Computational Statistics and Data Analysis 103 (2016) 139–150

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

Reduced rank regression with possibly non-smooth criterion functions: An empirical likelihood approach

Sanying Feng[a], Heng Lian[b,*], Fukang Zhu[c]

[a] School of Mathematics and Statistics, Zhengzhou University, Zhengzhou 450001, China

[b] School of Mathematics and Statistics, University of New South Wales, Sydney 2052, Australia

[c] School of Mathematics, Jilin University, Changchun 130012, China

ARTICLE INFO

Article history:
Received 22 December 2015
Received in revised form 29 April 2016
Accepted 29 April 2016
Available online 11 May 2016

Keywords:
Asymptotic normality
Empirical likelihood
Median regression
Quantile regression

ABSTRACT

Reduced rank regression is considered when the criterion function is possibly non-smooth, which includes the previously un-studied reduced rank quantile regression. The approach used is based on empirical likelihood with a rank constraint. Asymptotic properties of the maximum empirical likelihood estimator (MELE) are established using general results on over-parametrized models. Empirical likelihood leads to more efficient estimators than some existing estimators. Besides, in the framework of empirical likelihood, it is conceptually straightforward to test the rank of the unknown matrix. The proposed methods are illustrated by some simulation studies and real data analyses.

1. Introduction

Reduced rank regression was proposed in the 1950s (Anderson, 1951) with the purpose of introducing a more parsimonious model in cases with multiple responses. Let $\mathbf{y} = (y_1, \ldots, y_q)^{\mathrm{T}}$ be the q responses that are to be related to the p-dimensional predictor $\mathbf{x} = (x_1, \ldots, x_p)^{\mathrm{T}}$. The multivariate linear regression model is given by

$$\mathbf{y} = \boldsymbol{\alpha} + \mathbf{B}^{\mathrm{T}}\mathbf{x} + \mathbf{e}, \tag{1}$$

where $\boldsymbol{\alpha} \in R^q$ is the intercept, $\mathbf{B}$ is the $p \times q$ coefficient matrix and $\mathbf{e}$ is the q-vector of mean zero noises. Different columns of $\mathbf{B}$ are associated with different responses. If we assume the noises for different responses are independent and use ordinary least squares procedure to fit the model, the procedure is the same as linear regression for each response separately. Reduced rank regression aims at producing a more parsimonious model with fewer than $p \times q$ parameters by assuming that $\mathrm{rank}(\mathbf{B}) \le r$ for an integer $r < \min(p, q)$. Informally the number of parameters in reduced rank regression can be counted as follows. Assuming the first r columns of $\mathbf{B}$ are linearly independent, we have the parametrization $\mathbf{B} = (\mathbf{B}_1, \mathbf{B}_1\mathbf{B}_2^{\mathrm{T}})$ for a rank r matrix $\mathbf{B}$, where $\mathbf{B}_1$ is a $p \times r$ matrix and $\mathbf{B}_2$ is a $(q-r) \times r$ matrix, and the number of free parameters is thus $r(p+q-r) < pq$. With reduced degrees of freedom, the reduced rank regression has a potential to produce a more efficient estimator of $\mathbf{B}$, as shown in Anderson (1999). Some more recent references on reduced rank regression include Geweke (1996), Bunea et al. (2011, 2012), Chen and Huang (2012), Chen et al. (2012), Chen et al. (2013) and Lian and Ma (2013).

In this paper, we consider reduced rank regression in a more general framework, allowing also the criterion functions to be non-smooth, as for median or quantile regression, although the framework proposed here is more generally applicable

* Corresponding author.
E-mail address: heng.lian@unsw.edu.au (H. Lian).

http://dx.doi.org/10.1016/j.csda.2016.04.012

than the numerical examples provided later. For example, we can also deal with generalized linear model with multivariate binary or count responses. Let $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^T, \dots, \boldsymbol{\beta}_q^T)^T = \text{vec}(\mathbf{B})$ and let $\boldsymbol{\alpha} \in R^l$ be other parameters in the model (for example $\boldsymbol{\alpha}$ is the intercept in (1)). Let $\boldsymbol{\theta} = (\boldsymbol{\alpha}^T, \boldsymbol{\beta}^T)^T$ be all the parameters. We assume there are m estimating equations, with $m \geq qp + l$,

$$\mathbf{g}(\mathbf{z}, \boldsymbol{\theta}) = (g_1(\mathbf{z}, \boldsymbol{\theta}), \dots, g_m(\mathbf{z}, \boldsymbol{\theta}))^T,$$

where $\mathbf{z} = (\mathbf{x}^T, \mathbf{y}^T)^T$. For example, in the standard reduced rank regression for mean, we can use the estimating equations $x_j(y_k - \alpha_k - \mathbf{x}^T\boldsymbol{\beta}_k) = 0,\ j = 1, \dots, p,\ k = 1, \dots, q$ together with $y_k - \alpha_k - \mathbf{x}^T\boldsymbol{\beta}_k = 0,\ k = 1, \dots, q$. It is known that least squares procedure is extremely sensitive to outliers or heavy-tailed error distribution and does not have asymptotically normal distribution when the second moment of the error is not finite. An obvious alternative to least squares procedure is to replace the quadratic loss with absolute least deviation loss, resulting in median regression. More generally, suppose we want to estimate the τth quantiles of the q responses, we can use the estimating equations $x_j(\tau - I\{y_k \leq \alpha_k + \mathbf{x}^T\boldsymbol{\beta}_k\}) = 0,\ j = 1, \dots, p,\ k = 1, \dots, q$ together with $\tau - I\{y_k \leq \alpha_k + \mathbf{x}^T\boldsymbol{\beta}_k\} = 0,\ k = 1, \dots, q$. When $\tau = 1/2$ the above reduces to median regression. The flexibility of allowing more estimating equations than the number of parameters also means we can easily incorporate additional information on the parameters. For example, we can collect the estimating equations for mean regression and median regression together, if the error distribution is symmetric, to improve efficiency. Thus our approach is more general than previous reduced rank models which solely focused on least squares loss.

