Journal of the Korean Statistical Society 44 (2015) 134–149

Contents lists available at ScienceDirect

Journal of the Korean Statistical Society

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

Simultaneous variable selection and parametric estimation for quantile regression

Wei Xiong, Maozai Tian *

Center for Applied Statistics, The School of Statistics, Renmin University of China, Beijing, 100872, China

ARTICLE INFO

Article history:
Received 7 October 2013
Accepted 14 June 2014
Available online 16 July 2014

AMS 2000 subject classifications:
primary 62G05
62G08
secondary 62G35

Keywords:
Variable selection
Quantile regression
One-step estimator
Oracle property
BIC-like criterion

ABSTRACT

In this paper, variable selection techniques in the linear quantile regression model are mainly considered. Based on the penalized quantile regression model, a one-step procedure that can simultaneously perform variable selection and coefficient estimation is proposed. The proposed procedure has three distinctive features: (1) By considering quantile regression, the set of relevant variables can vary across quantiles, thus making it more flexible to model heterogeneous data; (2) The one-step estimator has nice properties in both theory and practice. By applying SCAD penalty (Fan and Li, 2001) and Adaptive-LASSO penalty (Zou, 2006), we establish the oracle property for the sparse quantile regression under mild conditions. Computationally, the one-step estimator is fast, dramatically reducing the computation cost; (3) We suggest a BIC-like tuning parameter selector for the penalized quantile regression and demonstrate the consistency of this criterion. That is to say the true model can be identified consistently based on the BIC-like criterion, making our one-step estimator more reliable practically. Monte Carlo simulation studies are conducted to examine the finite-sample performance of this procedure. Finally, we conclude with a real data analysis. The results are promising.

1. Introduction

In statistical modeling it is common that a large number of candidate predictors are available. Then a series of problems arise naturally such as whether all the candidate variables can be added into the model and if not, how to appropriately select variables to achieve the model sparsity. Thus in this sense variable selection plays a crucial role in the model building process to make an effective statistical inference. And a good model is indeed essential to statistical decision and prediction. Usually a large number of predictors are introduced initially to build a model. However, if the number of key variables selected is too small, an underfitted model could be produced, which always results in large bias and unreliable inference; On the contrary, if the model keeps many irrelevant variables, it leads to an overfitted model, which usually degrades the efficiency of the estimation and prediction. So variable selection also serves as a tool to balance between bias and variance.

Classical model selection criteria such as AIC, BIC and cross-validation are widely used and most of them have been shown to enjoy various nice theoretical properties, however there are also some drawbacks. When applied to handle high dimensional problems, these classical criteria could cause some troubles. For example, they are unstable (Breiman, 1996) and often bring about complicated stochastic properties (Fan & Li, 2001) etc. To overcome the deficiency of classical criteria, a number of new techniques have been developed in literature recent years, such as nonnegative garrote, Lasso, bridge

* Corresponding author.
E-mail address: mztian@ruc.edu.cn (M. Tian).

http://dx.doi.org/10.1016/j.jkss.2014.06.003
1226-3192/

regression, least angle regression (LARs) and boosting. Fan and Li (2001) proposed a new method via penalized least square regression which can simultaneously perform variable selection and coefficient estimation. In this context, various types of penalties have been introduced to achieve variable selection. See, for example smoothly clipped absolute deviation (SCAD), least absolute shrinkage and selection operator (Lasso), elastic net and adaptive lasso etc. Furthermore, Fan and Li (2001) demonstrated the oracle property of the SCAD estimator. Later Zou (2006) proposed the adaptive-lasso penalty function and also showed the oracle property of the related estimator.

