



Bayesian Adaptive Bridge Regression for Ordinal Models with an Application

Dhiyaa Hazem Qasem Al-Jabri*, Rahim Jabbar Al-Hamzawi

Department of Statistics, College of Administration and Economics, University of Al-Qadisiyah, Iraq

Received: 16/12/ 2019

Accepted: 15/ 3/2020

Abstract

In this article, we propose a Bayesian Adaptive bridge regression for ordinal model. We developed a new hierarchical model for ordinal regression in the Bayesian adaptive bridge. We consider a fully Bayesian approach that yields a new algorithm with tractable full conditional posteriors. All of the results in real data and simulation application indicate that our method is effective and performs very good compared to other methods. We can also observe that the estimator parameters in our proposed method, compared with other methods, are very close to the true parameter values.

Keywords: Bayesian Bridge, Variable Selection, Ordinal Model, Scale Mixture Of Uniform.

انحدار بيزين برج التكيفي للنماذج الترتيبية مع التطبيق

ضياء حازم قاسم الجابري*، رحيم جبار ظاهر الحمزاوي

قسم الاحصاء، كلية الادارة والاقتصاد، جامعة القادسيه، الديوانيه، العراق

الخلاصه

في هذه المقالة، نقترح طريقة انحدار البرج البيزي التكيفي للنموذج الانحدار الترتيبية (الرتبي). قمنا بتطوير نموذج هرمي جديد للانحدار الترتيبية في البرج البيزي التكيفي. النهج البيزي بالكامل ينتج عنه خوارزمية جديدة. في ورقتنا هذه جميع النتائج في التطبيق الحقيقي للبيانات وأيضاً من خلال المحاكاة تشير إلى أن طريقتنا فعالة وتؤدي أداءً جيداً مقارنة بالطرق الأخرى، ويمكننا ملاحظة المعلمات المقدره في طريقتنا المقترحة قريبة جداً مع قيم المعلمات الحقيقية مقارنة بالطرق الأخرى، وأيضاً المقدرات الأخرى تكون افضل في طريقتنا مقارنة مع طرق المقارنة.

1- Introduction

In statistical learning, there are two important goals; First, knowledge of important variables in the model. Second, ensuring the accuracy of the high prediction. Determination of the important predictors leads to reinforcing the prediction performance for the fitted model [1]. The linear regression (LR) model, where p is the predictors and n is the observations, is written as follows:

$$y_i = \mathbf{X}_i \boldsymbol{\beta} + \varepsilon_i. \quad (1)$$

Where $y_i = (y_i, \dots, y_n)^t$ is a vector of centered responses, $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_{p-1})^t$ is a vector of the coefficient to be estimated, $\mathbf{X} = (x_1, x_2, \dots, x_p)$ is the independent variable and $\varepsilon_i = (\varepsilon_1, \dots, \varepsilon_n)^t$ is a vector of error with a mean of zero and a variance of one. Mallick and Yi [2] proposed that the

*Email: thya223@gmail.com

Bayesian bridge regression (BBR) was better in estimating than each of Lasso, ridge regression (RR) and bridge regression, while the estimator of BBR can be written as follow:

$$\min_{\beta} (\mathbf{y} - \mathbf{X}\beta)^t (\mathbf{y} - \mathbf{X}\beta) + \lambda \sum_{j=1}^p |\beta_j|^\delta, \quad \lambda, \delta \geq 0. \quad (2)$$

The above formula contains special cases: RR when ($\delta = 2$), the Lasso estimator when ($\delta = 1$) and the best subset selection if ($\delta = 0$) [3,4] .

Recently, researches showed the regularization approaches which uses the variable selection (VS) and found that the simultaneous estimation will be effective. These methods improve the prediction accuracy in the regression. In our current paper, we propose the Bayesian adaptive bridge regression for ordinal model (BABROM). This method included the desired properties such as sparsity, oracle and unbiasedness when ($0 < \delta < 1$).

The ordinal data are naturally arranged categories and a type of the statistical data. This data is found in many fields such as climatology, political economy, economics, social sciences, psychology and medicine [5-8]. One of the examples of ordinal data is the intelligence level of students (weak, medium, high). The categories weak, medium and high take the values of the order 1, 2, and 3, respectively. The high level is not a multiplier of the medium level. The outcome variable y_i takes one value of s values, where $s = 1, \dots, S$. In this example, the last category equals $S = 3$. The ordinal data model has the general shape of:

$$pr(y_i = s) = F(\mathbf{X}_i^t \beta), \quad i = 1, \dots, n, \quad (3)$$

where F is the link function [9] . One of the main regression problems is when the number of covariates is increasing. The regression model may be containing many unimportant variables. Therefore, we will focus on the process of VS to get the appropriate model. The process of VS provides a perfect operator for selecting the effective variables and for estimating the parameters [10]. One of the most popular method's criteria used for VS is the Akaike information criterion (AIC) that can be written as follows:

$$AIC = 2p - 2 \log D, \quad (4)$$

where (D) is the probability function that is estimated in a maximum likelihood estimation, and p is the number of parameters in the model [11]. In this criterion, the best model of a set of candidate models is that with minimum AIC value [12]. When p size is large, the model will be inconsistent [13]. Therefore, we will use the Bayesian information criterion (BIC) that can be written as follows:

$$BIC = p \log n - 2 \log D, \quad (5)$$

we use this criterion to get the appropriate model with a probability of 1 [11,13].

