bqror: An R package for Bayesian Quantile Regression in Ordinal Models

by Prajual Maheshwari and Mohammad Arshad Rahman

Abstract This article describes an R package *bqror* that estimates Bayesian quantile regression in ordinal models introduced in Rahman (2016). The paper classifies ordinal models into two types and offers computationally efficient, yet simple, Markov chain Monte Carlo (MCMC) algorithms for estimating ordinal quantile regression. The generic ordinal model with 3 or more outcomes (labeled OR_I model) is estimated by a combination of Gibbs sampling and Metropolis-Hastings algorithm. Whereas an ordinal model with exactly 3 outcomes (labeled OR_{II} model) is estimated using a Gibbs sampling algorithm only. In line with the Bayesian literature, we suggest using the marginal likelihood for comparing alternative quantile regression models and explain how to compute the same. The models and their estimation procedures are illustrated via multiple simulation studies and implemented in two applications. The article also describes several functions contained within the *bqror* package, and illustrates their usage for estimation, inference, and assessing model fit.

1 Introduction

Quantile regression defines the conditional quantiles of a continuous dependent variable as a function of the covariates without assuming any error distribution (Koenker and Bassett 1978). The method is robust and offers several advantages over least squares regression such as desirable equivariance properties, invariance to monotone transformation of the dependent variable, and robustness against outliers (Koenker 2005; Davino, Furno, and Vistocco 2014; Furno and Vistocco 2018). However, quantile regression with discrete outcomes is more complex because quantiles of discrete data cannot be obtained through a simple inverse operation of the cumulative distribution function (cdf). Besides, discrete outcome (binary and ordinal) modeling requires location and scale restrictions to uniquely identify the parameters (see Section 2 for details). Kordas (2006) estimated quantile regression with binary outcomes using simulated annealing, while Benoit and Van den Poel (2010) proposed Bayesian binary quantile regression where a working likelihood for the latent variable was constructed by assuming the error follows an asymmetric Laplace (AL) distribution (Yu and Moyeed 2001). The estimation procedure for the latter is available in the **bayesQR** package of R software (Benoit and Van den Poel 2017). A couple of recent works on Bayesian quantile regression with binary longitudinal (panel) outcomes are Rahman and Vossmeyer (2019) and Bresson, Lacroix, and Rahman (2021). Extending the quantile framework to ordinal outcomes is more intricate due to the difficulty in satisfying the ordering of cut-points while sampling. Rahman (2016) introduced Bayesian quantile analysis of ordinal data and proposed two efficient MCMC algorithms. Since Rahman (2016), ordinal quantile regression has attracted some attention which includes Alhamzawi (2016), Alhamzawi and Ali (2018), Ghasemzadeh, Ganjali, and Baghfalaki (2018), Rahman and Karnawat (2019), Ghasemzadeh, Ganjali, and Baghfalaki (2020), and Tian et al. (2021).

Ordinal outcomes frequently occur in a wide class of applications in economics, finance, marketing, and the social sciences. Here, ordinal regression (e.g. ordinal probit, ordinal logit) is an important tool for modeling, analysis, and inference. Given the prevalence of ordinal models in applications and the recent theoretical developments surrounding ordinal quantile regression, an estimation package is essential so that applied econometricians and statisticians can benefit from a more comprehensive data analysis. At present, no statistical software (such as R, MATLAB, Python, Stata, SPSS, and SAS) have any package for estimating quantile regression with ordinal outcomes. The current paper fills this gap in the literature and describes the implementation of **bqror** package (version 1.6.0) for estimation and inference in Bayesian ordinal quantile regression.

The **bqror** package offers two MCMC algorithms. Ordinal model with 3 or more outcomes is estimated through a combination of Gibbs sampling (Casella and George 1992) and Metropolis-Hastings (MH) algorithm (S. Chib and Greenberg 1995). The method is implemented in the function quantregOR1. For ordinal models with exactly 3 outcomes, the package presents a Gibbs sampling algorithm that is implemented in the function quantregOR2. We recommend using this procedure for an ordinal model with 3 outcomes, since its simpler and faster. Both functions, quantregOR1 and quantregOR2, report the posterior mean, posterior standard deviation, 95% posterior credible interval, and inefficiency factor of the model parameters. To compare alternative quantile regression models, we recommend using the marginal likelihood over the deviance information criterion (DIC). This is because the "Bayesian approach" to compare models is via the marginal likelihood (Siddhartha Chib 1995; Siddhartha Chib and Jeliazkov 2001). So, the **bqror** package also provides the marginal

likelihood with technical details for computation described in this paper. Additionally, the package includes functions to calculate the covariate effects and example codes to produce trace plots of MCMC draws. Lastly, this paper utilizes the **bqror** package to demonstrate the estimation of quantile ordinal models on simulated data and real-life applications.

2 Quantile regression in ordinal models

Ordinal outcomes are common in a wide class of applications in economics, finance, marketing, social sciences, statistics in medicine, and transportation. In a typical study, the observed outcomes are ordered and categorical; so for the purpose of analysis scores/numbers are assigned to each outcome. For example, in a study on public opinion about offshore drilling (Mukherjee and Rahman 2016), responses may be recorded as follows: 1 for 'strongly oppose', 2 for 'somewhat oppose', 3 for 'somewhat support', and 4 for 'strongly support'. The numbers have an ordinal meaning but have no cardinal interpretation. We cannot interpret a score of 2 as twice the support compared to a score of 1, or the difference in support between 2 and 3 is the same as that between 3 and 4. With ordinal outcomes, the primary modeling objective is to express the probability of outcomes as a function of the covariates. Ordinal models that have been extensively studied and employed in applications include the ordinal probit and ordinal logit models (Johnson and Albert 2000; Greene and Hensher 2010), but they only give information about the average probability of outcomes conditional on the covariates.

Quantile regression with ordinal outcomes can be estimated using the monotone equivariance property and provides information on the probability of outcomes at different quantiles. In the spirit of Albert and Chib (1993), the ordinal quantile regression model can be presented in terms of an underlying latent (or unobserved) variable z_i as follows:

$$z_i = x_i' \beta_p + \epsilon_i, \qquad \forall\, i = 1, \cdots, n, \qquad (1)$$

where x_i' is a $1 \times k$ vector of covariates, β_p is a $k \times 1$ vector of unknown parameters at the p-th quantile, ϵ_i follows an AL distribution i.e., $\epsilon_i \sim AL(0, 1, p)$, and n denotes the number of observations. Note that unlike the Classical (or Frequentist) quantile regression, the error is assumed to follow an AL distribution in order to construct a (working) likelihood (Yu and Moyeed 2001). The latent variable z_i is related to the observed discrete response y_i through the following relationship,

$$\gamma_{p,j-1} < z_i \le \gamma_{p,j} \Rightarrow y_i = j, \qquad \forall\, i = 1, \cdots, n;\ j = 1, \cdots, J, \qquad (2)$$

where $\gamma_p = (\gamma_{p,0} = -\infty, \gamma_{p,1}, \ldots, \gamma_{p,J-1}, \gamma_{p,J} = \infty)$ is the cut-point vector and J denotes the number of outcomes or categories. Typically, the cut-point $\gamma_{p,1}$ is fixed at 0 to anchor the location of the distribution required for parameter identification (Jeliazkov and Rahman 2012). Given the observed data $y = (y_1, \cdots, y_n)'$, the joint density (or likelihood when viewed as a function of the parameters) for the ordinal quantile model can be written as,

$$f(y|\Theta_p) = \prod_{i=1}^{n} \prod_{j=1}^{J} P(y_i = j|\Theta_p)^{I(y_i=j)} \qquad (3)$$

where $\Theta_p = (\beta_p, \gamma_p)$, $F_{AL}(\cdot)$ denotes the *cdf* of an AL distribution and $I(y_i = j)$ is an indicator function, which equals 1 if $y_i = j$ and 0 otherwise.

Working directly with the AL likelihood (3) is not convenient for MCMC sampling. Therefore, the latent formulation of the ordinal quantile model (1), following Kozumi and Kobayashi (2011), is expressed in the normal-exponential mixture form as follows,

$$z_i = x_i' \beta_p + \theta w_i + \tau \sqrt{w_i}\, u_i, \qquad \forall\, i = 1, \cdots, n, \qquad (4)$$

where $\epsilon_i = \theta w_i + \tau \sqrt{w_i}\, u_i \sim AL(0, 1, p)$, $w_i \sim \mathcal{E}(1)$ is mutually independent of $u_i \sim N(0, 1)$, N and $\mathcal{E}$ denotes normal and exponential distributions, respectively; $\theta = (1 - 2p)/[p(1 - p)]$ and $\tau = \sqrt{2/[p(1 - p)]}$. Based on this formulation, we can write the conditional distribution of the latent variable as $z_i|\beta_p, w_i \sim N(x_i' \beta_p + \theta w_i, \tau^2 w_i)$ for $i = 1, \ldots, n$. This allows access to the properties of normal distribution which helps in constructing efficient MCMC algorithms.

OR_I model

The term "OR$_I$ model" is ascribed to an ordinal model in which the number of outcomes (J) is equal to or greater than 3, location restriction is imposed by setting $\gamma_{p,1} = 0$, and scale restriction is achieved via constant variance (See Rahman (2016); for a given value of p, variance of a standard AL distribution is constant). Note that in contrast to Rahman (2016), our definition of OR_I model includes an ordinal

model with exactly 3 outcomes. The location and scale restrictions are necessary to uniquely identify the parameters (see Jeliazkov, Graves, and Kutzbach (2008) and Jeliazkov and Rahman (2012) for further details and a pictorial representation).

During the MCMC sampling of the OR_I model, we need to preserve the ordering of cut-points $(\gamma_{p,0} = -\infty < \gamma_{p,1} < \gamma_{p,2} < \ldots < \gamma_{p,J-1} < \gamma_{p,J} = \infty)$. This is achieved by using a monotone transformation from a compact set to the real line. Many such transformations are available (e.g., log-ratios of category bin widths, arctan, arcsin), but the **bqror** package utilizes the logarithmic transformation (Albert and Chib 2001; Rahman 2016),

$$\delta_{p,j} = \ln(\gamma_{p,j} - \gamma_{p,j-1}), \qquad 2 \leq j \leq J - 1. \tag{5}$$

The cut-points $(\gamma_{p,1}, \gamma_{p,2}, \cdots, \gamma_{p,J-1})$ can be obtained from a one-to-one mapping to $(\delta_{p,2}, \cdots, \delta_{p,J-1})$.

