Computational Statistics and Data Analysis 64 (2013) 180–191

Contents lists available at SciVerse ScienceDirect

Computational Statistics and Data Analysis

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

Two step composite quantile regression for single-index models

Rong Jiang [a], Zhan-Gong Zhou [b], Wei-Min Qian [a,*], Yong Chen [c]

[a] *Department of Mathematics, Tongji University, Shanghai 200092, China*
[b] *Department of Statistics, Jiaxing University, Jiaxing 314001, China*
[c] *College of Architecture & Urban Planning, Tongji University, Shanghai 200092, China*

ARTICLE INFO

Article history:
Received 2 November 2012
Received in revised form 4 March 2013
Accepted 12 March 2013
Available online 21 March 2013

Keywords:
Linearity condition
Local polynomial regression
Composite quantile regression
Single-index model

ABSTRACT

This paper is concerned with composite quantile regression for single-index models. Under mild conditions, we show that the linear composite quantile regression offers a consistent estimate of the index parameter vector. With a root-n consistent estimate of the index vector, the unknown link function can be estimated by local composite quantile regression. This procedure enables us to reduce the computational cost and is also appealing in high-dimensional data analysis. We show that the resulting estimator of the composite quantile function performs asymptotically as efficiently as if the true value of the index vector is known. The simulation studies and real data applications are conducted to illustrate the finite sample performance of the proposed methods.

1. Introduction

Single-index models (SIMs) provide an efficient way of coping with high-dimensional nonparametric estimation problems and avoid the "curse of dimensionality" by assuming that the response is only related to a single linear combination of the covariates. Therefore, much effort has been devoted to studying its estimation and other relevant inference problems. Härdle and Stoker (1989) and Ichimura (1993) had given examples of classical regression, discrete regression, and censored regression that could all be classified as single-index models. Various strategies for estimating the index parameter vector have been proposed in the past two decades. Two most popular methods are the average derivative method (ADE) introduced in Powell et al. (1989) and Härdle and Stoker (1989), and the simultaneous minimization method of Härdle et al. (1993). Hristache et al. (2001) improved the ADE approach by lowering the dimension of the kernel gradually. In addition, Yu and Ruppert (2002) proposed the penalized spline estimation procedure, while Xia and Härdle (2006) applied the minimum average variance estimation (MAVE) method, which was originally introduced by Xia et al. (2002) for dimension reduction. However, efficient and stable estimation of SIMs is still a challenging problem and has inspired many recent works in this area (Wang et al., 2010; Liang et al., 2010). Recent works in this area can see in Wang and Xue (2011), Xu and Zhu (2012), Fan and Zhu (2013) and so on. The single-index model is of the following form:

$$Y = g(X^T\beta_0) + \varepsilon, \tag{1.1}$$

where Y is the univariate response and X is a vector of d-dimensional covariates, $g(\cdot)$ is the unknown univariate link function, β_0 is the unknown single-index vector coefficient, and for the sake of identifiability (Lin and Kulasekera, 2007), we assume that $\|\beta_0\| = 1$ and that the first component of β_0 is positive; here $\|\cdot\|$ denotes the Euclidean norm, and ε is random error.

* Corresponding author. Tel.: +86 02165983240.
E-mail addresses: jrtrying@126.com (R. Jiang), Zhouzg@mail.zjxu.edu.cn (Z.-G. Zhou), wmqian2003@yahoo.com.cn (W.-M. Qian).