We organize this paper as follows: we will show the Bayesian adaptive bridge for ordinal model in section Two. In section Three, we show the Bayesian inference, prior elicitation, hierarchical representation and full conditional distributions (FCD). We elucidate the computation in section Four. We apply the simulation study in section Five and real data application in section Six. The conclusions are presented in section Seven.

2-Bayesian adaptive bridge regression for ordinal model

In this section, we introduce the BABROM using different penalty parameters λ_j . The formula of Bayesian adaptive bridge regression can be written as follows:

$$\min_{\beta} (\mathbf{h} - \mathbf{X}\beta)^t (\mathbf{h} - \mathbf{X}\beta) + \sum_{j=1}^p \lambda_j |\beta_j|^\delta, \quad \lambda_j, \delta \geq 0, \quad (6)$$

where $\lambda_j = \lambda |\hat{\beta}_j^{\text{OLS}}|^{-c}$ for $j = 1, \dots, p$, c is a positive constant [14]. Here, λ_j will be different values of tuning parameters, where every regression parameter is a multiplication with different penalty parameter. We will give a large penalty parameter for the not important variables [1,15,16]. The value of δ in this paper is 0.5.

The ordinal data analysis is distinguished by relative simplicity in the frequentist approach. Although the approximate theory for ordinal regression (OR) model has been well studied, the Bayesian method enables exact estimation even when p is greater than n [17].

In a simple univariate case, the response variable y_i takes one of the s ordered values, where $s = 1, \dots, S$. If there is a normal distribution (ND) with the cumulative distribution function (cdf) Φ , we can write the probability (\mathbf{y}) to be equal to s category through the following [18]:

$$Pr(\mathbf{y} = s) = \Phi(\gamma_{s+1}) - \Phi(\gamma_s), \tag{7}$$

where $\gamma_0, \dots, \gamma_S$ are cut-points, which coordinates satisfy $-\infty = \gamma_0 < \gamma_1 < \dots < \gamma_{S-1} < \gamma_S = \infty$. Here γ_{s-1} is the lower bound and γ_s is the upper bound of the interval corresponding to response s [19]; [7,19]. We can motivate the problem when assuming that the latent variable h_i depends on a p -vector for covariates (\mathbf{x}_i) through the model $h_i = \mathbf{x}_i^t \boldsymbol{\beta} + \varepsilon_i$, and the response variable is:

$$y_i = s \text{ if } \gamma_{s-1} < h_i < \gamma_s. \tag{8}$$

The probability of $y_i = s$ is conditional on $\boldsymbol{\beta}$ and $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_S)^t$ and is given by $p_r(y_i = s | \boldsymbol{\beta}, \boldsymbol{\gamma}) = p_r([\gamma_{s-1} < h_i] \cap [h_i \leq \gamma_s])$. So, we obtain:

$$p_r(y_i = s | \boldsymbol{\beta}, \boldsymbol{\gamma}) = \Phi(\gamma_s - \mathbf{x}_i^t \boldsymbol{\beta}) - \Phi(\gamma_{s-1} - \mathbf{x}_i^t \boldsymbol{\beta}). \tag{9}$$

Given that $\mathbf{y} = (y_1, \dots, y_n)^t$, we can write the likelihood function for the model as:

$$f(\mathbf{y} | \boldsymbol{\beta}, \boldsymbol{\gamma}) = \prod_{i=1}^n \prod_{s=1}^S [\Phi(\gamma_s - \mathbf{x}_i^t \boldsymbol{\beta}) - \Phi(\gamma_{s-1} - \mathbf{x}_i^t \boldsymbol{\beta})]^{1\{y_i=s\}}, \tag{10}$$

The indicator function is $1\{y_i = s\}$ for the event $y_i = s$ [6,20]. When $\gamma_1 = 0$, it leads to the removal of the possibility of change distribution, with no shifting to the probability of observing y_i .

$$\text{where } y_i = \begin{cases} 1 & \text{if } \gamma_0 < h_i < \gamma_1 \\ s & \text{if } \gamma_{s-1} < h_i < \gamma_s, \quad s = 2, \dots, S-1. \\ S & \text{if } \gamma_{S-1} < h_i < \gamma_S \end{cases} \tag{11}$$

By fixing the cut-point and adding to $\gamma_0 = -\infty$ and $\gamma_S = \infty$, the identification problem is typically corrected simply. In Bayesian inference, the steps to determine prior distribution for parameters are very important[7,21]. The prior distribution plays an important role [22]. A prior must be selected with care, because some problems can occur if the prior distribution is used without care [23].

3- Bayesian inference

3-1 Prior elicitation

In this article, similar to Mallick and Yi [2], we consider the conditional generalized Gaussian (GG) prior specification:

$$\pi(\boldsymbol{\beta}) \propto \prod_{j=1}^p \exp\{-\lambda_j (|\beta_j|)^\delta\}. \quad j = 1, \dots, p, \text{ here } \lambda_j, \delta \geq 0. \tag{12}$$

We will solve the problem (formula 6) by using the Gibbs sampler (GS), that confirms construction of a Markov chain, which has a joint posterior for $(\boldsymbol{\beta})$ as a constant distribution [11].

