Open Journal of Statistics, **2013, 3, 65-73**
http://dx.doi.org/10.4236/ojs.2013.32009 Published Online April 2013 (http://www.scirp.org/journal/ojs)

Composite Quantile Regression for Nonparametric Model with Random Censored Data

Rong Jiang, Weimin Qian
Department of Mathematics, Tongji University, Shanghai, China
Email: jrtrying@126.com, wmqian2003@yahoo.com.cn

Received May 29, 2012; revised June 30, 2012; accepted July 15, 2012

Copyright © 2013 Rong Jiang, Weimin Qian. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

ABSTRACT

The composite quantile regression should provide estimation efficiency gain over a single quantile regression. In this paper, we extend composite quantile regression to nonparametric model with random censored data. The asymptotic normality of the proposed estimator is established. The proposed methods are applied to the lung cancer data. Extensive simulations are reported, showing that the proposed method works well in practical settings.

Keywords: Kaplan-Meier Estimator; Censored Data; Composite Quantile Regression; Kernel Estimator; Nonparametric Model

1. Introduction

Consider the following nonparametric regression model with random censored data:

$$T = m(U) + \sigma(U)\varepsilon, \tag{1}$$

where $m(\cdot)$ is an unknown smoothing function, $\sigma(\cdot)$ is a positive function representing the standard deviation and ε is the random error with mean 0 and variance 1. Let C denote the censoring variable, whose distribution may depend on U, where U is vector of observed covariates. In this paper, we focus on random right censoring, we only observe the triples (U, Y, δ), where $Y = \min(T, C)$ and $\delta = I(T \le C)$ are the observed response variable and the censoring indicator respectively, where T is the survival time.

Censored quantile regression was first studied by [1] for fixed censoring. [2] proposed an estimator for a conditional quantile assuming that the regression models at lower quantiles are all linear. A recursively weighted estimation procedure that can be regarded as a generalization of the Kaplan-Meier estimator to conditional quantiles was described in their paper. Afterward, [3] presented an alternative approach that is based on the Nelson-Aalen estimator of the cumulative hazard function but still requires the same global-linearity assumption as Portnoy's. Their method provides a more direct approach to the asymptotic theory and a simpler computation algorithm. More recent studies by [4], proposed to overcome the global-linearity assumption by directly estimating the conditional censoring distribution nonparametrically using the local Kaplan-Meier method. Their computational algorithm is more stable and simpler to implement than Portnoy's or Peng and Huang's. Moreover, the local nonparametric estimator on which the model is based performs best when the covariates can be assumed independent.

Intuitively, the composite quantile regression (CQR) should provide estimation efficiency gain over a single quantile regression; see [5]. A composite quantile regression 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. From a more general regression perspective, composite quantile regression seeks to model a set of parallel regression curves, and thus it can be viewed as a compromise between a set of quantile regression curves with different intercepts and slopes and a single summary regression curve. [6] proposed the local polynomial CQR estimators (LCQR) 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. Furthermore, [7] studied semiparametric CQR estimates for semiparametric varying-coefficient partially linear model. They compared CQR with least squares and quantile regression, and the results showed that CQR outperformed both least squares and quantile regression. [8] considered CQR

estimates for single-index models. Recently, [9] extended the CQR method to linear model with randomly censored data. This motivates us to extend the CQR method to nonparametric model with censored data (LCQRC).

The paper is organized as follows. In Section 2, local composite quantile regression for nonparametric model with censored data is introduced, and the main theoretical results are also given in this section. Both simulation examples and a real data application are given in Section 3 to illustrate the proposed procedures. Final remarks are given in Section 4. The technical proofs are deferred to the Appendix.