For the classical linear regression model $Y = X^T\beta+\varepsilon$, to achieve variable selection and make statistical inference possible, sparsity assumption should be made first, that is only a set of predictors contribute to the response, then coefficients β can be estimated via penalization

$$\hat{\beta} = \text{Arg}\min_{\beta} \sum_{i=1}^{n}(Y_i - \mathbf{X}_i^T\beta)^2 + n\sum_{j=1}^{p} p_\lambda(|\beta_j|),$$

where $p_\lambda(\cdot)$ is a specific penalty function, $\{(\mathbf{X}_i, Y_i)\}_{i=1}^n$ is a random sample. When $p_\lambda(\cdot)$ is a SCAD penalty, Fan and Li (2001) demonstrated the Oracle property of the SCAD estimator. Generally, Oracle property refers to two things: one is sparsity and the other is efficiency. That is relevant variables can be identified with probability one and coefficients can be estimated as efficient as if the set of variables were given a priori.

In recent years, quantile regression becomes popular in many scientific areas owing to its robustness and practical usefulness. This technique can insightfully explore the complete relationship between response and covariates, which makes it an efficient way in data analysis. And in quantile regression setting, variable selection procedures have also seen some developments. Koenker (2005) applied the LASSO penalty to the mixed-effect quantile regression model. Li and Zhu (2007) considered the L_1 norm penalized quantile regression. Wu and Liu (2009) studied penalized quantile regression by employing a Difference Convex Algorithm to deal with optimization problem for SCAD penalty and derived the oracle properties of penalized quantile regression. However though their asymptotic properties are well established, none of them demonstrated any properties of model selection criteria, which also play crucial part in variable selection. Thus in this paper, we pay our attention most on the tuning parameter selectors and show their properties. Actually, compared with the substantial literature on variable selection in mean regression setting, relative sparse attention has been paid to that in penalized quantile regression. And by considering quantile regression, the set of relevant variables can differ across quantiles, which makes the proposed estimator more flexible and reliable and enables us to explore the entire distribution of the response, giving us a more realistic picture. All of above motivate us to develop a simple method in penalized quantile regression. In addition, to make our proposed approach to be easy to implement, a one-step quantile estimator is introduced to further reduce the computation burden. We also show the oracle properties of the one-step estimator under mild conditions.

It is well known that the tuning parameter (or regularization parameter) plays a crucial part in variable selection. Properly choosing tuning parameters based on some criteria leads to both efficient estimation and valid inference. So far there have been many criteria for selection of tuning parameter, but not all of them are applicable. Traditional model selection criteria such as Akaike's information criterion (AIC, Akaike, 1973) and Bayesian information criterion (BIC, Schwarz, 1978) are easy to implement, but also suffer from some limitations. Their validity depend heavily on model assumptions, and generally they do not work well in finite sample case. Other criteria like cross-validation (CV, Allen, 1974; Stone, 1974) and Bootstrap methods (Efron, 1979) are often used today. They are model-free, relying on data re-sampling to determine the tuning parameter. But when the sample size is small, these criteria are challenged. Especially when applied to sparse regularization method, Lasso, CV can result in instability and lead to unreliable result. And for generalized cross-validation criterion (GCV), Wang, Li, and Tsai (2007) demonstrated that it suffered from a nonignorable overfitting effect even when the sample size is sufficiently large, while a BIC-based selector is consistent in this case. Their results are inspiring and motivate us to use BIC-like criterion (what we called in our paper) for penalized quantile regression model. The BIC-like criterion which has a similar form to the usual BIC criteria is much more suitable for the penalized quantile regression model. Further in this paper we show that the proposed BIC-like criterion is consistent in the sense that it is able to identify the true model with probability one.

The rest of paper is organized as follows. In Section 2, quantile regression model is considered and based on it a one-step procedure that conducts variable selection and coefficient estimation simultaneously is proposed. Oracle properties of relevant estimators are also shown. To choose a proper tuning parameter in quantile regression setting, BIC-like criteria is investigated in Section 3. Section 4 presents the results of simulation studies. Finally in Section 5 we use a real data to illustrate our proposed method. More details about Proofs and real data are presented in the Appendix A.

2. One-step procedure selection

In this paper, we will just restrict ourselves to the context of linear regression, but note that these methods can be easily generalized to the nonlinear regression model via basis expansion.

Consider the simple linear regression model

$$\mathbf{Y} = \mathbf{X}\beta + \varepsilon, \tag{2.1}$$