Statistics and Probability Letters 117 (2016) 209–215

Contents lists available at ScienceDirect

Statistics and Probability Letters

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

Local asymptotics for nonparametric quantile regression with regression splines

Weihua Zhao [a], Heng Lian [b,*]

[a] School of Science, Nantong University, Nantong, PR China

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

ARTICLE INFO

Article history:
Received 3 January 2016
Received in revised form 2 June 2016
Accepted 2 June 2016
Available online 16 June 2016

Keywords:
Asymptotic normality
B-splines
Check loss function
L_∞ norm

ABSTRACT

We consider nonparametric quantile regression using B-splines and derive local asymptotic properties of the estimator. As a by-product, we establish the convergence rate of the estimator in L_∞ norm which seems to be missing in the literature. Simulations are carried out to investigate the coverage of the pointwise confidence interval.

1. Introduction

In this paper, we consider the nonparametric quantile regression model with independent and identically distributed (i.i.d.) observations $Y_i = g(X_i) + e_i$, $i = 1, \ldots, n$, where $P(e_i \leq 0|X_i) = \tau$ and X_i, Y_i are one-dimensional covariate and response, respectively. The parametric quantile regression introduced by Koenker and Bassett (1978) has been well developed in the econometrics and statistics literature. When the distribution of the errors in the model is heavy-tailed or the data contain some outliers, it is well known that median regression, a special case of quantile regression, is more robust than mean regression. More importantly, it can be used to obtain a large collection of conditional quantiles to characterize the entire conditional distribution. To construct a richer class of regression models capturing flexibly the relationships between the covariates and the response distribution, nonparametric quantile estimation has been studied in Hendricks and Koenker (1992) and Yu and Jones (1998). For varying coefficient models, Kim (2007) studied quantile regression for independent data using splines, and Cai and Xu (2008) used local polynomial estimation method for time series data. Further extensions to partially linear varying-coefficient models are considered by Wang et al. (2009) and Cai and Xiao (2012).

The approach we consider for estimation is using spline approximation. For quantile regression, He and Shi (1994) is probably the first that studied the spline estimation and demonstrated the convergence rate $\|\widehat{g} - g\| = O_p(\sqrt{\frac{K}{n}} + K^{-d})$ under some assumptions, where $\|\cdot\|$ denotes the L_2 norm of functions, K is the number of spline basis functions used in the approximation, and d is the smoothness parameter of g. Roughly speaking, the term K^{-d} in the convergence rate is associated with the approximation error of splines, or bias, while the term $\sqrt{K/n}$ is the variance part.

The focus of the current paper is on the inference for regression splines in quantile model. More specifically, we derive the asymptotic normality of $(\widehat{g}(x) - g(x))/\sqrt{Var(\widehat{g}(x)|\mathbf{X})}$, where $\widehat{Var}(\widehat{g}(x)|\mathbf{X})$ is an estimate of the variance of $\widehat{g}(x)$ given

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

http://dx.doi.org/10.1016/j.spl.2016.06.002
0167-7152/

$\mathbf{X} = (X_1, \ldots, X_n)^{\mathrm{T}}$. Such local asymptotics has attracted attention in nonparametric mean regression problems using regression splines, as investigated in Zhou et al. (1998) and Huang (2003). However, the corresponding result for quantile regression is lacking and in particular, currently there is no theoretically proven method of constructing pointwise confidence intervals of g using regression splines. Compared to mean regression, there is no closed form expression for $\widehat{g}(x)$ for quantile regression, making it more difficult to derive asymptotic normality. In this paper, we set out to derive the local asymptotics of the spline estimator in quantile regression. Furthermore, we also derive the convergence rate for $\|\widehat{g} - g\|_\infty$ where $\|\cdot\|_\infty$ is the supremum norm for functions. Such rate is missing from the literature of regression splines, as far as we know. It is actually easy to see that $\|\widehat{g} - g\| = O_p(\sqrt{\frac{K}{n}} + K^{-d})$ implies that $\|\widehat{g} - g\|_\infty = O_p(\sqrt{K}(\sqrt{\frac{K}{n}} + K^{-d}))$. We show actually the much better rate $O_p(\sqrt{\frac{K \log n}{n}} + K^{-d})$. Since any rate in L_∞ would imply the same rate in L_2 norm, we see that the derived L_∞ rate cannot be improved except for a log n factor.

In a recent paper, Chen and Liao (2014) considered inferences for general sieve estimators, which included our asymptotic normality result obtained here as a special case. In Remark 2.2 of Chen and Liao (2014), the L_∞ convergence rate $K/\sqrt{n}+K^{-d}$ was obtained for general sieve estimators. We believe our faster rate is optimal. Note that Comment 4.5 in Belloni et al. (2015) presented the same L_∞ convergence rate for least squares estimators. Besides this improvement, our treatment of this special case also makes the bias term explicit in the asymptotic normality, which is hidden in Chen and Liao (2014). Furthermore, we study the empirical coverage in simulations using under-smoothing to remove bias, which was not performed in Chen and Liao (2014).

