Simultaneous Bayesian Estimation of Multiple Quantiles with an Extension to Hierarchical Models

By Karthik Sriram*, R. V. Ramamoorthi**, Pulak Ghosh*

**Indian Institute of Management Bangalore, **Michigan State University*

April 27, 2012

*Karthik Sriram is a Phd student and Dr. Pulak Ghosh is a Professor in the department of Quantitative Methods and Information Systems at Indian Institute of Management, Bangalore, India (Email: karthik.sriram09@iimb.ernet.in, pulak.ghosh@iimb.ernet.in). Dr. R.V. Ramamoorthi is a professor of Statistics and Probability at Michigan State University, East Lansing, Michigan, USA (Email: ramamoor@stt.msu.edu). This work is part of the first author's current dissertation.

Abstract

Simultaneously modeling multiple quantiles by possibly incorporating constraints across quantiles, in particular that of monotonicity, has been an important problem. While recent attempts to address this problem focus mostly on the monotonicity issue, we take a different route using a Bayesian approach. We propose a parametric pseudo-likelihood based approach for simultaneous Bayesian estimation of multiple quantiles that is computationally simple and has the flexibility to accomodate linear as well as nonlinear model forms along with different types of prior specifications. A unique feature of our method compared to existing approaches is the posterior consistency property for the case of linear quantile regression. Further, we develop a useful extension of our method to a hierarchical setting which is applicable in particular to the normal random effects model and binary regression. We demonstrate our methods using simulations and two real life examples. The first example demonstrates an alternative way to address heteroskedasticity issues in modeling worker's compensation claims. The second example provides a novel approach to flexibly model inefficiencies of firms in a stochastic frontier analysis applied to a dataset on hospital costs.

Key words: Asymmetric Laplace distribution; Bayesian quantile regression ; Crossing ; Hierarchical models; Pseudo likelihood; Simultaneous quantile regression

1 Introduction

Quantile Regression (Koenker and Bassett 1978; Koenker 2005) is a powerful methodology for modeling any conditional quantile of the response (Y_i) given some covariates ($\mathbf{X}_i$). Knowing all quantiles is equivalent to knowing the entire conditional distribution. Practically, this can be achieved by modeling quantiles corresponding to a sufficiently dense grid of probabilities. Let $Q(\mathbf{X}_i, \tau)$ denote the τ^{th} conditional quantile of Y_i conditional on $\mathbf{X}_i$ and $0 < \tau_1 < \tau_2 < .... < \tau_K < 1$ be a grid of K probability values. A naive approach to modeling multiple quantiles would be by modeling each quantile separately. In the classical set up, this can be achieved by solving the below problem (1) separately for each $\tau \in \{\tau_1, ..., \tau_K\}$.

$$\min_{Q \in \Theta} \sum_{i=1}^{N} \rho_\tau(Y_i - Q(\mathbf{X}_i, \tau)) \tag{1}$$
$$\text{where, } \rho_\tau(u) = u(\tau - I_{(u \leq 0)})$$

Here $\rho_\tau(\cdot)$ is usually referred to as the "check function" with $I_{(\cdot)}$ being the indicator function. The above formulation is called linear quantile regression when $Q(\mathbf{X}_i, \tau) = \mathbf{X}_i^T \boldsymbol{\beta}_\tau$ for some unknown parameter vector $\boldsymbol{\beta}_\tau$. However, such an approach would not adequately address two major issues. Firstly, the desired monotonicity of $Q(\mathbf{x}, \tau)$ w.r.t τ is not automatically guaranteed and secondly it is not easy to impose constraints involving multiple quantiles (e.g. fixing a parameter across quantiles or varying it according to a specific nonlinear function etc).

In this paper, we consider the longstanding problem of simultaneously estimating multiple quantiles. While recent attempts to address this problem focus mostly on the monotonicity issue (see e.g. Tokdar and Kadane 2012; Reich et al. 2011), we take a different route using a Bayesian approach. More specifically, we propose a parametric pseudo-likelihood based on the asymmetric Laplace distribution (ALD) that can be used as a simple tool for simultaneous Bayesian quantile estimation with significant advantages over existing approaches. These advantages include (a) flexibility to accomodate linear as well as nonlinear model forms, (b) ability to accomodate different types of prior specifications, including priors which may help ensure monotonicity and (c) computational simplicity. Further, we develop an useful extension of our method to a hierarchical setting. In particular, such an extension provides a new way to flexibly model random effects in a normal regression model and the link distribution function in a binary regression model. We demonstrate the working of our methods using simulations and two interesting real life examples. The first example demonstrates an alternative way to address heteroskedasticity issues in modeling worker's compensation claims, using the data originally presented in Klugman (1992). The second example provides a powerful approach to flexibly model inefficiencies of firms in a stochastic frontier analysis and is applied on the hospital cost data previously studied by Koop et al. (1997) and Griffin and Steel (2004).

We establish two important theoretical properties of our methodology for the case of simultaneous linear quantile regression. The first property is "posterior consistency" which is a desirable property in Bayesian estimation where the posterior distribution of the parameters, for increasing sample sizes, converges to the distribution degenerate at the true parameter values. We provide sufficient conditions under which the linear quantile regression parameters are posterior consistent for the true parameter values. It is interesting to note that the posterior consistency property holds inspite of the method being based on a pseudo-likelihood that is also a misspecification of the true likelihood. To the best of our knowledge, this is the first approach to the problem with a posterior consistency justification. The second property is that the method leads to a proper posterior even under an improper flat prior on the linear quantile regression parameters and the posterior consistency property holds even under such priors.

Recently, the problem of simultaneously modeling multiple quantiles has received much attention leading to the development of both classical as well as Bayesian approaches. Classical approaches have more commonly involved construction of algorithms to ensure monotonicity in the resulting estimators (e.g. He 1997; Takeuchi and Furuhashi 2004; Wu and Liu 2009). Chernozhukov et al. (2010) propose an elegant approach to ensure monotonicity via a post-processing step based on re-arrangement of the individually estimated quantiles. Such an approach can also be used in conjunction with our method as will be discussed later. Dette and Volgushev (2008) propose a nonparametric method to directly estimate the conditional density avoiding the use of the check function.

Bayesian methods provide an interesting alternative to this problem. However, specifying a likelihood for the data is necessary for a Bayesian approach. For modeling a single τ^{th} quantile ($0 < \tau < 1$), Yu and Moyeed (2001) proposed the idea of assuming asymmetric Laplace density (ALD) for the response, *i.e.* $Y_i \sim ALD(., \mu_i^{\tau}, \sigma, \tau)$, where

$$ALD(y; \mu_i^{\tau}, \sigma, \tau) = \frac{\tau(1-\tau)}{\sigma} exp\left\{-\frac{\rho_{\tau}(y-\mu_i^{\tau})}{\sigma}\right\}, \quad -\infty < y < \infty \tag{2}$$

Then μ_i^{τ} , which happens to be the τ^{th} quantile of the above density is modeled as a function of covariates as $\mu_i^{\tau} = Q(\mathbf{X}_i, \tau)$. Based on empirical findings, they argued that the use of ALD is satisfactory even if it is a misspecification of the true underlying distribution. Recently, Sriram et al. (2012) have provided a mathematical justification for this phenomenon by showing posterior consistency of the linear quantile regression parameters under the ALD misspecification. An alternative to ALD is using Bayesian nonparametric methods by relaxing the distributional assumption (e.g. Reich et al. 2010).

Among key Bayesian approaches for simultaneously modeling quantiles, Tokdar and Kadane (2012) propose a method for Bayesian linear quantile regression using Gaussian process prior for the case of a single covariate and an approximate extension to handle multiple covariates using a single-index formulation. Reich et al. (2011) formulate the linear quantile regression parameters as Bernstein polynomials and use it's properties to derive a prior that ensures monotonicity of quantiles. Both the approaches are specifically applicable to linear quantile regression and do not extend naturally to non-linear formulations. In contrast Taddy and Kottas (2010) propose a Bayesian nonparametric approach, where they estimate the joint distribution of $(Y, \mathbf{X})$ using a Dirichlet process prior. The joint distribution then enables the estimation of conditional quantile functions. However, this purely nonparametric approach can become computationally challenging for large number of covariates and is not easily amenable to non-iid covariates. In addition, such an approach cannot accomodate other semi-parametric specifications , such as a single index model with $Q(\mathbf{X}_i, \tau) = g(\mathbf{X}_i^T \boldsymbol{\beta})$, where both $g(\cdot)$ and $\boldsymbol{\beta}$ are unknown.

One way to overcome the afore mentioned challenges is to formulate the modeling of multiple quantiles using a suitable (but perhaps misspecified) likelihood on the data. Dunson and Taylor (2005) proposed a (pseudo) "substitution likelihood" motivated by the form of the joint density of sample quantiles from a set of i.i.d observations. This likelihood requires apriori that the parameters used to model the quantiles be ordered and further due to its non-conjugate nature, the computation of the posterior becomes challenging. Another limitation of the substitution-likelihood is that a flat improper prior does not lead to a proper posterior.

In this paper, we address the above challenges by proposing a Bayesian method based on an ALD-based-pseudo-density. In section 2, we present our methodolody and it's extension. In section 3, we discuss some theoretical properties. We share results from simulations and two empirical examples in section 4 and conclude in section 5.

2 Methodology and Extension

In this section, we present our proposed methodology for simultaneously modeling multiple quantiles and it's extension to hierarchical models. We also comment on some computational issues.

2.1 Proposed Methodology

We describe a new Bayesian approach to simultaneous estimation of multiple quantiles by constructing an "ALD-based-pseudo-density" (PALD) for the response as follows.

$$\begin{aligned} &PALD(y; \mu^{\tau_1}, \mu^{\tau_2}, ..., \mu^{\tau_K}, \sigma_{\tau_1}, \sigma_{\tau_2}, ..., \sigma_{\tau_K}) \\ = &\prod_{j=1}^{K} \frac{1}{\sigma_{\tau_j}}.exp\left\{-\sum_{j=1}^{K} \frac{\rho_{\tau_j}(y-\mu^{\tau_j})}{\sigma_{\tau_j}}\right\}, \quad \text{for } y \in \mathcal{X} \\ &\text{where}, \tau_0 = 0 < \tau_1 < \tau_2 < ... < \tau_K < 1,\ \sigma_{\tau_j} > 0\ \forall\ j \end{aligned} \tag{3}$$

where $\mathcal{X}$ denotes the support of Y and $\rho_\tau(\cdot)$ is as in equation (1). The (pseudo) likelihood for modeling multiple quantiles on the data $(Y_1, \mathbf{X}_1), (Y_2, \mathbf{X}_2)$, ..., $(Y_N, \mathbf{X}_N)$

then takes the form,

$$L(Q,\boldsymbol{\sigma}|Y,\mathbf{X}) = \prod_{j=1}^{K} \frac{1}{\sigma_{\tau_j}^{N}}.exp\left\{-\sum_{i=1}^{N}\sum_{j=1}^{K}\frac{\rho_{\tau_j}(Y_i-\mu_i^{\tau_j})}{\sigma_{\tau_j}}\right\}.I_{\{Y_1,..,Y_N\in\mathcal{X}\}} \tag{4}$$

where, $\tau_0 = 0 < \tau_1 < \tau_2 < ... < \tau_K < 1$,

$$\boldsymbol{\sigma} = (\sigma_{\tau_1},...,\sigma_{\tau_K}),\ \ \sigma_{\tau_j} > 0\ \forall\ j$$

and $\quad \mu_i^{\tau_j} = Q(\mathbf{X}_i,\tau_j)$

Given a prior $\Pi(Q,\boldsymbol{\sigma})$, simultaneous estimation of quantiles is equivalent to knowing the posterior distribution of $(Q,\boldsymbol{\sigma})$ given by,

$$\begin{aligned}&\Pi(Q,\boldsymbol{\sigma}|Y,\mathbf{X})\\ &\propto \prod_{j=1}^{K}\frac{1}{\sigma_{\tau_j}^{N}}.\ exp\left\{-\sum_{i=1}^{N}\sum_{j=1}^{K}\frac{\rho_{\tau_j}(Y_i-Q(\mathbf{X}_i,\tau_j))}{\sigma_{\tau_j}}\right\}.\ \pi(Q,\boldsymbol{\sigma}).I_{\{Y_1,..,Y_N\in\mathcal{X}\}}\end{aligned} \tag{5}$$

Formulation (3) to (5) provides some significant advantages.

(a) It has the flexibility to accomodate both linear as well as nonlinear models and different types of prior specifications, in particular those that ensure monotonicity of quantiles. It can also accomodate other constraints across quantiles such as keeping a parameter fixed across quantiles or forcing a specific nonlinear functional relation between quantiles. It is particularly advantageous when monotonicity ensuring priors are hard to construct as may be the case in complex nonlinear models. In such cases, this formulation still allows for an exploratory simultaneous estimation without requiring the monotonicity condition apriori.

(b) Posterior consistency holds for the case of linear quantile regression involving covariates that are independent but possibly non-identically distributed across observations. Interestingly, this is inspite of the "pseudo" nature of the density, which is also a mis-specification of the underlying true distribution. Further, it leads to a proper posterior even under a flat improper prior, while retaining the posterior consistency property.

(c) The computational scheme is but a simple extension of the scheme for the single quantile ALD formulation.

A couple of important observations regarding the ALD-based-pseudo-density (3) are due.

First, it is clearly motivated by the ALD density (2) for the single quantile case. Although the intention is to model the τ_j^{th} quantile via the μ^{τ_j} parameter, we do not impose the restriction of $\mu^{\tau_1} < \mu^{\tau_2} < ... < \mu^{\tau_K}$ through the density. Instead, we prefer to leave the density unrestricted w.r.t these parameters and impose such restrictions through the prior, if needed. There are several reasons for doing so. (i) Computations become much simpler without these conditions in the likelihood. (ii) If our interest is in estimating fewer and sparsely spaced quantiles, monotonicity of resulting quantiles is less likely to be an issue and the trouble of restricting μ^{τ_j}'s may not even be worth the effort. (iii) Monotonicity can also be ensured through an appropriate prior on the quantile regression parameters. (iv) In the absence of suitable priors ensuring monotonicity, which is likely while dealing with complex nonlinear models, estimation using equation (3) can be easily integrated with a post-processing step such as that in the lines of Chernozhukov et al. (2010) to ensure monotonicity.

