbol{\beta}_j + \sum_{\ell=1}^{q} \boldsymbol{P}_\ell\, \boldsymbol{f}_\ell,$$

$$\alpha \sim N(0, \delta_\alpha^{-1}), \;\; \boldsymbol{\beta}_j \mid s_j, v_{jk}, \delta_0 \sim N(\boldsymbol{0}, s_j\delta_0^{-1}\boldsymbol{\Omega}_j^{-1}),$$

$$s_j \mid \lambda^2 \sim \text{Gamma}\left(\frac{K_j+1}{2}, \frac{\lambda^2}{2}\right),\ \lambda^2 \sim \text{Gamma}(a_\lambda, b_\lambda),$$
$$v_{jk} \sim \text{Gamma}(0.5, 0.5),\ \ j = 1, \ldots, p,\ k = 1, \ldots, K_j,$$
$$f_\ell \sim N(\mathbf{0}, \delta_\ell^{-1} \boldsymbol{Q}_\ell^-),\ \delta_\ell \sim \text{Gamma}(a_\ell, b_\ell),\ \ell = 1, \ldots, q, \tag{3.1}$$

where $\boldsymbol{D}_w = \text{diag}(w_1, \ldots, w_n)$ and $\text{Exp}(x)$ denotes the exponential density function with mean x^{-1}.

To make Bayesian inference, we employ Gibbs sampling method to obtain the joint posterior distribution of model (3.1). More specifically, we derive the full conditional distributions and simulate samples from those distributions in turn until the Markov chain becomes stationary and enough samples are available. The algorithm is tractable and efficient, which works as follows:

1. Simulate

$$y_i^* \mid \cdot \sim y_i I(y_i > 0) + TN_{(-\infty,0]}(\eta_i + \xi w_i, \sigma^2\delta_0^{-1} w_i) I(y_i = 0),$$

   where $TN_{(a,b]}(\mu, \sigma^2)$ denotes a normal distribution with mean μ and variance σ^2 truncated on the interval $(a, b]$.
2. Simulate $w_i^{-1} \mid \cdot \sim$ Inverse Gaussian(μ', λ') for $i = 1, \ldots, n$, where

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

   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).
3. Instead to sample α and $\boldsymbol{\beta}_j$ separately, we reparameterize these parameters and sample the 'new' parameters as a block to speed up MCMC convergency. To be specific, we let $\tilde{\boldsymbol{\beta}}_j = \boldsymbol{U}_j\boldsymbol{\beta}_j$ and have the prior $\tilde{\boldsymbol{\beta}}_j \mid s_j, v_{jk}, \delta_0 \sim N(\mathbf{0}, s_j\delta_0^{-1}\tilde{\boldsymbol{\Omega}}_j^{-1})$, where $\tilde{\boldsymbol{\Omega}}_j = K_j\boldsymbol{V}_j$. Note that the transformation only applies to the $\boldsymbol{\beta}_j$ of ordinal variables. With $\tilde{\boldsymbol{X}}_j = \boldsymbol{X}_j\boldsymbol{U}_j^{-1}$, $\tilde{\boldsymbol{X}} = (\mathbf{1}, \tilde{\boldsymbol{X}}_1, \ldots, \tilde{\boldsymbol{X}}_p)$ and $\tilde{\boldsymbol{\beta}} = (\alpha, \tilde{\boldsymbol{\beta}}_1', \ldots, \tilde{\boldsymbol{\beta}}_p')'$, the model (2.3) becomes

$$\boldsymbol{\eta}_\tau = \tilde{\boldsymbol{X}}\tilde{\boldsymbol{\beta}} + \sum_{\ell=1}^{q} \boldsymbol{P}_\ell\, f_\ell,$$

and the prior on $\tilde{\boldsymbol{\beta}}$ is $N(\mathbf{0}, \delta_0^{-1}\tilde{\boldsymbol{\Omega}}^{-1})$, where

$$\tilde{\boldsymbol{\Omega}} = \operatorname{diag}\left(\frac{\delta_\alpha}{\delta_0}, \frac{\tilde{\boldsymbol{\Omega}}_1}{s_1}, \ldots, \frac{\tilde{\boldsymbol{\Omega}}_p}{s_p}\right).$$

We then simulate $\tilde{\boldsymbol{\beta}} \mid \cdot \sim N(\boldsymbol{\mu}_{\boldsymbol{\beta}}, \sigma^2\delta_0^{-1}\boldsymbol{\Sigma}_{\boldsymbol{\beta}})$, where

$$\boldsymbol{\mu}_{\boldsymbol{\beta}} = \boldsymbol{\Sigma}_{\boldsymbol{\beta}}\tilde{X}'D_w^{-1}\left(\boldsymbol{y}^* - \boldsymbol{\eta}_\tau + \tilde{X}\tilde{\boldsymbol{\beta}} - \xi\boldsymbol{w}\right), \quad \boldsymbol{\Sigma}_{\boldsymbol{\beta}} = \left(\tilde{X}'D_w^{-1}\tilde{X} + \sigma^2\tilde{\boldsymbol{\Omega}}\right)^{-1}.$$

Note that $\tilde{X}$ is a sparse matrix, and $\boldsymbol{D_w}$ and $\tilde{\boldsymbol{\Omega}}$ are diagonal matrices. Making use of those sparsity features, it is fairly efficient to sample $\tilde{\boldsymbol{\beta}}$ from its full conditional. Finally, we obtain the original dummy coefficients by back-transformation $\boldsymbol{\beta}_j = \boldsymbol{U}_j^{-1}\tilde{\boldsymbol{\beta}}_j$.