We show a new practice, namely the BABROM, by using Scale Mixture of Uniform (SMU). To proceed with Bayesian analysis, we assume the generalized Gaussian distribution (GGD) which can be written as SMU. Following Mallick and Yi [2], the GGD can be adapted as:

$$\frac{\lambda_j^{\frac{1}{\delta}}}{2\Gamma(\frac{1}{\delta} + 1)} \exp\{-\lambda_j |\beta|^\delta\} = \int_{k > |\beta|^\delta} \frac{\lambda_j^{\frac{1}{\delta} + 1}}{2k^{\frac{1}{\delta}} \Gamma(\frac{1}{\delta} + 1)} k^{\frac{1}{\delta}} e^{-\lambda_j k} dk. \tag{13}$$

$$\text{Let } g = \lambda_j k \Rightarrow k > |\beta|^\delta \Rightarrow \frac{g}{\lambda_j} > |\beta|^\delta \Rightarrow g > \lambda_j |\beta|^\delta,$$

$$\int_{-\frac{g_j}{\lambda_j} < \beta < \frac{g_j}{\lambda_j}} \frac{\lambda_j^{\frac{1}{\delta}}}{2 g^{\frac{1}{\delta}} \Gamma(\frac{1}{\delta} + 1)} g^{(\frac{1}{\delta} + 1) - 1} e^{-g} dg. \tag{14}$$

In practice, we have found that the mixture representation (14) performs better than (13) in sampling the regression coefficients in terms of prediction accuracy.

Albert and Chib (2001) proposed the logarithmic transformation in order sampling of the cut points $\boldsymbol{\gamma}$.

$$\boldsymbol{\psi} = \ln(\gamma_s, \dots, \gamma_{s-1}), \quad 2 \leq s \leq S-1.$$

To eliminate high autocorrelation for cut-points, we use the transformation $\boldsymbol{\psi} = (\psi_2, \dots, \psi_{S-1})^t$ instead of $\boldsymbol{\gamma} = (\gamma_2, \dots, \gamma_{S-1})^t$ to obtain the parameters of the tailored proposal in Metropolis-Hastings

(MH) for $\boldsymbol{\psi}$. We assume that the prior of $\boldsymbol{\psi}$ is $\boldsymbol{\psi} \sim N(\boldsymbol{\psi}_0, \boldsymbol{\zeta}_0)$. The prior of δ is assumed as $\delta \sim Be(\tau, \varphi)$ where Be is a beta distribution (BD).

3-2 Hierarchical Representation

The hierarchical representation for BABROM can be written as follows:

$$\begin{aligned}
 & y_i = s \text{ if } \gamma_{s-1} < h_i \leq \gamma_s \\
 & h_i | \boldsymbol{\beta} \sim N(\mathbf{x}_i^t \boldsymbol{\beta}, 1), \\
 & \boldsymbol{\beta} | \mathbf{g}, \delta \sim \prod_{j=1}^p \text{uniform} \left(-\frac{g_j}{\lambda_j^\delta}, \frac{g_j}{\lambda_j^\delta} \right), \\
 & \mathbf{g} | \lambda, \delta \sim \prod_{j=1}^p \text{Gamma} \left(\frac{1}{\delta} + 1, \lambda_j \right), \\
 & \lambda_j \sim \text{Gamma}(c, d), \\
 & \boldsymbol{\psi} \sim N(\boldsymbol{\psi}_0, \boldsymbol{\zeta}_0), \\
 & \delta \sim Be(\tau, \varphi).
 \end{aligned} \tag{15}$$

Where $\boldsymbol{\lambda}_j = (\lambda_1, \lambda_2, \dots, \lambda_p)$ and $\mathbf{g}_j = (g_1, g_2, \dots, g_p)$.

3-3 Full Conditional Distributions

Under the hierarchical representation in (15), we write the (FCD) for the parameters as follows:

The FCD of \mathbf{h}_i as follows:

$$\begin{aligned}
 & pr(h_i | \boldsymbol{\beta}, \boldsymbol{\gamma}) \\
 & pr(h_i | \boldsymbol{\beta}, \boldsymbol{\gamma}) \propto pr(y_i | h_i, \boldsymbol{\gamma}) pr(h_i | \boldsymbol{\beta}), \\
 & \propto 1\{\gamma_{s-1} < h_i < \gamma_s\} N(h_i; \mathbf{x}_i^t \boldsymbol{\beta}, 1).
 \end{aligned} \tag{16}$$

The (FCD) of $\boldsymbol{\beta}$ is given by:

$$\boldsymbol{\beta} | \mathbf{y}, \mathbf{X}, \mathbf{g}, \lambda \sim N_p(\hat{\boldsymbol{\beta}}_{OLS}(\mathbf{X}^t \mathbf{X})^{-1}) \prod_{j=1}^p I\left\{ |\beta_j|^\delta < \frac{g_j}{\lambda_j} \right\}. \tag{17}$$