Second, it is a "pseudo" likelihood as it does not integrate to 1 w.r.t y. It is not hard to evaluate the (albeit cumbersome) normalizing constant. Equation (16) in Appendix B gives the constant when $\mathcal{X} = (-\infty, \infty)$. We do not include the normalizing constant in our methodology as it would lead to similar problems as in Dunson and Taylor (2005) (viz. non-conjugate form) making the computations difficult. Further, posterior consistency property of our methodology is more readily established without the normalizing constant.

2.2 Extension to Hierarchical Models

We propose an extension of the above methodology to hierarchical models. Let $C_{ij},\ i = 1, 2, .., s$ and $j = 1, 2, .., T_i$, be possibly non-identically distributed observations and let $V_1, ..., V_s$ be independent latent variables. Here i can be thought as an index for one out of s subjects containing a fixed and known T_i number of observations. We assume that conditional on $(V_1, ..., V_s)$, the observations C_{ij} are independent with density function

$f_{ij}(\cdot|V_i, \boldsymbol{\theta})$, i.e,

$$C_{ij}|V_i \sim f_{ij}(\cdot|V_i, \boldsymbol{\theta}) \text{ independent for all i, j} \tag{6}$$

Our goal is to flexibly model the distribution of latent variable V_i as a function of covariates $\mathbf{X}_i$ by using our simultaneous quantile modeling method. Examples of (6) include commonly used models such as,

(i) Linear Regression Model with Random Effects

$$C_{ij}|V_i \sim f_{ij}(\cdot|V_i, \boldsymbol{\theta} = (\boldsymbol{\gamma}, \nu)) = Normal(\mathbf{Z}_{ij}^T\boldsymbol{\gamma} + V_i,\ \nu^2) \tag{7}$$

Here, V_i is a random effect which is typically modeled by assuming a known distributional form. Our method is a way to relax any specific distributional assumption. A further interesting feature is that the distribution of the random effect (V_i) can be modeled as a function of other subject specific characteristics (denoted $\mathbf{X}_i$). We later provide an application of this idea to a Stochastic Frontier Analysis study.

(ii) Binary Regression Model

$$C_{ij}|V_i \sim f_{ij}(\cdot|V_i, \boldsymbol{\theta} = \boldsymbol{\gamma}) = Bernoulli(p_{ij}) \tag{8}$$

where,

$$p_{ij} = \begin{cases} 1, & \text{if } V_i \leq \mathbf{Z}_{ij}^T\boldsymbol{\gamma} \\ 0, & \text{otherwise} \end{cases}$$

We propose to model the latent variable V_i in (6) by estimating a suitably chosen grid of it's quantiles. An interesting feature of our method is that the random effect distribution can be modeled as a function of other subject specific characteristics. Suppose $\mathbf{X}_i$ is a vector of characteristics of the i^{th} subject. Then, our model includes the following additional specification for the conditional quantile function of V_i.

$$Q(\mathbf{X}_i, \tau) = \mathbf{X}_i^T\boldsymbol{\beta}_\tau \tag{9}$$

The (pseudo) likelihood for a Bayesian estimation of the above model based on our proposed method in section 2.1 would take the form,

$$
\begin{aligned}
&L(\boldsymbol{\beta}, \boldsymbol{\sigma}, \boldsymbol{\theta}|C, \mathbf{X}, \mathbf{Z}) \\
&= \int_{\mathcal{X}} \prod_{i=1}^{s} \prod_{j=1}^{T_i} f_{ij}(C_{ij}|V_i, \boldsymbol{\theta}). \prod_{i=1}^{s} g(V_i|\boldsymbol{\beta}, \boldsymbol{\sigma}) dV_1..dV_s \\
&\text{where,} \\
&g(V_i|\boldsymbol{\beta}, \boldsymbol{\sigma}) = \prod_{l=1}^{K} \frac{1}{\sigma_{\tau_l}^{s}}.exp\left\{ -\sum_{i=1}^{s} \sum_{l=1}^{K} \frac{\rho_{\tau_l}(V_i - \mathbf{X}_i^T \boldsymbol{\beta}_{\tau_l})}{\sigma_{\tau_l}} \right\} \\
&\tau_0 = 0 < \tau_1 < \tau_2 < ... < \tau_K < 1, , \ \sigma_{\tau_j} > 0 \ \forall \ j \qquad (10) \\
&\mathcal{X} \text{ is the intended support of } V_i
\end{aligned}
$$

Let $\boldsymbol{\beta} = (\boldsymbol{\beta}_{\tau_1}, ...\boldsymbol{\beta}_{\tau_K})$ and $\boldsymbol{\sigma} = (\boldsymbol{\sigma}_{\tau_1}, ...\boldsymbol{\sigma}_{\tau_K})$. Bayesian estimation is carried out by specifying a prior $\Pi(\cdot)$ on the parameter space Θ of $(\boldsymbol{\beta}, \boldsymbol{\sigma}, \boldsymbol{\theta})$. It is important to note that the intended support $\mathcal{X}$ of V_i and the prior on parameters need to be chosen carefully to avoid non-identifiability issues in the model. For example, in the stochastic frontier efficiency study that we see later in section 4.3, it makes contextual sense to constrain the random effect V_i to be supported on $(0, \infty)$ which also helps ensure identifiability of the model.

2.3 Computational Aspects

Here, we comment on some computational aspects relating to the methods described above.

- In the single quantile case, a Markov Chain Monte Carlo (MCMC) scheme can be designed by considering the mixture normal representation of ALD (see Yue and Rue 2011; Tsionas 2003). A similar method can be adapted here by writing the pseudo likelihood in (4) as follows

$$
\begin{aligned}
&L(Q|Y,X) \\
&= \prod_{j=1}^{K}\prod_{i=1}^{N}\left(\frac{1}{\epsilon_j\sqrt{\sigma_{\tau_j}W_{ij}}}\phi\left(\frac{Y_i - Q(X_i,\tau_j) - \xi_j W_{ij}}{\epsilon_j\sqrt{\sigma_{\tau_j}W_{ij}}}\right)\right) \qquad (11)\\
&\text{where ,}\\
&\phi(\cdot) \text{ is the standard normal density}\\
&W_{ij} \overset{ind}{\sim} exponential(mean = \sigma_{\tau_j}),\ \forall i,j\\
&\epsilon_j^2 = \frac{2}{\tau_j(1-\tau_j)},\ \xi_j = \frac{1-2\tau_j}{\tau_j(1-\tau_j)}
\end{aligned}
$$

- The above representation turns out to be particularly useful in the case of linear quantile regression with normal priors for the regression parameters, which leads to conjugacy and hence efficient gibbs-sampling mechanism. The details of the MCMC algorithms for the specific models we estimate in this paper are provided in Appendix C.
- Computations for the hierarchical extension of our method in section 2.2 can be particularly tricky as we are modeling quantiles of a latent variable that is not directly observed. In this case, some tweaks to the MCMC scheme may help expedite the convergence of the algorithm. For example, in the normal random effects model, the MCMC may be very slow in identifying the split of variance attributable to the normal error term and the latent variable V_i. As a result it may keep attributing most variation to the error term leaving almost no room for variation in the latent variables, which would result in very slow convergence. In such cases, we found it beneficial to raise the normal density to some power by taking $f_{ij}^{K_0}$ for some fixed $K_0 > 1$, while writing the likelihood (10). This seems to have the effect of artificially deflating the variance attributable to normal error in (7) to $\frac{\nu^2}{K_0}$ and helps the convergence.
- As mentioned in the introduction, monotonicity of quantiles using our method can be ensured in two ways. One way is to choose a prior on parameters that already incorporates the monotonicity property (e.g. Gaussian process based prior as in Tokdar and Kadane (2012) or Bernstein polynomial based prior as in Re-

ich et al. (2011)). Another perhaps computationally more attractive way is to include a post-processing step in the lines of Chernozhukov et al. (2010) as follows.

For a given value $\mathbf{X} = \mathbf{x}$ of the covariate, suppose $\hat{\mathbf{Q}}_j(\mathbf{x}) = (\hat{Q}_j(\mathbf{x}, \tau_1), ..., \hat{Q}_j(\mathbf{x}, \tau_K))$ (for $j = 1, 2, ..., M$) be samples for the quantile vector $\mathbf{Q}(\mathbf{x}) = (Q(\mathbf{x}, \tau_1), ..., Q(\mathbf{x}, \tau_K))$ from the MCMC scheme. The conditional quantile vector can then be estimated by

$$\hat{\mathbf{Q}}(\mathbf{x}) = \left(\frac{1}{M} \sum_{j=1}^{M} \hat{Q}_j(\mathbf{x}, \tau_1), ..., \frac{1}{M} \sum_{j=1}^{M} \hat{Q}_j(\mathbf{x}, \tau_K) \right)$$

Monotonized version of quantile vector (denoted by $\hat{\mathbf{Q}}_{mon}(\mathbf{x})$) is obtained by computing and inverting the transformation suggested by Chernozhukov et al. (2010) as follows.

$$\begin{aligned} H_{\mathbf{x}}(y) &= \int_0^1 I_{\left\{\frac{1}{M}\sum_{j=1}^{M} \hat{Q}_j(\mathbf{x},\tau) \le y\right\}} d\tau \\ \hat{\mathbf{Q}}_{mon}(\mathbf{x}) &= \left(H_{\mathbf{x}}^{-1}(\tau_1), ..., H_{\mathbf{x}}^{-1}(\tau_K)\right) \end{aligned}$$

Infact, the same approach can be used to monotonize other quantities of interest such as the 95% credible interval bounds, median etc. Such an approach is more preferable in exploratory studies, especially involving nonlinear models where constructing monotonicity ensuring priors may not be straight forward.

3 Theoretical Properties

In this section, we establish posterior consistency of the linear quantile regression parameters. We also show how it leads to a proper posterior even under an improper flat prior while retaining the posterior consistency property.

3.1 Posterior Consistency

Here we present the theoretical result in support of our proposed methodology for the case of linear quantile regression. To keep the notation simple, we take $\mathbf{X}_i = (X_{1i}, X_{2i})$, where X_{1i} and X_{2i} are univariate covariates. The case of more than two covariates is similar. Let $\{Y_i,\ i = 1, 2, ..., n\}$ be independent observations of a univariate response and let $\{\mathbf{X}_i,\ i = 1, 2, ..., n\}$ be 2-dimensional vectors of covariates that are independent but not necessarily identically distributed across observations. Let P_{0i} denote the true (but unknown) probability distribution of $(Y_i, \mathbf{X}_i)$, with the true τ^{th} conditional quantile given by $Q_0(\mathbf{X}_i, \tau) = \alpha_{0\tau} + \mathbf{X}_i^T \boldsymbol{\beta}_{0\tau}$. Let P be the corresponding true product measure. Suppose however that the specified model for Y_i is as in equation (4), where $Q(\mathbf{X}_i, \tau) = \alpha_\tau + \mathbf{X}_i^T \boldsymbol{\beta}_\tau$ for $\tau \in \{\tau_1, ..., \tau_K\}$. Let $\boldsymbol{\alpha} = (\alpha_{\tau_1}, ..., \alpha_{\tau_K})$ and $\boldsymbol{\beta} = (\boldsymbol{\beta}_{\tau_1}, ..., \boldsymbol{\beta}_{\tau_K})$. Similarly, the true parameter values are denoted by $\boldsymbol{\alpha}_0 = (\alpha_{0\tau_1}, ..., \alpha_{0\tau_K})$ and $\boldsymbol{\beta}_0 = (\boldsymbol{\beta}_{0\tau_1}, ..., \boldsymbol{\beta}_{0\tau_K})$. Let $\Pi(.)$ be a prior on the parameters $(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\sigma}) \in \Theta \times \Theta_{\boldsymbol{\sigma}}$, where $(\boldsymbol{\alpha}, \boldsymbol{\beta}) \in \Theta \subseteq \Re^{3K}$ and $\boldsymbol{\sigma} \in \Theta_\sigma = \Theta_{\sigma_{\tau_1}} \times ... \times \Theta_{\sigma_{\tau_K}} \subseteq (0, \infty)^K$. Expectation $E(\cdot)$ will always be w.r.t the true underlying probability P. It turns out that posterior consistency for $\boldsymbol{\sigma}$ is achieved at $\boldsymbol{\sigma}_0 = (\sigma_{0\tau_1}, ..., \sigma_{0\tau_K})$, where

$$\sigma_{0\tau} = \arg \max_{\sigma_\tau \in \Theta_{\sigma_\tau}} \left\{ \lim_{m \to \infty} \frac{1}{m} \sum_{i=1}^{m} E\left(log f_{i,\alpha_{0\tau},\beta_{0\tau},\sigma_\tau,\tau}(Y_i)\right) \right\} \tag{12}$$

It will be seen later that $\boldsymbol{\sigma}_0$ is well defined under the assumptions we make.

We would like to show that the posterior probability of any set that is a neighborhood of $(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0, \boldsymbol{\sigma}_0)$ computed using the possibly misspecified pseudo-likelihood (4) tends to 1 for large sample sizes. Our result is an extension of the corresponding result in Sriram et al. (2012) for the single quantile case. Hence, we state the assumptions and result without much elaboration and provide a sketch of the proof in the appendix by only highlighting details specific to the multiple quantile case. Let

$$f_{i,\alpha_\tau,\beta_\tau,\sigma_\tau,\tau}(Y_i) = \frac{1}{\sigma_\tau} exp\left\{ -\frac{1}{\sigma_\tau} \rho_\tau (Y_i - \alpha_\tau - \mathbf{X}_i^T \beta_\tau) \right\} \tag{13}$$

Our assumptions are given below. Assumption 1 is on the prior and assumption 2 is on the covariates. The rest of the assumptions involve the true underlying probability. Asssumption 3, in a way, ensures that the quantiles being estimated are unique. Assumption 4 is a technical condition to enable the application of Kolmogorov's strong law of large numbers for non i.i.d random variables and is also needed to ensure that $\boldsymbol{\sigma}_0$ is well defined. Assumption 5 is on the finiteness of a quantity similar to the Kullback-Liebler (KL) divergence of the specified model from the true model. It is not exactly KL since the specified model is not a true probability density. Interestingly, this assumption is not needed in the proof of the main theorem, but comes into play while extending our result to the case of improper priors.

