Statistics and Probability Letters 118 (2016) 177–181

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

Bayesian variable selection in binary quantile regression

Man-Suk Oh [a,*], Eun Sug Park [b,1], Beong-Soo So [a,2]

[a] Department of Statistics, Ewha Womans University, Seo-dae-moon Gu, Seoul 120-750, Republic of Korea

[b] Texas A&M Transportation Institute, 3135 TAMU, College Station, TX 77843-3135, USA

ARTICLE INFO

Article history:
Received 16 October 2015
Received in revised form 4 July 2016
Accepted 4 July 2016
Available online 11 July 2016

Keywords:
Bayesian model selection
Bayes factor
Quantile regression
Markov chain Monte Carlo

ABSTRACT

We propose a simple Bayesian variable selection method in binary quantile regression. Our method computes the Bayes factors of all candidate models *simultaneously* based on a *single* set of MCMC samples from a model that encompasses all candidate models. The method deals with multicollinearity problems and variable selection under constraints.

1. Introduction

Since the seminal work of Koenker and Bassett (1978) quantile regression (QR) has been rapidly expanding over recent years in various application areas such as econometrics, finance, survival analysis, social sciences, and microarray research. See Kottas and Krnjajic (2009), Drovancli and Pettitt (2011) and Taddy and Kottas (2011) among others. QR extends the classical mean regression to conditional quantiles of the response variable. It is more robust to non-normal error distributions and outliers, and provides more comprehensive information on the relationship between the response variable and the covariates than classical mean regression.

As with any regression problem, selection of appropriate covariates is important in QR. Excluding important covariates may yield biased estimators whereas including spurious covariates may lead to loss in estimation efficiency. Bayesian approaches for variable selection in QR have received considerable attention in recent literature because Bayesian methods are often more competitive for small or moderate data sets with a low signal-to-noise ratio (Antoniadis et al., 2009). Li et al. (2010) studied regularization, e.g. lasso, in quantile regression from a Bayesian perspective. Alhamzawi and Yu (2012), Ji et al. (2012), and Yu et al. (2013) extended stochastic search variable selection (SSVS, George and McCulloh, 1993) methods in mean regression to quantile regression by introducing latent variables into QR. Oh et al. (2016) proposed an alternative Bayesian variable selection method in QR using the Savage–Dickey density ratio.

Binary quantile regression is first introduced by Manski (1985). Kordas (2006) has shown that QR leads to much more comprehensive view on how the predictor variables influence the response even in binary case. Especially, Kordas (2006) showed that binary quantile regression can be very useful for unbalanced data in which there are excessive zeros/ones in data. However, most studies in binary quantile regression focused on median binary regression. To the best of our knowledge, the only exception is Ji et al. (2012) who extended a SSVS in mean regression to binary quantile regression.

* Corresponding author. Fax: +82 2 3277 3607.
E-mail addresses: msoh@ewha.ac.kr (M.-S. Oh), e-park@tamu.edu (E.S. Park), bsso@ewha.ac.kr (B.-S. So).
1 Fax: +1 979 845 6008.
2 Fax: +82 2 3277 3607.

http://dx.doi.org/10.1016/j.spl.2016.07.001
0167-7152/

In this paper we extend the method of Oh et al. (2016) to deal with binary response variables in QR. The rest of paper is organized as follows. Section 2 presents the posterior inference on parameters in binary quantile regression using the Gibbs sampling algorithm. Section 3 describes simultaneous estimation of Bayes factors of all candidate models. The proposed method is illustrated with both simulated and real data sets in Section 4. We end with concluding remarks in Section 5.

2. Posterior inference

Let y_i denote the response variable and $\mathbf{x}_i$ the $q \times 1$ vector of covariates for the ith observation. Following Benoit and Poel (2012), binary QR with quantile p, $0 < p < 1$, is given as

$$y_i^* = \mathbf{x}_i'\boldsymbol{\beta}_p + \varepsilon_i, \qquad y_i = I(y_i^* > 0), \quad i = 1, \ldots, n,$$

where ε_i is an error term of which the pth quantile is equal to zero, $\boldsymbol{\beta}_p$ is an unknown vector of coefficients that depends on p, and I is the indicator function. In this model, binary QR is considered as a linear QR with a continuous latent response variable which is not fully observed.

The QR coefficient $\boldsymbol{\beta}_p$ can be estimated by minimizing

$$\sum_{i=1}^{n} \rho_p(y_i - I(\mathbf{x}_i'\boldsymbol{\beta}_p > 0)), \tag{1}$$

where ρ_p is the check loss function, given by $\rho_p(u) = u(p - I(u < 0))$. If y_i^* is given, the function $\rho_p(y_i - I(\mathbf{x}_i'\boldsymbol{\beta}_p > 0))$ in (1) can be replaced by $\rho_p(y_i^* - \mathbf{x}_i'\boldsymbol{\beta}_p)$ and the binary QR with binary observations $\mathbf{y} = (y_1, .., y_n)$ is replaced by the linear QR with continuous observations $\mathbf{y}^* = (y_1^*, .., y_n^*)$. From here on we suppress p on $\boldsymbol{\beta}_p$ to simplify notation.

