Computational Statistics and Data Analysis 103 (2016) 68–78

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

Bayesian model selection in ordinal quantile regression

Rahim Alhamzawi

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

ARTICLE INFO

Article history:
Received 27 October 2015
Received in revised form 23 April 2016
Accepted 30 April 2016
Available online 11 May 2016

Keywords:
Bayesian inference
MCMC
Quantile regression
Ordinal models
SSVS

ABSTRACT

A Bayesian stochastic search variable selection (BSSVS) method is presented for variable selection in quantile regression (QReg) for ordinal models. A Markov Chain Monte Carlo (MCMC) method is adopted to draw the unknown quantities from the full posteriors. Through simulations and analysis of an educational attainment dataset, the performance of the proposed approach is compared with some existing approaches, showing that the proposed approach performs quite good in comparison to some other methods.

1. Introduction

QReg models provide more extensive statistical models than standard mean regression models and have drawn enormous interest in the modern literature. QReg has been used in many different areas, including biomedical studies, ecology, economics, political economy, growth charts, microarray studies, the social sciences, and survival analysis. A comprehensive review of some recent applications can be seen in Alhamzawi (2013), Koenker (2005) and Yu et al. (2003).

Although the asymptotic theory for QReg models has been well studied, inference for these models is difficult, especially for ordinal responses. In contrast, a Bayesian formulation for QReg enables exact inference, even when the number of observations is small, and is well suited to incorporate ordinal responses. Since QReg does not assume any parametric distribution for the errors, different Bayesian QReg techniques have been suggested by Yu and Moyeed (2001), Kottas and Gelfand (2001), Dunson et al. (2007), Kottas and Krnjajić (2009), Reich et al. (2010), Taddy and Kottas (2010) and Noufaily and Jones (2013). However, a convenient parametric distribution choice is the skewed Laplace distribution (SLD) reported in Yu and Moyeed (2001), because the mode of the resulting posterior under a flat prior for the unknown quantity $\boldsymbol{\beta}$ is the usual QReg estimates. Although the assumption of the SLD for the errors may cause some apprehension due to its vitiating the nonparametric modality of a QReg, there have been many extensive numerical studies showing that Bayesian approaches are quite insensitive to the assumptions of the error distribution, and even if the implied distribution does not follow a SLD, the outcomes would be acceptable (see, for example, Li et al., 2010; Yuan and Yin, 2010; Ji et al., 2012; Lum and Gelfand, 2012). In addition, Koenker and Machado (1999). In addition, Koenker and Machado (1999) presented a goodness of fit for QReg depends on the SLD and demonstrate that the asymptotic features work well even if the error does not follow a SLD. Furthermore, there are mathematical justifications for using the SLD, which can be found in Sriram et al. (2013), Alhamzawi and Yu (2013) and Komunjer (2005).

Model selection plays a significant role in constructing QReg models. In particular, the prediction accuracy can be increased by selecting the active variables in the regression. Since the seminal work of Breiman (1995), several shrinkage methods have been applied to model selection in classical mean regression models. See, Lasso (Tibshirani, 1996), SCAD (Fan and Li, 2001), the least-angle regression (LARS; Efron et al., 2004), fused Lasso (Tibshirani et al., 2005), adaptive Lasso (Zou,

E-mail address: ralhamzawi@yahoo.com.

http://dx.doi.org/10.1016/j.csda.2016.04.014
0167-9473/

2006), and group Lasso (Yuan and Lin, 2006). A comprehensive review of some recent shrinkage methods can be found in Tibshirani (2011). A natural interpretation for shrinkage methods is to follow the Bayesian paradigm as shown by the Bayesian Lasso of Hans (2009), Park and Casella (2008) and Bae and Mallick (2004).

With regard to QReg, frequentist statisticians have applied shrinkage approaches to subset selection in QReg by automatically identifying the subset of predictors having nonzero coefficients. See, for example, Lasso (Koenker, 2004; Li and Zhu, 2008), adaptive Lasso (Wang et al., 2007; Bang and Jhun, 2012; Wang et al., 2013), the elastic net (Slawski, 2012), group Lasso (Kato, 2011), and SCAD (Wu and Liu, 2009; Jiang et al., 2012). Similarly, from a Bayesian view, Li et al. (2010) suggested a Bayesian regularized method for QReg and considered different shrinkage techniques. Alhamzawi (2013) consider Bayesian adaptive Lasso QReg and show that Bayesian QReg with the adaptive Lasso penalty generally performs better than with Bayesian Lasso, Bayesian elastic net, or non-Bayesian regularized QReg methods. Alhamzawi (in press) proposed Bayesian elastic net Tobit QReg and Benoit et al. (2013) studied Bayesian Lasso binary QReg. However, as no point mass at 0 is allocated in the Bayesian regularized techniques, updates of insignificant coefficients from the conditional distribution are never fixed precisely to zero. Therefore, some ad hoc methods could be applied to identify the significant coefficients. Alternatively, several BSSVS methods for identifying the active predictors in a QReg model have been suggested recently (see, Yu et al., 2013 and Alhamzawi and Yu, 2013). In this paper, the BSSVS framework is extended to QReg for ordinal models. This is the first paper to discuss variable selection in QReg for ordinal models.