We propose to use empirical likelihood (Owen, 1988, 1990) which naturally deals with a general set of estimating equations. As demonstrated in Qin and Lawless (1994) and Newey and Smith (2004), it requires less restrictive distributional assumptions for statistical inferences. Empirical likelihood involving non-smooth criterion functions has previously been treated in Lopez et al. (2009). Notably, when there are more estimating equations than the number of parameters, which is typically the case when using reduced rank regression, Qin and Lawless (1994) showed that the empirical likelihood approach is fully efficient in the sense that the estimator has the same asymptotic variance as the optimal estimator obtained from the class of estimating equations that are linear combinations of $g_1, \dots, g_m$ (see Corollary 2 in Qin and Lawless (1994)). Another advantage of using empirical likelihood is that it is easy to develop a test statistic for the coefficient matrix rank based on empirical likelihood ratio, which is an important but difficult problem.

The rest of the paper is organized as follows. In Section 2, we present the general framework for reduced rank regression based on estimating equations using empirical likelihood. Section 3 specializes the general results to the case of mean regression and quantile regression, verifying the maximum empirical likelihood estimator is more efficient than some naive alternative estimation approaches. Section 4 reports our numerical results, using both simulations and two data examples. We conclude with some discussions in Section 5. The technical assumptions and theoretical proofs are relegated to the supplementary material (Appendix A).

2. Empirical likelihood for reduced rank regression

Assume we are given i.i.d. observations $\mathbf{z}_i = (\mathbf{x}_i^T, \mathbf{y}_i^T)^T,\ i = 1, \dots, n$ from some distribution F with unknown parameters $(\boldsymbol{\alpha}, \mathbf{B})$ with $\boldsymbol{\alpha} \in R^l$ and $\mathbf{B} \in R^{p\times q}$. Let $\boldsymbol{\beta} = \text{vec}(\mathbf{B})$ and $\boldsymbol{\theta} = (\boldsymbol{\alpha}^T, \boldsymbol{\beta}^T)^T$. With m estimating equations $\mathbf{g}(\mathbf{z}, \boldsymbol{\theta})$, which satisfy $E[\mathbf{g}(\mathbf{z}, \boldsymbol{\theta})] = 0$ if $\boldsymbol{\theta}$ is the true value, the empirical likelihood ratio is defined as

$$EL(\boldsymbol{\theta}) = \sup\left\{\prod_{i=1}^n np_i \middle| p_i \geq 0, \sum_i p_i = 1, \sum_i p_i\mathbf{g}(\mathbf{z}_i, \boldsymbol{\theta}) = \mathbf{0}\right\},$$

and the maximum empirical likelihood estimator (MELE) is defined to be the maximizer of $EL(\boldsymbol{\theta})$, which exists and is unique under mild regularity assumptions. The reduced rank estimator is naturally defined as the maximizer of $EL(\boldsymbol{\theta})$ under the constraint $\text{rank}(\mathbf{B}) \leq r$, for a given integer r. When the maximizer is well defined, the Lagrange multiplier method provides the optimal weights to be

$$p_i = \frac{1}{n}\frac{1}{1 + \boldsymbol{\lambda}^T\mathbf{g}(\mathbf{z}_i, \boldsymbol{\theta})}, \tag{2}$$

where the multiplier $\boldsymbol{\lambda}$ is determined by

$$\sum_i \frac{1}{n}\frac{\mathbf{g}(\mathbf{z}_i, \boldsymbol{\theta})}{1 + \boldsymbol{\lambda}^T\mathbf{g}(\mathbf{z}_i, \boldsymbol{\theta})} = \mathbf{0}, \tag{3}$$

and thus $\boldsymbol{\lambda} = \boldsymbol{\lambda}(\boldsymbol{\theta})$ is considered as a function of $\boldsymbol{\theta}$.

The following matrix plays a role in the proof of our main result:

$$\mathbf{V} = \begin{pmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{12}^T & \mathbf{0} \end{pmatrix},$$

where $\mathbf{V}_{11} = E\mathbf{g}(\mathbf{z}, \boldsymbol{\theta})\mathbf{g}^T(\mathbf{z}, \boldsymbol{\theta})$ is an $m \times m$ symmetric matrix, and $\mathbf{V}_{12} = -\frac{\partial}{\partial\boldsymbol{\theta}^T}E[\mathbf{g}(\mathbf{z}, \boldsymbol{\theta})]$ is an $m \times (pq + l)$ matrix. Note we do not assume that $\mathbf{g}(\mathbf{z}, \boldsymbol{\theta})$ is differentiable in $\boldsymbol{\theta}$, only that its expectation is.