Journal of Statistical Planning and Inference 143 (2013) 1049–1063

Contents lists available at SciVerse ScienceDirect

Journal of Statistical Planning and Inference

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

Weighted local linear composite quantile estimation for the case of general error distributions

Jing Sun [a], Yujie Gai [b], Lu Lin [a,*]

[a] Shandong University Qilu Securities Institute for Financial Studies, and School of Mathematics, Shandong University, China
[b] School of Economics, Central University of Finance and Economics, China

ARTICLE INFO

Article history:
Received 6 January 2012
Received in revised form
11 October 2012
Accepted 7 January 2013
Available online 17 January 2013

Keywords:
Nonparametric regression
Local linear composite quantile regression
Asymmetric distribution
Consistency
Asymptotic efficiency

ABSTRACT

It is known that for nonparametric regression, local linear composite quantile regression (local linear CQR) is a more competitive technique than classical local linear regression since it can significantly improve estimation efficiency under a class of non-normal and symmetric error distributions. However, this method only applies to symmetric errors because, without symmetric condition, the estimation bias is non-negligible and therefore the resulting estimator is inconsistent. In this paper, we propose a *weighted* local linear CQR method for general error conditions. This method applies to both symmetric and asymmetric random errors. Because of the use of weights, the estimation bias is eliminated asymptotically and the asymptotic normality is established. Furthermore, by minimizing asymptotic variance, the optimal weights are computed and consequently the optimal estimate (the most efficient estimate) is obtained. By comparing relative efficiency theoretically or numerically, we can ensure that the new estimation outperforms the local linear CQR estimation. Finite sample behaviors conducted by simulation studies further illustrate the theoretical findings.

1. Introduction

It is known that many smoothing methods, such as kernel regression, spline smoothing, orthogonal series approximation and local polynomial regression, have been proposed for nonparametric regression. Among the above linear smoothers, local polynomial regression, which has been thoroughly studied in the literature (Fan and Gijbels, 1996), is the best linear smoother in terms of minimax efficiency. However, local polynomial regression is not always the best choice because it can lose much efficiency and be much worse than local least absolute deviation polynomial regression when the error follows a Laplacian distribution (Fan et al., 1994; Welsh, 1996).

In contrast to the above estimation approaches for nonparametric regression, quantile regression is a robust alternative (Koenker and Bassett, 1978). This method has been deeply investigated in the literature and widely applied in econometrics, social sciences and biomedical studies; for comprehensive treatments see for example Koenker (2005). Recently, Zou and Yuan (2008) proposed a new regression technique called composite quantile regression (CQR) for linear models. The CQR model assumes that there exist common covariate effects in a range of quantiles such that the quantile levels only differ in terms of the intercept. The CQR enjoys great advantages in terms of estimation efficiency whether the variance of error is finite or not. It could be much more efficient than least squares. Subsequently, Kai et al. (2010)

* Corresponding author.
E-mail addresses: almasdu2010@gmail.com, almasdu@163.com (J. Sun), gaiyujie83@gmail.com (Y. Gai), linlu@sdu.edu.cn (L. Lin).

0378-3758/$ - see front matter
http://dx.doi.org/10.1016/j.jspi.2013.01.002

introduced local composite quantile regression (local CQR) for general nonparametric models. Such a method is valid for a wide class of symmetric error distributions. As a kind of nonlinear smoothers, local CQR does not require finite error variance, and hence can work well even when the error distribution has infinite variance. Moreover, Kai et al. (2010) have proved that local linear CQR could significantly improve the estimation efficiency of local linear least squares for the case of non-normal and symmetric error distributions.

Although local linear CQR is an efficient and robust alternative to local linear least squares, it is based on the assumption of symmetric random errors. Such an assumption condition is an indispensable prerequisite for estimation consistency since it can remove the non-vanishing term from the asymptotic representation of the estimation bias (see the proof of Theorem 1 in Kai et al., 2010). This implies that local linear CQR becomes invalid for asymmetric errors. In practice, however, there indeed exist many asymmetric error distributions. This motivates us to put forward a unified method to construct unbiased local linear CQR estimation for cases of both asymmetric and symmetric errors. In this paper, we propose a general method called *weighted* local linear CQR to estimate the nonparametric regression function. Our method, containing local linear CQR as a special case, not only inherits good properties that local linear CQR owns for symmetric errors but also is applicable to asymmetric error distributions. Because of the use of weights, the estimation bias is eliminated asymptotically and the asymptotic normality is established. Furthermore, via minimizing asymptotic variance, the optimal weights are computed and consequently the optimal estimate (the most efficient estimate) is obtained. Under some criteria such as relative efficiency and average squared errors, the new estimation outperforms local linear CQR estimation.

