Journal of Multivariate Analysis 148 (2016) 34–48

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

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

# Weighted composite quantile regression for single-index models

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

[a] Department of Mathematics, College of Sciences, Donghua University, Shanghai 201620, China
[b] Department of Mathematics, Tongji University, Shanghai 200092, China
[c] Nanhu College, Jiaxing University, Jiaxing 314001, China

ARTICLE INFO

*Article history:*
Received 5 February 2015
Available online 2 March 2016

*AMS subject classifications:*
primary 60G08
secondary 62G20

*Keywords:*
Single-index model
Weighted composite quantile regression
Adaptive LASSO

ABSTRACT

In this paper we propose a weighted composite quantile regression (WCQR) estimation for single-index models. For parametric part, the WCQR is augmented using a data-driven weighting scheme. With the error distribution unspecified, the proposed estimators share robustness from quantile regression and achieve nearly the same efficiency as the semi-parametric maximum likelihood estimator for a variety of error distributions including the Normal, Student's t, Cauchy distributions, etc. Furthermore, based on the proposed WCQR, we use the adaptive-LASSO to study variable selection for parametric part in the single-index models. For nonparametric part, the WCQR is augmented combining the equal weighted estimators with possibly different weights. Because of the use of weights, the estimation bias is eliminated asymptotically. By comparing asymptotic relative efficiency theoretically and numerically, WCQR estimation all outperforms the CQR estimation and some other estimate methods. Under regularity conditions, the asymptotic properties of the proposed estimations are established. The simulation studies and two real data applications are conducted to illustrate the finite sample performance of the proposed methods.


## 1. Introduction

Single-index models 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 [6] proposed the average derivative method (ADE). Ichimura [7] studied the properties of a semiparametric least-squares estimator in a general single-index model. Yu and Ruppert [34] considered the penalized spline estimation procedure, while Xia and Härdle [32] applied the minimum average variance estimation method, which was originally introduced by Xia, et al. [33] for dimension reduction. Wu, et al. [31] studied single-index quantile regression. Feng, et al. [5] proposed the rank-based outer product of gradients estimator for parametric part in the single-index model. Liu, et al. [20] applied the local linear model regression estimator method to single-index model. In this paper, we consider the following heteroscedastic single-index model:

$$\mathbf{Y} = g_0(\mathbf{X}^\top \gamma_0) + \sigma(\mathbf{X}^\top \gamma_0)\varepsilon, \tag{1.1}$$

* Corresponding author.
*E-mail addresses:* jrtrying@dhu.edu.cn (R. Jiang), wmqian2003@aliyun.com (W.M. Qian), zhouzg@mail.zjxu.edu.cn (Z.G. Zhou).

http://dx.doi.org/10.1016/j.jmva.2016.02.015

where $\mathbf{Y}$ is the univariate response and $\mathbf{X}$ is a vector of $p$-dimensional covariates. The functions $g_0(\cdot)$ and $\sigma(\cdot)$ are unspecified, nonparametric smoothing functions. $\gamma_0$ is the unknown single-index vector coefficient, and for the sake of identifiability, see [19], we assume that $\|\gamma_0\| = 1$ and that the first component of $\gamma_0$ is positive, here $\|\cdot\|$ denotes the Euclidean norm, and the error term $\varepsilon$ is assumed to be independent of $\mathbf{X}$ with $\mathrm{E}(\varepsilon) = 0$ and $\mathrm{Var}(\varepsilon) = 1$. The model (1.1) was also studied by Zhu and Zhu [36] and Zhu et al. [35].

The composite quantile regression (CQR) was first proposed by Zou and Yuan [39] for estimating the regression coefficients in the classical linear regression model. Zou and Yuan [39] showed that the relative efficiency of the CQR estimator compared with the least squares estimator is greater than 70% regardless of the error distribution. Based on CQR, Kai, et al. [16] 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. [17] studied semiparametric CQR estimates for semiparametric varying-coefficient partially linear model. Jiang et al. [10] and Tang et al. [28] considered CQR method for random censored data. Jiang et al. [15] extended CQR method to single-index model. Jiang et al. [14] proposed a computationally efficient two-step composite quantile regression for single-index model. Jiang et al. [13] considered CQR estimates for single-index models with heteroscedasticity and general error distributions. Other references about CQR method can see Jiang et al. [9], Jiang et al. [11], Ning and Tang [21], Jiang et al. [12] and so on. However, the CQR method is a sum of different quantile regression with equal weights. Intuitively, equal weights are not optimal in general, and hence we propose a weighted CQR (WCQR) estimation method for single-index model. By comparing asymptotic relative efficiency theoretically and numerically, WCQR estimation all outperforms the CQR estimation and some other estimate methods.