Assumption 1: Every open neighbourhood of $(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0, \boldsymbol{\sigma}_0)$ has positive $\Pi(\cdot)$ measure. i.e. $(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0, \boldsymbol{\sigma}_0)$ is in the support of Π.

Assumption 2: $\exists\ M > 0$, such that $E|X_{1i}| \leq M$, $E|X_{2i}| \leq M$ and $EX_{1i}^2 \leq M$, $EX_{2i}^2 \leq M \quad \forall\ i \geq 1$.

Assumption 3: $\exists\ \epsilon_0 > 0$ such that, for any $\Delta > 0$ and $\tau \in \{\tau_1, ..., \tau_K\}$

$$\liminf_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} E\left\{I_{\{0<Y_i-\alpha_{0\tau}-\beta_{01\tau}X_{1i}-\beta_{02\tau}X_{2i}<\Delta\}}.I_{S_i}\right\} \ > 0$$

holds for each of the following possiblities for the set S_i, viz, $S_i = \{X_{1i} > \epsilon_0, X_{2i} > \epsilon_0\}$ or $S_i = \{X_{1i} > \epsilon_0, X_{2i} < -\epsilon_0\}$ or $S_i = \{X_{1i} < -\epsilon_0, X_{2i} > \epsilon_0\}$ or $S_i = \{X_{1i} < -\epsilon_0, X_{2i} < -\epsilon_0\}$.

Assumption 4: Both limits $\lim_{m\to\infty} \frac{1}{m} \sum_{i=1}^{m} E\left(|Z_{\tau i}|\right)$, $\lim_{m\to\infty} \frac{1}{m} \sum_{i=1}^{m} E\left(Z_{\tau i}\right)$ exist and are finite. $\sum_{i=1}^{\infty} \frac{E\left(|Z_{\tau i}|^2\right)}{i^2} < \infty$, where $Z_{\tau i} = Y_i - \alpha_{0\tau} - \beta_{01\tau}X_{1i} - \beta_{02\tau}X_{2i}$, for $\tau \in \{\tau_1, ..., \tau_K\}$.

Assumption 5:

$$E\left(\log \frac{p_{0i}(Y_i)}{\prod_{j=1}^{K} f_{i,\alpha_{0\tau_j},\beta_{0\tau_j},\sigma_{0\tau_j},\tau_j}(Y_i)}\right) < \infty, \forall\ i$$

Below, we state the main result.

Theorem 1.
Let $\Theta_{\boldsymbol{\sigma}} = [\sigma_1, \sigma_2]^K$ *with* $0 < \sigma_1 \leq \sigma_2 < \infty$. *Let* U *be an open neighborhood of* $(\boldsymbol{\alpha_0}, \boldsymbol{\beta_0}, \boldsymbol{\sigma_0})$. *Then, under assumptions* **1** *to* **4**, *we have*

$$\Pi(U^c/(Y_1, \mathbf{X}_1), (Y_2, \mathbf{X}_2), .., (Y_n, \mathbf{X}_n)) \to 0\ a.s.\ [P]$$

A sketch of the proof of the theorem is given in Appendix A.

Remark 1. *Using an argument similar to that in Sriram et al. (2012), the theorem can be extended to the case* $\Theta_\sigma = (0, \infty)^K$. *The idea of the proof is to choose a large enough compact region* $[\sigma_1, \sigma_2]^K$ *that contains* $\boldsymbol{\sigma}_0 = (\boldsymbol{\sigma}_{0\tau_1}, ..., \boldsymbol{\sigma}_{0\tau_K})$ *so that the posterior probability on it's complement goes to 0.*

3.2 Posterior Propriety and Consistency under Improper Priors

If we further make assumption **5**, the theorem above generalizes to the case when the prior Π is improper but has a formal posterior. We say that a formal posterior exists if the denominator in the posterior probability is finite, i.e.,

$$\int_{\Theta \times \Theta_\sigma} \prod_{j=1}^{K} f_{i,\alpha_{\tau_j},\beta_{\tau_j},\sigma_{\tau_j},\tau_j}(y) d\Pi(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\sigma}) < \infty, \ \forall \ y$$

Existence of a formal posterior would imply that $\Pi(\cdot|Y_1)$ is proper. Then the idea is to show that $\Pi(\cdot|Y_1)$ satisfies assumption **1**, which follows using arguments similar to that in Sriram et al. (2012). This result is particularly interesting in view of theorem 1 of Yu and Moyeed (2001) where it is shown that the posterior based on ALD is always well defined for a flat prior. Their argument can be extended even to work for our proposed pseudo density which would imply in particular that theorem 1 will hold when the prior $\Pi(\cdot)$ is flat w.r.t $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ (i.e. when $\Pi(\boldsymbol{\alpha}, \boldsymbol{\beta}|\boldsymbol{\sigma}) \propto 1$) and proper w.r.t $\boldsymbol{\sigma}$. As mentioned in the introduction, this is an advantage over the pseudo-likelihood approach of Dunson and Taylor (2005).

4 Simulations and Empirical Examples

In this section, we assess the performance of our methods through simulation of a linear, nonlinear and a hierarchical model. Using two real datasets, we further demonstrate the usefulness of our approach.

4.1 Simulations

We simulate three models. The first two models demonstrate the simultaneous quantile modeling methodology on a linear and a nonlinear model respectively. The third simulation demonstrates the working of our extension to modeling latent variables in a hierarchical set up using conditional quantiles. The models are described below. The priors used and the MCMC scheme for the models are provided in Appendix C.

Model 1: There are three types of covariates. For $\mathbf{X} = (X_1, X_2)$, where $X_1 \sim N^2(3, 1)$ (where $N^2(\mu, \sigma^2)$ denotes square of a normal random variable with mean μ and variance σ^2), X_2 is binary (1/0) with 30% values being 1, the coefficients vary by quantile and not by subject. For $Z \sim N^2(0, .5)$, the coefficient is fixed across quantiles and also across subjects, which illustrates a constraint that includes multiple quantiles. For $P \sim N^2(2, .3)$, the coefficients $\boldsymbol{\nu}$ vary by subject as well as by quantile. Let S be the vector that maps observations $\{1, 2, ..., N\}$ to the respective subject number $\{1, 2, .., s\}$.. Ofcourse, the true quantile functions completely determine the conditional distribution of the response Y which is then simulated. The true quantile function which is linear in covariates but with it's coefficients being non-linear in τ, is given by,

$$Q_0(X_i, Z_i, \tau) = \alpha_{0\tau} + \beta_{01\tau}X_{1i} + \beta_{02\tau}X_{2i} + \gamma_0 Z_i + \nu_{0\tau}[S[i]] \times P[i]$$

where,

$$\alpha_{0\tau} = .01 + .02\tau,\ \beta_{01\tau} = .03 + .04\tau + .09\tau^2,\ \beta_{02\tau} = .02 + .07\tau + .11\tau^3$$

$$\gamma_0 = .4,\ s = 10, \nu_{0\tau} = (.01\tau, .02\tau, .., .1\tau),\ N = 2000$$

$$S[1 : \lfloor .1N \rfloor] = 1, ...S[(9 \times \lfloor .1N \rfloor + 1) : N] = 10, (\lfloor q \rfloor = \text{largest integer } \leq q)$$

The model is specified as follows.

$$Q(X_i, Z_i, \tau) = \alpha_\tau + \beta_{1\tau}X_{1i} + \beta_{2\tau}X_{2i} + \gamma Z_i + \nu_\tau[S[i]] \times P[i]$$

where $\alpha_\tau, \beta_{1\tau}, \beta_{2\tau}, \gamma, \nu_\tau$ are unknown

Figure 1 shows the estimated conditional quantiles versus actual values for a given vector of covariates. We can see that the actual quantile value (green solid line)

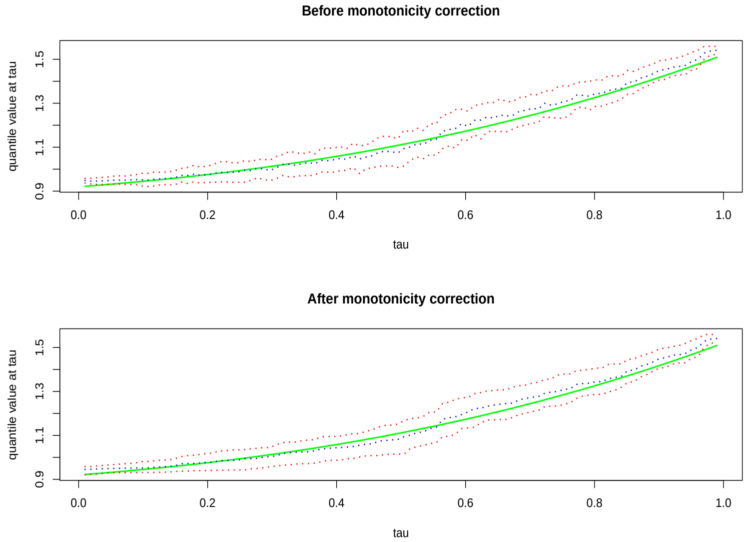

Figure 1: Simulation 1- Conditional Quantiles given fixed covariates. Green solid line is actual quantile curve and blue dotted line is the estimated using posterior mean. Red dotted lines mark the 95% credible region.

is captured by the 95% credible interval (red dotted lines). The chart in the top row is without the monotonicity constraint. The jaggedness of the dotted lines in the graph is indicative of the lack of monotonicity in the estimated quantiles, which is corrected in the graph in the second row. We use the approach described in section 2.3 for doing the monotonicity correction as a post-processing step. It is also seen that lack of monotonicity is more of a problem for closely spaced quantiles which is easily corrected in the second chart.

Model 2: This example demonstrates that incorporating a nonlinear model formulation is relatively easy with our proposed approach. The covariates (X, P) and the vector S are obtained in the same way as in model 1. Here, we consider the true quantile function which is non-linear in it's parameters.

$$Q_0(X_i, \tau) = \beta_{01\tau} X_1 + \tau \left(\beta_{02\tau} X_2 + \beta_{03} P\right)^{\nu_0[S[i]]}$$

where,

$$\beta_{01\tau} = .03 + .04\tau + .09\tau^2,\ \beta_{02\tau} = .02 + .07\tau + .11\tau^3,\ \beta_{03} = .5,$$

$$s = 4, \nu_0 = (.25, .5, 1, 2) \text{ varies by subject}$$

We specify the model by treating the nonlinear dependence as unknown using a single-index formulation as follows.

$$Q(X_i, Z_i, P_i, \tau) = \beta_{1\tau} X_{1i} + \gamma_1 Z_i + g_{\tau S[i]}(\omega_{1\tau} X_{2i} + \omega_{2\tau} P_i)$$

where $\beta_{1\tau}, \gamma_1, \omega_{1\tau}, \omega_{2\tau},\ g_{\tau j}(\cdot)$ are unknown

Following, Ruppert et al. (2003), we specify $g_{\tau i}(\cdot)$ by using a spline formulation based on a piecewise truncated polynomial of degree L as follows.

$$g_{\tau i}(x) = \nu_{\tau.i1} + \nu_{\tau.i2} x + \cdots + \nu_{\tau.i(L+1)} x^L + \sum_{d=1}^{D} \nu_{\tau.i,L+d+1} (x - \eta_{\tau d})_+^L \qquad (14)$$

where $(x)_+ = x$ if $x > 0$, and 0 otherwise, and $\eta_{\tau 1} < \eta_{\tau 2} < \cdots < \eta_{\tau D}$ are the fixed knots, which are typically placed at quantiles of the distribution of values of $V_\tau = \omega_{1\tau} X_2 + \omega_{2\tau} P$. For our example, we find it convenient to work with L=1, i.e linear splines. Note that both the function $g_{\tau j}(\cdot)$ and the parameters $(\omega_{1\tau}, \omega_{2\tau})$ in it's argument are unknown. Such a formulation is called a single-index formulation. See Hardle et al. (1993); Ichimura (1993); Yu and Ruppert (2002); Wu et al. (2010); Antoniadis et al. (2004); Wang (2009) and references therein for some applications and key developments in the analysis of single-index models. For identifiability, the intercept is included as part of the unknown function $g_{\tau j}(\cdot)$ and the condition $\omega_{1\tau}^2 + \omega_{2\tau}^2 = 1$ is imposed. We estimate ten quantiles simultaneously by taking 10 equally spaced values of τ between 0.1 and 0.9. Figure 2 shows the estimated nonlinear relationship involving the variables X_2 and P for the four different groups. Each chart within the figure corresponds to a particular group as indicated by the respective titles. The plotted curves are for a fixed value of X_2 and for a range of values for P. Different quantile curves are shown

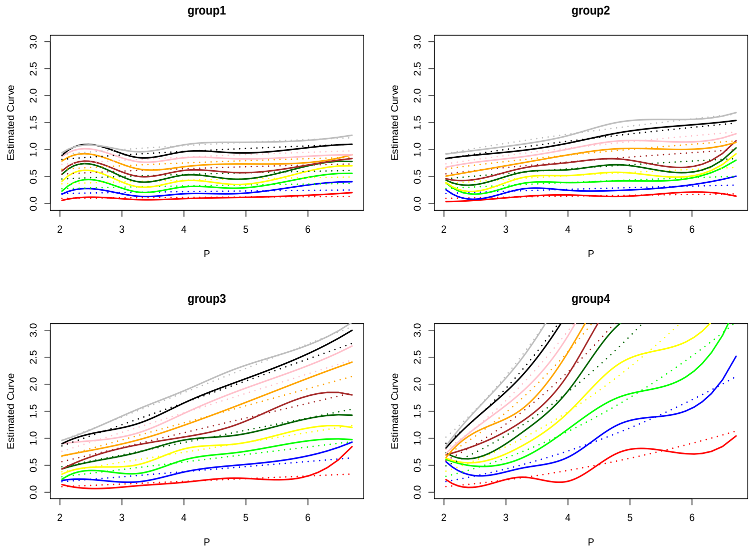

Figure 2: Simulation 2- Single Index Curve versus P for fixed X_2. Quantiles shown in different colors. Dotted lines are actual curves; Solid lines are estimated.

in different color. The lower most curve corresponds to $\tau = .1$ and the upper most $\tau = .9$. The actual curves are shown as dotted lines and the estimated curves are shown as solid lines of the same color. Although the actual curves take on different shapes across groups as well as across quantiles within groups, the estimated curves reasonably approximate them.