With all the modeling ingredients in place, we employ the Bayes' theorem and express the joint posterior distribution as proportional to the product of the likelihood and the prior distributions. As in Rahman (2016), we employ independent normal priors: $\beta_p \sim N(\beta_{p0}, B_{p0})$, $\delta_p \sim N(\delta_{p0}, D_{p0})$ in the **bqror** package. The augmented joint posterior distribution for the OR_I model can thus be written as,

$$
\begin{aligned}
\pi(z, \beta_p, \delta_p, w | y) &\propto f(y | z, \beta_p, \delta_p, w) \, \pi(z | \beta_p, w) \, \pi(w) \, \pi(\beta_p) \, \pi(\delta_p), \\
&\propto \left\{ \prod_{i=1}^{n} f(y_i | z_i, \delta_p) \right\} \pi(z | \beta_p, w) \, \pi(w) \, \pi(\beta_p) \, \pi(\delta_p), \\
&\propto \prod_{i=1}^{n} \left\{ \prod_{j=1}^{J} 1\{\gamma_{p,j-1} < z_i \leq \gamma_{p,j}\} \, N(z_i | x_i' \beta_p + \theta w_i, \tau^2 w_i) \, \mathcal{E}(w_i | 1) \right\} \\
&\quad \times \, N(\beta_p | \beta_{p0}, B_{p0}) \, N(\delta_p | \delta_{p0}, D_{p0}).
\end{aligned}
\tag{6}
$$

where in the likelihood function of the second line, we use the fact that the observed y_i is independent of (β_p, w) given (z_i, δ_p). This follows from equation (2) which shows that y_i given (z_i, δ_p) is determined with probability 1. In the third line, we specify the conditional distribution of the latent variable and the prior distribution on the parameters.

The conditional posterior distributions are derived from the augmented joint posterior distribution (6), and the parameters are sampled as per Algorithm~1. This algorithm is implemented in the **bqror** package. The parameter β_p is sampled from an updated multivariate normal distribution and the latent weight w is sampled element-wise from a generalized inverse Gaussian (GIG) distribution. The cut-point vector δ_p is sampled marginally of (z, w) using a random-walk MH algorithm. Lastly, the latent variable z is sampled element-wise from a truncated normal (TN) distribution.

Algorithm 1: Sampling in OR_I model.

- Sample $\beta_p | z, w \sim N(\tilde{\beta}_p, \tilde{B}_p)$, where,

  - $\tilde{B}_p^{-1} = \left(\sum_{i=1}^{n} \frac{x_i x_i'}{\tau^2 w_i} + B_{p0}^{-1} \right)$ and $\tilde{\beta}_p = \tilde{B}_p \left(\sum_{i=1}^{n} \frac{x_i (z_i - \theta w_i)}{\tau^2 w_i} + B_{p0}^{-1} \beta_{p0} \right)$.

- Sample $w_i | \beta_p, z_i \sim GIG(0.5, \tilde{\lambda}_i, \tilde{\eta})$, for $i = 1, \cdots, n$, where,

  - $\tilde{\lambda}_i = \left(\frac{z_i - x_i' \beta_p}{\tau} \right)^2$ and $\tilde{\eta} = \left(\frac{\theta^2}{\tau^2} + 2 \right)$.

- Sample $\delta_p | y, \beta$ marginally of w (latent weight) and z (latent data), by generating δ_p' using a random-walk chain $\delta_p' = \delta_p + u$, where $u \sim N(0_{J-2}, \iota^2 \hat{D})$, ι is a tuning parameter and $\hat{D}$ denotes the negative inverse Hessian, obtained by maximizing the log-likelihood with respect to δ_p. Given the current value of δ_p and the proposed draw δ_p', return δ_p' with probability,

$$\alpha_{MH}(\delta_p, \delta_p') = \min \left\{ 1, \frac{f(y | \beta_p, \delta_p') \, \pi(\beta_p, \delta_p')}{f(y | \beta_p, \delta_p) \, \pi(\beta_p, \delta_p)} \right\};$$

  otherwise repeat the old value δ_p. The variance of u may be tuned as needed for appropriate step size and acceptance rate.
- Sample $z_i | y, \beta_p, \gamma_p, w \sim TN_{(\gamma_{p,j-1}, \gamma_{p,j})}(x_i' \beta_p + \theta w_i, \tau^2 w_i)$ for $i = 1, \cdots, n$, where TN denotes a truncated normal distribution and γ_p is obtained via δ_p using equation (5)

OR$_{II}$ **model**

The term "OR$_{II}$ model" is used for an ordinal model with exactly 3 outcomes (i.e., $J = 3$) where both location and scale restrictions are imposed by fixing the cut-points. Since there are only 2 cut-points and both are fixed at some values, the scale of the distribution needs to be free. Therefore, a scale parameter σ_p is introduced and the quantile ordinal model is rewritten as follows:

$$
\begin{aligned}
z_i &= x_i'\beta_p + \sigma_p \epsilon_i = x_i'\beta_p + \sigma_p \theta w_i + \sigma_p \tau \sqrt{w_i}\, u_i, && \forall\, i = 1, \cdots, n,\\
\gamma_{j-1} &< z_i \le \gamma_j \;\Rightarrow\; y_i = j, && \forall\, i = 1, \cdots, n;\ j = 1, 2, 3,
\end{aligned}
\tag{7}
$$

where $\sigma_p \epsilon_i \sim AL(0, \sigma_p, p)$, (γ_1, γ_2) are fixed, and recall $\gamma_0 = -\infty$ and $\gamma_3 = \infty$. In this formulation, the conditional mean of z_i is dependent on σ_p which is problematic for Gibbs sampling. So, we define a new variable $\nu_i = \sigma_p w_i \sim \mathcal{E}(\sigma_p)$ and rewrite the model in terms of ν_i. In this representation, $z_i | \beta_p, \sigma_p, \nu_i \sim N(x_i'\beta_p + \theta\nu_i, \tau^2 \sigma_p \nu_i)$, the conditional mean is free of σ_p and the model is conducive to Gibbs sampling.

The next step is to specify the prior distributions required for Bayesian inference. We follow Rahman (2016) and assume $\beta_p \sim N(\beta_{p0}, B_{p0})$ and $\sigma_p \sim IG(n_0/2, d_0/2)$; where IG stands for an inverse-gamma distribution. These are the default prior distributions in the **bqror** package. Employing the Bayes' theorem, the augmented joint posterior distribution can be expressed as,

$$
\begin{aligned}
\pi(z, \beta_p, \nu, \sigma_p | y) &\propto f(y|z, \beta_p, \nu, \sigma_p)\, \pi(z|\beta_p, \nu, \sigma_p)\, \pi(\nu|\sigma_p)\, \pi(\beta_p)\, \pi(\sigma_p),\\
&\propto \left\{ \prod_{i=1}^{n} f(y_i|z_i, \sigma_p) \right\} \pi(z|\beta_p, \nu, \sigma_p)\, \pi(\nu|\sigma_p)\, \pi(\beta_p)\, \pi(\sigma_p),\\
&\propto \left\{ \prod_{i=1}^{n} \prod_{j=1}^{3} 1(\gamma_{j-1} < z_i \le \gamma_j)\, N(z_i|x_i'\beta_p + \theta\nu_i, \tau^2 \sigma_p \nu_i)\, \mathcal{E}(\nu_i|\sigma_p) \right\}\\
&\quad \times N(\beta_p|\beta_{p0}, B_{p0})\, IG(\sigma_p|n_0/2, d_0/2),
\end{aligned}
\tag{8}
$$

where the derivations in each step are analogous to those for the OR_I model.

The augmented joint posterior distribution, given by equation (8), can be utilized to derive the conditional posterior distributions and the parameters are sampled as presented in Algorithm 2. This involves sampling β_p from an updated multivariate normal distribution and sampling σ_p from an updated IG distribution. The latent weight ν is sampled element-wise from a GIG distribution and similarly, the latent variable z is sampled element-wise from a truncated normal distribution.

Algorithm 2: Sampling in OR$_{II}$ model.

- Sample $\beta_p | z, \sigma_p, \nu \sim N(\tilde{\beta}_p, \tilde{B}_p)$, where,

  - $\tilde{B}_p^{-1} = \left(\sum_{i=1}^{n} \frac{x_i x_i'}{\tau^2 \sigma_p \nu_i} + B_{p0}^{-1} \right)$ and $\tilde{\beta}_p = \tilde{B}_p \left(\sum_{i=1}^{n} \frac{x_i(z_i - \theta\nu_i)}{\tau^2 \sigma_p \nu_i} + B_{p0}^{-1} \beta_{p0} \right)$

- Sample $\sigma_p | z, \beta_p, \nu \sim IG(\tilde{n}/2, \tilde{d}/2)$, where,

  - $\tilde{n} = (n_0 + 3n)$ and $\tilde{d} = \sum_{i=1}^{n} (z_i - x_i'\beta_p - \theta\nu_i)^2 / \tau^2 \nu_i + d_0 + 2\sum_{i=1}^{n} \nu_i$.

- Sample $\nu_i | z_i, \beta_p, \sigma_p \sim GIG(0.5, \tilde{\lambda}_i, \tilde{\eta})$, for $i = 1, \cdots, n$, where,

  - $\tilde{\lambda}_i = \frac{(z_i - x_i'\beta_p)^2}{\tau^2 \sigma_p}$ and $\tilde{\eta} = \left(\frac{\theta^2}{\tau^2 \sigma_p} + \frac{2}{\sigma_p} \right)$

- Sample $z_i | y, \beta_p, \sigma_p, \nu_i \sim TN_{(\gamma_{j-1}, \gamma_j)}(x_i'\beta_p + \theta\nu_i, \tau^2 \sigma_p \nu_i)$ for $i = 1, \cdots, n$, and $j = 1, 2, 3$.