4. Simulate for $j = 1, \ldots, p$,

$$s_j^{-1} \mid \cdot \sim \text{Inverse Gaussian}\left(\sqrt{\frac{\lambda^2}{\delta_0\boldsymbol{\beta}_j'\boldsymbol{\Omega}_j\boldsymbol{\beta}_j}},\ \lambda^2\right).$$

5. Simulate for $j = 1, \ldots, p$ and $k = 1, \ldots, K_j$

$$v_{jk} \mid \cdot \sim \text{Gamma}\left(1, \frac{\delta_0 K_j}{2s_j}(\beta_{jk} - \beta_{j,k-1})^2 + \frac{1}{2}\right),$$

if $\boldsymbol{x}_j$ is an ordinal predictor and

$$v_{jk} \mid \cdot \sim \text{Gamma}\left(1, \frac{\delta_0 K_j}{2s_j}\beta_{jk}^2 + \frac{1}{2}\right),$$

if $\boldsymbol{x}_j$ is a nominal predictor.

6. Simulate

$$\lambda^2 \mid \cdot \sim \text{Gamma}\left(\frac{1}{2}\sum_{j=1}^{p} K_j - \frac{p}{2} + a_\lambda, \frac{1}{2}\sum_{j=1}^{p} s_j + b_\lambda\right).$$

7. Simulate $\boldsymbol{f}_\ell \mid \cdot \sim N(\boldsymbol{\mu}_\ell, \sigma^2\delta_0^{-1}\boldsymbol{\Sigma}_\ell)$ for $\ell = 1, \ldots, q$, where

$$\boldsymbol{\mu}_\ell = \boldsymbol{\Sigma}_\ell\boldsymbol{P}_\ell'\boldsymbol{D}_w^{-1}(\boldsymbol{y}^* - \boldsymbol{\eta}_\tau + \boldsymbol{f}_\ell - \xi\boldsymbol{w}), \quad \boldsymbol{\Sigma}_\ell = (\boldsymbol{P}_\ell'\boldsymbol{D}_w^{-1}\boldsymbol{P}_\ell + \sigma^2\phi_\ell\boldsymbol{Q}_\ell)^{-1}$$

and $\phi_\ell = \delta_\ell/\delta_0$. For identifiability, we add sum-to-zero constraint to $\boldsymbol{f}_\ell$ by computing $\boldsymbol{f}_\ell^* = \boldsymbol{f}_\ell - \boldsymbol{\Sigma}_\ell\mathbf{1}(\mathbf{1}'\boldsymbol{\Sigma}_\ell\mathbf{1})^{-1}\mathbf{1}'\boldsymbol{f}_\ell$ (see Rue and Held, 2005, Section 2.3.3). Since $\boldsymbol{\Sigma}_\ell$ is a banded matrix, we can efficiently sample $\boldsymbol{f}_\ell$ using banded Cholesky decomposition algorithm (see e.g., Rue and Held, 2005, Section 2.4).

8. Simulate for $\ell = 1, \ldots, q$,

$$\delta_\ell \mid \cdot \sim \text{Gamma}\left(a_\ell + \frac{m_\ell - 2}{2}, b_\ell + \frac{1}{2} f_\ell' \mathbf{Q}_\ell f_\ell\right).$$

9. Simulate

$$\delta_0 \mid \cdot \sim \text{Gamma}\left(\frac{3n}{2} + \frac{1}{2}\sum_{j=1}^{p} K_j,\right.$$
$$\left.\frac{1}{2\sigma^2}\sum_{i=1}^{n} w_i^{-1}(y_i^* - \eta_{\tau i} - \xi w_i)^2 + \sum_{j=1}^{p}\frac{\boldsymbol{\beta}_j'\boldsymbol{\Omega}_j\boldsymbol{\beta}_j}{2s_j} + \sum_{i=1}^{n} w_i\right).$$

Note that all the full conditionals above are the regular distributions (e.g., truncated normal and gamma distributions), and can be simulated easily using R software package. Repeat the above steps until the Markov chains converge for a few thousands iterations. The Bayesian inference can then be made based on the samples of the posterior distributions. For instance, the posterior means are often used to obtain point estimates and the posterior quantiles are used to build credible intervals to count for uncertainties.

4 Results

In the 2007 MEPS data, we limit the subjects whose age are at least 19 years old since many covariates used in our model are inapplicable for children. Participants with missing values are also excluded from the study. Finally, among 30 964 subjects in the original survey, 15 890 participants are considered in the analysis. As for explanatory variables, we use socioeconomic factors, e.g., *education*, *poverty level* and *region* and personal characteristics/conditions, e.g., *self-rated health* and *seat-belt use*. Table 1 shows a complete list of the 11 categorical variables and two continuous variables in use and their descriptions. These predictors are often selected in econometrical analysis of healthcare costs (e.g., Dominici and Zeger, 2005; Mullahy, 2009). Note that our method can generously include many categorical predictors because it performs a variable selection using the group Lasso prior.