Model 3: Here we look at a simulated stochastic frontier model which is the type of model we study in a real application in section 4.3. The simulation is done using the model description (7) by taking $V_i \overset{iid}{\sim}$ exponential with mean 5, $\nu^2 = 6.25$ and the $\mathbf{Z}$ containing two covariates obtained as squares of N(3,2) and N(0,3) random variables respectively, with coefficients 1 and 0.5 respectively. In the model specification although the distribution of V_i is treated as unknown, the support is assumed to be $(0, \infty)$. Figure 3 shows that actual versus modeled density for the simulated efficiency defined by e^{-V_i}. The estimated pdf is close to the actual pdf which assumes an exponential distribution for V_i, thus supporting the extension of our simultaneous quantile modeling method to a hierarchical setting as described in section 2.2.

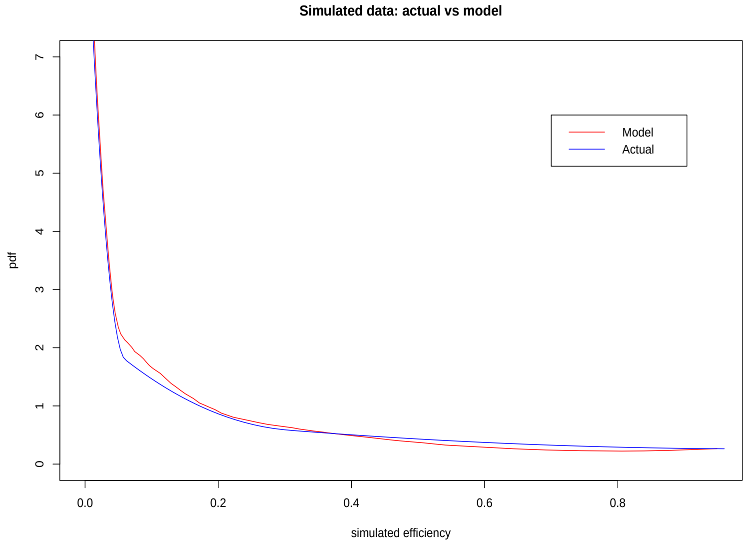

Figure 3: Simulation 3- Modeling Random Effects. Red solid line denotes the actual pdf based on exponentially distributed random effects and blue solid line is based on the hierarchical extension of our simultaneous quantile modeling approach.

4.2 Empirical Example 1

Here we use data on workers compensation claims to demonstrate an application of our methodology developed in section 2.1. This data is originally from Klugman (1992) and has been analyzed by Frees et al. (2001), where a more detailed description is available. The data contains information on workers compensation claims(or losses) for 121 risk groups over 7 years. The goal is to model the loss per dollar of payroll, also refered to as pure premium (PP). A challenge is the heteroskedasticity in PP that seems to be related to the amount of exposure measured by dollars of payroll (E). Frees et al. (2001) consider different model formulations and conclude that the following model which assumes error variance as being proportional to the exposure is the most reasonable in terms of simplicity and performance. They take

$$PP_{it} = \alpha_{1i} + \beta_1 + \epsilon_{it}\sqrt{E_{it}}$$

where i denotes the risk group, t denotes time and dependencies between observations from the same firm are modeled through a random effect α_{1i}. We show how simultaneous modeling of quantiles can be another way to address the issue of heteroskedasticity. Our model formulation involves modeling different conditional quantiles of the response PP as a function of the covariate E. More formally, we take

$$Q(E_{it}, \tau_k) = \alpha_{1i} + g_{\tau_k}(log(E_{it})),\ \ g(\cdot) \text{ unknown}$$

where the function $g(\cdot)$ is treated as unknown and modeled using a spline as in equation (14). Two important aspects in our formulation demonstrate the flexibility and strength of our approach. First, the random effect that captures dependencies between observations for the same firm, is held fixed across quantiles, but varies by firm. Such a constraint on the random effect involves multiple quantiles and is easily handled by our simultaneous quantile approach. Second, the function $g(\cdot)$ is modeled nonlinearly using splines and varied across quantiles. Thus different quantile curves can be of different shapes, which is also conveniently modeled through our approach. Knowing all conditional quantiles at every exposure essentially amounts to knowing the entire conditional distribution of PP and hence is a natural way to account for heteroskedasticity. The MCMC scheme used here is same as given for model 2 in Appendix C. Figure 4 shows the pure premium (black dots) plotted against log(exposure). The curves drawn in different colors are the quantile curves with the lowest curve corresponding to 1^{st} percentile and the highest corresponding to the 99^{th} percentile. The quantiles as marked by the curves capture the underlying heteroskedasticity in the data. For example, the quantile curves at the lower and higher values of exposure are more closely spaced versus those at the middle that are more separated, thus indicating larger variance in the middle than extremes. At any fixed value of exposure that can be marked on the x axis, the points on the different quantile curves help approximate the entire conditional distribution of pure premium.

4.3 Empirical Example 2

In this example, we show how the hierarchical extension presented in section 2.2 can be used to flexibly model inefficiencies of firms in Stochastic Frontier Analysis. In particular, we use the data on hospital costs which has been previously studied by Koop et al. (1997) and Griffin and Steel (2004). The data contains information on costs, inputs and outputs of 382 relatively homogeneous sample of non-teaching hospitals over 1987-1991. Estimating the cost frontier involves modeling log(cost) as a function of five different outputs (viz. Y_1 =number of cases, Y_2 =number of inpatient days, Y_3 =number of beds, Y_4 =number of outpatient visits and Y_5 =a case mix index), an aggregate wage index (P) as a measure of input, a variable to measure the capital stock or total fixed assets(K) and a linear as well as quadratic time trend variable to capture time dynamics. The cost frontier model is then formulated as

$$\log(cost_{it}) = Z_{it}^T \gamma + V_i + e_{it}$$

where, $e_{it} \overset{iid}{\sim} N(0, \nu^2)$

V_i is the firm specific inefficiency term

$$\begin{aligned} &Z^T\gamma \\ &= \sum_{j=1}^{5} \gamma_i \log Y_j + \gamma_6 \log P + \gamma_7 \log P^2 + \sum_{j=1}^{5} \gamma_{7+j} \log Y_j \log P + \gamma_{13} \log K \\ &+ \sum_{j=1}^{5} \gamma_{13+j} \log Y_j \log K + \gamma_{19} \log(P) \log K + \gamma_{20} (\log K)^2 + \gamma_{21} t + \gamma_{22} t^2 \\ &+ \sum_{j=1}^{5} \sum_{l=i}^{5} \gamma_{22+5(j-1)+l} \log Y_j \log Y_l \end{aligned}$$

Koop et al. (1997) modeled the inefficiency term V_i by treating it as a random effect with a known parameteric distributional form supported on $(0, \infty)$. Griffin and Steel (2004) relaxed this assumption by treating the distribution as completely unknown and modeled it using a Dirichlet process prior (Ferguson 1973). Further, they varied the inefficiencies by firm characteristics by dividing the data into different segments, each with a separate Dirichlet process prior. We propose a novel alternative approach to flexibly model the inefficiencies using the hierarchical extension of our simultaneous quantile modeling approach. We use the same model as above except that we formulate the model for inefficiency term V_i as a function of firm characteristics (X_i) using quantiles as follows.

$$Q(\mathbf{X}_i, \tau) = \mathbf{X}_i^T \boldsymbol{\beta}(\tau)$$

A particular advantage of our approach apart from computational simplicity is that it can easily incorporate firm characteristics that are continuous without having to group or segment the population. In our particular example however, the firm characteristics happen to be binary variables indicating whether the firm is "for profit", "non profit" or "government run". Consistent with the afore mentioned papers, we define efficiency as $exp(-V_i)$. The MCMC scheme used here is same as given for model 3 in Appendix C. Figure 5 shows the estimated pdf of efficiency for the three types of firms in the hospital data. Consistent with the results in Griffin and Steel (2004), the modes of the distributions are around 0.7. Also, we see that profit firms show more variation than non profit firms and the distribution of government run firms is supported more on the higher values of efficiency compared to the other two firm types. Although the modes of the efficiency distributions for the three groups are not exactly the same as that in Griffin and Steel (2004), our analysis is in agreement with their observation that the non-profit and government run hospitals have modes that are close to each other and higher than that of for-profit group. Notwithstanding some minor differences from prior studies, this example does help demonstrate that our method is a novel approach to flexibly model random effects in regression models and in particular can be used in stochastic frontier efficiency studies.

5 Conclusion

The problem of simultaneously modeling multiple quantiles is a topic of active research. We propose a new Bayesian approach to address this issue and develop a novel extension to hierarchical models. Theoretically, we derive the posterior consistency property of our method. We further illustrate it's usefulness through simulations and empirical examples. There can be several applications of our methodology to other econometric problems particularly those involving flexible modeling of distributions.

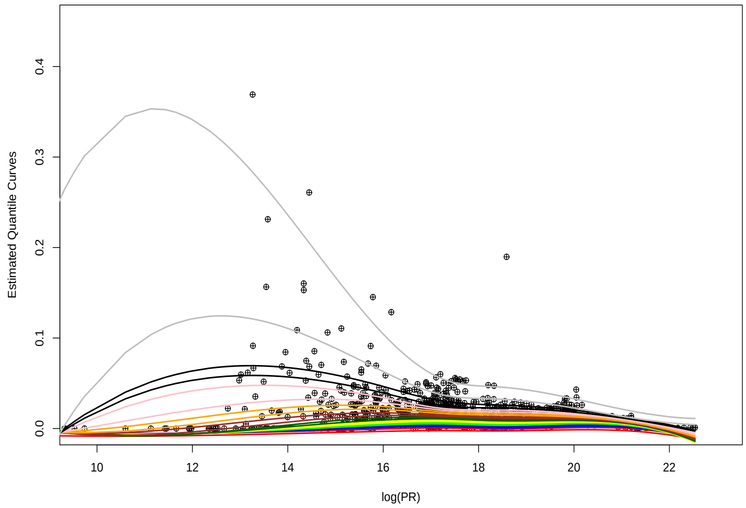

Figure 4: Workers Compensation Claims: Conditional Quantile Curves

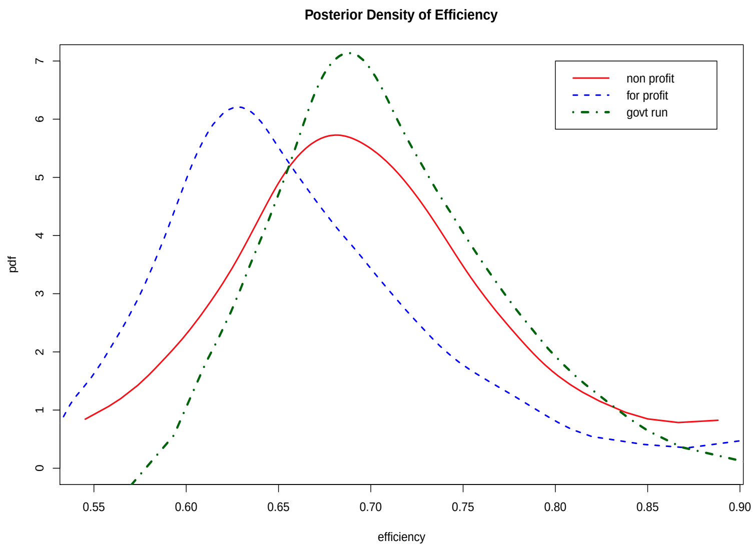

Figure 5: Efficiency: Hospital Cost Data. The pdf of efficiencies estimated for three different types of firms (viz. non-profit, profit and government run) in the hospital cost data

Appendix A Proof Sketch of Theorem 1

The proof of our theorem 1 is an extension of the single quantile case in Sriram et al. (2012). The key is in adapting the lemmas and propositions in their result to accomodate multiple quantiles. Hence, we sketch the proof by highlighting only those areas requiring special attention while accomodate multiple quantiles. The result is established by writing

$$\begin{aligned}
&\Pi\left(U^c|(Y_1,\mathbf{X}_1),(Y_2,\mathbf{X}_2),...,(Y_n,\mathbf{X}_n)\right)\\
&= \frac{\int_{U^c}\prod_{i=1}^{n}\prod_{j=1}^{K}\frac{f_{(i,\alpha_{\tau_j},\boldsymbol{\beta}_{\tau_j},\sigma_{\tau_j},\tau_j)}(Y_i)}{f_{(i,\alpha_{0\tau_j},\boldsymbol{\beta}_{0\tau_j},\sigma_{0\tau_j},\tau_j)}(Y_i)}d\Pi(\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\sigma})}{\int_{\Theta\times\Theta_\sigma}\prod_{i=1}^{n}\prod_{j=1}^{K}\frac{f_{(i,\alpha_{\tau_j},\boldsymbol{\beta}_{\tau_j},\sigma_{\tau_j},\tau_j)}(Y_i)}{f_{(i,\alpha_{0\tau_j},\boldsymbol{\beta}_{0\tau_j},\sigma_{0\tau_j},\tau_j)}(Y_i)}d\Pi(\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\sigma})} = \frac{I_{1n}}{I_{2n}}
\end{aligned} \tag{15}$$

and showing that $\exists\ d_0 > 0$ such that $e^{nd_0}I_{1n} \rightarrow\ 0\ a.s.[P]$ and $\forall\ d > 0\ e^{nd}I_{2n} \rightarrow \infty\ a.s.[P]$

Lemma 1 gives some basic equalities and inequalities for the log ratio of ALD likelihoods, that are useful for the proof. Lemma 2, proposition 2 and lemma 3 help show that $\forall\ d > 0\ e^{nd}I_{2n} \rightarrow\ \infty\ a.s.[P]$, which takes care of the denominator in the posterior probability. Proposition 1 is an interesting observation. Typically, in misspecified models posterior consistency holds at parameter values that minimize the Kullback-Liebler(KL) divergence of the specified model from the true model. However, here the specified model is a pseudo likelihood and hence the expectation in proposition 1 given by $E\left\{log\left(\frac{p_{0i}(Y_i)}{f_{(i,\alpha_\tau,\beta_\tau,\sigma_\tau,\tau)}(Y_i)}\right)\right\}$ is not exactly a KL divergence. Nevertheless, the property that $(\alpha_{0\tau},\beta_{0\tau})$ minimize this "KL-divergence-like" expectation is useful to establish our result. Similarly, in proposition 2, the set V_δ is similar to but not exactly a KL-neighborhood.