---

3 Marginal likelihood

The article by Rahman (2016) employed DIC (Spiegelhalter et al. 2002; Gelman et al. 2013) for model comparison. However, in the Bayesian framework alternative models are typically compared using the marginal likelihood or the Bayes factor (Poirier 1995; Greenberg 2012). Therefore, we prefer using the marginal likelihood (or the Bayes factor) for comparing two or more regression models at a given quantile.

Consider a model $\mathcal{M}_s$ with parameter vector Θ_s. Let $f(y|\mathcal{M}_s, \Theta_s)$ be its sampling density and $\pi(\Theta_s|\mathcal{M}_s)$ be the prior distribution; where $s = 1, \ldots, S$. Then, the marginal likelihood for the model $\mathcal{M}_s$ is given by the expression, $m(y|\mathcal{M}_s) = \int f(y|\Theta_s, \mathcal{M}_s)\pi(\Theta_s|\mathcal{M}_s)\, d\Theta_s$. The Bayes factor is the ratio of marginal likelihoods. So, for any two models $\mathcal{M}_q$ versus $\mathcal{M}_r$, the Bayes factor, $B_{qr} = \frac{m(y|\mathcal{M}_q)}{m(y|\mathcal{M}_r)} = \frac{\int f(y|\mathcal{M}_q, \Theta_q)\,\pi(\Theta_q|\mathcal{M}_q)\, d\Theta_q}{\int f(y|\mathcal{M}_r, \Theta_r)\,\pi(\Theta_r|\mathcal{M}_r)\, d\Theta_r}$, can be easily computed once we have the marginal likelihoods.

Siddhartha Chib (1995) and later Siddhartha Chib and Jeliazkov (2001) showed that a simple and stable estimate of marginal likelihood can be obtained from the MCMC outputs. The approach is based on the recognition that the marginal likelihood can be written as the product of likelihood function and prior density over the posterior density. So, the marginal likelihood $m(y|\mathcal{M}_s)$ for model $\mathcal{M}_s$ is expressed as,

$$m(y|\mathcal{M}_s) = \frac{f(y|\mathcal{M}_s, \Theta_s)\,\pi(\Theta_s|\mathcal{M}_s)}{\pi(\Theta_s|\mathcal{M}_s, y)}. \tag{9}$$

Siddhartha Chib (1995) refers to equation (9) as the *basic marginal likelihood identity* since it holds for all values in the parameter space, but typically computed at a high density point (e.g., mean, mode) denoted Θ^* to minimize estimation variability. The likelihood ordinate $f(y|\mathcal{M}_s, \Theta^*)$ is directly available from the model and the prior density $\pi(\Theta^*|\mathcal{M}_s)$ is assumed by the researcher. The novel part is the computation of posterior ordinate $\pi(\Theta^*|y, \mathcal{M}_s)$, which is estimated using the MCMC outputs. Since the marginal likelihood is often a large number, it is convenient to express it on the logarithmic scale. An estimate of the logarithm of marginal likelihood is given by,

$$\ln \hat{m}(y) = \ln f(y|\Theta^*) + \ln \pi(\Theta^*) - \ln \hat{\pi}(\Theta^*|y), \tag{10}$$

where we have dropped the conditioning on $\mathcal{M}_s$ for notational simplicity. The next two subsections explain the computation of marginal likelihood for the OR_I and OR_{II} quantile regression models.

Marginal likelihood for OR_I model

We know from Section 2.1 that the MCMC algorithm for estimating the OR_I model is defined by the following conditional posterior densities: $\pi(\beta_p|z, w)$, $\pi(\delta_p|\beta_p, y)$, $\pi(w|\beta_p, z)$, and $\pi(z|\beta_p, \delta_p, w, y)$. The conditional posteriors for β_p, w, and z have a known form, but that of δ_p is not tractable and is sampled using an MH algorithm. Consequently, we adopt the approach of Siddhartha Chib and Jeliazkov (2001) to calculate the marginal likelihood for the OR_I model.

To simplify the computational process (specifically, to keep the computation over a reasonable dimension), we estimate the marginal likelihood marginally of the latent variables (w, z). Moreover, we decompose the posterior ordinate as,

$$\pi(\beta_p^*, \delta_p^*|y) = \pi(\delta_p^*|y)\pi(\beta_p^*|\delta_p^*, y),$$

where $\Theta^* = (\beta_p^*, \delta_p^*)$ denotes a high density point. By placing the intractable posterior ordinate first, we avoid the MH step in the *reduced run* – the process of running an MCMC sampler with one or more parameters fixed at some value – of the MCMC sampler. We first estimate $\pi(\delta_p^*|y)$ and then the reduced conditional posterior ordinate $\pi(\beta_p^*|\delta_p^*, y)$.

To obtain an estimate of $\pi(\delta_p^*|y)$, we need to express it in a computationally convenient form. The parameter δ_p is sampled using an MH step, which requires specifying a proposal density. Let $q(\delta_p, \delta_p'|\beta_p, w, z, y)$ denote the proposal density for the transition from δ_p to δ_p', and let,

$$\alpha_{MH}(\delta_p, \delta_p') = \min\left\{1, \frac{f(y|\beta_p, \delta_p')\,\pi(\beta_p)\pi(\delta_p')}{f(y|\beta_p, \delta_p)\,\pi(\beta_p)\pi(\delta_p)} \times \frac{q(\delta_p', \delta_p|\beta_p, w, z, y)}{q(\delta_p, \delta_p'|\beta_p, w, z, y)}\right\}, \tag{11}$$

denote the probability of making the move. In the context of the model, $f(y|\beta_p, \delta_p)$ is the likelihood given by equation (3), $\pi(\beta_p)$ and $\pi(\delta_p)$ are normal prior distributions (i.e., $\beta_p \sim N(\beta_{p0}, B_{p0})$ and $\delta_p \sim N(\delta_{p0}, D_{p0})$ as specified in Section 2.1, and the proposal density $q(\delta_p, \delta_p'|\beta_p, w, z, y)$ is normal given by $f_N(\delta_p'|\delta_p, \iota^2\hat{D})$ (see Algorithm~1 in Section 2.1). There are two points to be noted about the proposal density. First, the conditioning on (β_p, w, z, y) is only for generality and not necessary as illustrated by the use of a random-walk proposal density. Second, since our MCMC sampler utilizes a random-walk proposal density, the second ratio on the right hand side of equation (11) reduces to 1.

We closely follow the derivation in Siddhartha Chib and Jeliazkov (2001) and arrive at the following expression of the posterior ordinate,

$$\pi(\delta_p^*|y) = \frac{E_1\{\alpha_{MH}(\delta_p, \delta_p^*|\beta_p, w, z, y)\, q(\delta_p, \delta_p^*|\beta_p, w, z, y)\}}{E_2\{\alpha_{MH}(\delta_p^*, \delta_p|\beta_p, w, z, y)\}}, \tag{12}$$

where E_1 represents expectation with respect to the distribution $\pi(\beta_p, \delta_p, w, z|y)$ and E_2 represents expectation with respect to the distribution $\pi(\beta_p, w, z|\delta_p^*, y) \times q(\delta_p^*, \delta_p|\beta_p, w, z, y)$. The quantities in equation (12) can be estimated using MCMC techniques. To estimate the numerator, we take the draw $\{\beta_p^{(m)}, \delta_p^{(m)}, w^{(m)}, z^{(m)}\}_{m=1}^M$ from the *complete MCMC run* and take an average of the quantity $\alpha_{MH}(\delta_p, \delta_p^*|\beta_p, w, z, y) \, q(\delta_p, \delta_p^*|\beta_p, w, z, y)$, where $\alpha_{MH}(\cdot)$ is given by equation (11) with δ_p' replaced by δ_p^*, and $q(\delta_p, \delta_p^*|\beta_p, w, z, y)$ is given by the normal density $f_N(\delta_p^*|\delta_p, \iota^2 \hat{D})$.

The estimation of the quantity in the denominator is tricky! This requires generating an additional sample (say of H iterations) from the reduced conditional densities: $\pi(\beta_p|w, z)$, $\pi(w|\beta_p, z)$, and $\pi(z|\beta_p, \delta_p^*, w, y)$, where note that δ_p is fixed at δ_p^* in the MCMC sampling, and thus corresponds to a *reduced run*. Moreover, at each iteration, we generate

$$\delta_p^{(h)} \sim q(\delta_p^*, \delta_p|\beta_p^{(h)}, w^{(h)}, z^{(h)}, y) \equiv f_N(\delta_p|\delta_p^*, \iota^2 \hat{D}).$$

The resulting quadruplet of draws $\{\beta_p^{(h)}, \delta_p^{(h)}, w^{(h)}, z^{(h)}\}$, as required, is a sample from the distribution $\pi(\beta_p, w, z|\delta_p^*, y) \times q(\delta_p^*, \delta_p|\beta_p, w, z, y)$. With the numerator and denominator now available, the posterior ordinate $\pi(\delta_p^*|y)$ is estimated as,

$$\hat{\pi}(\delta_p^*|y) = \frac{M^{-1} \sum_{m=1}^M \{\alpha_{MH}(\delta_p^{(m)}, \delta_p^*|\Lambda^{(m)}, y) \, q(\delta_p^{(m)}, \delta_p^*|\Lambda^{(m)}, y)\}}{H^{-1} \sum_{h=1}^H \{\alpha_{MH}(\delta_p^*, \delta_p^{(h)}|\Lambda^{(h)}, y)\}}. \tag{13}$$

where $\Lambda^{(m)} = (\beta_p^{(m)}, w^{(m)}, z^{(m)})$ and $\Lambda^{(h)} = (\beta_p^{(h)}, w^{(h)}, z^{(h)})$.

The computation of the posterior ordinate $\pi(\beta_p^*|\delta_p^*, y)$ is trivial. We have the sample of H draws $\{w^{(h)}, z^{(h)}\}$ from the reduced run, which are marginally of β_p from the distribution $\pi(w, z|\delta_p^*, y)$. These draws are utilized to estimate the posterior ordinate as,

$$\hat{\pi}(\beta_p^*|\delta_p^*, y) = H^{-1} \sum_{h=1}^H \pi(\beta_p^*|\delta_p^*, w^{(h)}, z^{(h)}, y). \tag{14}$$

Substituting the two density estimates given by equations (13) and (14) into equation (10), an estimate of the logarithm of marginal likelihood for OR_I model is obtained as,

$$\ln \hat{m}(y) = \ln f(y|\beta_p^*, \delta_p^*) + \ln \left[\pi(\beta_p^*) \pi(\delta_p^*) \right] - \ln \left[\hat{\pi}(\delta_p^*|y) \, \hat{\pi}(\beta_p^*|\delta_p^*, y) \right], \tag{15}$$

where the likelihood $f(y|\beta_p^*, \delta_p^*)$ and prior densities are evaluated at $\Theta^* = (\beta_p^*, \delta_p^*)$.