The FCD of \mathbf{g} is given by:

$$\mathbf{g} | \mathbf{y}, \mathbf{X}, \boldsymbol{\beta}, \lambda \sim \prod_{j=1}^p \text{Gamma} \left(\frac{1}{\delta} + 1, \lambda_j \right) I\{g_i > \lambda_j |\beta_j|^\delta\}. \tag{18}$$

The FCD of λ_j is given by:

$$\lambda_j | \beta_j \sim \text{Gamma}(c + 1, d) I\left\{ \lambda_j < \frac{g_j}{|\beta_j|} \right\}. \tag{19}$$

The FCD of $\boldsymbol{\psi}$ by:

$$\begin{aligned}
 & p(\boldsymbol{\psi}_s | \mathbf{y}, \mathbf{h}) \propto p(\mathbf{y} | \mathbf{h}, \boldsymbol{\psi}) p(\boldsymbol{\psi}) \\
 & \propto \prod_{i=1}^n \sum_{s=1}^S 1(y_i = s) 1(\psi_{s-1} < h_i < \psi_s) \mathbf{1}(\boldsymbol{\psi} \in T).
 \end{aligned} \tag{20}$$

Where $T = \{(\psi_{\min}, \psi_1, \dots, \psi_{\max}) | \psi_{\min} < \psi_1 < \dots < \psi_{s-1} < \psi_{\max}\}$,

Where $I(\cdot)$ is an indicator function.

To calculate δ , we assign a beta prior (τ, φ) and the FCD of δ is given by:

$$\pi(\delta | \mathbf{h}, \mathbf{X}, \boldsymbol{\beta}, \mathbf{g}, \lambda) \propto \delta^{\tau-1} (1 - \delta)^{\varphi-1} \frac{\lambda^{\frac{p}{\delta}}}{\{\Gamma(\frac{1}{\delta} + 1)\}^p} \exp\left\{ -\lambda \sum_{j=1}^p |\beta_j|^\delta \right\}. \tag{21}$$

4- Computation

Firstly, we construct the Gibbs sampler for BABROM procedure by initiating the initial valuations for parameters $\boldsymbol{\beta}, \boldsymbol{\lambda}, \boldsymbol{\psi}, \delta$ and \mathbf{w} . Then we execute the algorithm as follows:

- a- Sampling the latent variable \mathbf{h}_i from the TND, $NT(\mathbf{x}_i^t \boldsymbol{\beta}, 1)$, determined by (γ_{s-1}, γ_s) , for $i = 1, \dots, n$ and $j = 1, \dots, p$.
- b- Updating $\boldsymbol{\beta}$ from a truncated multivariate normal distribution, where

$$\boldsymbol{\beta} | \mathbf{h}, \mathbf{X}, \mathbf{g}, \lambda \sim N_n(\widehat{\boldsymbol{\beta}}_{OLS}, (X^t X)^{-1}) \prod_{j=1}^p I\left\{-\frac{g_j}{\lambda_j} < \beta_j < \frac{g_j}{\lambda_j}\right\}.$$

- c- Sampling g_j from the truncated gamma distribution, with a shape parameter $(\frac{1}{\delta} + 1)$ and a rate parameter (λ) .
- d- Sampling $\boldsymbol{\psi} | \mathbf{y}, \boldsymbol{\beta}$ marginally out of h . Similar to Jeliazkov *et al.*, [20], we generated the $\boldsymbol{\psi}$. Here $\boldsymbol{\psi}^t \sim ft(\boldsymbol{\psi} | \widehat{\boldsymbol{\psi}}, \widehat{\boldsymbol{\zeta}}, u)$ and ft is a t distribution, where $\widehat{\boldsymbol{\psi}} = \arg \max f(\mathbf{y} | \boldsymbol{\beta}, \boldsymbol{\psi}) f(\boldsymbol{\psi})$, u refers to the degree of freedom and $\widehat{\boldsymbol{\zeta}}$ is the negative inverse Hessian. Given the current values of $\boldsymbol{\psi}$ and the proposed draw for values of $\boldsymbol{\psi}^t$, we return \mathbf{x} with the probability of:

$$g_{MH} = \min\left\{1, \frac{f(\mathbf{y} | \boldsymbol{\beta}, \boldsymbol{\psi}^t) f(\boldsymbol{\beta}, \boldsymbol{\psi}^t) ft(\boldsymbol{\psi} | \widehat{\boldsymbol{\psi}}, \widehat{\boldsymbol{\zeta}}, u)}{f(\mathbf{y} | \boldsymbol{\beta}, \boldsymbol{\psi}) f(\boldsymbol{\beta}, \boldsymbol{\psi}) ft(\boldsymbol{\psi}^t | \widehat{\boldsymbol{\psi}}, \widehat{\boldsymbol{\zeta}}, u)}\right\}$$

- e- Generating λ from the truncated gamma distribution with an $p + c$ shape parameter and a rate parameter d .

$$\pi(\lambda | \mathbf{h}, \mathbf{X}, \boldsymbol{\beta}, \mathbf{g}, \delta) \propto \lambda_j^{(p+q+\frac{p}{\delta})-1} \exp\{-\lambda_j(l + \sum_{j=1}^p |\beta_j|^\delta)\} I\left\{\lambda_j < \frac{g_j}{|\beta_j|^\delta}\right\}.$$

- f- Sampling δ by forming a BD with the parameters (τ, ϕ) , as follows:

$$\pi(\delta | \mathbf{h}, \mathbf{X}, \boldsymbol{\beta}, \mathbf{g}, \lambda) \propto \delta^{\tau-1} (1 - \delta)^{\phi-1} \frac{\lambda_j^{\frac{p}{\delta}}}{\{\Gamma(\frac{1}{\delta} + 1)\}^p} \exp\left\{-\lambda_j \sum_{j=1}^p |\beta_j|^\delta\right\}.$$

The efficient GS is based on this full condition to extract samples from every full (cpd). The process of sampling will continue until all chains converge.