Lemma 1. *The following identities and inequalities hold true.*

(a) $log\left(\frac{f_{(i,\alpha_\tau,\beta_\tau,\sigma_\tau,\tau)}(Y_i)}{f_{(i,\alpha_{0\tau},\beta_{0\tau},\sigma_\tau,\tau)}(Y_i)}\right)$

$$= \frac{1}{\sigma_\tau}.\begin{cases} -b_{i\tau}(1-\tau), & if\ Y_i \leq \min(\alpha_\tau + \mathbf{X}_i^T\beta_\tau, \alpha_{0\tau} + \mathbf{X}_i^T\beta_{0\tau}) \\ (Y_i - \alpha_{0\tau} - \mathbf{X}_i^T\beta_{0\tau}) - b_{i\tau}(1-\tau), & if\ \alpha_{0\tau} + \mathbf{X}_i^T\beta_{0\tau} < Y_i \leq \alpha_\tau + \mathbf{X}_i^T\beta_\tau \\ b_{i\tau}\tau - (Y_i - \alpha_{0\tau} - \mathbf{X}_i^T\beta_{0\tau}), & if\ \alpha_\tau + \mathbf{X}_i^T\beta_\tau < Y_i \leq \alpha_{0\tau} + \mathbf{X}_i^T\beta_{0\tau} \\ b_{i\tau}\tau, & if\ Y_i \geq \max(\alpha_\tau + \mathbf{X}_i^T\beta_\tau, \alpha_{0\tau} + \mathbf{X}_i^T\beta_{0\tau}) \end{cases}$$

(b) $\left|log\left(\frac{f_{(i,\alpha_\tau,\beta_\tau,\sigma_\tau,\tau)}(Y_i)}{f_{(i,\alpha_{0\tau},\beta_{0\tau},\sigma_\tau,\tau)}(Y_i)}\right)\right| \leq \frac{\max(\tau,1-\tau)}{\sigma_\tau}(|\alpha_\tau - \alpha_{0\tau}| + |\beta_{1\tau} - \beta_{10\tau}||X_{1i}| + |\beta_{2\tau} - \beta_{20\tau}||X_{2i}|)$

(c) $log\left(\frac{f_{(i,\alpha_\tau,\beta_\tau,\sigma_\tau,\tau)}(Y_i)}{f_{(i,\alpha_{0\tau},\beta_{0\tau},\sigma_\tau,\tau)}(Y_i)}\right) \leq |Y_i - \alpha_{0\tau} - \mathbf{X}_i^T\beta_{0\tau}|/\sigma_\tau$

(d) *If* $E|X_i| <= M$ *then* $E\left\{log\left(\frac{f_{(i,\alpha_\tau,\beta_\tau,\sigma_\tau,\tau)}(Y_i)}{f_{(i,\alpha_{0\tau},\beta_{0\tau},\sigma_\tau,\tau)}(Y_i)}\right)\right\} \leq \max(\tau, 1-\tau).(|\alpha_\tau - \alpha_{0\tau}| + |\beta_{1\tau} - \beta_{10\tau}|M + |\beta_{2\tau} - \beta_{20\tau}|M)/\sigma_\tau$

(e) $log\left(\frac{f_{(i,\alpha_\tau,\beta_\tau,\sigma_\tau,\tau)}(Y_i)}{f_{(i,\alpha_{0\tau},\beta_{0\tau},\sigma_\tau,\tau)}(Y_i)}\right) = \frac{1}{\sigma}.\begin{cases} -b_{i\tau}(1-\tau) + \min(Z_{i\tau}^+, b_{i\tau}), & if\ b_{i\tau} > 0 \\ b_{i\tau}\tau + \min(Z_{i\tau}^-, -b_{i\tau}), & if\ b_{i\tau} \leq 0 \end{cases}$

(f) $\left|log\left(\frac{f_{(i,\alpha_\tau,\beta_\tau,\sigma_\tau,\tau)}(Y_i)}{f_{(i,\alpha_{0\tau},\beta_{0\tau},\sigma_{0\tau},\tau)}(Y_i)}\right)\right| \leq |log(\sigma_\tau) - log(\sigma_{0\tau})| + |Z_{i\tau}|.\left|\frac{1}{\sigma_\tau} - \frac{1}{\sigma_{0\tau}}\right|$

where $b_{i\tau} = (\alpha_\tau - \alpha_{0\tau}) + \mathbf{X}_i^T(\beta_\tau - \beta_{0\tau})$, $Z_{i\tau} = Y_i - \alpha_{0\tau} - \mathbf{X}_i^T\beta_{0\tau}$, $Z_{i\tau}^+ = \max(Z_{i\tau}, 0)$ *and* $Z_{i\tau}^- = \max(-Z_{i\tau}, 0)$.

Proof. The lemma follows easily with a bit of algebra and hence the proof is omitted. □

Lemma 2. *The following identities and inequalities hold true.*

(a) $E\left\{log\left(\frac{f_{(i,\alpha_\tau,\beta_\tau,\sigma_\tau,\tau)}(Y_i)}{f_{(i,\alpha_{0\tau},\beta_{0\tau},\sigma_\tau,\tau)}(Y_i)}\right)\right\}$
$= E\left\{\frac{(Y_i - \alpha_\tau - \mathbf{X}_i^T\beta_\tau)}{\sigma_\tau}.I_{(\alpha_{0\tau} + \mathbf{X}_i^T\beta_{0\tau} < Y < \alpha_\tau + \mathbf{X}_i^T\beta_\tau)}\right\}$
$+ E\left\{\frac{(\alpha_\tau + \mathbf{X}_i^T\beta_\tau - Y_i)}{\sigma_\tau}.I_{(\alpha_\tau + \mathbf{X}_i^T\beta_\tau < Y_i < \alpha_{0\tau} + \mathbf{X}_i^T\beta_{0\tau})}\right\}$

(b) $E\left\{log\left(\frac{f_{(i,\alpha_\tau,\beta_\tau,\sigma_\tau)}(Y_i)}{f_{(i,\alpha_{0\tau},\beta_{0\tau},\sigma_\tau)}(Y_i)}\right)\right\} \leq 0$

Further, in (b) equality is achieved if $\alpha = \alpha_{0\tau}$ *and* $\beta = \beta_{0\tau}$.

Proof. The lemma follows by a little algebra and by noting that $(\alpha_{0\tau} + \mathbf{X}_i^T \beta_{0\tau})$ is the true τ^{th} quantile of Y_i given $\mathbf{X}_i$. $\square$

Proposition 1. *If assumption* **5** *holds then,*

$$\inf_{(\alpha,\beta)\in\Theta,\ \sigma_\tau\in\Theta_{\sigma_\tau}} E\left\{log\left(\frac{p_{0i}(Y_i)}{f_{(i,\alpha_\tau,\beta_\tau,\sigma_\tau,\tau)}(Y_i)}\right)\right\} \geq E\left\{log\left(\frac{p_{0i}(Y_i)}{f_{(i,\alpha_{0\tau},\beta_{0\tau},\sigma_{0\tau i},\tau)}(Y_i)}\right)\right\}$$
$$where\ \sigma_{0\tau i} = arg \max_{\sigma_\tau\in\Theta_{\sigma_\tau}} E\left\{f_{(i,\alpha_{0\tau},\beta_{0\tau},\sigma_\tau,\tau)}(Y_i)\right\}$$

Proof. This proposition is a consequence of lemma 2. $\square$

Proposition 2. *Suppose* $\Theta_\sigma = [\sigma_1, \sigma_2]$ *such that* $0 < \sigma_1 \leq \sigma_2 < \infty$ *and* $\forall\ \delta > 0,\ \Pi(V_\delta) > 0$ *, where,*

$$V_\delta = \left\{(\alpha,\beta,\sigma) \in \Theta\times\Theta_\sigma : \limsup_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} E\left[\sum_{j=1}^{K}\left\{log\left(\frac{f_{(i,\alpha_{0\tau_j},\beta_{0\tau_j},\sigma_{0\tau_j},\tau_j)}(Y_i)}{f_{(i,\alpha_{\tau_j},\beta_{\tau_j},\sigma_{\tau_j},\tau_j)}(Y_i)}\right)\right\}\right] < \delta\right\}$$
$$\bigcap \left\{(\alpha,\beta,\sigma) \in \Theta\times\Theta_\sigma : \sum_{i=1}^{\infty} \frac{1}{i^2} E\left[\left\{\sum_{j=1}^{K}\left\{log\left(\frac{f_{(i,\alpha_{0\tau_j},\beta_{0\tau_j},\sigma_{0\tau_j},\tau_j)}(Y_i)}{f_{(i,\alpha_{\tau_j},\beta_{\tau_j},\sigma_{\tau_j},\tau_j)}(Y_i)}\right)\right\}\right\}^2\right] < \infty\right\}$$

then $\forall\ d > 0,\quad e^{nd} I_{2n} \to \infty\ a.s\ [P]$

Proof. The proof of the proposition is in the same lines as theorem 4.4.1 of Ghosh and Ramamoorthi (2003). $\square$

Lemma 3. *Suppose* $\Theta_\sigma = [\sigma_1, \sigma_2]$ *with* $0 < \sigma_1 \leq \sigma_2 < \infty$ *. If assumptions* **1** *and* **2** *hold, then* $\forall\ d > 0,\quad e^{nd} I_{2n} \to \infty\ a.s\ [P]$

Proof. The idea is to verify that conditions of proposition 2 are satisfied. This follows by using parts (b), (d) and (f) of lemma 1 along with assumptions **1** and **2**. $\square$

The above lemma helps take care of the denominator in the posterior probability. In order to handle the numerator, without loss of generality consider neighborhood around true parameter of the form $U = W \times V$ where,

$$\begin{aligned}
W &= \{|(\alpha_{\tau_1} - \alpha_{0\tau_1}| < \Delta_{1\tau_1}, |\beta_{1\tau_1} - \beta_{01\tau_1}| < \Delta_{2\tau_1}, |\beta_{2\tau_1} - \beta_{02\tau_1}| < \Delta_{3\tau_1}) \\
&\quad |(\alpha_{\tau_2} - \alpha_{0\tau_2}| < \Delta_{1\tau_2}, |\beta_{1\tau_2} - \beta_{01\tau_2}| < \Delta_{2\tau_2}, |\beta_{2\tau_3} - \beta_{02\tau_3}| < \Delta_{3\tau_2})\} \\
V &= \{|\sigma_{\tau_1} - \sigma_{0\tau_1}| < \Delta_{4\tau_1}, ..., |\sigma_{\tau_K} - \sigma_{0\tau_K}| < \Delta_{4\tau_K}\}
\end{aligned}$$

Now, it is convenient to split the complement of the neighbourhood W^c into number of subregions each part of a separate quadrant in the euclidean space with dimension equal to the number of parameters. More precisely,

$$W^c = \bigcup_{j=1}^{J} W_j$$

where J is the total number of quadrants and the sets W_j are of following form.

$$\begin{aligned}
W_1 &= \{(\alpha_{\tau_1}, \beta_{1\tau_1}, \beta_{2\tau_1}, \alpha_{\tau_2}, \beta_{1\tau_2}, \beta_{2\tau_2}) : \\
&\qquad \alpha_{\tau_1} - \alpha_{0\tau_1} \geq \Delta_{1\tau_1}, \beta_{1\tau_j} \geq \beta_{10\tau_1}, \beta_{2\tau_1} \geq \beta_{20\tau_1} \\
&\qquad \alpha_{\tau_2} \geq \alpha_{0\tau_1}, \beta_{1\tau_2} \geq \beta_{10\tau_2}, \beta_{2\tau_2} \geq \beta_{20\tau_2}\} \\
W_2 &= \{(\alpha_{\tau_1}, \beta_{1\tau_1}, \beta_{2\tau_1}, \alpha_{\tau_2}, \beta_{1\tau_2}, \beta_{2\tau_2}) : \\
&\qquad \alpha_{\tau_1} - \alpha_{0\tau_1} \geq \Delta_{1\tau_1}, \beta_{1\tau_j} \geq \beta_{10\tau_1}, \beta_{2\tau_1} \geq \beta_{20\tau_1} \\
&\qquad \alpha_{\tau_2} \geq \alpha_{0\tau_1}, \beta_{1\tau_2} \geq \beta_{10\tau_2}, \beta_{2\tau_2} < \beta_{20\tau_2}\} \\
&\vdots \\
&\quad \text{and so on} \\
&\vdots
\end{aligned}$$

We handle the numerator by splitting the parameter space of $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ into two parts as $\Theta = G \cup G^c$, where G is a compact set such that the integral over G^c decays in a exponential manner. The proposition below is analogous to theorem 1.3.3 in Ghosh and Ramamoorthi (2003) and gives a uniform strong law for independent non-identically distributed random variables.