2. Methodology

2.1. Local Composite Quantile Regression with Censored Data

We first consider an ideal situation where $F_0\left(t|U_i\right)$, the conditional cumulative distribution function of the survival time T_i given U_i, is assumed to be known. In this case, we define the following weight function:

$$\omega_i\left(F_0,\tau\right) = \begin{cases} 1, & \delta_i = 1 \text{ or } F_0\left(C_i|U_i\right) > \tau, \\ \dfrac{\tau - F_0\left(C_i|U_i\right)}{1 - F_0\left(C_i|U_i\right)}, & \delta_i = 0 \text{ and } F_0\left(C_i|U_i\right) < \tau, \end{cases} \tag{2}$$

$i = 1,\cdots,n$. In reality, $F_0\left(t|U_i\right)$ is unknown and has to be estimated. We propose to estimate $F_0\left(\cdot|U\right)$ nonparametrically using the local Kaplan-Meier estimator

$$\hat{F}\left(t|U\right) = 1 - \prod_{j=1}^{n}\left\{1 - \frac{B_{nj}\left(U\right)}{\sum_{s=1}^{n} I\left(Y_s \geq Y_j\right) B_{ns}\left(U\right)}\right\}^{\xi_j(t)},$$

where $\xi_j\left(t\right) = I\left(Y_j \geq t, \delta_j = 1\right)$ and $B_{nk}\left(U\right) = K\left(\left[U - U_k\right]/h^*\right) \Big/ \sum_{i=1}^{n} K\left(\left[U - U_i\right]/h^*\right)$, where $K\left(\cdot\right)$ is a smooth kernel function, $h^* \in R^+$ is the bandwidth converging to zero as $n \to \infty$. By plugging $\hat{F}\left(\cdot|U\right)$ into (2), we obtain the estimated local weights $\omega_i\left(\hat{F},\tau\right)$. Consider estimating the value of $m\left(U\right)$ at u_0. The LCQRC procedure estimates $m\left(u_0\right)$, defined by $\hat{m}\left(u_0\right) = \frac{1}{q}\sum_{k=1}^{q}\hat{a}_k$, via minimizing the locally weighted objective function

$$\sum_{k=1}^{q}\sum_{i=1}^{n}\Big[\omega_i\left(\hat{F},\tau_k\right)\rho_{\tau_k}\left\{Y_i - a_k - b\left(U_i - u_0\right)\right\} + \left(1 - \omega_i\left(\hat{F},\tau_k\right)\right)\rho_{\tau_k}\left\{Y_i^{+\infty} - a_k - b\left(U_i - u_0\right)\right\}\Big] K\left(\frac{U_i - u_0}{h}\right)$$

where $\rho_{\tau_k}\left(r\right) = \tau_k r - rI\left(r < 0\right), k = 1,2,\cdots,q$, be q check loss functions at q quantile positions: $\tau_k = k/\left(q+1\right)$ and $Y^{+\infty}$ is any value sufficiently large to exceed all $m\left(U_i\right)$.

Remark 1. The detail explant of $\omega_i\left(F_0,\tau\right)$ can see Remark 1 of [4].

2.2. Asymptotic Properties

Denote by $f_U\left(\cdot\right)$ the marginal density function of the covariate U, $\mu_j = \int u^j K\left(u\right)\mathrm{d}u$ and $\nu_j = \int u^j K^2\left(u\right)\mathrm{d}u$. To prove main results in this paper, the following technical conditions are imposed.

A1. The functions $F_0\left(t|U\right)$ and $G\left(t|U\right)$ have first derivatives with respect to t, denoted as $f_0\left(t|U\right)$ and $g\left(t|U\right)$, which are uniformly bounded away from infinity. In addition, $F_0\left(t|U\right)$ and $G\left(t|U\right)$ have bounded second order partial derivatives with respect to U.

A2. Σ is positive definite matrix.

A3. $m\left(U\right)$ has a continuous second derivative in the neighborhood of u_0.

A4. $f_U\left(\cdot\right)$ is differentiable and positive in the neighborhood of u_0.

A5. The conditional variance $\sigma^2\left(u\right)$ is continuous in the neighborhood of u_0.

A6. Assume that the error has a symmetric distribution with a positive density $f_0\left(\cdot\right)$.