Marginal likelihood for OR_{II} model

We know from Section 2.2 that the OR_{II} model is estimated by Gibbs sampling and hence we follow Siddhartha Chib (1995) to compute the marginal likelihood. The Gibbs sampler consists of four conditional posterior densities given by $\pi(\beta_p|\sigma_p, \nu, z)$, $\pi(\sigma_p|\beta_p, \nu, z)$, $\pi(\nu|\beta_p, \sigma_p, z)$, and $\pi(z|\beta_p, \sigma_p, \nu, y)$. However, the variables (ν, z) are latent. So, we integrate them out and write the posterior ordinate as $\pi(\beta_p^*, \sigma_p^*|y) = \pi(\beta_p^*|y) \pi(\sigma_p^*|\beta_p^*, y)$, where the terms on the right hand side can be written as,

$$\pi(\beta_p^*|y) = \int \pi(\beta_p^*|\sigma_p, \nu, z, y) \, \pi(\sigma_p, \nu, z|y) \, d\sigma_p \, d\nu \, dz,$$

$$\pi(\sigma_p^*|\beta_p^*, y) = \int \pi(\sigma_p^*|\beta_p^*, \nu, z, y) \, \pi(\nu, z|\beta_p^*, y) \, d\nu \, dz,$$

and $\Theta^* = (\beta_p^*, \sigma_p^*)$ denotes a high density point, such as the mean or the median.

The posterior ordinate $\pi(\beta_p^*|y)$ can be estimated as the ergodic average of the conditional posterior density with the posterior draws of (σ_p, ν, z). Therefore, $\pi(\beta_p^*|y)$ is estimated as,

$$\hat{\pi}(\beta_p^*|y) = G^{-1} \sum_{g=1}^G \pi(\beta_p^*|\sigma_p^{(g)}, \nu^{(g)}, z^{(g)}, y). \tag{16}$$

The term $\pi(\sigma_p^*|\beta_p^*, y)$ is a reduced conditional density ordinate and can be estimated with the help of a *reduce run*. So, we generate an additional sample (say another G iterations) of $\{\nu^{(g)}, z^{(g)}\}$ from $\pi(\nu, z|\beta_p^*, y)$ by successively sampling from $\pi(\sigma_p|\beta_p^*, \nu, z)$, $\pi(\nu|\beta_p^*, \sigma_p, z)$, and $\pi(z|\beta_p^*, \sigma_p, \nu, y)$, where

note that β_p is fixed at β_p^* in each conditional density. Next, we use the draws $\{v^{(g)}, z^{(g)}\}$ to compute,

$$\hat{\pi}(\sigma_p^*|\beta_p^*, y) = G^{-1} \sum_{g=1}^{G} \pi(\sigma_p^*|\beta_p^*, v^{(g)}, z^{(g)}, y). \tag{17}$$

which is a simulation consistent estimate of $\pi(\sigma_p^*|\beta_p^*, y)$.

Substituting the two density estimates given by equations (16) and (17) into equation (10), we obtain an estimate of the logarithm of marginal likelihood,

$$\ln \hat{m}(y) = \ln f(y|\beta_p^*, \sigma_p^*) + \ln \left[\pi(\beta_p^*) \pi(\sigma_p^*) \right] - \ln \left[\hat{\pi}(\beta_p^*|y) \, \hat{\pi}(\sigma_p^*|\beta_p^*, y) \right], \tag{18}$$

where the likelihood function and prior densities are evaluated at $\Theta^* = (\beta_p^*, \sigma_p^*)$. Here, the likelihood function has the expression,

$$f(y|\beta_p^*, \sigma_p^*) = \prod_{i=1}^{n} \prod_{j=1}^{3} \left[F_{AL}\left(\frac{\gamma_j - x_i'\beta_p^*}{\sigma_p^*} \right) - F_{AL}\left(\frac{\gamma_{j-1} - x_i'\beta_p^*}{\sigma_p^*} \right) \right]^{I(y_i=j)},$$

where the cut-points γ are known and fixed for identification reasons as explained in Section 2.2.

4 Simulation studies

This section explains the data generating process for simulation studies, the functions offered in the **bqror** package, and usage of the functions for estimation and inference in ordinal quantile models.

OR_I model: data, functions, and outputs

Data Generation: The data for simulation study in OR_I model is generated from the regression: $z_i = x_i'\beta + \epsilon_i$, where $\beta = (-4, 5, 6)$, $(x_2, x_3) \sim U(0, 1)$, and $\epsilon_i \sim AL(0, \sigma = 1, p)$ for $i = 1, \ldots, n$. Here, U and AL denote a uniform distribution and an asymmetric Laplace distribution, respectively. The z values are continuous and are classified into 4 categories based on the cut-points $(0, 2, 4)$ to generate ordinal values of y, the outcome variable. We follow the above procedure to generate 3 data sets each with 500 observations (i.e., $n = 500$). The 3 data sets correspond to the quantile p equaling 0.25, 0.50, and 0.75, and are stored as data25j4, data50j4, and data75j4, respectively. Note that the last two letters in the name of the data (i.e., j4) denote the number of unique outcomes in the y variable.

We now describe the major functions for Bayesian quantile estimation of OR_I model, demonstrate their usage, and note the inputs and outputs of each function.

quantregOR1: The quantregOR1 is the primary function for estimating Bayesian quantile regression in ordinal models with 3 or more outcomes (i.e., OR_I model) and implements Algorithm~1. In the code snippet below, we first read the data and then do the following: define the ordinal response variable (y) and covariate matrix (xMat), specify the number of covariates (k) and number of outcomes (J), and set the prior means and covariances for β_p and δ_p. We then call the quantregOR1 function and specify the inputs: ordinal outcome variable (y), covariate matrix including a column of ones (xMat), prior mean (b0) and prior covariance matrix (B0) for the regression coefficients, prior mean (d0) and prior covariance matrix (D0) for the transformed cut-points, burn-in size (burn), post burn-in size (mcmc), quantile (p), the tuning factor (tune) to adjust the MH acceptance rate, and the auto correlation cutoff value (accutoff). The last input verbose, when set to TRUE (FALSE) will (not) print the summary output.

In the code below, we use a diffuse normal prior $\beta_p \sim N(0_k, 10 * I_k)$ where 0_k and I_k are matrices of dimension $k \times 1$ and $k \times k$, respectively. The prior distribution can be further diffused (i.e., made less informative) by increasing 10 to say 100. Besides, the prior variance for δ_p should be small, such as $0.25 * I_{J-2}$, since the distribution is on the transformed cut-points, which is in the logarithmic scale. If there is a need for prior elicitation, they can be designed from previous subject based knowledge or by estimating the model on a training sample and then using the results to form prior distributions (See Greenberg (2012), for examples.)

```
library('bqror')
data("data25j4")
y <- data25j4$y
xMat <- data25j4$x
k <- dim(xMat)[2]
```

```
J <- dim(as.array(unique(y)))[1]
b0 <- array(rep(0, k), dim = c(k, 1))
B0 <- 10*diag(k)
d0 <- array(0, dim = c(J-2, 1))
D0 <- 0.25*diag(J - 2)
modelORI <- quantregOR1(y = y, x = xMat, b0, B0, d0, D0, burn = 1125,
                  mcmc = 4500, p = 0.25, tune = 1, accutoff = 0.5,
                  verbose = TRUE)
```

[1] Summary of MCMC draws:

	Post Mean	Post Std	Upper Credible	Lower Credible	Inef Factor
beta_1	-3.6434	0.4293	-2.8369	-4.5073	2.3272
beta_2	4.8283	0.5597	5.9577	3.7624	2.4529
beta_3	5.9929	0.5996	7.2474	4.8819	2.7491
delta_1	0.7152	0.1110	0.9616	0.5004	3.2261
delta_2	0.7456	0.0940	0.9281	0.5543	2.1497

```
[1] MH acceptance rate: 31.82%
[1] Log of Marginal Likelihood: -545.72
[1] DIC: 1060.56
```

The outputs from the `quantregOR1` function are the following quantities: `summary.bqrorOR1`, `postMeanbeta`, `postMeandelta`, `postStdbeta`, `postStddelta`, `gammacp`, `acceptancerate`, `logMargLike`, `dicQuant`, `ineffactor`, `betadraws`, and `deltadraws`. A detailed description of each output is presented in the **bqror** package help file. In the summary, we report the posterior mean, posterior standard deviation, 95% posterior credible (or probability) interval, and the inefficiency factor of the quantile regression coefficients β_p and transformed cut-points δ_p. These quantities are presented in the last five columns and labeled appropriately.

The posterior means of (β_p, δ_p) are close to the true values used to generate the data with small standard deviations. So, `quantregOR1` function is successful in recovering the true values of the parameters. The inefficiency factor is computed from the MCMC samples using the batch-means method (Greenberg 2012). They indicate the cost of working with correlated samples. For example, an inefficiency factor of 3 implies that it takes 3 correlated draws to get one independent draw. As such, low inefficiency factor indicates better mixing and a more efficient MCMC algorithm. Inefficiency factor also bears a direct relationship with effective sample size, where the latter can be obtained as the total number of (post burn-in) MCMC draws divided by the inefficiency factor (Siddhartha Chib 2012). The inefficiency factors for (β_p, δ_p) are stored in the object `modelOR1` of class `bqrorOR1` and can be obtained by calling `modelORI$ineffactor`.

The third last row displays the random-walk MH acceptance rate for δ_p, for which the preferred acceptance rate is around 30 percent. The last two rows present the model comparison measures, the logarithm of marginal likelihood and the DIC. The logarithm of marginal likelihood is computed using the MCMC outputs from the complete and reduced runs as explained in Section 3, while the principle for computing the DIC is presented in Gelman et al. (2013). For any two competing models at the same quantile, the model with a higher (lower) marginal likelihood (DIC) provides a better model fit.