The rest of the article is organized as follows. In Section 2, we detail the estimation approach and our theoretical results. Section 3 contains some simulation results showing the performance of the constructed pointwise confidence interval. We finally conclude with a discussion in Section 4.

2. Methodology and theory

We assume that the support of X is $[0, 1]$. Such a compactness assumption is always used in nonparametric regression spline estimation. We use polynomial splines to approximate the nonparametric function. Let $t_0 = 0 < t_1 < \cdots < t_{K'} < 1 = t_{K'+1}$ be a partition of $[0, 1]$ into subintervals $[t_k, t_{k+1})$, $k = 0, \ldots, K'$ with K' internal knots. For simplicity, we assume the knots are equally spaced on $[0, 1]$. A polynomial spline of order s is a function whose restriction to each subinterval is a polynomial of degree $s-1$ and globally $s-2$ times continuously differentiable on $[0, 1]$. The collection of splines with a fixed sequence of knots has a B-spline basis $\{B_1(x), \ldots, B_K(x)\}$ with $K = K'+s$. We assume the B-spline basis is normalized to have $\sum_{k=1}^K B_k(x) = \sqrt{K}$, although any scaling can be used without changing the theoretical results. Let $\mathbf{B}(.) = (B_1(.), \ldots, B_K(.))^{\mathrm{T}}$. Using spline estimator, writing $g(.) \approx \mathbf{B}^{\mathrm{T}}(.)\boldsymbol{\theta}$, we minimize

$$\sum_{i=1}^n \rho_\tau(Y_i - \mathbf{B}^{\mathrm{T}}(X_i)\boldsymbol{\theta}),$$

with the minimizer denoted by $\widehat{\boldsymbol{\theta}}$. The estimator of $g(x)$ is thus $\widehat{g}(x) = \mathbf{B}^{\mathrm{T}}(x)\widehat{\boldsymbol{\theta}}$.

We impose the following assumptions.

(A1) The covariate X has a density that is bounded and bounded away from zero on $[0, 1]$.

(A2) Let $f(.|X)$ be the conditional density of e. We assume $f(.|X)$ is bounded and bounded away from zero in a neighborhood of zero, uniformly over the support of X. The derivative of $f(.|X)$ is uniformly bounded in a neighborhood of zero over the support of X.

(A3) The function g is in the Hölder space of order $d > 1$. That is $|g^{(m)}(x) - g^{(m)}(y)| \le C|x - y|^r$ for $d = m + r$ and m is the largest integer strictly smaller than d, where $g^{(m)}$ is the mth derivative of g. The order of the spline is larger than $d + 1/2$.

All the above conditions are standard in the literature, as in He and Shi (1994).

Let $\boldsymbol{\theta}_0$ be spline coefficients in the best spline approximation of g with $\sup_t |g(t) - \mathbf{B}^{\mathrm{T}}(t)\boldsymbol{\theta}_0| \le CK^{-d}$ (De Boor, 2001). Define $R(x) = \mathbf{B}^{\mathrm{T}}(x)\boldsymbol{\theta}_0 - g(x)$, $R_i = \mathbf{B}^{\mathrm{T}}(X_i)\boldsymbol{\theta}_0 - g(X_i)$ and $\mathbf{R} = (R_1, \ldots, R_n)^{\mathrm{T}}$.

Theorem 1. *Under conditions* (A1)–(A3) *and that* $K \to \infty$, $K^{1-d}\log n \to 0$, $K^3 \log^3 n/n \to 0$, *we have*

$$\|\widehat{g} - g\|_\infty = O_p\left(\sqrt{\frac{K \log n}{n}} + K^{-d}\right),$$

and

$$\frac{\widehat{g}(x) - g(x) - [\mathbf{B}^{\mathrm{T}}(x)\boldsymbol{\theta}_0 - g(x) - \mathbf{B}^{\mathrm{T}}(x)(\mathbf{Z}^{\mathrm{T}}\mathbf{f}\mathbf{Z})^{-1}\mathbf{Z}^{\mathrm{T}}\mathbf{f}\mathbf{R}]}{\sqrt{\widehat{Var}(\widehat{g}(x)|\mathbf{X})}} \xrightarrow{d} N(0, 1),$$

where $\widehat{Var}(\widehat{g}(x)|\mathbf{X}) = \tau(1-\tau)\mathbf{B}^{\mathrm{T}}(x)(\mathbf{Z}^{\mathrm{T}}\mathbf{f}\mathbf{Z})^{-1}(\mathbf{Z}^{\mathrm{T}}\mathbf{Z})(\mathbf{Z}^{\mathrm{T}}\mathbf{f}\mathbf{Z})^{-1}\mathbf{B}(x)$, $\mathbf{Z} = [\mathbf{B}(X_1), \ldots, \mathbf{B}(X_n)]^{\mathrm{T}}$ *and* $\mathbf{f} = \mathrm{diag}\{f(0|X_1), \ldots, f(0|X_n)\}$.