5- Simulation study

In this section, the performance of the proposed method is illustrated by simulations. The proposed method is compared with Bayesian lasso median regression, ordered logistic regression and ordinal probit Regression, as in Jeliazkov *et al.* (2008). These methods are evaluated based on the median of mean absolute deviations (MMAD) over 100 replications and standard deviation (SD). The Bayesian estimates are posterior mean estimates using 11,000 samples of the Gibbs sampler after burn-in the first 1000 samples. In this simulation, we set $\delta = 0.5$.

5-1 Simulation

In this simulation study, we generate 200 observations from the model $w_i = \mathbf{x}_i^t \boldsymbol{\beta} + \boldsymbol{\varepsilon}_i$, where \mathbf{x}_i is a vector of 10 covariates that are simulated from standard multivariate normal distribution, and $\boldsymbol{\varepsilon}_i$ is simulated from standard normal distribution. We set $\boldsymbol{\gamma} = (0, 0.5, 1.5, 2)$ and $\boldsymbol{\beta} = (3, 2, 0, 0, 3, 0, 0, 0, 0, 0)$. In the next simulation, we will consider the following values:

Table 1-Summarized the MMAD and SD results for simulation.

Method	MMAD	SD
BABROM	0.44688	0.35022
ORD (logistic)	3.68581	1.34067
ORD (probit)	0.54191	0.56730

Table 2-List parameter estimates for simulation.

Method	B1	B2	B3	B4	B5	B6	B7	B8	B9	B10
B (True)	3	2	0	0	3	0	0	0	0	0
BABROM	3.110	2.110	0.033	0.022	3.012	-	-	-	0.051	0.004
ORD (logistic)	6.2	4.066	0.051	0.027	6.243	-	-	-	0.147	0.02
ORD (probit)	3.507	2.299	0.042	0.02	3.523	-	-	-	0.08	0.004

We can see from the results of our proposed method that BABROM has smallest MMAD and SD. The estimates of our proposed method are close to the B (True) compared with the other methods. In Table- 3, we list cut-point estimates for simulation.

Method	γ_1	γ_2	γ_3	γ_4
γ (True)	0	0.5	1.5	2
BABROM	-0.161	0.476	1.423	2.052
ORD (logistic)	0.007	0.967	2.984	4.012
ORD (probit)	0.008	0.551	1.685	2.258

The estimates of BABROM are very close to γ (True) compared with the other methods in 3 cut-points of 4.

Table 4-MSE for the cut-points of simulation 2

Method	γ_1	γ_2	γ_3	γ_4
BABROM	0.05611	0.04311	0.0622	0.10011
ORD (logistic)	0.1816	0.40203	2.52907	4.57999
ORD (probit)	0.05629	0.06157	0.13191	0.22358

From Table-4, we can see that our proposed method, BBROMR, has the smallest MSE in all cut-points.

Figure-1 shows that the sampler is moving from a point to another in relatively few steps. Figure-2 illustrates that the conditional posteriors (CP) are in fact the desired stationary truncated univariate normal distributions.