We apply the proposed Tobit quantile regression model to the MEPS data. We here only present the results from median to upper quantiles because our primary interest lies in the empirical implications of high costs. Moreover, 17% of individuals did not have any medical expenditure, making it meaningless to consider lower quantiles. The Bayesian estimates are obtained based on 20 000 MCMC iterations with 5 000 burn-in and three thining. We implement our model in R software interface and use spam package (Furrer and Sain, 2010) to take advantage of the sparse matrices in the full conditionals. Figures 3–6 show the fitted coefficients at 0.50th, 0.75th

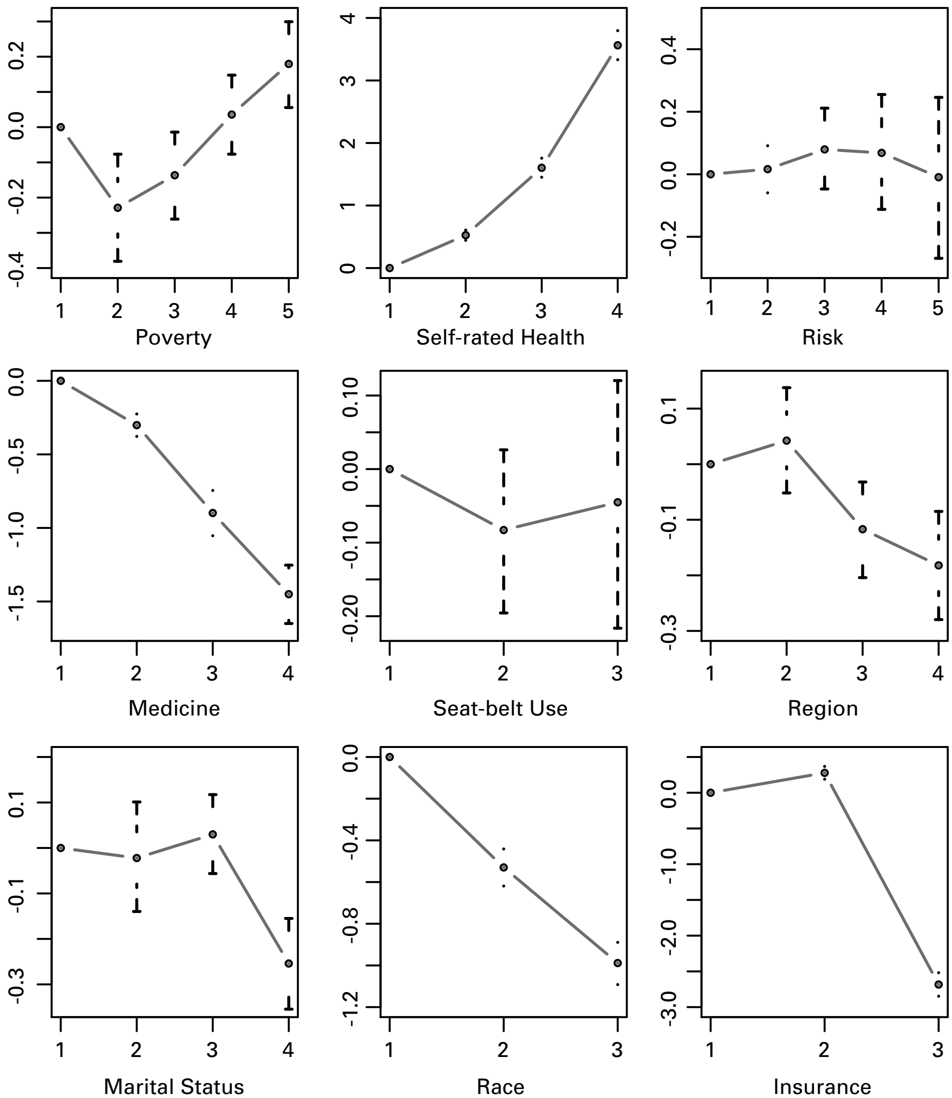

Figure 3 Dummy coefficients with 95% credible intervals at $\tau = 0.50$. First row: poverty status, self-rated health and risk attitude; second row: over illness without medicine attitude, seat-belt use and region; third row: marital status, race and insurance type (from the left to the right direction). The reference levels are summarized in Table 2.

Table 2 Background of variables in MEPS

Variable	Description	Reference level
Expenditure	$ spent (*continuous*)	
Age	Age of the respondent (*continuous*)	
Region	1 = NE, 2 = MW, 3 = S, 4 = W	NE
Education	Completed years of education (*continuous*)	
Sex	0 = male, 1 = female	male
Medicine	Over illness without medicine attitude: 1 = disagree strongly, 2 = disagree somewhat, 3 = uncertain, 4 = agree somewhat/agree strongly	disagree strongly
Insurance	1 = private, 2 = public, 3 = uninsured	private
Marital status	1 = married, 2 = widowed, 3 = separated, 4 = never married	married
Poverty	1 = poor/negative, 2 = near poor, 3 = low income, 4 = middle income, 5 = high income	poor/negative
Race	1 = non-Hispanic white, 2 = non-Hispanic black, 3 = Hispanic	non-Hispanic white
Risk	General risk taking attitude: 1 = disagree strongly, 2 = disagree somewhat, 3 = uncertain, 4 = agree somewhat, 5 = agree strongly	disagree strongly
Seat-belt use	Reported seat-belt use: 1 = always, 2 = nearly always, 3 = sometimes/seldom/never	always
Self-rated health	4-point self-rating of health status: 1 = excellent, 2 = very good, 3 = good, 4 = fair/poor	excellent
Smoking	0 = smoker, 1 = non-smoker	smoker

and 0.95th quantiles along with their 95% credible intervals for ordinal and nominal predictors. Figure 7 presents the fitted functions for continuous explanatory variables *age* and *education*, together with 95% pointwise credible intervals.