The paper is organized as follows. In Section 2, we introduce the weighted composite quantile procedure for parametric part in model (1.1). In Section 3, a variable selection method is developed. We propose the weighted local composite quantile procedure for nonparametric part in Section 4. Both simulation examples and the application of two real data are given in Section 5 to illustrate the proposed procedures. Final remarks are given in Section 6. All the conditions and technical proofs are deferred to the Appendix.

## 2. Weighted composite quantile regression for $\gamma_0$

Let $\{\mathbf{X}_i, Y_i\}_{i=1}^n$ be an independent identically distributed (i.i.d.) sample from $(\mathbf{X}, \mathbf{Y})$. For $\mathbf{X}_i^\top\gamma$ "close to" $u$, the $\tau$th conditional quantile at $\mathbf{X}_i^\top\gamma$ can be approximated linearly by

$$g(\mathbf{X}_i^\top\gamma) \approx g(u) + g'(u)(\mathbf{X}_i^\top\gamma - u) = a + b(\mathbf{X}_i^\top\gamma - u),$$

where $a \triangleq g(u)$ and $b \triangleq g'(u)$. Let $\rho_{\tau_k}(r) = \tau_k r - rI(r < 0)$, $k = 1, \ldots, q$, be $q$ check loss functions with $0 < \tau_1 < \cdots < \tau_q < 1$. Typically, we use the equally spaced quantiles: $\tau_k = k/(q+1)$ for $k = 1, \ldots, q$. Let $K(\cdot)$ be the kernel weight function and $h$ is the bandwidth.

In order to define the weighted composite quantile regression, let us briefly review the quantile regression (QR) and composite quantile regression (CQR) method to estimate $\gamma_0$ in the single-index model.

By Wu et al. [31], the $\tau$th QR estimate of $\gamma_0$ can be obtained as follows:

Step 1.0 (Initialization step): Obtain initial $\hat{\gamma}^{(0)}$ from average derivative estimate (ADE) of Chaudhuri et al. [1]. Standardize the initial estimate such that $\|\hat{\gamma}^{QR}\| = 1$ and $\hat{\gamma}_1^{QR} > 0$.

Step 1.1: Given $\hat{\gamma}^{QR}$, obtain $\{\hat{a}_j, \hat{b}_j\}_{j=1}^n$ by solving a series of the following

$$\min_{(a_j,b_j)} \sum_{i=1}^n \rho_\tau\{Y_i - a_j - b_j(\mathbf{X}_i - \mathbf{X}_j)^\top\hat{\gamma}^{QR}\}\omega_{ij},$$

where $\omega_{ij} = K\left\{\left(\mathbf{X}_i^\top\hat{\gamma}^{QR} - \mathbf{X}_j^\top\hat{\gamma}^{QR}\right)/h\right\} / \sum_{l=1}^n K\left\{\left(\mathbf{X}_l^\top\hat{\gamma}^{QR} - \mathbf{X}_j^\top\hat{\gamma}^{QR}\right)/h\right\}$ and with the bandwidth $h$ chosen optimally.

Step 1.2: Given $\{\hat{a}_j, \hat{b}_j\}_{j=1}^n$, obtain $\hat{\gamma}^{QR}$ by solving

$$\min_{\gamma} \sum_{j=1}^n \sum_{i=1}^n \rho_\tau\{Y_i - \hat{a}_j - \hat{b}_j(\mathbf{X}_i - \mathbf{X}_j)^\top\gamma\}\omega_{ij},$$

with $\omega_{ij}$ evaluated at $\gamma$ and $h$ from step 1.1.

Step 1.3: Repeat Steps 1.1 and 1.2 until convergence.

Composite several quantile information, Jiang et al. [13] proposed CQR method to estimate $\gamma_0$ as follows:

Step 2.0 (Initialization step): Obtain initial $\hat{\gamma}^{(0)}$ from the minimum average variance estimation (MAVE) in Xia and Härdle [32]. Standardize the initial estimate such that $\|\hat{\gamma}^{CQR}\| = 1$ and $\hat{\gamma}_1^{CQR} > 0$.