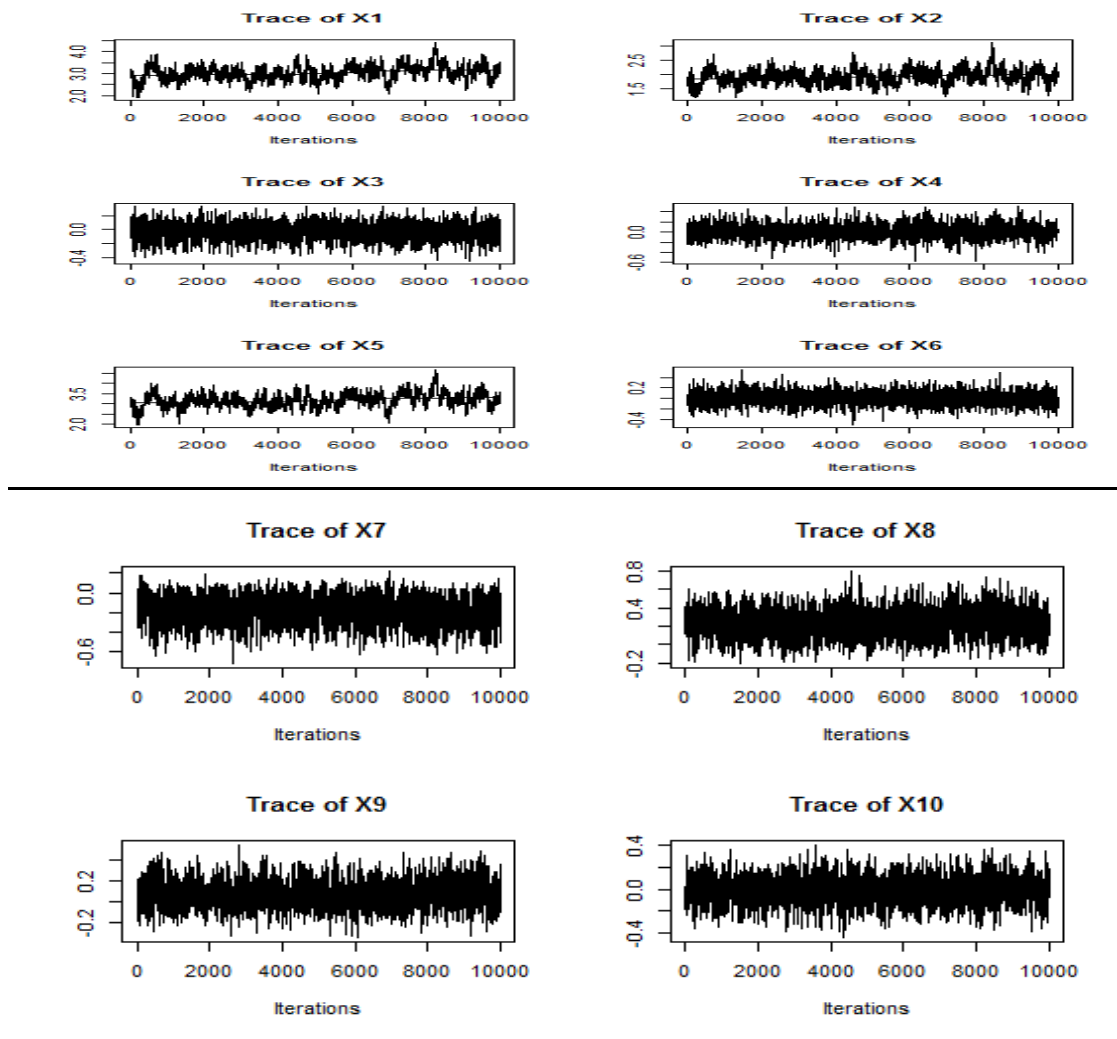


Figure 1-The trace plot in simulation.

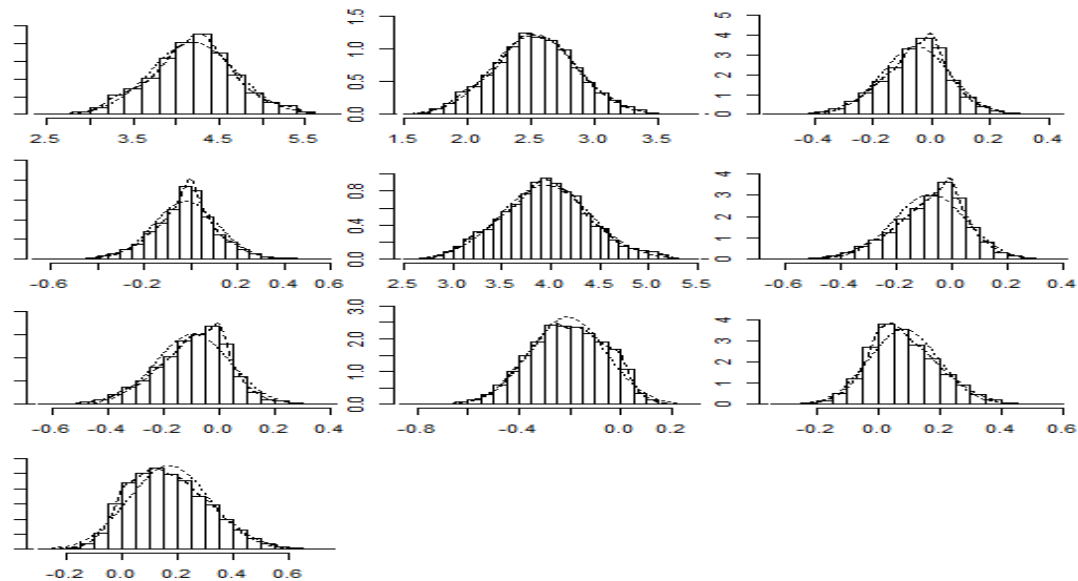


Figure 2-Posterior histograms for the variables in simulation.

6- Real data application

In this section, we apply our proposed method by using real data. The real data was collected through a questionnaire for employees and workers of the Oil Products Distribution Company in Thi-Qar- IRAQ. The questionnaire was about the management of the company and how to deal with employees. Sample size is 150 observations and the number of the covariates is 17.

For real data, we estimated the parameters for our proposed method and the comparison methods. The results indicated that our proposed method is better than the other comparison methods.