For Bayesian inference, we need a likelihood function. Since minimizing (1) is equivalent to maximizing a likelihood function under an asymmetric Laplace (AL) error distribution (Koenker and Machado, 1999), we assume AL distribution for ε_i. Also, a location-scale mixture representation of the asymmetric Laplace distribution allows convenient Bayesian inference using the Gibbs sampling algorithm of Gelfand and Smith (1990). See Yu and Moyeed (2001), Yu and Stander (2007), Reed and Yu (2009), Kozumi and Kobayashi (2011) and Yu et al. (2013) among others.

We use the mixture representation given in Yu et al. (2013). Let w_i follow an exponential distribution with rate $p(1-p)$ and assume that, given w_i, ε_i follows $N((1-2p)w_i, 2w_i)$. Then the marginal distribution of ε_i is the AL distribution. Thus, introducing latent variables $\mathbf{y}^*$ and $\mathbf{w} = (w_1, \ldots, w_n)'$ into the model enables convenient posterior sample generation using the Gibbs sampler. The likelihood function of $(\boldsymbol{\beta}, \mathbf{w}, \mathbf{y}^*)$ is proportional to

$$\prod_{i=1}^{n} w_i^{-1/2} \exp^{-\frac{1}{4}\sum_{i=1}^{n}\frac{(y_i^*-\mathbf{x}_i'\beta-(1-2p)w_i)^2}{w_i} - \sum_{i=1}^{n} p(1-p)w_i} \cdot [I(y_i^* > 0, y_i = 1) + I(y_i^* \le 0, y_i = 0)].$$

With $N(\boldsymbol{\beta}_0, \Sigma_0)$ prior for $\boldsymbol{\beta}$, the conditional posterior distribution of $\boldsymbol{\beta}$ is given by $N(\boldsymbol{\mu}_\beta, \Sigma_\beta)$, where $\Sigma_\beta = (\mathbf{X}_w'\mathbf{X}_w + \Sigma_0^{-1})^{-1}$, $\mu_\beta = \Sigma_\beta(\mathbf{X}_w'\mathbf{y}_w^* + \Sigma_0^{-1}\boldsymbol{\beta}_0)$, $\mathbf{X}_w = \left(\frac{\mathbf{x}_1}{\sqrt{2w_1}}, \ldots, \frac{\mathbf{x}_n}{\sqrt{2w_n}}\right)'$, $\mathbf{y}_w = \left(\frac{y_1^*-(1-2p)w_1}{\sqrt{2w_1}}, \ldots, \frac{y_n^*-(1-2p)w_n}{\sqrt{2w_n}}\right)'$. The conditional posterior distribution of w_i^{-1} is given by $IG\left(\frac{1}{|y_i^*-\mathbf{x}_i'\beta|}, \frac{1}{2}\right)$, where IG denotes the inverse Gaussian distribution. Finally, the conditional posterior distribution of y_i^* is given by $N(\mathbf{x}_i\boldsymbol{\beta} + (1-2p)w_i,\ 2w_i) \cdot [I(y_i^* > 0, y_i = 1) + I(y_i^* \le 0, y_i = 0)]$.

3. Bayes factor

Let $\{M_J\}$ be a set of candidate models where M_J is a binary QR model in which only a subset $\boldsymbol{\beta}_J$ of $\boldsymbol{\beta}$ is non-zero. Let M_F be a model which encompasses all the candidate models, i.e., any candidate model M_J is a special case of M_F. For a prior of $\boldsymbol{\beta}_J$ under M_J, we assume an encompassing prior, i.e., $\pi_J(\boldsymbol{\beta}_J) = \pi_F(\boldsymbol{\beta}_J|\boldsymbol{\beta}_J^c = 0)$, where $\pi_J(\boldsymbol{\beta}_J)$ is the prior density of $\boldsymbol{\beta}_J$ under M_J and $\pi_F(\boldsymbol{\beta}_J|\boldsymbol{\beta}_J^c = 0)$ is the conditional prior density of $\boldsymbol{\beta}_J$ given $\boldsymbol{\beta}_J^c = 0$ under M_F, and $\boldsymbol{\beta}_J^c$ is a complementary set of $\boldsymbol{\beta}_J$. Under the above prior assumption, the Bayes factor (BF) of M_J against M_F is given as

$$BF_J = \frac{P(M_J|\mathbf{y})/P(M_J)}{P(M_F|\mathbf{y})/P(M_F)} = \frac{\pi_F(\boldsymbol{\beta}_J^c = 0|\mathbf{y})}{\pi_F(\boldsymbol{\beta}_J^c = 0)}, \tag{2}$$

where $P(M_J|\mathbf{y})$ and $P(M_J)$ are the posterior and the prior probability of M_J, respectively, and $\pi_F(\boldsymbol{\beta}_J = 0|\mathbf{y})$ and $\pi_F(\boldsymbol{\beta}_J = 0)$ are the marginal posterior and prior density of $\boldsymbol{\beta}_J$, respectively, at zero under the encompassing model. Eq. (2) is called the Savage–Dickey density ratio (Dickey and Lientz, 1970; Dickey, 1971, 1976). A generalized version of the Savage–Dickey density ratio is given by Verdinelli and Wasserman (1995) which can be used when non-encompassing prior is used. Marin and Robert (2010) raised some fundamental issues on the Savage–Dickey density ratio and proposed an alternative approach.

Under a normal prior for $\boldsymbol{\beta}$, the marginal prior density of $\boldsymbol{\beta}_J^c$ is given in a closed form. The marginal posterior density function of $\boldsymbol{\beta}_J^c$ is not given in a closed form, but the conditional distribution of $\boldsymbol{\beta}_J^c$ is given as a multivariate normal density.