0167-9473/$ – see front matter
http://dx.doi.org/10.1016/j.csda.2013.03.014

Quantile regression was first studied in linear quantile regression by Koenker and Bassett (1978) in their seminal work. Since then quantile regression has experienced deep and exciting developments in theory, methodology and applications. For example, Hendricks and Koenker (1992) and Yu and Jones (1998), among others, studied nonparametric quantile regression for independent observations. Cai and Xu (2008) considered local polynomial estimators for time series data. Honda (2004) and Kim (2007) studied varying coefficient models for independent data using local polynomials and splines, respectively. For a complete review on quantile regression, see Koenker (2005). Recent works can see Zhou et al. (2011), Tang et al. (2012a, 2013), among others. Recently, quantile regression for single index model has been studied, Chaudhuri et al. (1997) proposed an estimation procedure for β_0 using the ADE method, which takes partial derivatives of the conditional quantile with respect to the covariates. However the involvement of a high-dimensional kernel in ADE hinders its popularity. Another embarrassment of the ADE is that if the expectation of the derivative of the conditional quantile function with respect to β_0 is zero, it fails to provide consistent estimate of β_0 in theory. Wu et al. (2010) proposed a back-fitting algorithm which was shown to be more efficient than the ADE, to serve the same purpose in single-index quantile regression. Kong and Xia (2012) proposed an adaptive estimation procedure and an iterative algorithm for single-index quantile regression, which was also more efficient than the ADE method. Zhu et al. (2012) proposed a computationally efficient two-step estimation procedure to estimate the parameters involved in the quantile regression function. Hua et al. (in press) developed a fully Bayesian approach to fitting single-index models in conditional quantile regression. They established a Gaussian process prior proposed by Gramacy and Lian (2012) for the unknown nonparametric link function and a Laplace distribution on the index vector, with the latter motivated by the recent popularity of the Bayesian lasso idea.

Intuitively, the composite quantile regression (CQR) should provide estimation efficiency gain over a single quantile regression; see Zou and Yuan (2008). Kai et al. (2010) proposed the local polynomial CQR estimators for estimating the nonparametric regression function and its derivative. It is shown that the local CQR method can significantly improve the estimation efficiency of the local least squares estimator for commonly-used non-normal error distributions. Kai et al. (2011) studied semiparametric CQR estimates for varying-coefficient partially linear models. Guo et al. (2012) considered CQR estimates for varying-coefficient models with heteroscedasticity. Jiang et al. (2012a) and Tang et al. (2012b) extended the CQR method to censored data. Tang et al. (2012c) studied CQR estimates for infinite variance autoregressive models. Jiang et al. (2012c) suggested a back-fitting CQR algorithm for single-index model. However, their algorithm is computationally expensive when p is large. In this paper, we propose a two-step CQR (TCQR) estimation procedure for SIMs. We first estimate the index parameter vector using a linear composite quantile regression (Zou and Yuan, 2008) and show that the resulting estimate is consistent. This property enables us to substantially reduce the computational cost of the back-fitting algorithm (Jiang et al., 2012c); this is appealing in high-dimensional data analysis. Next, we employ local composite regression (Kai et al., 2010) to estimate the nonparametric part. Moreover, asymptotic properties of the proposed procedures are studied.

The paper is organized as follows. In Section 2, we introduce the two-step composite quantile regression methods for model (1.1), and the main theoretical results are also given in that section. Both simulation examples and the application of two real data are given in Section 3 to illustrate the proposed procedures. Final remarks are given in Section 4. All the conditions and technical proofs are deferred to the Appendix.

2. Methodology

2.1. Estimation of β_0

Let

$$L(b_1, \dots, b_q, \beta) = \sum_{k=1}^{q} E\{\rho_{\tau_k}(Y - b_k - \beta^T X)\},$$

where $\rho_{\tau_k}(r) = \tau_k r - rI(r < 0)$, $k = 1, 2, \dots, q$, are q check loss functions at k quantile positions: $\tau_k = k/(q+1)$. Take

$$(\bar{b}_1, \dots, \bar{b}_q, \bar{\beta}) = \arg\min_{b_1,\dots,b_q,\beta} L(b_1, \dots, b_q, \beta)$$

Theorem 1. *If the covariate vector X at (1.1) satisfies*

$$E(X|\beta_0^T X) = var(X)\beta_0\{\beta_0^T var(X)\beta_0\}^{-1}\beta_0^T X, \tag{2.1}$$

then $\bar{\beta} = c\beta_0$ for some constant c.

Remark 1. With the linearity condition (2.1), we are able to obtain a root n consistent estimator of the direction of index parameter β by the linear composite quantile regression; see Theorem 2. The linearity condition (2.1) is widely assumed in the context of sufficient dimension reduction. Li (1991) pointed out that it is satisfied when X follows an elliptically contoured distribution, and Hall and Li (1993) proved that it always holds to a good approximation in single-index models of the form (1.1) when the dimension p of the covariates becomes large. Thus, the linearity condition is typically regarded as mild, particularly when p is fairly large. Zhu et al. (2012) considered linearity condition for quantile regression.