There are several interesting findings. First of all, *self-rated health*, *medicine*, *insurance*, *marital status*, *smoking* and *sex* are consistently significant through all three quantiles. More specifically, individuals with poor self-rated health status and the attitude that one should heavily depend on medicine to overcome illness clearly show higher level of medical costs than others. Also, individuals with public insurance spend a little more on healthcare than those with private insurance, and the insurance holders have much higher expenditure than non-insurance holders. Furthermore, individuals who have never been married tend to have lower medical expenditure than those with other marital status. Additionally, females and smokers are likely to spend more on medical expenditure than males and non-smokers.

Secondly, the regression coefficients of *race, poverty, region* and *risk* change significantly over the quantiles. The disparities in medical costs between (non-Hispanic) black and (non-Hispanic) white and between Hispanic and (non-Hispanic) white diminish as the quantile increases, but expenditures for Hispanics remain significantly lower than for whites and blacks throughout all quantiles. At median, higher

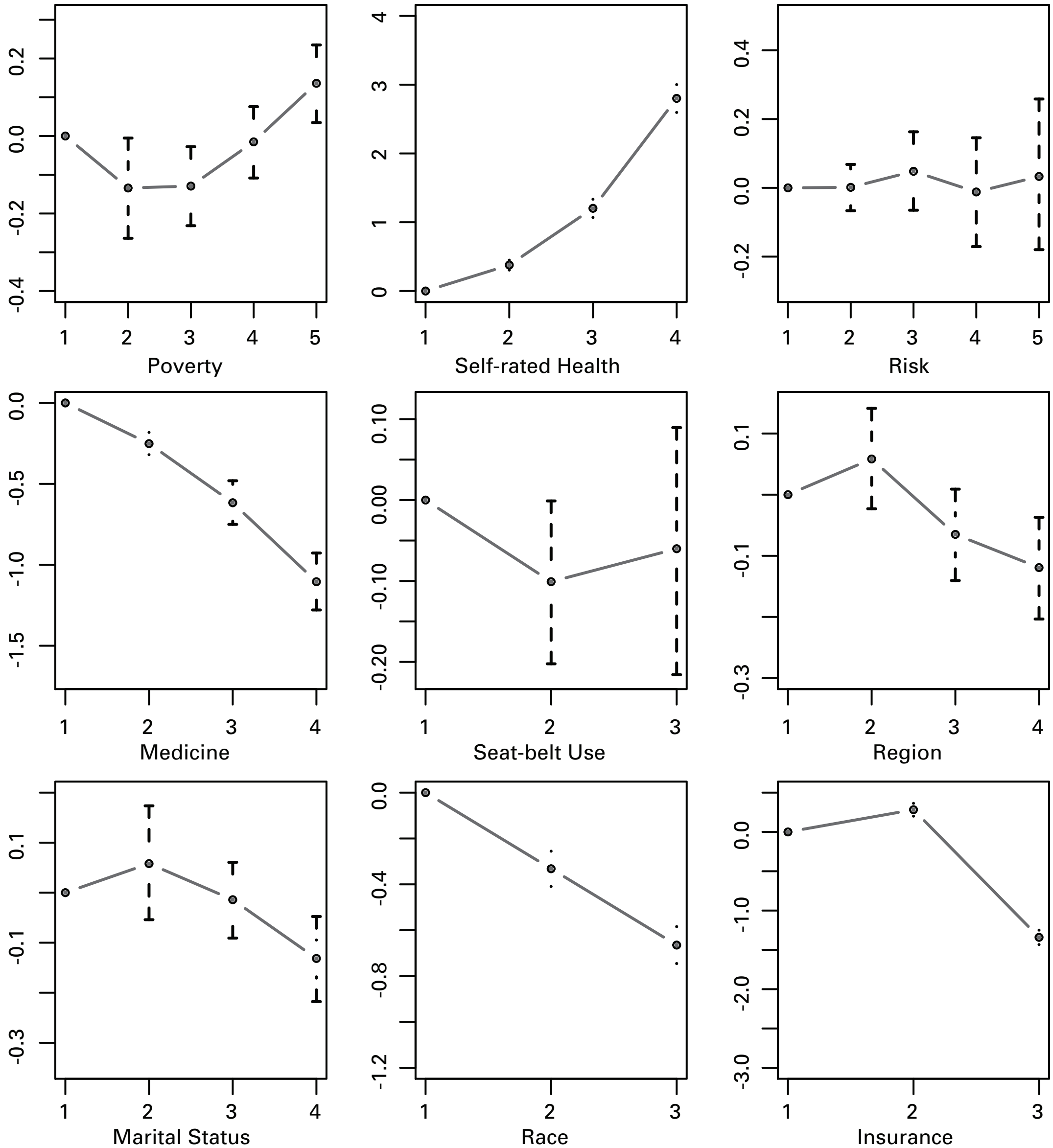

Figure 4 Dummy coefficients with 95% credible intervals at $\tau = 0.75$. First row: poverty status, self-rated health and risk attitude; second row: over illness without medicine attitude, seat-belt use and region; third row: marital status, race and insurance type (from the left to the right direction). The reference levels are summarized in Table 2.