The rest of the paper is organized as follows. In Section 2, we first review the local linear CQR method (Kai et al., 2010) and then present our *weighted* local linear CQR method for general nonparametric regression. In Section 3, we study the asymptotic properties of the proposed method. The main theoretical results such as asymptotic normality and the optimal weights are presented in this section. Section 4 contains a comprehensive simulation study that indicates our method performs comparably to and in most cases better than both local linear CQR and local linear least squares in the sense of estimation bias and estimation efficiency. Also a real data analysis is reported in Section 4. All the conditions and technical proofs are deferred to the Appendix.

2. Methodology

Throughout this paper, we restrict our attention to the general nonparametric regression model

$$Y = m(T) + \sigma(T)\varepsilon,$$

where Y is the response variable, T is a scalar covariate independent of ε, $m(T) = \mathrm{E}(Y|T)$, which is assumed to be a smooth nonparametric function, and $\sigma(T)$ is a positive function representing the standard deviation of ε. Assume that $\mathrm{E}(\varepsilon) = 0$ and $\mathrm{var}(\varepsilon) = 1$. Let $F(\cdot)$ and $f(\cdot)$ denote the cumulative distribution function and density function of the error ε, respectively. Denote by $f_T(\cdot)$ the marginal density function of the covariate T. Suppose that $(t_i, y_i), i = 1, \ldots, n$, is an independent and identically distributed random sample of the population (T, Y). We consider estimating the value of $m(T)$ at t_0.

2.1. Local linear CQR

We first briefly recall the local linear CQR method. Let $\rho_{\tau_k}(r) = \tau_k r - rI(r<0), k = 1, 2, \ldots, q$, be q check loss functions at q quantile positions: $\tau_k = k/(q+1)$. Kai et al. (2010) considered minimizing the local CQR loss

$$\sum_{k=1}^{q}\left[\sum_{i=1}^{n}\rho_{\tau_k}\{y_i - a_k - b(t_i - t_0)\}K\left(\frac{t_i - t_0}{h}\right)\right], \tag{2.1}$$

where $K(\cdot)$ and h denote the smooth kernel function and the smoothing parameter, respectively. We adopt the majorization-and-minimization algorithm that was proposed by Hunter and Lange (2000) for solving the local linear CQR smoothing estimator. Denote the minimizer of (2.1) by $(\hat{a}_1, \ldots, \hat{a}_q, \hat{b})$ and let $\hat{a} = (\hat{a}_1, \hat{a}_2, \ldots, \hat{a}_q)'$. Then the local linear CQR of $m(t_0)$ is defined as

$$\hat{m}_{cqr}(t_0) = \frac{1}{q}\sum_{k=1}^{q}\hat{a}_k = \frac{1}{q}\mathbf{1}'\hat{a}, \tag{2.2}$$

where $\mathbf{1}$ is a q dimensional column vector with all elements 1.

It is worth pointing out that the local linear CQR estimator $\hat{m}_{cqr}(t_0)$ could be regarded as a weighted sum of initial estimators $\{\hat{a}_k, k = 1, \ldots, q\}$ with uniform weights $\{\omega_k = 1/q, k = 1, \ldots, q\}$. In addition, the local CQR loss (2.1) could be thought of as a weighted sum of check functions with uniform weights and uniform quantiles $\{\tau_k = k/(q+1), k = 1, \ldots, q\}$. By checking the proof procedures of their theoretical results, we find that the combination between uniform weights in $\hat{m}_{cqr}(t_0)$ and symmetric condition of the error distribution can remove the estimation bias successfully. As a result, the local linear CQR estimator $\hat{m}_{cqr}(t_0)$ is consistent.