Remark 2. Assumption A1 is needed for the local Kaplan-Meier estimator. It allows us to obtain the local expansions of $F_0\left(t|U\right)$ and $G\left(t|U\right)$ in the neighborhood of $m\left(U\right)$, and to obtain the uniform consistency and the linear representation of $\hat{F}\left(t|U\right)$, which are needed for deriving the asymptotic normality result. Assumption A2 ensures that the expectation of the estimating function has a unique zero, and it is needed to establish the asymptotic distribution. Assumptions A3-A6 are the same conditions for establishing the asymptotic normality of local composite quantile regression ([6]).

We state the asymptotic normality for $\hat{m}\left(u_0\right)$ in the following theorem.

Theorem 1. Assume that the triples $\left(U_i, Y_i, \delta_i\right)$ constitute and i.i.d. multivariate random sample, and that the censoring variable C_i is independent of T_i conditional on the covariate U_i. Suppose that u_0 is an interior of the support of $f_U\left(\cdot\right)$. Under the regularity conditions A1-A6, if $h \to 0$ and $nh \to \infty$, then

$$\sqrt{nh}\left[\hat{m}\left(u_0\right) - m\left(u_0\right) - \frac{1}{2}m''\left(u_0\right)\mu_2 h^2\right] \xrightarrow{L} N\left(0,\Sigma\right),$$

where $\xrightarrow{L}$ stands for convergence in distribution and $\Sigma = \dfrac{\nu_0}{f_U\left(u_0\right)}\dfrac{1}{q^2}\sum_{k,k'=1}^{q}\Gamma_k^{-1}\Lambda_{kk'}\Gamma_{k'}^{-1}$, where

Copyright © 2013 SciRes.

$$\Gamma_k = E \cdot\left[\left(1-G\left(m(U)+\sigma(U)c_k|U\right)\right) f_0\left(m(U)+\sigma(U)c_k|U\right)\right],$$

$$\Lambda_{kk'} = E_{(U,C)} \left[I\left(C \le m(U)+\sigma(U)c_k \wedge c_{k'}\right)\cdot \frac{F_0\left(C|U\right)\left(1-\tau_k\right)\left(1-\tau_{k'}\right)}{1-F_0\left(C|U\right)} + I\left(C > m(U)+\sigma(U)c_k \wedge c_{k'}\right)\cdot\left(\tau_k \wedge \tau_{k'} - \tau_k\tau_{k'}\right)\right]$$

and $c_k = F^{-1}\left(\tau_k\right)$.

3. Numerical Studies

In this section, we conduct simulation studies to assess the finite sample performance of the proposed procedures and illustrate the proposed methodology on a lung cancer data set. Moreover, we compare the performance of the newly proposed method with LCQR ([6]) and nonparametric quantile regression with censored data (NQRC) that was proposed by [10].

In the proposed compute process, we take $Y^{+\infty} = 100\max\left\{Y_1,\cdots,Y_n\right\}$ and $K = \frac{15}{16}\left(1-U^2\right)^2 I\left(|U| \le 1\right)$. The bandwidth h^* can be obtained by 10-fold cross-validation method (see [4]), and we use the short-cut strategy method to select h (see [6]).

3.1. Example 1

The data are generated from the following model

$$T_i = 10U_i \sin\left(2\pi U_i\right) + \varepsilon_i, i = 1,\cdots,n,$$

where U_i is uniformly distributed on $[0,1]$ and ε is i.i.d. standard normal random variables. The censoring variable $C_i \sim U\left(0,c\right)$ and $Y_i = \min\left(T_i, C_i\right)$. The value of the constant c in the model determines the censoring proportion. In our simulations, we consider three censoring rates (CR): 20%, 30% and 40%. For each censoring rate, the sample sizes are taken to be 100 and 200. To evaluate the finite sample performance of our estimator. Two distance measures are approximated, the first one the mean absolute deviation error (MADE) is given by $n^{-1}\sum_{i=1}^{n}\left|\hat{m}\left(U_i\right) - m\left(U_i\right)\right|$, and the second one the mean squared error (MSE) defines as $n^{-1}\sum_{i=1}^{n}\left[\hat{m}\left(U_i\right) - m\left(U_i\right)\right]^2$. Furthermore, we define the rate of MADE and MSE which are $\mathrm{RMADE} = \mathrm{MADE}/\mathrm{MADE}_{\mathrm{LCQR}}$ and $\mathrm{RMSE} = \mathrm{MSE}/\mathrm{MSE}_{\mathrm{LCQR}}$.