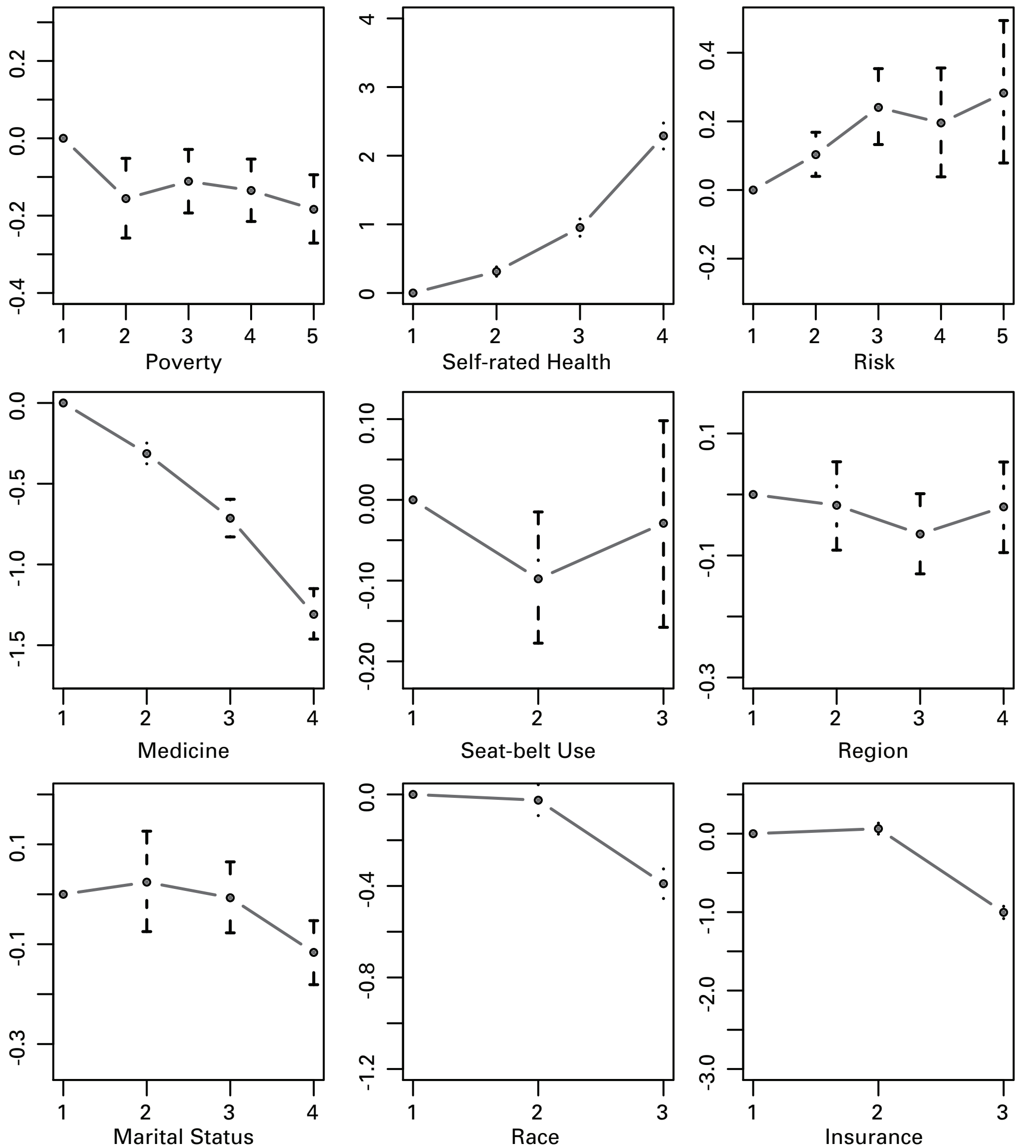

Figure 5 Dummy coefficients with 95% credible intervals at $\tau = 0.95$. First row: poverty status, self-rated health and risk attitude; second row: over illness without medicine attitude, seat-belt use and region; third row: marital status, race and insurance type (from the left to the right direction). The reference levels are summarized in Table 2.