While the two model comparison measures are printed as part of the summary output, they can also be called individually. For example, the logarithm of marginal likelihood can be obtained by calling `modelORI$logMargLike`. Whereas, the DIC can be obtained by calling `modelORI$dicQuant$DIC`. Two more quantities that are part of the object `modelORI$dicQuant` are effective number of parameters denoted p_D and the deviance computed at the posterior mean. They can be obtained by calling `modelORI$dicQuant$pd` and `modelORI$dicQuant$dev`, respectively. Besides, post estimation, one may also use the command `modelORI$summary` or `summary.bqrorOR1(modelORI)` to extract and print the summary output.

covEffectOR1: The function `covEffectOR1` computes the average covariate effect for different outcomes of OR_I model at a specified quantile, marginally of the parameters and the remaining covariates. While a demonstration of this function is best understood in a real-life study and is presented in the application section, here we present the mechanics behind the computation of average covariate effect.

Suppose, we want to compute the average covariate effect when the l-th covariate $\{x_{i,l}\}$ is set to the values a and b, denoted as $\{x_{i,l}^a\}$ and $\{x_{i,l}^b\}$, respectively. We split the covariate and parameter vectors as follows: $x_i^a = (x_{i,l}^a, x_{i,-l})$, $x_i^b = (x_{i,l}^b, x_{i,-l})$, and $\beta_p = (\beta_{p,l}, \beta_{p,-l})$, where $-l$ in the subscript denotes all covariates (parameters) except the l-th covariate (parameter). We are interested in the distribution of the difference $\{\Pr(y_i = j | x_{i,l}^b) - \Pr(y_i = j | x_{i,l}^a)\}$ for $1 \leq j \leq J$, marginalized over $\{x_{i,-l}\}$

and (β_p, δ_p), given the data $y = (y_1, \cdots, y_n)'$. We marginalize the covariates using their empirical distribution and the parameters based on the posterior distribution.

To obtain draws from the distribution $\{\Pr(y_i = j|x_{i,l}^b) - \Pr(y_i = j|x_{i,l}^a)\}$, we use the method of composition (See Siddhartha Chib and Jeliazkov (2006), Rahman and Vossmeyer (2019), and Bresson, Lacroix, and Rahman (2021), for additional details). In this process, we randomly select an individual, extract the corresponding sequence of covariate values, draw a value (β_p, δ_p) from the posterior distributions, and evaluate $\{\Pr(y_i = j|x_i^b, \beta_p, \delta_p) - \Pr(y_i = j|x_i^a, \beta_p, \delta_p)\}$, where,

$$\Pr(y_i = j|x_i^q, \beta_p, \delta_p)$$
$$= F_{AL}(\gamma_{p,j} - x_{i,l}^q \beta_{p,l} - x_{i,-l}' \beta_{p,-l}) - F_{AL}(\gamma_{p,j-1} - x_{i,l}^q \beta_{p,l} - x_{i,-l}' \beta_{p,-l}),$$

for $q = b, a$ and $1 \le j \le J$. This process is repeated for all remaining individuals and other MCMC draws from the posterior distribution. Finally, the average covariate effect (ACE) for outcome j (for $1 \le j \le J$) is calculated as the mean of the difference in pointwise probabilities,

$$\frac{1}{M}\frac{1}{n} \sum_{m=1}^{M} \sum_{i=1}^{n} \left[\Pr(y_i = j|x_i^b, \beta_p^{(m)}, \delta_p^{(m)}) - \Pr(y_i = j|x_i^a, \beta_p^{(m)}, \delta_p^{(m)}) \right], \tag{19}$$

where $(\beta_p^{(m)}, \delta_p^{(m)})$ is an MCMC draw of (β_p, δ_p), and M is the number of post burn-in MCMC draws.

OR_{II} model: data, function, and outputs

Data Generation: The data generating process for the OR_{II} model closely resembles that of OR_I model. In particular, 500 observations are generated for each value of p from the regression model: $z_i = x_i'\beta + \epsilon_i$, where $\beta = (-4, 6, 5)$, $(x_2, x_3) \sim U(0,1)$ and $\epsilon_i \sim AL(0, \sigma = 1, p)$ for $i = 1, \ldots, n$. The continuous values of z are classified based on the cut-points $(0,3)$ to generate 3 ordinal values for y, the outcome variable. Once again, we choose p equal to 0.25, 0.50, and 0.75 to generate three samples from the model, which are referred to as data25j3, data50j3, and data75j3, respectively. The last two letters in the data names (i.e., j3) denote the number of unique outcomes in the y variable.

quantregOR2: The function quantregOR2 implements Algorithm~2 and is the main function and for estimating Bayesian quantile regression in OR_{II} model i.e., an ordinal model with exactly 3 outcomes. In the code snippet below, we first read the data, define the required quantities, and then call the quantregOR2 for estimating the quantile model. The function inputs are as follows: the ordinal outcome variable (y), covariate matrix including a column of ones (xMat), prior mean (b0) and prior covariance matrix (B0) for β_p, prior shape (n0) and scale (d0) parameters for σ_p, second cut-point (gammacp2), burn-in size (burn), post burn-in size (mcmc), quantile (p), auto correlation cutoff value (accutoff), and the verbose option which when set to TRUE (FALSE) will (not) print the outputs. We use a relatively diffuse prior distributions on (β_p, σ_p) to allow the data to speak for itself.

```
library('bqror')
data("data25j3")
y <- data25j3$y
xMat <- data25j3$x
k <- dim(xMat)[2]
b0 <- array(rep(0, k), dim = c(k, 1))
B0 <- 10*diag(k)
n0 <- 5
d0 <- 8
modelORII <- quantregOR2(y = y, x = xMat, b0, B0, n0, d0, gammacp2 = 3,
                         burn = 1125, mcmc = 4500, p = 0.25, accutoff = 0.5,
                         verbose = TRUE)

[1] Summary of MCMC draws:


       Post Mean Post Std Upper Credible Lower Credible  Inef Factor
beta_1   -3.8900   0.4560        -3.0578        -4.8346      2.4784
beta_2    5.8257   0.5341         6.9263         4.8243      2.1231
beta_3    4.7194   0.5227         5.7502         3.7194      2.2693
sigma     0.8968   0.0763         1.0587         0.7626      2.4079

[1] Log of Marginal Likelihood: -404.34
[1] DIC: 790.72
```

The outputs from the `quantregOR2` function are the following: `summary.bqrorOR2`, `postMeanbeta`, `postMeansigma`, `postStdbeta`, `postStdsigma`, `logMargLike`, `dicQuant`, `ineffactor`, `betadraws`, and `sigmadraws`. A detailed description of each output is presented in the **bqror** package help file. Once again, we summarize the MCMC draws by reporting the posterior mean, posterior standard deviation, 95% posterior credible (or probability) interval, and the inefficiency factor of the quantile regression coefficients β_p and scale parameter σ_p. These quantities are presented in the last five columns and labeled appropriately.

The output also exhibits the logarithm of marginal likelihood and the DIC for OR_{II} model. For two or more models at the same quantile, the model with a higher (lower) marginal likelihood (DIC) indicates a better model fit. The logarithm of marginal likelihood is computed using the Gibbs output as explained in Section 3, while the computation of DIC follows Gelman et al. (2013).

The two model comparison measures can also be printed individually. For example, the logarithm of marginal likelihood can be obtained by calling `modelORII$logMargLike` and the DIC by calling `modelORII$dicQuant$DIC`. The effective number of parameters p_D and the deviance computed at the posterior mean can obtained by calling `modelORII$dicQuant$pd` and `modelORII$dicQuant$dev`, respectively. Lastly, one may use the command `modelORII$summary` or `summary.bqrorOR2(modelORII)` to extract and print the summary output.

covEffectOR2: The function `covEffectOR2` computes the average covariate effect for the 3 outcomes of OR_{II} model at a specified quantile, marginally of the parameters and remaining covariates. The principle underlying the computation is analogous to that of OR_I model and explained below. An implementation of the function is presented in the tax policy application.

Suppose, we are interested in computing the average covariate effect for the l-th covariate $\{x_{i,l}\}$ for two different values a and b, and split the covariate and parameter vectors as: $x_i^a = (x_{i,l}^a, x_{i,-l})$, $x_i^b = (x_{i,l}^b, x_{i,-l})$, $\beta = (\beta_{p,l}, \beta_{p,-l})$. We are interested in the distribution of the difference $\{\Pr(y_i = j|x_{i,l}^b) - \Pr(y_i = j|x_{i,l}^a)\}$ for $1 \leq j \leq J = 3$, marginalized over $\{x_{i,-l}\}$ and (β_p, σ_p), given the data $y = (y_1, \cdots, y_n)'$. We again employ the method of composition i.e., randomly select an individual, extract the corresponding sequence of covariate values, draw a value (β_p, σ_p) from their posterior distributions, and lastly evaluate $\{\Pr(y_i = j|x_i^b, \beta_p, \sigma_p) - \Pr(y_i = j|x_i^a, \beta_p, \sigma_p)\}$, where

$$\Pr(y_i = j|x_i^q, \beta_p, \sigma_p)$$
$$= F_{AL}\left(\frac{\gamma_{p,j} - x_{i,l}^q \beta_{p,l} - x_{i,-l}' \beta_{p,-l}}{\sigma_p}\right) - F_{AL}\left(\frac{\gamma_{p,j-1} - x_{i,l}^q \beta_{p,l} - x_{i,-l}' \beta_{p,-l}}{\sigma_p}\right),$$

for $q = b, a$, and $1 \leq j \leq J = 3$. This process is repeated for other individuals and the remaining Gibbs draws to compute the ACE for outcome $j (= 1, 2, 3)$ as below,

$$\frac{1}{G}\frac{1}{n}\sum_{g=1}^{G}\sum_{i=1}^{n}\left[\Pr(y_i = j|x_i^b, \beta_p^{(g)}, \sigma_p^{(g)}) - \Pr(y_i = j|x_i^a, \beta_p^{(g)}, \sigma_p^{(g)})\right], \tag{20}$$

where $(\beta_p^{(g)}, \sigma_p^{(g)})$ is a Gibbs draw of (β_p, σ_p) and G is the number of post burn-in Gibbs draws.