For right censored data, quantile functions with τ close to 1 may not be identifiable due to censorship. In our similations, we consider $q = 5$ for LCQR and LCQRC estimators. The means and standard deviations of MADE, MSE, RMADE and RMSE are respectively reported in **Table 1** and **Table 2**. From **Tables 1** and **2**, we can make the following observations: the performance of proposal method is better than that of LCRQ and NQRC. Moreover, LCQRC estimators are much more accurate when sample sizes increase. **Figure 1** summarize the Curve estimates for three censoring rates of 20%, 30% and 40% with different sample sizes. It shows that the performance of LCQRC is very close to the true value.

3.2. Example 2

It is necessary to investigate the effect of heteroscedastic errors. The observations $\left(U_i, T_i, C_i\right), i = 1,\cdots,n$, are generated from following model

$$T_i = 2\exp\left(-16U_i^2\right) + \sin\left(\pi U_i\right) + 0.2U_i\varepsilon_i, i = 1,\cdots,n,$$

where U_i, ε_i, and C_i are generated following the same way as in Example 1. The means and standard deviations of MADE, MSE, RMADE and RMSE are respectively reported in **Table 3** and **Table 4**. The

Table 1. Simulation results of $\hat{m}(\cdot)$ with n = 100 for Example 1.

CR	Method	RMADE	MADE	RMSE	MSE
20%	LCQR5	-	0.7169 (0.5047)	-	0.7804 (0.9798)
	LCQRC5	0.7946 (0.1195)	0.5643 (0.4679)	0.7297 (0.2452)	0.5481 (0.8718)
	NQRC0.5	0.9655 (0.1837)	0.5919 (0.4155)	0.9338 (0.3355)	0.5476 (0.6646)
30%	LCQR5	-	0.8284 (0.6031)	-	1.0578 (1.2602)
	LCQRC5	0.6607 (0.1187)	0.5431 (0.4468)	0.4903 (0.1639)	0.5060 (0.7765)
	NQRC0.5	0.8378 (0.3501)	0.6960 (0.4994)	0.7920 (0.6298)	0.8650 (1.0371)
40%	LCQR5	-	1.0102 (0.7520)	-	1.6054 (1.8037)
	LCQRC5	0.5501 (0.1348)	0.5507 (0.5448)	0.4332 (0.3251)	0.6810 (1.9258)
	NQRC0.5	0.7222 (0.1882)	0.7169 (0.5424)	0.5548 (0.2795)	0.8530 (1.1186)

Copyright © 2013 SciRes.

Table 2. Simulation results of $\hat{m}(\cdot)$ with n = 200 for Example 1.

CR	Method	RMADE	MADE	RMSE	MSE
20%	LCQR5	-	0.5374 (0.3750)	-	0.4386 (0.5481)
	LCQRC5	0.7327 (0.1089)	0.3893 (0.3443)	0.6612 (0.2327)	0.2737 (0.5262)
	NQRC0.5	0.8356 (0.2114)	0.5443 (0.3769)	0.8254 (0.2910)	0.4732 (0.5346)
30%	LCQR5	-	0.7342 (0.5154)	-	0.8121 (0.9382)
	LCQRC5	0.5768 (0.1191)	0.4191 (0.3379)	0.3815 (0.1658)	0.2988 (0.4677)
	NQRC0.5	0.7539 (0.1344)	0.5484 (0.4015)	0.5985 (0.1898)	0.4729 (0.5870)
40%	LCQR5	-	0.9755 (0.7164)	-	1.4825 (1.5772)
	LCQRC5	0.4231 (0.0771)	0.4084 (0.3504)	0.2084 (0.0839)	0.2957 (0.5196)
	NQRC0.5	0.7689 (0.2634)	0.7355 (0.4962)	0.6059 (0.4493)	0.8477 (0.9701)

Table 3. Simulation results of $\hat{m}(\cdot)$ with n = 100 for Example 2.