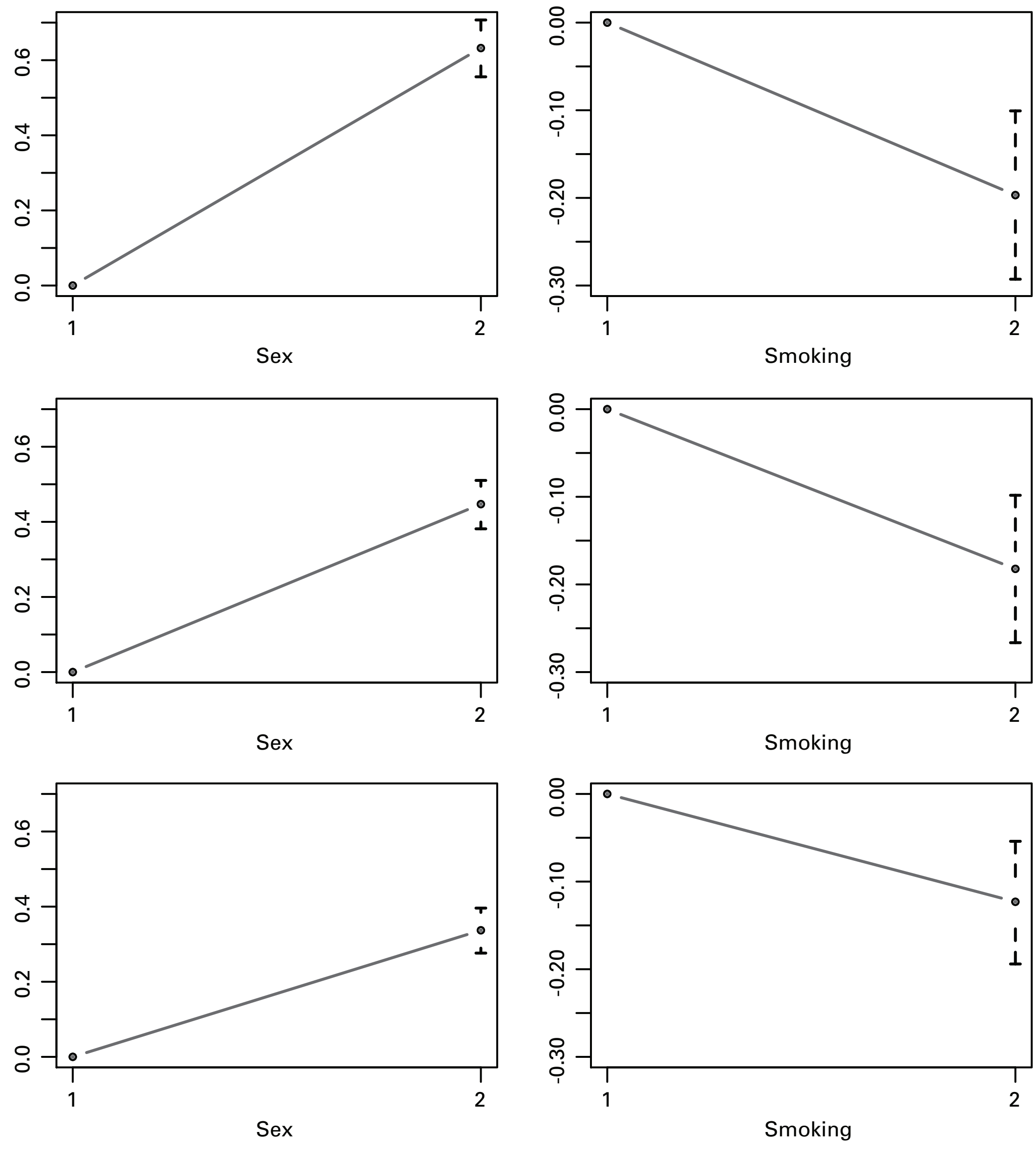

Figure 6 Dummy coefficients of gender (left) and smoking status (right) with 95% credible intervals at $\tau = 0.50$ (top), 0.75 (middle) and 0.95 (bottom). The reference levels are summarized in Table 2.