5 Applications

In this section, we consider the educational attainment and tax policy applications from Rahman (2016) to demonstrate the real data applications of the proposed **bqror** package. While the educational attainment study displays the implementation of ordinal quantile regression in OR_I model, the tax policy study highlights the use of ordinal quantile regression in OR_{II} model. Data for both the applications are included as a part of the **bqror** package.

Educational attainment

In this application, the goal is to study the effect of family background, individual level variables, and age cohort on educational attainment of 3923 individuals using data from the National Longitudinal Study of Youth (NLSY, 1979) (Jeliazkov, Graves, and Kutzbach 2008; Rahman 2016). The dependent variable in the model, education degrees, has four categories: *(i) Less than high school, (ii) High school degree, (iii) Some college or associate's degree*, and *(iv) College or graduate degree*. A bar chart of the four categories is presented in Figure 1. The independent variables in the model include intercept, square root of family income, mother's education, father's education, mother's working status, gender, race, indicator variables to point whether the youth lived in an urban area or South at the age of 14, and three
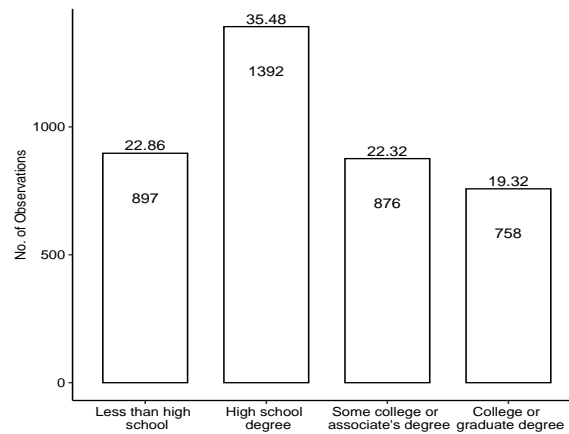
Figure 1: Bar chart showing the different categories of educational attainment. The number of responses (percentage) for each category are shown inside (at the top) of each bar.

indicator variables to indicate the individual's age in 1979 (serves as a control for age cohort effects).

To estimate the Bayesian ordinal quantile model on educational attainment data, we load the **bqror** package, prepare the required inputs and feed them into the quantregOR1 function. Specifically, we specify the outcome variable, covariate matrix (with covariates in order as in Rahman 2016), prior distributions for (β_p, δ_p), burn-in size, number of post burn-in MCMC iterations, quantile value ($p = 0.5$ for this illustration), and the values for tuning factor and autocorrelation cutoff.

```
library('bqror')
data("Educational_Attainment")
data <- na.omit(Educational_Attainment)
data$fam_income_sqrt <- sqrt(data$fam_income)
cols <- c("mother_work","urban","south", "father_educ","mother_educ",
          "fam_income_sqrt","female", "black","age_cohort_2","age_cohort_3",
          "age_cohort_4")
x <- data[cols]
x$intercept <- 1
xMat <- x[,c(12,6,5,4,1,7,8,2,3,9,10,11)]
yOrd <- data$dep_edu_level
k <- dim(xMat)[2]
J <- dim(as.array(unique(yOrd)))[1]
b0 <- array(rep(0, k), dim = c(k, 1))
B0 <- 1*diag(k)
d0 <- array(0, dim = c(J-2, 1))
D0 <- 0.25*diag(J - 2)
p <- 0.5

EducAtt <- quantregOR1(y = yOrd, x = xMat, b0, B0, d0, D0, burn = 1125,
                mcmc = 4500, p, tune=1, accutoff = 0.5, TRUE)

[1] Summary of MCMC draws:
```

	Post Mean	Post Std	Upper Credible	Lower Credible	Inef Factor
intercept	-3.2546	0.2175	-2.8350	-3.6798	2.3143
fam_income_sqrt	0.2788	0.0230	0.3254	0.2337	2.1396
mother_educ	0.1242	0.0190	0.1619	0.0878	1.9499
father_educ	0.1866	0.0154	0.2165	0.1578	2.3062
mother_work	0.0664	0.0821	0.2287	-0.0930	1.9349
female	0.3492	0.0786	0.5070	0.2022	1.9086
black	0.4413	0.0997	0.6400	0.2506	1.9270
urban	-0.0777	0.0971	0.1104	-0.2712	1.9019
south	0.0842	0.0880	0.2529	-0.0895	1.9153
age_cohort_2	-0.0345	0.1192	0.1963	-0.2660	1.7502
age_cohort_3	-0.0426	0.1223	0.2033	-0.2849	1.9053
age_cohort_4	0.4938	0.1212	0.7256	0.2570	1.7512
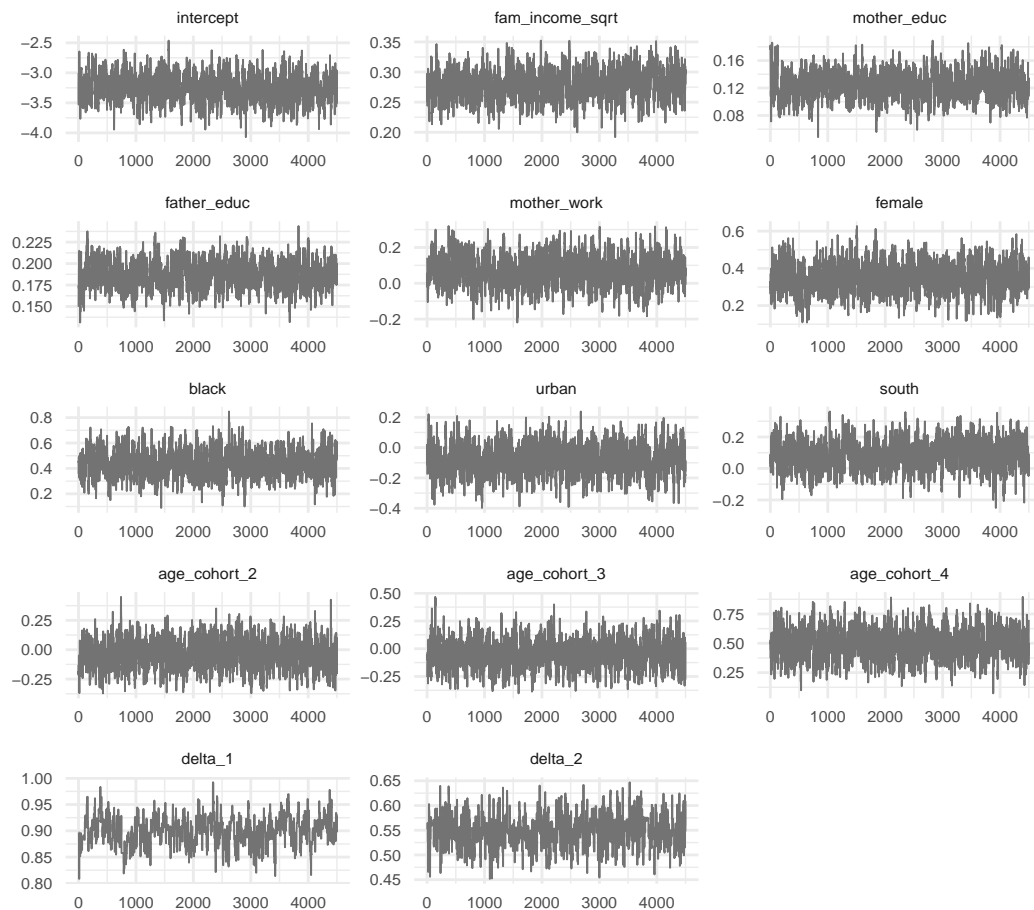
Figure 2: Trace plots of the MCMC draws in the educational attainment study.

```
delta_1              0.8988   0.0276        0.9534        0.8461        4.6186
delta_2              0.5481   0.0313        0.6146        0.4890        3.6003

[1] MH acceptance rate: 26.8%
[1] Log of Marginal Likelihood: -4923.48
[1] DIC: 9781.91
```

The posterior results[1] from the MCMC draws are summarized above. In the summary, we report the posterior mean and posterior standard deviation of the parameters (β_p, δ_p), 95% posterior credible interval, and the inefficiency factors. Additionally, the summary displays the MH acceptance rate of δ_p, the logarithm of marginal likelihood, and the DIC.

```
mcmc <- 4500
burn <- round(0.25*mcmc)
nsim <- mcmc + burn
mcmcDraws <- cbind(t(EducAtt$betadraws), t(EducAtt$deltadraws))
color_scheme_set('darkgray')
bayesplot_theme_set(theme_minimal())
mcmc_trace(mcmcDraws[(burn+1):nsim, ], facet_args = list(ncol = 3))
```

Figure 2 presents the trace plots of the MCMC draws, which can be generated by loading the **bayesplot** package and using the codes presented above. The purpose of trace plots is to show that

---

[1]The results reported here are slightly different from those presented in Rahman (2016). This difference in results, aside from lesser number of MCMC draws, is due to a different approach in sampling from the GIG distribution. Rahman (2016) employed the ratio of uniforms method to sample from the GIG distribution (Dagpunar 2007), while the current paper utilizes the rgig function in the **GIGrvg** package that overcomes the disadvantages associated with sampling using the ratio of uniforms method (see **GIGrvg** documentation for further details). Also, see Devroye (2014) for an efficient sampling technique from a GIG distribution.

the Markov chains have converged to the joint posterior distribution, such as shown in Figure 2. The idea is that the trace plots should show variation around a central value if the chain has converged, rather than drift without settling down.

Next, we utilize the covEffectOR1 function to compute the average covariate effect for a $10,000 increase in family income on the four categories of educational attainment. In general, the calculation of average covariate effect requires creation of either one or two new covariate matrices depending whether the covariate is continuous or indicator (binary), respectively. Since income is a continuous variable, a modified covariate matrix is created by adding $10,000 to each observation of family income. This is xMod2 in the code below and the increased income variable corresponds to $x_{i,l}^b$ for $i = 1, \cdots, n$ in Section 4.1. The second covariate matrix xMod1 (in the code below) is simply the covariate or design matrix xMat; so with reference to Section 4.1, $x_{i,l}^a = x_{i,l}$ for $i = 1, \cdots, n$. When the covariate of interest is an indicator variable, then xMod1 also requires modification as illustrated in the tax policy application.