Ordinal outcome data are often collected in many different areas, including behavioral research, ecology, economics, education, geology, medicine, the social sciences, and psychology. For example, in a survey regarding one's educational experience (EX), outcomes of interest may be recorded as follows (Jeliazkov et al., 2008): 1 for "elementary school graduate (ESG)", 2 for "high school graduate (HSG)", 3 for "some college (SC)", and 4 for "college graduate (CG)". Although I could order the people by 1, 2, 3 and 4 according to these four types of EX, the size of the difference between the types of EX is inconsistent. In other words, the score 2 implies more EX (HSG) than 1 (ESG), but this result does not imply that 2 is twice as much EX as 1. It is of enormous interest here to apply a model that incorporates the ordinal nature of the outcome of interest. Standard ordinal regression models have been used to describe the relationship between an ordinal outcome and a set of regressors. As a useful supplement to classical ordinal regression, QReg for ordinal models provides an efficient statistical model than classical ordinal regression. However, ordinal QReg has been addressed only in the last few years, for example, see, Hong and Zhou (2013), Goffe et al. (1994), Hong and He (2010), and Kirkpatrick and Vecchi (1983). Ordinal QReg can be described as a linear QReg using a latent continuous dependent variable which is incompletely seen. Following Rahman (2016), in this paper the pth quantile for the latent variable w_i is simulated according to the regression model

$$w_i = \boldsymbol{x}_i'\boldsymbol{\beta} + \varepsilon_i, \quad i = 1, 2, \ldots, n, \tag{1}$$

where $\boldsymbol{x}_i$ is a $k \times 1$ vector of regressors for the ith unobserved continuous latent random variable w_i, $\boldsymbol{\beta}$ is a $k \times 1$ vector of parameters, and ε_i is the error term. Then, the observed outcome of interest y_i for the ith observation is described by the classification of the unobserved outcome w_i according to

$$y_i = j \quad \text{if } \delta_{j-1} < w_i \le \delta_j,\ j = 1, \ldots, J, \tag{2}$$

where $\delta_0, \ldots, \delta_J$ are cut-points whose coordinates satisfy

$$-\infty = \delta_0 < \delta_1 < \cdots < \delta_{J-1} < \delta_J = +\infty. \tag{3}$$

From a Bayesian framework, Rahman (2016) proposed a Bayesian hierarchical model for ordinal QReg using the SLD for the errors and sampling the QReg coefficients $\boldsymbol{\beta}$ from its posterior using the Gibbs sampling method.

BSSVS was first introduced by George and McCulloch (1993) for regressor selection in linear regression model, and now has widespread regression applications, such as multivariate Bayesian models (Lee et al., 2003; Ai-Jun and Xin-Yuan, 2010), gene selection (Yi et al., 2003), regression mixture modeling (Gupta and Ibrahim, 2007), logistic mixed models (Kinney and Dunson, 2007), linear QReg (Yu et al., 2013; Alhamzawi and Yu, 2012), and Tobit and binary QReg (Ji et al., 2012). In QReg, BSSVS searches for models having high posterior probabilities and searches for regressors having high marginal inclusion probabilities (MIPs) by (1) fixing the quantile level and the total number of iterations; (2) starting with the full model of regressors; (3) selecting mixture prior distributions to removing the inactive regressors in the regression by zeroing their coefficients; (4) sampling the parameters from the posteriors; (5) estimating the MIPs for each regressor using the amount of BSSVS samples containing each regressor; and (6) estimating the posterior model probabilities (PMPs) for each model using the amount of BSSVS samples spent in each subset (model).

Section 2 introduces the structure of the proposed hierarchical Bayesian ordinal QReg model and discusses the prior specifications. Section 3 presents an MCMC-based computation technique for variable selection in ordinal QReg. Section 4 introduces the results of three simulation scenarios to investigate the performance of the proposed method, and Section 5 illustrates the proposed method using educational attainment dataset. Section 6 provides the conclusions of this work.

2. Methods

2.1. Quantile regression

Let y_i denote the response variable. Then, the pth QReg takes the form

$$Q_{y_i|\boldsymbol{x}_i}(p) = \boldsymbol{x}_i'\boldsymbol{\beta}, \tag{4}$$