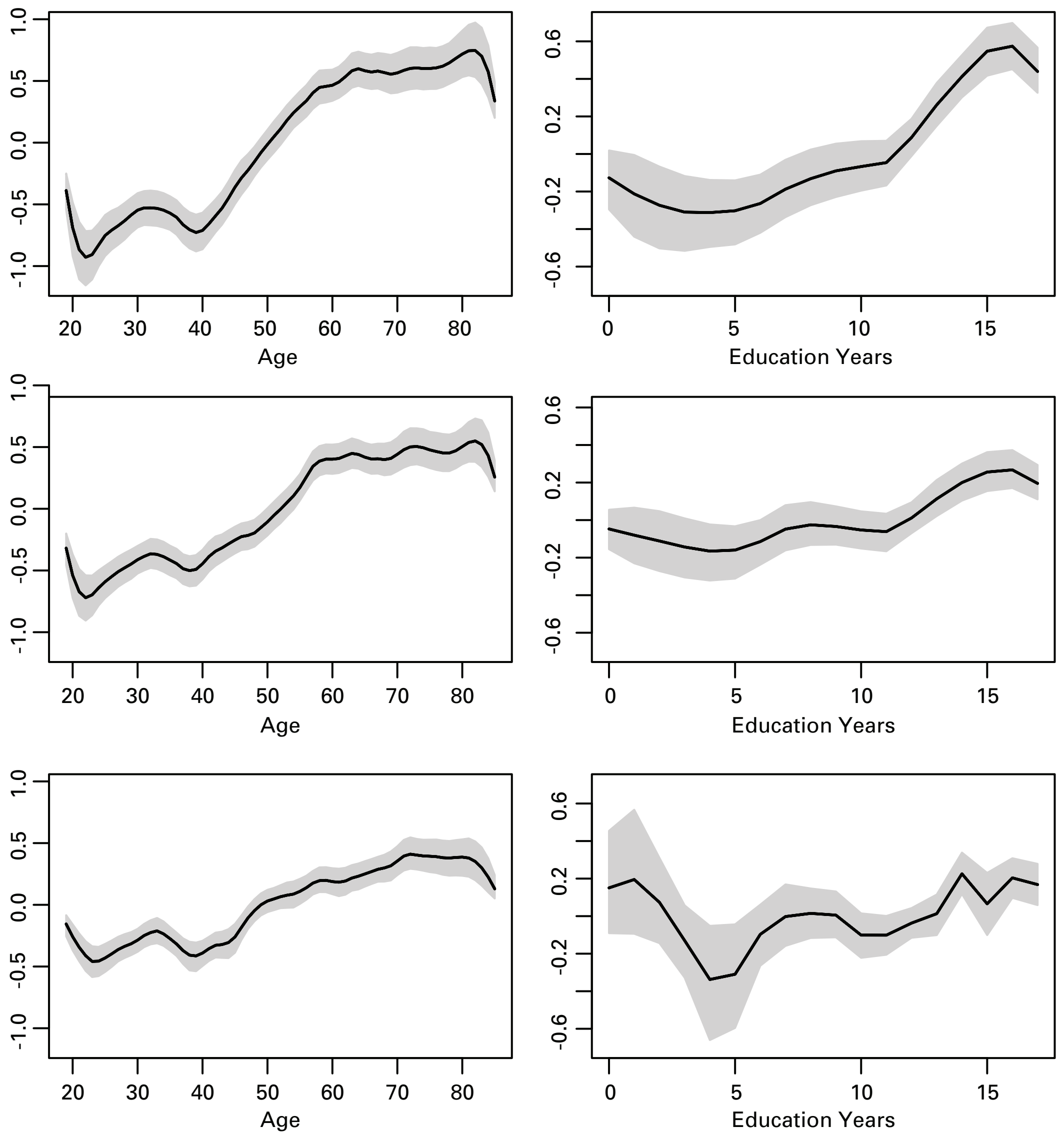

Figure 7 Quantile curves and 95% bands of age and education years: from the first to the third row: $\tau = \{0.50, 0.75, 0.95\}$.

incomes bring higher medical costs, but this income effect is not observed at higher quantiles. A slight regional effect appears for median and third quartiles: people from northeast and midwest tend to spend more on healthcare than those from south and west, but the effect disappears at 95% quantile. Unlike poverty status and region, the risk attitude is only significant at 95% quantile, so that people who go strongly against risk have lower medical costs than those who do not.

Finally, variable *age* has an apparent nonlinear relationship with the medical expenditure. The costs slightly increase from 20 to 30, followed by a little drop from 30 to 40, and then increase dramatically between 40 and 60, and finally turn flat after 60. Such pattern, however, becomes less clear as the quantiles go up, indicating that *age* might be no longer an important factor for very high expenditures. The nonlinear effect of *education* is significant at median: people who received more education, especially between 10 and 15 years, tend to spend more money on healthcare. Beyond median, the education level does not seem to be associated with the medical expenditure.

5 Conclusion

In this paper, we proposed a Bayesian Tobit quantile regression approach to analyze the MEPS data. Not only does the method accommodate the messy attributes of the medical expenditure response but also provides a complete picture of the covariate effects on the distribution of the expenditures. Moreover, it successfully selects the important categorical predictors and models the nonlinear relationships. Here is a summary of our findings. First, while confirming with earlier reports that the uninsured, poor health, and old people spend more on healthcare, the *age* seems not to be an important factor claimed for very high medical costs. Second, the low income and educational attainment make people less likely to afford the out-of-pocket costs of care, even if the costs of care is not very high. These factors have become the barriers for some groups of people in the United States to receiving healthcare services. For instance, it is not so surprising that Hispanics have obstacles to receiving the timely and appropriate services due to their low average income and education (e.g., Escarce and Kapur, 2006). This indicates that future changes in healthcare policy or system for uninsured, illegal immigrants, lower income people can create a whole new distribution of medical expenditures. Third, an attitude to overcome the illness without medicine appears to be a very important indicator of medical costs. It is even comparable to the indicator of being uninsured, which is known to discourage people greatly from receiving a quality healthcare service.

As mentioned, the Tobit model assumes a single decision-making process, which is a strong assumption. It requires that before making contact with the health services, the individual has full information on the costs of alternative courses of treatment. It also rules out the possibility that the initial decision to seek treatment is made solely by the individual, while both the patient and the doctor influence the decision about the amount of treatment. An alternative approach is to use SSM, which allows for two interdependent decisions. The decision to seek medical care and the choice of how much to spend can be influenced by distinct but correlated observable and unobservable factors. Chib *et al.* (2009) and van Hasselt (2011) have considered methods of Bayesian inference in an SSM. It would be interesting to generalize their methods to quantile regression context.

Acknowledgements

We thank two referees for their extremely helpful and insightful comments on our manuscript, which has significantly improved its quality.

References

Alemayehu B and Warner KE (2004) The lifetime distribution of health care costs. *Health Services Research*, **39**, 627–42.

Andrews DF and Mallows, CL (1974) Scale mixtures of normal distributions. *Journal of the Royal Statistical Society, Series B: Methodological*, **36**, 99–102.

Bell JF, Zimmerman FJ, Arterburn DE and Maciejewski ML (2011) Health-care expenditures of overweight and obese males and females in the medical expenditures panel survey by age cohort. *Obesity*, **19**, 228–32.

Cameron AC and Trivedi PK (2010) *Microeconometrics using stata*. Revised edition. College Station, Texas: Stata Press.

Carter CK and Kohn R (1996) Markov chain Monte Carlo in conditionally Gaussian state space models. *Biometrika*, **83**, 589–601.