We now call the covEffectOR1 function and supply the inputs to get the results.

```
xMat1 <- xMat
xMat2 <- xMat
xMat2$fam_income_sqrt <- sqrt((xMat1$fam_income_sqrt)^2 + 10)
EducAttCE <- covEffectOR1(EducAtt, yOrd, xMat1, xMat2, p = 0.5, verbose = TRUE)

[1] Summary of Covariate Effect:

           Covariate Effect
Category_1          -0.0314
Category_2          -0.0129
Category_3           0.0193
Category_4           0.0250
```

The results shows that at the 50th quantile and for a $10,000 increase in family income, the probability of obtaining *less than high school* (*high school degree*) decreases by 3.14 (1.29) percent, while the probability of achieving *some college or associate's degree* (*college or graduate degree*) increases by 1.93 (2.50) percent.

Tax Policy

Here, the objective is to analyze the factors that affect public opinion on the proposal to raise federal taxes for couples (individuals) earning more than $250,000 ($200,000) per year in the United States (US). The proposal was designed to extend the Bush Tax cuts for the lower and middle income classes, but restore higher rates for the richer class. Such a policy is considered pro-growth, since it is aimed to promote economic growth in the US by augmenting consumption among the low-middle income families. After extensive debate, the proposed policy received a two year extension and formed a part of the "Tax Relief, Unemployment Insurance Reauthorization, and Job Creation Act of 2010".
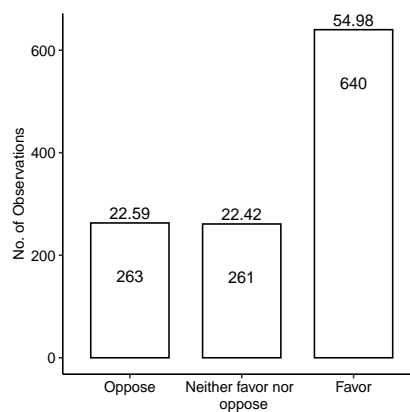
Figure 3: Bar chart for public opinion on tax increase. The number of responses (percentage) for each category are shown inside (at the top) of each bar.

The data for the study was taken from the 2010-2012 American National Election Studies (ANES) on the Evaluations of Government and Society Study 1 (EGSS 1) and contains 1,164 observations. The dependent variable in the model, individual's opinion on tax increase, has 3 categories: *Oppose*, *Neither favor nor oppose*, and *Favor* (See Figure 3). The covariates included in the model are the intercept, indicator variables for employment status, income above $75,000, bachelors' degree, post-bachelors' degree, computer ownership, cell phone ownership, and white race.

To estimate the quantile model on public opinion about federal tax increase, we load the **bqror** package, prepare the data, and provide the inputs into the `quantregOR2` function. Specifically, we define the outcome variable, covariate matrix (with covariates in order as in Rahman 2016), specify the prior distributions for (β_p, σ_p), choose the second cut-off value, burn-in size, number of post burn-in MCMC iterations, and values for quantile ($p = 0.5$ for this illustration) and autocorrelation cutoff.

```
library(bqror)
data("Policy_Opinion")
data <- na.omit(Policy_Opinion)
cols <- c("Intercept","EmpCat","IncomeCat","Bachelors","Post.Bachelors",
          "Computers","CellPhone", "White")
x <- data[cols]
xMat <- x[,c(1,2,3,4,5,6,7,8)]
yOrd <- data$y
k <- dim(x)[2]
b0 <- array(rep(0, k), dim = c(k, 1))
B0 = 1*diag(k)
n0 <- 5
d0 <- 8
FedTax <- quantregOR2(y = yOrd, x = xMat, b0, B0, n0, d0, gammacp2 = 3,
                   burn = 1125, mcmc = 4500, p = 0.5, accutoff = 0.5, TRUE)

[1] Summary of MCMC draws :

                Post Mean Post Std Upper Credible Lower Credible  Inef Factor
Intercept          2.0142   0.4553          2.9071         1.0959       1.4473
EmpCat             0.2496   0.2953          0.8294        -0.3270       1.6760
IncomeCat         -0.5083   0.3323          0.1329        -1.1580       1.6700
Bachelors          0.0809   0.3744          0.8569        -0.6324       1.6726
Post.Bachelors     0.5082   0.4406          1.3964        -0.3435       1.5053
Computers          0.7167   0.3509          1.4078         0.0219       1.4975
CellPhone          0.8464   0.4027          1.6191         0.0444       1.4524
White              0.0627   0.3659          0.7579        -0.6502       1.4931
sigma              2.2205   0.1442          2.5300         1.9552       2.3161

[1] Log of Marginal Likelihood: -1174.11
[1] DIC: 2334.58
```

The results (See Footnote 1) from the MCMC draws are summarized above, where we report the posterior mean, posterior standard deviation, 95% posterior credible interval, and the inefficiency factor of the parameters (β_p, σ_p). Additionally, the summary displays the logarithm of marginal likelihood and the DIC. Figure 4 presents the trace plots of the Gibbs draws, which can be generated by loading the **bayesplot** package and using the codes below.

```
mcmc <- 500
burn <- round(0.25*mcmc)
nsim <- mcmc + burn
mcmcDraws <- cbind(t(FedTax$betadraws), t(FedTax$sigmadraws))
color_scheme_set('darkgray')
bayesplot_theme_set(theme_minimal())
mcmc_trace(mcmcDraws[(burn+1):nsim, ], facet_args = list(ncol = 3))
```

Finally, we utilize the `covEffectOR2` function to demonstrate the calculation of average covariate effect within the OR_{II} framework. Below, we compute the average covariate effect for computer ownership (assume this is the *l*-th variable) on the 3 categories of public opinion about the tax policy. Here, the covariate is an indicator variable since you may either own a computer (coded as 1) or not (coded as 0). So, we need to create two modified covariate matrices. In the code snippet below, the first matrix xMat1 and the second matrix xMat2 are created by replacing the column on computer
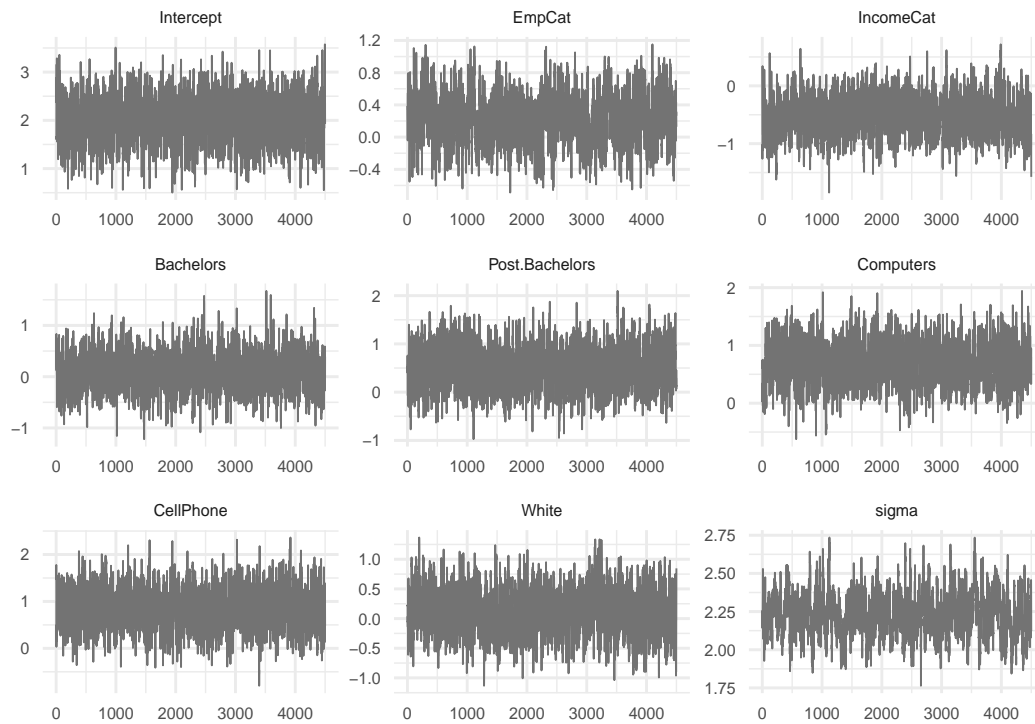
Figure 4: Trace plots of the MCMC draws in the tax policy study.

ownership with a column of zeros and ones, respectively. With reference to notations in Section 4.1 and Section 4.2, $x_{i,l}^a = 0$ and $x_{i,l}^b = 1$ for $i = 1, \cdots, n$. We then call the `covEffectOR2` function and supply the inputs to get the results.

```
xMat1 <- xMat
xMat1$Computers <- 0
xMat2 <- xMat
xMat2$Computers <- 1
FedTaxCE <- covEffectOR2(FedTax, yOrd, xMat1, xMat2, gammacp2 = 3, p = 0.5,
                         verbose = TRUE)

[1] Summary of Covariate Effect:

            Covariate Effect
Category_1          -0.0396
Category_2          -0.0331
Category_3           0.0726
```

The result on covariate effect shows that at the 50th quantile, ownership of computer decreases probability for the first category *Oppose* (*Neither favor nor oppose*) by 3.96 (3.31) percent, and increases the probability for the third category *Favor* by 7.26 percent.

6 Conclusion

A wide class of applications in economics, finance, marketing, and the social sciences have dependent variables which are ordinal in nature (i.e., they are discrete and ordered) and are characterized by an underlying continuous variable. Modeling and analysis of such variables have been typically confined to ordinal probit or ordinal logit models, which offers information on the average probability of outcome variable given the covariates. However, a recently proposed method by Rahman (2016) allows Bayesian quantile modeling of ordinal data and thus presents the tool for a more comprehensive analysis and inference. The prevalence of ordinal responses in applications is well known and hence a software package that allows Bayesian quantile analysis with ordinal data will be of immense interest to applied researchers from different fields, including economics and statistics.