Table 5-Parameters estimation

covariate	Method		
	BABROD	ORD (logistic)	ORD (probit)
X_1	0	0	0
X_2	0.02	0.49	0.26
X_3	0	-0.05	-0.03
X_4	0.04	0.63	0.31
X_5	-0.07	-0.53	-0.31
X_6	0	0.12	0.05
X_7	0	0.03	0.02
X_8	-0.04	-0.09	-0.07
X_9	0.09	0.31	0.19
X_{10}	0.06	0.42	0.31
X_{11}	0.03	0.06	0.02
X_{12}	-0.04	-0.03	-0.01
X_{13}	0.08	0.08	0.06
X_{14}	0.07	0.13	0.07
X_{15}	0.18	0.6	0.32
X_{16}	-0.03	-0.32	-0.17
X_{17}	0.17	0.98	0.53

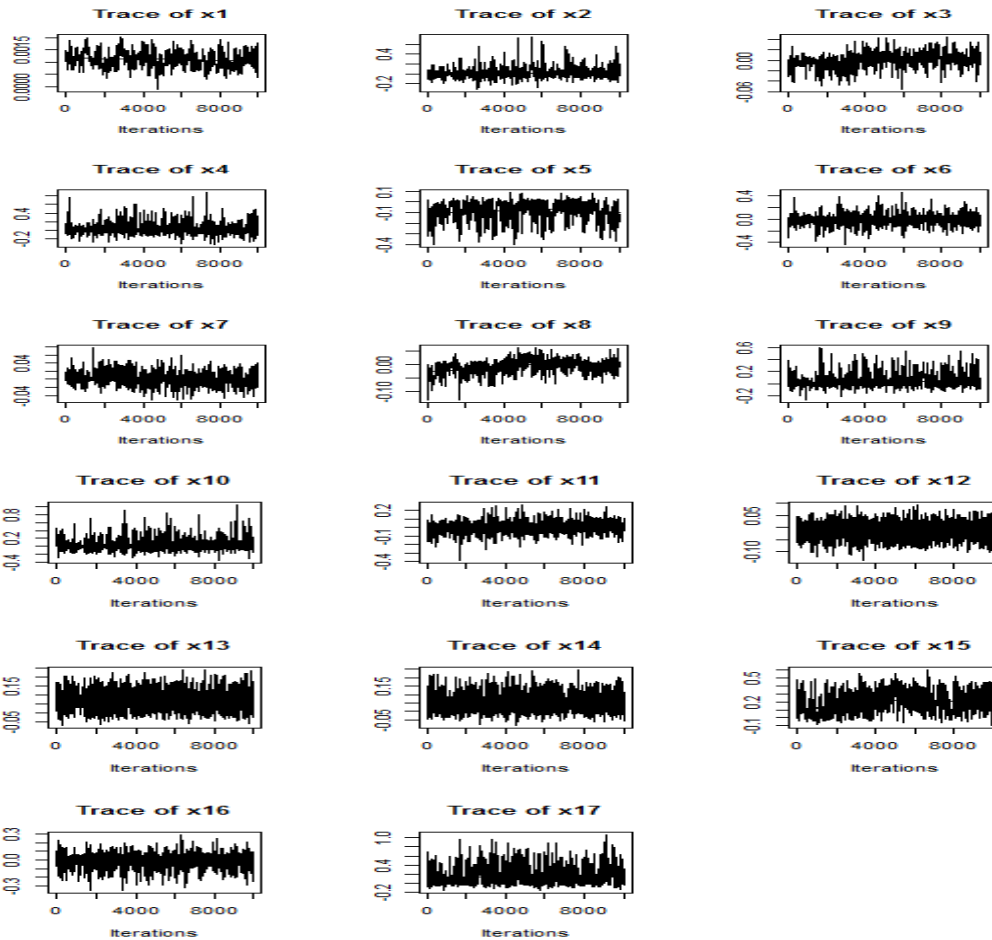


Figure 3-Trace plots of the real data.

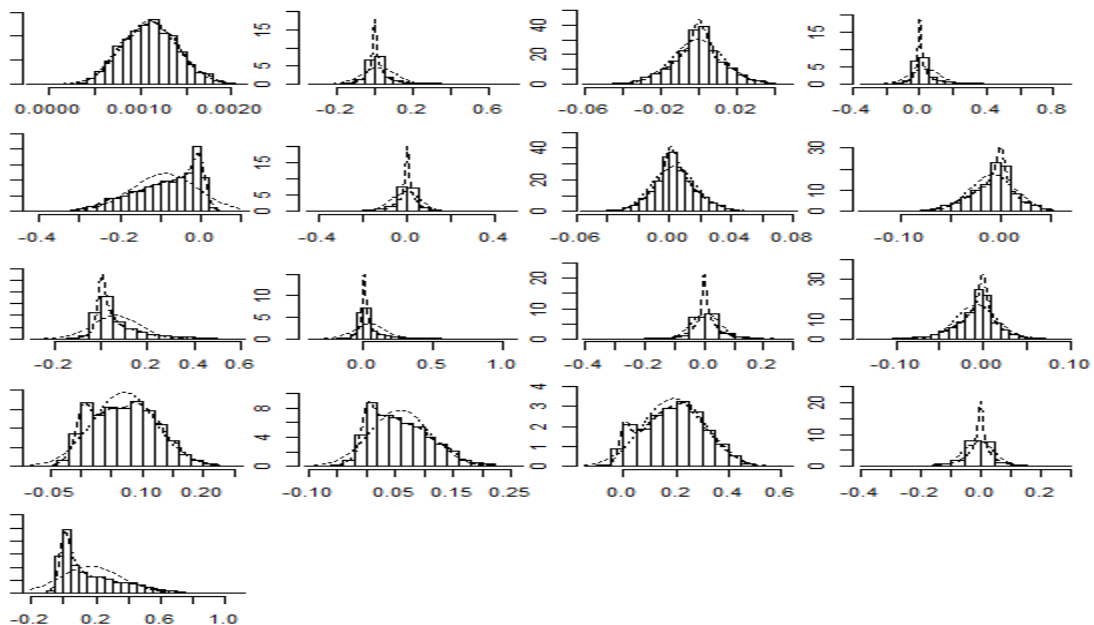


Figure 4-Histograms based on posterior samples of the real data.