Chhikara RS and Folks L (1989) *The inverse gaussian distribution: theory, methodology, and applications*. New York: Marcel Dekker.

Chib S (1992) Bayes inference in the tobit censored regression model. *Journal of Econometrics*, **51**, 79–99.

Chib S, Greenberg E and Jeliazkov I (2009) Estimation of semiparametric models in the presence of endogeneity and sample selection. *Journal of Computational and Graphical Statistics*, **18**, 321–48.

Clements MP and Hendry DF (eds) (2011) *The Oxford handbook of economic forecasting*. New York: Oxford University Press.

Dominici F and Zeger SL (2005) Smooth quantile ratio estimation with regression: estimating medical expenditures for smoking-attributable diseases. *Biostatistics*, **6**, 505–19.

Duan N (1983) Smearing estimate: A nonparametric retransformation method. *Journal of the American Statistical Association*, **78**, 605–10.

Escarce JJ and Kapur K (2006) Access to and quality of health care. In Tienda M and Mitchell F (eds), *Hispanics and the future of America*, Chapter 10. Washington, DC: National Academies Press, pp. 410–46.

Fahrmeir L and Lang S (2001) Bayesian inference for generalized additive mixed models based on Markov random field priors. *Journal of the Royal Statistical Society, Series C: Applied Statistics*, **50**, 201–20.

Fahrmeir L and Tutz G (2001) *Multivariate statistical modeling based on generalized linear models*. Berlin: Springer.

Furrer R and Sain SR (2010) spam: A sparse matrix R package with emphasis on mcmc methods for gaussian markov random fields. *Journal of Statistical Software*, **36**, 1–25.

Gertheiss J, Hogger S, Oberhauser C and Tutz G (2011) Selection of ordinally scaled independent variables with applications to international classification of functioning core sets. *Journal of the Royal Statistical Society, Series C: Applied Statistics*, **60**, 377–95.

Gertheiss J and Tutz G (2009) Penalized regression with ordinal predictors. *International Statistical Review*, **77**, 345–65.

Green PJ and Silverman BW (1994) *Nonparametric regression and generalized linear models: a roughness penalty approach*. London: Chapman & Hall Ltd.

Jones AM (2000) Health econometrics. In Culyer AJ and Newhouse JP (eds), *Handbook of health economics*, vol. 1, 1 edn, Chapter 6, pp. 265–344. New York: Elsevier.

Jung J and Tran C (2010) *Medical consumption over the life cycle: facts from a U.S. medical expenditure panel survey*. Working paper series, Department of Economics: Towson University.

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

Kotz S, Podgorski K, Kozubowski TJ and Podgórski K (2001) *The Laplace distribution and generalizations: a revisit with applications to communications, economics, engineering, and finance*. Boston: Birkhäuser.

Kozumi H and Kobayashi G (2011) Gibbs sampling methods for Bayesian quantile regression. *Journal of Statistical Computation and Simulation*, **81**, 1565–78.

Li Q, Xi R and Lin N (2010) Bayesian regularized quantile regression. *Bayesian Analysis*, **5**, 533–56.

Lindgren F and Rue H (2008) On the second-order random walk model for irregular locations. *Scandinavian Journal of Statistics*, **35**, 691–700.

Mullahy J (2009) Econometric modelling of health care costs and expenditures: a survey of analytical issues and related policy considerations. *Medical Care*, **47**. S104–S108.

O'Donnell O, van Doorslaer E, Wagstaff A and Lindelow M (2008) *Analyzing health equity using household survey data: a guide to techniques and their implementation*. Washington, D.C.: The World Bank.

Park T and Casella G (2008) The Bayesian Lasso. *Journal of the American Statistical Association*, **103**, 681–86.

Rue H and Held L (2005) *Gaussian Markov random fields: theory and applications*, volume 104 of *Monographs on Statistics and Applied Probability*. London: Chapman & Hall.

Taddy M and Kottas A (2010) A Bayesian nonparametric approach to inference for quantile regression. *Journal of Business and Economic Statistics*, **28**, 357–69.

Tibshirani R (1996) Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B: Methodological*, **58**, 267–88.

Tobin J (1958) Estimation of relationships for limited dependent variables. *Econometrica*, **26**, 24–36.

Tsionas EG (2003) Bayesian quantile inference. *Journal of Statistical Computation and Simulation*, **73**, 659–74.

van Hasselt M (2011) Bayesian inference in a sample selection model. *Journal of Econometrics*, **165**, 221–32.

Wahba G (1990) *Spline models for observational data*. Philadelphia: SIAM [Society for Industrial and Applied Mathematics].

Yu K and Moyeed RA (2001) Bayesian quantile regression. *Statistics & Probability Letters*, **54**, 437–47.

Yu K and Stander J (2007) Bayesian analysis of a Tobit quantile regression model. *Journal of Econometrics*, **137**, 260–76.

Yuan M and Lin Y (2006) Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society, Series B: Statistical Methodology*, **68**, 49–67.
Yue YR and Rue H (2011) Bayesian inference for additive mixed quantile regression models. *Computational Statistics and Data Analysis*, **55**, 84–96.
Yue YR, Speckman PL and Sun D (2011) Priors for Bayesian adaptive spline smoothing. *Annals of the Institute of Statistical Mathematics*, **64**(3), 577–613 doi: 10.1007/s10463-010-0321-6.