The current paper presents an implementation of the **bqror** package – the only package available for estimation and inference of Bayesian quantile regression in ordinal models. The package offers two MCMC algorithms for estimating ordinal quantile models. An ordinal quantile model with 3 or more outcomes is estimated by a combination of Gibbs sampling and MH algorithm, while estimation of an ordinal quantile model with exactly 3 outcomes utilizes a simpler and computationally faster algorithm that relies solely on Gibbs sampling. For both forms of ordinal quantile models, the **bqror** package also provides functions to calculate the covariate effects (for continuous as well as binary regressors) and measures for model comparison – marginal likelihood and the DIC. The paper explains how to compute the marginal likelihood from the MCMC outputs and recommends its use over the DIC for model comparison. Additionally, this paper demonstrates the usage of functions for estimation and analysis of Bayesian quantile regression with ordinal data on simulation studies and two applications related to educational attainment and tax policy. In the future, the current package will be extended to include ordinal quantile regression with longitudinal data and variable selection in ordinal quantile regression with cross section and/or longitudinal data.

References

Albert, James, and Siddhartha Chib. 1993. "Bayesian Analysis of Binary and Polychotomous Response Data." *Journal of the American Statistical Association* 88 (422): 669–79. https://doi.org/10.1080/01621459.1993.10476321.

———. 2001. "Sequential Ordinal Modeling with Applications to Survival Data." *Biometrics* 57: 829–36. https://www.jstor.org/stable/3068422.

Alhamzawi, Rahim. 2016. "Bayesian Model Selection in Ordinal Quantile Regression." *Computational Statistics and Data Analysis* 103: 68–78. https://doi.org/10.1016/j.csda.2016.04.014.

Alhamzawi, Rahim, and Haithem Taha Mohammad Ali. 2018. "Bayesian Quantile Regression for Ordinal Longitudinal Data." *Journal of Applied Statistics* 45 (5): 815–28. https://doi.org/10.1080/02664763.2017.1315059.

Benoit, Dries F., and Dirk Van den Poel. 2010. "Binary Quantile Regression: A Bayesian Approach Based on the Asymmetric Laplace Distribution." *Journal of Applied Econometrics* 27 (7): 1174–88. https://doi.org/10.1002/jae.1216.

———. 2017. "BayesQR: A Bayesian Approach to Quantile Regression." *Journal of Statistical Software* 76 (7). https://cran.r-project.org/web/packages/bayesQR/index.html.

Bresson, Goerges, Guy Lacroix, and Mohammad Arshad Rahman. 2021. "Bayesian Panel Quantile Regression for Binary Outcomes with Correlated Random Effects: An Application on Crime Recidivism in Canada." *Empirical Economics* 60 (1): 227–59. https://doi.org/10.1007/s00181-020-01893-5.

Casella, George, and Edward I George. 1992. "Explaining the Gibbs Sampler." *The American Statistician* 46 (3): 167–74. https://doi.org/10.1080/00031305.1992.10475878.

Chib, S., and E. Greenberg. 1995. "Understanding the Metropolis-Hastings Algorithm." *The American Statistician* 49: 327–35. https://doi.org/10.1080/00031305.1995.10476177.

Chib, Siddhartha. 1995. "Marginal Likelihood from the Gibbs Output." *Journal of the American Statistical Association* 90 (432): 1313–21. https://doi.org/10.1080/01621459.1995.10476635.

———. 2012. "Introduction to Simulation and MCMC Methods." In *The Oxford Handbook of Bayesian Econometrics*, edited by John Geweke, Gary Koop, and Herman Van Dijk, 183–218. Oxford University Press, Oxford. https://doi.org/10.1093/oxfordhb/9780199559084.013.0006.

Chib, Siddhartha, and Ivan Jeliazkov. 2001. "Marginal Likelihood from the Metropolis-Hastings Output." *Journal of the American Statistical Association* 96 (453): 270–81. https://doi.org/10.1198/016214501750332848.

———. 2006. "Inference in Semiparametric Dynamic Models for Binary Longitudinal Data." *Journal of the American Statistical Association* 101 (474): 685–700. https://doi.org/10.1198/016214505000000871.

Dagpunar, John. 2007. *Simulations and Monte Carlo: With Applications in Finance and MCMC*. John Wiley & Sons Ltd., UK. https://onlinelibrary.wiley.com/doi/book/10.1002/9780470061336.

Davino, Cristino, Marileno Furno, and Domenico Vistocco. 2014. *Quantile Regression: Theory and Applications*. John Wiley & Sons, Chichester. https://doi.org/10.1002/9781118752685.

Devroye, Luc. 2014. "Random Variate Generation for the Generalized Inverse Gaussian Distribution." *Statistics and Computing* 24 (2): 239–46. https://doi.org/10.1007/s11222-012-9367-z.

Furno, Marilena, and Domenico Vistocco. 2018. *Quantile Regression: Estimation and Simulation*. John Wiley & Sons, New Jersey. https://doi.org/10.1002/9781118863718.

Gelman, Andrew, John B. Carlin, Hal S. Stern, David B. Dunson, Aki Vehtari, and Donald B. Rubin. 2013. *Bayesian Data Analysis*. Chapman & Hall, New York. https://doi.org/10.1201/b16018.

Ghasemzadeh, Siamak, Mojtaba Ganjali, and Taban Baghfalaki. 2018. "Bayesian Quantile Regression for Analyzing Ordinal Longitudinal Responses in the Presence of Non-Ignorable Missingness." *METRON* 76 (3): 321–48. https://doi.org/10.1007/s40300-018-0136-4.

———. 2020. "Bayesian Quantile Regression for Joint Modeling of Longitudinal Mixed Ordinal Continuous Data." *Communications in Statistics − Simulation and Computation* 49 (2): 1375–95. https://doi.org/10.1080/03610918.2018.1484482.

Greenberg, Edward. 2012. *Introduction to Bayesian Econometrics*. 2nd Edition, Cambridge University Press, New York. https://doi.org/10.1017/CBO9780511808920.

Greene, William H., and David A. Hensher. 2010. *Modeling Ordered Choices: A Primer*. Cambridge University Press, Cambridge. https://doi.org/10.1017/CBO9780511845062.

Jeliazkov, Ivan, Jennifer Graves, and Mark Kutzbach. 2008. "Fitting and Comparison of Models for Multivariate Ordinal Outcomes." *Advances in Econometrics: Bayesian Econometrics* 23: 115–56. https://doi.org/10.1016/S0731-9053(08)23004-5.

Jeliazkov, Ivan, and Mohammad Arshad Rahman. 2012. "Binary and Ordinal Data Analysis in Economics: Modeling and Estimation." In *Mathematical Modeling with Multidisciplinary Applications*, edited by Xin She Yang, 123–50. John Wiley & Sons Inc., New Jersey. https://doi.org/10.1002/9781118462706.ch6.

Johnson, Valen E., and James H. Albert. 2000. *Ordinal Data Modeling*. Springer, New York. https://doi.org/10.1007/b98832.

Koenker, Roger. 2005. *Quantile Regression*. Cambridge University Press, Cambridge. https://doi.org/10.1257/jep.15.4.143.

Koenker, Roger, and G. W. Bassett. 1978. "Regression Quantiles." *Econometrica* 46 (1): 33–50. https://doi.org/10.2307/1913643.

Kordas, Gregory. 2006. "Smoothed Binary Regression Quantiles." *Journal of Applied Econometrics* 21 (3): 387–407. https://doi.org/10.1002/jae.843.

Kozumi, Hideo, and Genya Kobayashi. 2011. "Gibbs Sampling Methods for Bayesian Quantile Regression." *Journal of Statistical Computation and Simulation* 81 (11): 1565–78. https://doi.org/10.1080/00949655.2010.496117.

Mukherjee, Deep, and Mohammad Arshad Rahman. 2016. "To Drill or Not to Drill? An Econometric Analysis of US Public Opinion." *Energy Policy* 91: 341–51. https://doi.org/10.1016/j.enpol.2015.11.023.

Poirier, Dale J. 1995. *Intermediate Statistics and Econometrics*. The MIT Press, Cambridge. https://mitpress.mit.edu/9780262660945/intermediate-statistics-and-econometrics/.

Rahman, Mohammad Arshad. 2016. "Bayesian Quantile Regression for Ordinal Models." *Bayesian Analysis* 11 (1): 1–24. https://doi.org/10.1214/15-BA939.

Rahman, Mohammad Arshad, and Shubham Karnawat. 2019. "Flexible Bayesian Quantile Regression in Ordinal Models." *Advances in Econometrics* 40B: 211–51. https://doi.org/10.1108/S0731-90532019000040B011.

Rahman, Mohammad Arshad, and Angela Vossmeyer. 2019. "Estimation and Applications of Quantile Regression for Binary Longitudinal Data." *Advances in Econometrics* 40B: 157–91. https://doi.org/10.1108/S0731-90532019000040B009.

Spiegelhalter, David J., Nicola G. Best, Bradley P. Carlin, and Angelika Van Der Linde. 2002. "Bayesian Measures of Model Complexity and Fit." *Journal of the Royal Statistical Society − Series B* 64 (4): 583–639. https://doi.org/10.1111/1467-9868.00353.

Tian, Yu-Zhu, Man-Lai Tang, Wai-Sum Chan, and Mao-Zai Tian. 2021. "Bayesian Bridge-Randomized Penalized Quantile Regression for for Ordinal Longitudinal Data, with Application to Firm's Bond Ratings." *Computational Statistics* 36: 1289–319. https://doi.org/10.1007/s00180-020-01037-4.

Yu, Keming, and Rana A. Moyeed. 2001. "Bayesian Quantile Regression." *Statistics and Probability Letters* 54 (4): 437–47. https://doi.org/10.1016/S0167-7152(01)00124-9.

Prajual Maheshwari
Quantitative Researcher, Ogha Research
2123, 14th Main Road, HAL 3rd Stage, Kodihalli, Bengaluru, Karnataka, India
https://prajual.netlify.app
prajual1391@gmail.com

Mohammad Arshad Rahman
Department of Economic Sciences, Indian Institute of Technology Kanpur
Room 672, Faculty Building, Indian Institute of Technology Kanpur, India
https://www.arshadrahman.com
ORCiD: 0000-0001-8434-0042
marshad@iitk.ac.in, arshadrahman25@gmail.com