Proposition 3.

Let $Y_i \sim P_{0i},\ i = 1, 2, ...,$ be a sequence of independent random variables and P denote the corresponding product measure. Let *(i)* Θ *be a compact parameter space*

(ii) $T_i(\theta, Y_i)$ *be measurable and for any compact set* $B \subset \Theta$, $\exists\ M > 0$ *such that* $E\left(\sup_{\theta\in B} |T_i(\theta, y)|^2\right) \leq M$

(iii) For any $\theta_0 \in \Theta$,

$$\lim_{\delta\to 0} \sup_{n\geq 1} E\left[\frac{1}{n}\sum_{i=1}^{n} \sup_{\{\theta:\|\theta-\theta_0\|<\delta\}} |T_i(\theta, Y_i) - E\{T_i(\theta, Y_i)\}\right.$$
$$\left. - T_i(\theta_0, Y_i) + E\{T_i(\theta_0, Y_i)\}|\right] = 0$$

Then,

$$\lim_{n\to\infty} \sup_{\theta\in\Theta} \left|\frac{1}{n}\sum_{i=1}^{n} T_i(\theta, Y_i) - \frac{1}{n}\sum_{i=1}^{n} E\{T_i(\theta, Y_i)\}\right| = 0 \ a.s.[P]$$

Proof. Proof of the proposition is in the lines of theorem 1.3.3 in Ghosh and Ramamoorthi (2003). □

Lemma 4. *If assumption* **2** *holds, then for any compact set* $G \subset \Theta$,

$$\sup_{(\alpha,\beta)\in\ G} \left|\frac{1}{n}\sum_{i=1}^{n}\sum_{l=1}^{K}\left[log\left(\frac{f_{(i,\alpha_{\tau_l},\beta_{\tau_l},\sigma_{\tau_l},\tau_l)}(Y_i)}{f_{(i,\alpha_{0\tau_l},\beta_{0\tau_l},\sigma_{\tau_l},\tau_l)}(Y_i)}\right) - E\left\{log\left(\frac{f_{(i,\alpha_{\tau_l},\beta_{\tau_l},\sigma_{\tau_l},\tau_l)}(Y_i)}{f_{(i,\alpha_{0\tau_l},\beta_{0\tau_l},\sigma_{\tau_l})}(Y_i)}\right)\right\}\right]\right|$$

$\rightarrow\ 0\ a.s\ [P]($ *uniformly in* $\boldsymbol{\sigma})$

Proof. The result follows by first noting the inequality below and then applying the same argument as in lemma A.1 of Sriram et al. (2012)

$$\sup_{(\alpha,\beta)\in\ G} \left|\frac{1}{n}\sum_{i=1}^{n}\sum_{l=1}^{K}\left[log\left(\frac{f_{(i,\alpha_{\tau_l},\beta_{\tau_l},\sigma_{\tau_l},\tau_l)}(Y_i)}{f_{(i,\alpha_{0\tau_l},\beta_{0\tau_l},\sigma_{\tau_l},\tau_l)}(Y_i)}\right) - E\left\{log\left(\frac{f_{(i,\alpha_{\tau_l},\beta_{\tau_l},\sigma_{\tau_l},\tau_l)}(Y_i)}{f_{(i,\alpha_{0\tau_l},\beta_{0\tau_l},\sigma_{\tau_l})}(Y_i)}\right)\right\}\right]\right|$$
$$\leq \sum_{l=1}^{K} \sup_{(\alpha,\beta)\in\ G} \left|\frac{1}{n}\sum_{i=1}^{n}\left[log\left(\frac{f_{(i,\alpha_{\tau_l},\beta_{\tau_l},\sigma_{\tau_l},\tau_l)}(Y_i)}{f_{(i,\alpha_{0\tau_l},\beta_{0\tau_l},\sigma_{\tau_l},\tau_l)}(Y_i)}\right) - E\left\{log\left(\frac{f_{(i,\alpha_{\tau_l},\beta_{\tau_l},\sigma_{\tau_l},\tau_l)}(Y_i)}{f_{(i,\alpha_{0\tau_l},\beta_{0\tau_l},\sigma_{\tau_l})}(Y_i)}\right)\right\}\right]\right|$$

□

Lemma 5. *Let assumptions* **1** *and* **3** *hold. Then, for any compact set* G *and for* $j = 1, 2, .., J$*, there exists* N_j^* *and* $K_j > 0$ *(independent of* $\boldsymbol{\sigma}$*) such that for* $n > N_j^*$*,*

$$I_{1n}^j(\sigma) < e^{-n\frac{K_j}{2\max\{\sigma_{\tau_1},..,\sigma_{\tau_K}\}}}$$

where,

$$I_{1n}^j(\sigma) = \int_{W_j \bigcap G} e^{\sum_{i=1}^n \sum_{l=1}^K log\left(\frac{f_{(i,\alpha_{\tau_l},\beta_{\tau_l},\sigma_{\tau_l})}(Y_i)}{f_{(i,\alpha_{0\tau_l},\beta_{0\tau_l},\sigma_{\tau_l})}(Y_i)}\right)} d\Pi(\alpha, \beta|\sigma)$$

Proof. The proof of the result is similar for each W_j. For the case of W_1, first lemma 4 essentially helps replace the log likelihood ratio within the integral with it's expectation and then using lemma 2, we note that

$$\begin{aligned}
& E\left\{log\left(\frac{f_{(i,\alpha_{\tau_1},\beta_{\tau_1},\sigma_{\tau_1})}(Y_i)}{f_{(i,\alpha_{0\tau_1},\beta_{0\tau_1},\sigma_{\tau_1})}(Y_i)}\right)\right\} \\
\leq\ & E\left\{\frac{(Y_i - \alpha_{\tau_1} - \mathbf{X}_i^T\beta_{\tau_1})}{\sigma} I_{(\alpha_{0\tau_1}+\mathbf{X}_i^T\beta_{0\tau_1}<Y_i<\alpha_{\tau_1}+\mathbf{X}_i^T\beta_{\tau_1})}.I_{\{X_{1i}>\epsilon_0,X_{2i}>\epsilon_0\}}\right\} \\
\leq\ & -\frac{\Delta_{1\tau_1}}{2\sigma_{\tau_1}} E\left\{I_{\left(0<Y_i-\alpha_{0\tau_1}-\beta_{0\tau_1}X_i<\frac{\Delta_{1\tau_1}}{2}\right)}.I_{\{X_{1i}>\epsilon_0,X_{2i}>\epsilon_0\}}\right\}
\end{aligned}$$

Also, for other $l \neq 1$, $E\left\{log\left(\frac{f_{(i,\alpha_{\tau_l},\beta_{\tau_l},\sigma_{\tau_1})}(Y_i)}{f_{(i,\alpha_{0\tau_l},\beta_{0\tau_l},\sigma_{\tau_l})}(Y_i)}\right)\right\} \leq 0$. Now result follows by taking

$$K_1 = \Delta_{1\tau_1}. \liminf_{m\to\infty} \frac{1}{m}\sum_{i=1}^m E\left\{I_{\left(0<Y_i-\alpha_{0\tau_1}-\beta_{0\tau_1}X_i<\frac{\Delta_{1\tau_1}}{2}\right)}.I_{\{X_{1i}>\epsilon_0,X_{2i}>\epsilon_0\}}\right\}$$

□

The next proposition essentially establishes the result in theorem 1 when the parameter space is compact.

Proposition 4. *Let assumptions* **1**, **3** *and* **4** *hold. Suppose* $G \subset \Theta$ *is compact and* $\Theta_\sigma = [\sigma_1, \sigma_2]^K$ *with* $0 < \sigma_1 \leq \sigma_2 < \infty$*. Then,* $\exists\ C' > 0$, $\delta_1 > 0$ *and* N^{**} *such that* $\forall n \geq N^{**}$

$$\int_{(W\times V)^c\cap(G\times\Theta_\sigma)} e^{\sum_{i=1}^n \sum_{j=1}^K log\left(\frac{f_{(i,\alpha_{\tau_j},\beta_{\tau_j},\sigma_{\tau_j})}(Y_i)}{f_{(i,\alpha_{0\tau_j},\beta_{0\tau_j},\sigma_{0\tau_j})}(Y_i)}\right)} d\Pi(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\sigma}) \leq C'e^{-n\delta_1}$$

Proof. First, we note that uniform SLLN holds for the log likelihood ratio, for every $j \in \{1, 2, ..., K\}$.

$$\frac{1}{n}\sum_{i=1}^{n} log\left(\frac{f_{(i,\alpha_{\tau_j},\beta_{\tau_j},\sigma_{\tau_j})}(Y_i)}{f_{(i,\alpha_{0\tau_j},\beta_{0\tau_j},\sigma_{0\tau_j})}(Y_i)}\right)$$
$$= \frac{1}{n}\sum_{i=1}^{n} log\left(\frac{f_{(i,\alpha_{\tau_j},\beta_{\tau_j},\sigma_{\tau_j})}(Y_i)}{f_{(i,\alpha_{0\tau_j},\beta_{0\tau_j},\sigma_{\tau_j})}(Y_i)}\right) + \frac{1}{n}\sum_{i=1}^{n} log\left(\frac{f_{(i,\alpha_{0\tau_j},\beta_{0\tau_j},\sigma_{\tau_j})}(Y_i)}{f_{(i,\alpha_{0\tau_j},\beta_{0\tau_j},\sigma_{0\tau_j})}(Y_i)}\right)$$

For the first term, uniform SLLN is implied by lemma 4. For second term it follows by noting that it is $= \log\left(\frac{\sigma_{0\tau_j}}{\sigma_{\tau_j}}\right) + \left(\frac{1}{\sigma_{0\tau_j}} - \frac{1}{\sigma_{\tau_j}}\right) Z_{i\tau_j}(\tau_j - I_{Z_{i\tau_j}<0})$. This would essentially help replace the exponent in the integrand with it's expectation. Then the integral over the set $(W \times V)^c \cap (G \times \Theta_\sigma)$ can be bounded by a sum of two terms, first term being an integral over the set $((W^c \cap G) \times \Theta_\sigma)$ and second term over the set $(G \times V^c)$. For the first integral, the result follows using lemma 5 and result for the second integral follows by using the definition of $\sigma_{0\tau}$ and arguing that for any $\tau \in \{\tau_1, ..., \tau_K\}$ with $|\sigma_\tau - \sigma_{0\tau}| > \Delta_{4\tau}$, $\lim \sum_{i=1}^{n} \frac{1}{n} E\left(log\frac{f_{i,\alpha_{0\tau},\beta_{0\tau},\sigma_\tau}(Y_i)}{f_{i,\alpha_{0\tau},\beta_{0\tau},\sigma_{0\tau}}(Y_i)}\right) < -\delta_1$ for some $\delta_1 > 0$. □

Lemma 6. *If assumptions* **1, 3** *and* **4** *hold, then for $j = 1, 2, .., J$, $\exists$ a compact set $G_j \subset W_j$, $b_j > 0$ and N_j^{**} (all not depending on $\boldsymbol{\sigma}$) such that*

$$\int_{G_j^c \cap W_j} e^{\sum_{i=1}^{n}\sum_{l=1}^{K} log\frac{f_{(i,\alpha_{\tau_l},\beta_{\tau_l},\sigma_{\tau_l})}(Y_i)}{f_{(i,\alpha_{0\tau_l},\beta_{0\tau_l},\sigma_{\tau_l})}(Y_i)}} d\Pi(\boldsymbol{\alpha}, \boldsymbol{\beta}|\boldsymbol{\sigma}) \leq e^{-nb_j/\max\{\sigma_{\tau_1},...,\sigma_{\tau_K}\}}$$
$$\forall\ n \geq N_j^{**}$$

Proof. We will prove the result for the set W_1 and for the case of two quantiles (i.e K=2). The argument is similar for other sets W_j for j=2,...,8 and for $K \geq 3$. Let ϵ_0 be as in assumption **3** and $Z_{i\tau} = Y_i - \alpha_{0\tau} - \mathbf{X}_i\beta_{0\tau}$. For j=1,2, define

$$C_0 = \frac{2\limsup_{m\to\infty} \frac{1}{m}\sum_{i=1}^{m} E(|Z_{i\tau_1}|) + 2\limsup_{m\to\infty} \frac{1}{m}\sum_{i=1}^{m} E(|Z_{i\tau_2}|)}{\liminf_{m\to\infty} \frac{1}{m}\sum_{i=1}^{m} E\left\{I_{X_{i1}>\epsilon_0, X_{i2}>\epsilon_0}\right\}}$$

Note that assumption **3** in particular implies that the denominator is well defined and assumption **4** ensures that the numerator is well defined. Now let $A_j = B_j\epsilon_0 = \frac{2C_0\sigma_{\tau_j}}{(1-\tau_j)\min(\sigma_{\tau_1},\sigma_{\tau_2})}$ and define

$$G_1 \;=\; \cap_{j=1}^{2}\{(\alpha, \beta) \in W_1 : \alpha_{\tau_j} - \alpha_{0\tau_j} \leq A_j\ ,\ \beta_{1\tau_j} - \beta_{10\tau_j} \leq B_j,\ \beta_{2\tau_j} - \beta_{20\tau_j} \leq B_j\}$$