We computed the Deviance Information Criteria (DIC) for the three models (BABROM, ORD (logistic), and ORD (probit)). The values were 350.8743, 389.2391 and 378.2251, respectively. The results of deviance information criteria show that our proposed methods perform better than the ORD (logistic) and ORD (probit) method.

7- Conclusion

In this paper, we developed a Bayesian adaptive bridge regression for ordinal model in the univariate case. Our method is based on a conditional conjugate prior distribution for regression parameters. We developed a new hierarchical representation for our method. To estimate the ordinal regression parameters, we introduced a Gibbs sampler for generating samples from the posterior distribution. In a simulation study, we found that the **MMAD**, **SD**, **MSE**, the **cut-point** and **DIC** indicated that our proposed method is better than the comparison methods. The studies showed that, in comparison with existing ORD (logistic) and ORD (probit), the Bayesian ordinal regression method by using the conjugate prior distribution generally performs better than the other methods.

References

1. Zou, H. **2006**. The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, **101**(476): 1418–1429.
2. Mallick, H. and Yi, N. **2018**. Bayesian bridge regression. *Journal of Applied Statistics*, **45**(6): 988–1008.
3. Xu, Z., Zhang, H., Wang, Y., Chang, X. and Liang, Y. **2010**. L 1/2 regularization. *Science China Information Sciences*, **53**(6): 1159–1169.
4. Polson, N. G., Scott, J. G. and Windle, J. **2014**. The bayesian bridge. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **76**(4): 713–733.
5. Rahman, M. A. **2016**. Bayesian quantile regression for ordinal models. *Bayesian Analysis*, **11**(1): 1–24.
6. Alhamzawi, R. **2016**. Bayesian model selection in ordinal quantile regression. *Computational Statistics & Data Analysis*, **103**: 68–78.
7. Alhamzawi, R. **2015**. Model selection in quantile regression models. *Journal of Applied Statistics*, **42**(2): 445–458.
8. Alhamzawi, R. and Ali, H. T. M. **2018a**. Bayesian quantile regression for ordinal longitudinal data. *Journal of Applied Statistics*, **45**(5): 815–828.
9. Zhou, X. **2006**. *Bayesian inference for ordinal data*. Rice University.
10. Griffin, J. E. and Brown, P. J. **2010**. Inference with normal-gamma prior distributions in regression problems. *Bayesian Analysis*, **5**(1): 171–188.
11. Mallick, H. **2015**. *Some contributions to Bayesian regularization methods with applications to genetics and clinical trials*. The University of Alabama at Birmingham.
12. Mallick, H. and Yi, N. **2013**. Bayesian methods for high dimensional linear models. *Journal of Biometrics & Biostatistics*, **1**(5).
13. Javed, F. and Mantalos, P. **2013**. GARCH-type models and performance of information criteria. *Communications in Statistics-Simulation and Computation*, **42**(8): 1917–1933.
14. Alhamzawi, R. and Ali, H. T. M. **2018b**. The Bayesian adaptive lasso regression. *Mathematical Biosciences*, **303**: 75–82.
15. Leng, C., Tran, M.-N. and Nott, D. **2014**. Bayesian adaptive lasso. *Annals of the Institute of Statistical Mathematics*, **66**(2): 221–244.
16. Wang, H. and Leng, C. **2008**. A note on adaptive group lasso. *Computational Statistics & Data Analysis*, **52**(12): 5277–5286.
17. Agresti, A. **2010**. *Analysis of ordinal categorical data* (Vol. 656). John Wiley & Sons.
18. Bürkner, P.-C. and Vuorre, M. **2018**. Ordinal Regression Models in Psychology: A Tutorial. *Advances in Methods and Practices in Psychological Science*, 2515245918823199.
19. Feng X, Wu H, X. S. X. **2017**. Bayesian adaptive lasso for ordinal regression with latent variables. *Sociological Methods & Research*, **46**(4): 926–953.
20. Jeliaskov, I., Graves, J. and Kutzbach, M. **2008**. Fitting and comparison of models for multivariate ordinal outcomes. In *Bayesian econometrics* (pp. 115–156). Emerald Group Publishing Limited.
21. Walli, G. M. **2010**. *Bayesian variable selection in Normal regression models*. na.
22. Alhamzawi, R. and Yu, K. **2012**. Variable selection in quantile regression via Gibbs sampling. *Journal of Applied Statistics*, **39**(4): 799–813.
23. Kinney, S. K. and Dunson, D. B. **2007**. Fixed and random effects selection in linear and logistic models. *Biometrics*, **63**(3): 690–698.