Clearly G_1 is compact. Now if $(\alpha, \beta) \in G_1^c \cap W_1$ then either $(\alpha_{\tau_j} - \alpha_{0\tau_j}) > A_j$ or $(\beta_{1\tau_j} - \beta_{10\tau_j}) > B_j$ or $(\beta_{2\tau_j} - \beta_{20\tau_j}) > B_j$ for some $j \in \{1, 2\}$. Further if $X_{i1} > \epsilon_0$ and $X_{i2} > \epsilon_0$ then in the first case we have $b_{i\tau_j} = (\alpha_{\tau_j} - \alpha_{0\tau_j}) + (\beta_{1\tau_j} - \beta_{10\tau_j})X_{1i} + (\beta_{2\tau_j} - \beta_{20\tau_j})X_{2i} > A_j$ for some j and in the other two cases, we would have $b_{i\tau_j} > B_j\epsilon_0$ for some j. So, in any case when $X_{1i} > \epsilon_0$ and $X_{2i} > \epsilon_0$, we have $b_{i\tau_j} > \frac{2C_0\sigma_{\tau_j}}{(1-\tau_j)\min(\sigma_{\tau_1},\sigma_{\tau_2})}$ for some $j \in \{1, 2\}$. So, without loss of generality assume $b_{i\tau_1} > \frac{2C_0\sigma_{\tau_1}}{(1-\tau_1)\min(\sigma_{\tau_1},\sigma_{\tau_2})}$ when $X_{1i} > \epsilon_0$ holds and $X_{2i} > \epsilon_0$, in which case we also have $b_{i\tau_2} \geq 0$. Now, we can write

$$\begin{aligned}
&\sum_{i=1}^{n}\sum_{j=1}^{2} log\left(\frac{f_{(i,\alpha_{\tau_j},\beta_{\tau_j},\sigma_{\tau_j},\tau_j)}(Y_i)}{f_{(i,\alpha_{0\tau_j},\beta_{0\tau_j},\sigma_{\tau_j},\tau_j)}(Y_i)}\right)\\
&= \sum_{j=1}^{2}\sum_{i=1}^{n} log\left(\frac{f_{(i,\alpha_{\tau_j},\beta_{\tau_j},\sigma_{\tau_j},\tau_j)}(Y_i)}{f_{(i,\alpha_{0\tau_j},\beta_{0\tau_j},\sigma_{\tau_j},\tau_j)}(Y_i)}\right) I_{\{X_{1i}>\epsilon_0, X_{2i}>\epsilon_0\}}\\
&+ \sum_{j=1}^{2}\sum_{i=1}^{n} log\left(\frac{f_{(i,\alpha_{\tau_j},\beta_{\tau_j},\sigma_{\tau_j},\tau_j)}(Y_i)}{f_{(i,\alpha_{0\tau_j},\beta_{0\tau_j},\sigma_{\tau_j},\tau_j)}(Y_i)}\right) I_{\{X_{1i}>\epsilon_0, X_{2i}>\epsilon_0\}^c}
\end{aligned}$$

Now, applying part(e) of lemma 1 to the first term in R.H.S and part (d) to the second term (for $(\alpha, \beta) \in G_1^c \cap W_1$), for sufficiently large n (say $\forall\ n \geq N_1^{**}$) we have,

$$\begin{aligned}
&\sum_{i=1}^{n} log\left(\frac{f_{(i,\alpha,\beta,\sigma)}(Y_i)}{f_{(i,\alpha_0,\beta_0,\sigma)}(Y_i)}\right)\\
&\leq -\frac{2C_0\sigma_{\tau_1}}{(1-\tau_1)\min(\sigma_{\tau_1},\sigma_{\tau_2})}(1-\tau_1)\frac{\sum_{i=1}^{n} I_{\{X_{1i}>\epsilon_0, X_{2i}>\epsilon_0\}}}{\sigma_{\tau_1}}\\
&+ \sum_{j=1}^{2} \frac{\sum_{i=1}^{n} Z_{i\tau_j}^{+} I_{\{X_{1i}>\epsilon_0, X_{2i}>\epsilon_0\}}}{\sigma_{\tau_j}} + \sum_{j=1}^{2} \frac{\sum_{i=1}^{n} |Z_{i\tau_j}| I_{\{X_{1i}>\epsilon_0, X_{2i}>\epsilon_0\}^c}}{\sigma_{\tau_j}}\\
&\leq -\frac{2nC_0}{\min(\sigma_{\tau_1},\sigma_{\tau_2})}\liminf_{m\to\infty}\frac{1}{m}\sum_{i=1}^{m} E\left\{I_{\{X_{1i}>\epsilon_0, X_{2i}>\epsilon_0\}}\right\} + n\sum_{j=1}^{2}\frac{\limsup_{m\to\infty}\frac{1}{m}\sum_{i=1}^{n} E|Z_{i\tau_j}|}{\sigma_{\tau_j}}\\
&\leq -\frac{nC_0}{\min(\sigma_{\tau_1},\sigma_{\tau_2})}\liminf_{m\to\infty}\frac{1}{m}\sum_{i=1}^{m} E\left\{I_{\{X_{1i}>\epsilon_0, X_{2i}>\epsilon_0\}}\right\}\\
&\leq -\frac{nC_0}{\max(\sigma_{\tau_1},\sigma_{\tau_2})}\liminf_{m\to\infty}\frac{1}{m}\sum_{i=1}^{m} E\left\{I_{\{X_{1i}>\epsilon_0, X_{2i}>\epsilon_0\}}\right\}
\end{aligned}$$

The steps in the inequality above use assumption **4**, which allows the application of SLLN on the sequence $\{|Z_{i\tau_j}|\}$. Now, the result follows by using propriety of prior from assumption **1** and taking $b_1 = C_0 \liminf_{m\to\infty} \frac{1}{m}\sum_{i=1}^{m} E\left\{I_{\{X_{1i}>\epsilon_0, X_{2i}>\epsilon_0\}}\right\}$ □

Proof. (of theorem 1). Theorem 1 is a direct consequence of the lemmas and propositions discussed above. Lemma 3 helps handle the denominator in the posterior probability showing that $e^{nd} I_{2n} \to \infty$ $a.s.$ $[P] \forall\ d > 0$. For numerator, it needs to be shows that $\exists\ d_0 > 0$ such that $e^{nd_0} I_{1n} \to 0$ $a.s$ $[P]$. The integral in the numerator is split over $(G \times \Theta_\sigma)$ and $(G^c \times \Theta_\sigma)$ where $G = \cap_{j=1}^{J} G_j$ with G_j as in lemma 6. Proposition 4 helps establish the result for the first term. Lemma 6 helps establish the convergence of second term, thus completing the proof. □

The next lemma helps extend theorem 1 to the case when $\Theta_\sigma = (0, \infty)^K$.

Lemma 7. *If assumptions* **1, 3** *and* **4** *hold, then for* $\exists\ C' > 0,\ d_0 > 0,\ N_0(\omega)$ *(all not depending on* σ*) such that* $\forall\ n \geq N_0,\ \boldsymbol{\sigma} \in \Theta_\sigma$,

$$I_{1n}(\sigma) = \int_{W^c} \prod_{i=1}^{n} \prod_{l=1}^{K} \frac{f_{(i, \alpha_{\tau_l}, \beta_{\tau_l}, \sigma_{\tau_l})}(Y_i)}{f_{(i, \alpha_{0\tau_l}, \beta_{0\tau_l}, \sigma_{\tau_l})}(Y_i)} d\Pi(\boldsymbol{\alpha}, \boldsymbol{\beta} | \boldsymbol{\sigma}) \leq C' e^{-2nd_0 / \max\{\sigma_{\tau_1}, \ldots, \sigma_{\tau_K}\}}$$

Proof. The lemma is an immediate consequence of lemma 5 and 6. □

Without delving into the details which would be similar to that in Sriram et al. (2012), we just note that lemma 7 is particularly useful in the proof of remark 1.

Appendix B Normalizing constant

The normalizing constant for the pseudo density in equation (3)for $y \in \mathcal{X} = (-\infty, \infty)$ is given by

$$C = \int_{-\infty}^{\infty} \prod_{j=1}^{K} \frac{1}{\sigma_{\tau_j}} exp\left\{ -\sum_{j=1}^{K} \frac{\rho_{\tau_j}(y - \mu^{\tau_j})}{\sigma_{\tau_j}} \right\} dy \tag{16}$$

$$= \sum_{l=1}^{K} (A_l B_l - log(\sigma_{\tau_j}))$$

where,

$$A_l = \begin{cases} \left(exp\left\{-b_l \mu_{(l)}\right\} - exp\left\{-b_l \mu_{(l-1)}\right\}\right)/b_l & , if\ b_l \neq 0 \\ \mu_{(l)} - \mu_{(l-1)} & , if\ b_l = 0 \end{cases}$$

$$b_l = \sum_{j=1}^{l-1} \frac{\tau_j}{\sigma_j} + \sum_{j=l}^{K} \frac{(\tau_j - 1)}{\sigma_j}$$

$$B_l = exp\left\{ \sum_{j=1}^{l-1} \frac{\mu_{(j)} \times \tau_j}{\sigma_j} + \sum_{j=l}^{K} \frac{\mu_{(j)} \times (\tau_j - 1)}{\sigma_j} \right\}$$

$(\mu_{(1)}, ..., \mu_{(K)})$ are ordered values of $(\mu^{\tau_1}, ..., \mu^{\tau_K})$ and $\mu_{(0)} = 0,\ \mu_{(K+1)} = \infty$

Appendix C Details of MCMC algorithms

MCMC for Simulated Model 1

- Data: $(Y_i, X_i, Z_i, P_i),\ i = 1, 2, ...N$
- Model Specification $Q(X_i, \tau) = X_i^T \boldsymbol{\beta}(\tau) + Z_i^T \boldsymbol{\gamma} + P_i^T \boldsymbol{\nu}_{S_i}(\tau_k)$
- Model estimation is done by considering a dense grid $\tau_1, ..., \tau_K$
- Using the scale mixture representation of $ALD(, X_i^T \boldsymbol{\beta}(\tau_k) + Z_i^T \boldsymbol{\gamma} + P_i^T \boldsymbol{\nu}_{S_i}(\tau_k), \sigma_k, \tau_k)$

for each $k = 1, 2.., K$, the likelihood becomes

$$L(Y|X, \beta(\cdot))$$
$$= \prod_{i=1}^{N} \prod_{k=1}^{K} \left(\sqrt{2\pi\sigma_k W_{ik}}\ \epsilon_k\right)^{-1}$$
$$\times exp\left\{-\sum_{k=1}^{K}\sum_{i=1}^{N}(Y_i - X_i^T\boldsymbol{\beta}(\tau_k) - Z_i^T\boldsymbol{\gamma} - P_i^T\boldsymbol{\nu}_{S_i}(\tau_k) - \xi_k W_{ik})^2/(2\epsilon_k^2\sigma_k W_{ik})\right\}$$

$W_{1k}, W_{2k}, ..., W_{Nk} \overset{iid}{\sim}$ exponential (mean=σ_k) and independent for k=1,2,..,K

$\xi_k = (1 - 2\tau_k)/(\tau_k(1 - \tau_k))$ and $\epsilon^2 = 2/(\tau_k(1 - \tau_k))$

- Let $\boldsymbol{\beta} = (\boldsymbol{\beta}(\tau_1), ..., \boldsymbol{\beta}(\tau_K))$, $\boldsymbol{\nu} = (\nu_1, .., \nu_s)$
- **Priors**

$$\boldsymbol{\beta}|\Sigma_0 \sim N(\boldsymbol{\beta_0},\ \boldsymbol{\Sigma_0})$$
$$\boldsymbol{\gamma}|\Sigma_\gamma \sim N(\gamma_0, \Sigma_\gamma)$$
$$\sigma_1^{-1}, ..., \sigma_K^{-1} \overset{iid}{\sim} Gamma(a, b) \text{ with density} g(s) \propto s^{a-1}e^{-bs}$$
$$\boldsymbol{\nu_1}, ..., \boldsymbol{\nu_s}|\Sigma_{nu} \overset{iid}{\sim} Multi - N(\boldsymbol{\nu_0}, \Sigma_{nu})$$
$$\boldsymbol{\Sigma_\nu^{-1}} \sim Wishart(D_\nu, d_\nu)$$
$$\left[V_{p\times p} \sim Wishart(D_0, d_0) \text{ has the density } \frac{|V|^{(d_0-p-1)/2)}}{|D_0|^{(d_0/2)}}e^{-Tr(D_0^{-1}V)/2}\right]$$

- Posterior distribution of $\boldsymbol{\beta}$

$$\beta|Y, ... \sim N(A_\beta^{-1}M_\beta, A_\beta^{-1})$$
$$A_\beta = \text{diagonal}\left[\sum_{i=1}^{N}\frac{X_iX_i^T}{\epsilon_k^2\sigma_k W_{ik}}, k = 1, 2.., K\right] + \Sigma_0^{-1}$$
$$M_\beta = \text{column vector } \left[\sum_{i=1}^{N}\left(\frac{Y_i - \xi_k W_{ik} - Z_i^T\boldsymbol{\gamma} - P_i^T\boldsymbol{\nu}_{S_i}(\tau_k)}{\epsilon_k^2\sigma_k W_{ik}}X_i\right),\ k = 1, 2.., K\right] + \Sigma_0^{-1}\boldsymbol{\beta}_0$$

- Posterior distribution of $\boldsymbol{\gamma}$

$$\gamma|Y, ... \sim N(A_\gamma^{-1}M_\gamma, A_\gamma^{-1})$$
$$A_\gamma = \sum_{k=1}^{K}\sum_{i=1}^{N}\frac{Z_iZ_i^T}{\epsilon_k^2\sigma_k W_{ik}} + \Sigma_\gamma^{-1}$$
$$M_\gamma = \sum_{k=1}^{K}\sum_{i=1}^{N}\left(\frac{Y_i - \xi_k W_{ik} - X_i^T\boldsymbol{\beta}(\tau_k) - P_i^T\boldsymbol{\nu}_{S_i}(\tau_k)}{\epsilon_k^2\sigma_k W_{ik}}Z_i\right) + \Sigma_\gamma^{-1}\gamma_0$$

- For each j=1,..,s, posterior distribution of $\boldsymbol{\nu}_j$

$$\begin{aligned}
&\nu_j|Y,... \sim N(A_{\nu_j}^{-1}M_{\nu_j}, A_{\nu_j}^{-1}) \\
&A_{\nu_j} = \sum_{k=1}^{K}\sum_{i:S_i=j}\frac{P_iP_i^T}{\epsilon_k^2\sigma_k W_{ik}} + \Sigma_\nu^{-1} \\
&M_{\nu_j} = \sum_{k=1}^{K}\sum_{i:S_i=j}\left(\frac{Y_i - \xi_k W_{ik} - X_i^T\boldsymbol{\beta}(\tau_k) - Z_i^T\boldsymbol{\gamma}}{\epsilon_k^2\sigma_k W_{ik}}Z_i\right) + \Sigma_\nu^{-1}\boldsymbol{\nu}_0
\end{aligned}$$

- For a given k, posterior distribution of σ_k

$$\begin{aligned}
&\sigma_k^{-1}|Y,... \sim Gamma(a_{\sigma_k}, b_{\sigma_k}) \\
&a_{\sigma_k} = a + N + N/2 \\
&b_{\sigma_k} = b + \sum_{i=1}^{N}(Y_i - X_i^T\boldsymbol{\beta}(\tau_k) - Z_i^T\boldsymbol{\gamma} - P_i^T\boldsymbol{\nu}_{S_i}(\tau_k) - \xi_k W_{ik})^2/(2\epsilon_k^2 W_{ik}) + \sum_{i=1}^{N} W_{ik}
\end{aligned}$$

- Posterior distribution of W_{ik}^{-1} for each i, k is inverse gaussian. The inverse gaussian density with parameters $(\lambda', \mu')$is given by

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

It can be seen that

$$\begin{aligned}
&W_{ik}^{-1}|Y,... \sim Inverse\ gaussian(\lambda', \mu') \\
&\text{where,} \\
&\lambda' = (\xi_k^2 + 2\epsilon_k^2)/(\sigma_k\epsilon_k) \\
&\mu' = \sqrt{(\xi_k^2 + 2\epsilon_k^2)/\left(Y_i - X_i^T\boldsymbol{\beta}(\tau_k) - Z_i^T\boldsymbol{\gamma} - P_i^T\boldsymbol{\nu}_{S_i}(\tau_k)\right)^2}
\end{aligned}$$

- Posterior for Σ_0, Σ_γ and Σ_ν

$$\Sigma_\nu^{-1}|Y,... \sim Wishart\left(\left(D_\nu^{-1} + \sum_{j=1}^{s}\boldsymbol{\nu}_j\boldsymbol{\nu}_j^T\right)^{-1}, d_\nu + s\right)$$

MCMC for Simulated Model 2

The MCMC scheme for the simulation model 2 proceeds on the same lines that of model 1 except for a couple of changes.

Firstly, we replace the $P_i^T \nu_{S[i]}(\tau_k)$ by $R_{i\tau_k}^T \nu_{S[i]}(\tau_k)$, where $R_{i\tau_k} = (1, V_i, (V_i - \eta_1)^+, ..., (V_i - \eta_D)^+)$, where V_i is the i^{th} element of the vector $V = (X_2 \times \omega_1 + P \times \omega_2)$ and $(\eta_1, ..., \eta_D)$ are D equally spaced quantiles of the element of the vector V (essentially the knots for the spline). Then the simulation of ν proceeds as in model 1.

Secondly, we add a step for simulating $\boldsymbol{\omega}(\tau_k) = (\omega_1(\tau_k), \omega_2(\tau_k))$ for $k = 1, 2, ..., K$ in the lines of Karabatsos (2009) using an Adaptive Random-Walk Metropolis (ARWM) algorithm. This uses a Metropolis-Hastings procedure with a symmetric distribution on the unit sphere as the proposal distribution and modifies the proposal distribution by introducing a scalar parameters λ so as to optimize the acceptance rate of the algorithm. Hence, we also introduce the parameter λ in our sampling. Suppose $\boldsymbol{\omega}^s$ is the current value. Then $\boldsymbol{\omega}^{(s+1)}(\tau_k)$ is obtained after generating a proposal $\boldsymbol{\omega}^*(\tau_k)$ as follows

- For each k, compute $\boldsymbol{\omega}(\tau_k) \sim N_v\left(\boldsymbol{\omega}^{(s)}(\tau_k) \times 2 \times {\lambda^{(s)}}^2, \mathbf{I}\right)$, where $\mathbf{I}$ is the identity matrix of appropriate order and $\boldsymbol{\omega}^*(\tau_k) = \boldsymbol{\omega}(\tau_k)/(\boldsymbol{\omega}(\tau_k)^T \boldsymbol{\omega}(\tau_k))$. Denote $(\omega) = (\boldsymbol{\omega}(\tau_1), ..., \boldsymbol{\omega}(\tau_K))$ and $(\omega^*) = (\boldsymbol{\omega}^*(\tau_1), ..., \boldsymbol{\omega}^*(\tau_K))$.
- Compute $\rho = \min\left(1, L(\mathbf{Y}|\mathbf{X}, ..., \boldsymbol{\omega}^*, ...)/L(\mathbf{Y}|\mathbf{X}, ..., \boldsymbol{\omega}^{(s)}, ...)\right)$, where parameters other than $\boldsymbol{\omega}$ in computing likelihood $L(\cdot)$ are held fixed at the values obtained at the s^{th} step.
- $\boldsymbol{\omega}^{(s+1)} = \boldsymbol{\omega}^*$ with probability ρ and $\boldsymbol{\omega}^{(s+1)} = \boldsymbol{\omega}^{(s)}$ with probability $(1 - \rho)$.
- $\lambda^{(s+1)} = max\left(0, \lambda^{(s)} + (s+1)^{-1/2}(.234 - \rho)\right)$.

MCMC for Simulated Model 3

- Recall that the model is as follows.

$$\text{Data: } (C_{ij}, Z_{ij}, X_i),\ i = 1, 2, .., s.\ j = 1, 2, .., T_i$$

$$C_{ij}|V_i \sim Normal(\mathbf{Z}_{ij}^T \boldsymbol{\gamma} + V_i,\ \nu^2)$$

$$Q(V_i|\mathbf{X}_i, \tau) = \mathbf{X}_i^T \boldsymbol{\beta}_\tau$$

$$j = 1, ..., T_i \quad i = 1, 2, ..., s$$

The MCMC scheme becomes simpler when we also simulate the latent variable V_i at every step along with the parameters $\boldsymbol{\beta}_\tau$ and $\boldsymbol{\gamma}$

- Let $\boldsymbol{\beta} = (\boldsymbol{\beta}(\tau_1), ..., \boldsymbol{\beta}(\tau_K))$
- **Priors**

$$\boldsymbol{\beta}|\Sigma_0 \sim \text{Truncated } N(\boldsymbol{\beta_0},\ \boldsymbol{\Sigma_0}) I_{\{\boldsymbol{\beta}>0\}}$$

$$\boldsymbol{\gamma}|\Sigma_\gamma \sim N(\gamma_0, \Sigma_\gamma)$$

$$\nu^2 \sim \text{ Inverge Gamma}(a_\nu, b_\nu)$$

$$\sigma_1^{-1}, ..., \sigma_K^{-1} \overset{iid}{\sim} Gamma(a, b) \text{ with density} g(s) \propto s^{a-1} e^{-bs}$$

- The posterior for σ_k and W_{ij} are obtained in the same manner as in model 1.
- Posterior distribution of $\boldsymbol{\beta}$

$$\beta|C, ... \sim N(A_\beta^{-1} M_\beta, A_\beta^{-1})$$

$$A_\beta = \text{diagonal}\left[\sum_{i=1}^{s} \frac{X_i X_i^T}{\epsilon_k^2 \sigma_k W_{ik}}, k = 1, 2.., K\right] + \Sigma_0^{-1}$$

$$M_\beta = \text{column vector } \left[\sum_{i=1}^{s} \left(\frac{V_i - \xi_k W_{ik}}{\epsilon_k^2 \sigma_k W_{ik}} X_i\right),\ k = 1, 2.., K\right] + \Sigma_0^{-1} \boldsymbol{\beta}_0$$

- Posterior distribution of $\boldsymbol{\gamma}$

$$\gamma|C, ... \sim N(A_\gamma^{-1} M_\gamma, A_\gamma^{-1})$$

$$A_\gamma = \mathbf{Z}^T \mathbf{Z}/\nu^2 + \Sigma_\gamma^{-1}$$

$$M_\gamma = \mathbf{Z}^T (C - V) + \Sigma_\gamma^{-1} \boldsymbol{\gamma}_0$$

where the elements of the vector (C-V) are given by $C_{ij} - V_i$

- Posterior distribution of $V = (V_1, .., V_s)$

$$V_i \sim \text{Truncated } N\left(\frac{MV_i}{AV_i}, \frac{1}{AV_i}\right)$$
$$AV_i = \frac{T_i}{\nu^2} + \sum_{l=1}^{K} \frac{1}{\epsilon_l^2 W_{il}\sigma_l}$$
$$MV_i = \sum_{j=1}^{T_i} \frac{C_{ij} - \mathbf{X}_{ij}^T\boldsymbol{\gamma}}{\nu^2} + \sum_{l=1}^{K} \frac{\mathbf{X}_i^T\boldsymbol{\beta} + \xi_l W_{il}}{\epsilon_l^2 W_{il}\sigma_l}$$

- Posterior for ν^2

$$\nu^2 \sim \text{Inverse Gamma}\left(a_\nu + \sum_{i=1}^{s} T_i,\ b_\nu + \sum_{i=1}^{s}\sum_{j=1}^{T_i}(C_{ij} - V_i - \mathbf{Z}_{ij}^T\gamma)^2/2\right)$$

References

Antoniadis, A., Grégoire, G., and Mckeague, I. W. (2004), "Bayesian Estimation in Single-index Models," *Statistica Sinica*, 14, 1147–1164.

Chernozhukov, V., Fernández-Val, I., and Galichon, A. (2010), "Quantile and Probability Curves Without Crossing," *Econometrica*, 78, 1093–1125.

Dette, H. and Volgushev, S. (2008), "Non-crossing Non-parametric Estimates of Quantile Curves," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70, 609–627.

Dunson, D. B. and Taylor, J. A. (2005), "Approximate Bayesian Inference for Quantiles," *Journal of Nonparametric Statistics*, 17, 385–400.

Ferguson, T. S. (1973), "A Bayesian Analysis of some Nonparametric Problems," *The Annals of Statistics*, 1, 209–230.

Frees, E. W., Young, V. R., and Luo, Y. (2001), "Case Studies using Panel Data Models," *North American Actuarial Journal*, 5, 24–42.

Ghosh, J. K. and Ramamoorthi, R. V. (2003), *Bayesian Nonparametrics*, Springer Verlag.

Griffin, J. E. and Steel, M. F. J. (2004), "Semiparametric Bayesian Inference for Stochastic Frontier Models," *Journal of Econometrics*, 121–152.

Hardle, W., Hall, P., and Ichimura, H. (1993), "Optimal Smoothing in Single-Index Models," *The Annals of Statistics*, 21, 157–178.

He, X. (1997), "Quantile Curves without Crossing," *The American Statistician*, 51, 186–192.

Ichimura, H. (1993), "Semiparametric Least Squares (SLS) and Weighted SLS Estimation of Single-Index Models," *Journal of Econometrics*, 58, 71–120.

Karabatsos, G. (2009), "Modeling Heteroscedasticity in the Single-index Model with the Dirichlet Process," *Advances and Applications in Statistical Sciences*, 1, 83–104.

Klugman, S. (1992), *Bayesian Statistics in Actuarial Science.*, Boston: Kluwer.

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

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

Koop, G., Osiewalski, J., and Steel, M. F. (1997), "Bayesian Efficiency Analysis Through Individual Effects: Hospital Cost Frontiers," *Journal of Econometrics*, 76, 77 – 105.

Reich, B. J., Bondell, H. D., and Wang, H. J. (2010), "Flexible Bayesian Quantile Regression for Independent and Clustered data," *Biostatistics*, 11, 337–352.

Reich, B. J., Fuentes, M., and Dunson, D. B. (2011), "Bayesian Spatial Quantile Regression," *Journal of the American Statistical Association*, 106, 6–20.

Ruppert, D., Wand, M. P., and Carroll, R. J. (2003), *Semiparametric Regression*, Cambridge University Press.

Sriram, K., Ramamoorthi, R. V., and Ghosh, P. (2012), "Posterior Consistency of Bayesian Quantile Regression Based on the Misspecified Asymmetric Laplace Density," *(Submitted)*.

Taddy, M. A. and Kottas, A. (2010), "A Bayesian Nonparametric Approach to Inference for Quantile Regression," *Journal of Business and Economic Statistics*, 28, 357–369.

Takeuchi, I. and Furuhashi, T. (2004), "Non-crossing Quantile Regressions by SVM," in *Neural Networks, 2004. Proceedings. 2004 IEEE International Joint Conference on*, vol. 1, pp. 4 vol. (xlvii+3302).

Tokdar, S. T. and Kadane, J. B. (2012), "Simultaneous Linear Quantile Regression: A Semiparametric Bayesian Approach," *Bayesian Analysis*, 7, 51–72.

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

Wang, H.-B. (2009), "Bayesian Estimation and Variable Selection for Single Index Models," *Compututational Statistics and Data Analysis*, 53, 2617–2627.

Wu, T. Z., Yu, K., and Yu, Y. (2010), "Single-Index Quantile Regression," *Journal of Multivariate Analysis*, 101, 1607–1621.

Wu, Y. and Liu, Y. (2009), "Stepwise Multiple Quantile Regression Estimation using Non-crossing Constraints," *Statistics and it's Interface*, 2, 299–310.

Yu, K. and Moyeed, R. A. (2001), "Bayesian Quantile Regression," *Statistics and Probability Letters*, 54, 437 – 447.

Yu, Y. and Ruppert, D. (2002), "Penalized Spline Estimation for Partially Linear Single-Index Models," *Journal of the American Statistical Association*, 97, 1042–1054.

Yue, Y. R. and Rue, H. (2011), "Bayesian Inference for Additive Mixed Quantile Regression Models," *Compututational Statistics and Data Analysis*, 55, 84–96.