CR	Method	RMADE	MADE	RMSE	MSE
20%	LCQR5	-	0.0639 (0.0538)	-	0.0074 (0.0175)
	LCQRC5	0.9620 (0.1600)	0.0603 (0.0451)	0.9230 (0.4398)	0.0059 (0.0123)
	NQRC0.5	1.2231 (0.3356)	0.0750 (0.0500)	1.5544 (1.0808)	0.0085 (0.0110)
30%	LCQR5	-	0.0936 (0.0856)	-	0.0191 (0.0337)
	LCQRC5	0.8380 (0.2094)	0.0791 (0.1318)	0.6470 (0.4510)	0.0065 (0.0028)
	NQRC0.5	0.8502 (0.3245)	0.0714 (0.0556)	0.7788 (0.6206)	0.0089 (0.0137)
40%	LCQR5	-	0.1564 (0.1549)	-	0.0560 (0.0957)
	LCQRC5	0.5269 (0.1720)	0.0676 (0.0763)	0.3816 (0.5250)	0.0167 (0.0962)
	NQRC0.5	0.5836 (0.3207)	0.0796 (0.0613)	0.3874 (0.5181)	0.0117 (0.0181)

Table 4. Simulation results of $\hat{m}(\cdot)$ with n = 200 for Example 2.

CR	Method	RMADE	MADE	RMSE	MSE
20%	LCQR5	-	0.0540 (0.0439)	-	0.0051 (0.0103)
	LCQRC5	0.9244 (0.2066)	0.0488 (0.0339)	0.8220 (0.3890)	0.0036 (0.0068)
	NQRC0.5	1.1134 (0.2889)	0.0599 (0.0433)	1.2714 (0.7178)	0.0059 (0.0082)
30%	LCQR5	-	0.0811 (0.0812)	-	0.0134 (0.0248)
	LCQRC5	0.6377 (0.1167)	0.0506 (0.0378)	0.3222 (0.1032)	0.0040 (0.0074)
	NQRC0.5	0.8157 (0.2516)	0.0638 (0.0493)	0.5476 (0.2898)	0.0067 (0.0095)
40%	LCQR5	-	0.1492 (0.1297)	-	0.0432 (0.0678)
	LCQRC5	0.4230 (0.1894)	0.0554 (0.0451)	0.1769 (0.1246)	0.0054 (0.0163)
	NQRC0.5	0.5235 (0.1757)	0.0714 (0.0474)	0.2591 (0.1744)	0.0077 (0.0104)

Copyright © 2013 SciRes.

performance of LCQRC is presented in **Figure 2**. The results of Example 1 and Example 2 show very similar messages.

3.3. Example 3

As an illustration, we now apply the proposed LCQRC to the lung cancer data. The data contain 228 observations on ten variables. The censoring percentage is 27%, so the estimators are expected to perform well. More details about the study can be found in [11], and the dataset is included in the R package *lung*. We are interested in estimating the conditional of survival time (in days) given age (in years). Here, we use model (1) to fit the lung cancer data, where Y is the $\log_{10}$ (survival time) and U is the age/100. To evaluate the performance of our estimator. Two distance measures were approximated, the first one the mean absolute deviation error $\left(\text{MADE}_y\right)$ given by $n^{-1}\sum_{i=1}^{n}\left|\hat{y}_i - y_i\right|$, and the second one the mean squared error $\left(\text{MSE}_y\right)$ defined as $n^{-1}\sum_{i=1}^{n}\left[\hat{y}_i - y_i\right]^2$, where $n = 228$. Furthermore, we define the rate of MADE_y and MSE_y which are $\text{RMADE}_y = \text{MADE}_y / \text{MADE}_{y_{\text{LCQR5}}}$ and $\text{RMSE}_y = \text{MSE}_y / \text{MSE}_{y_{\text{LCQR5}}}$. Next, we report and compare results with LCQR and NQRC for estimating the survival time. The simulation results for the LCQR, LCQRC and NQRC are given in **Table 5**. It shows that LCQRC is better than that of LCRQ and NQRC. **Figure 3** summarize the simulation results for LCQRC5. It

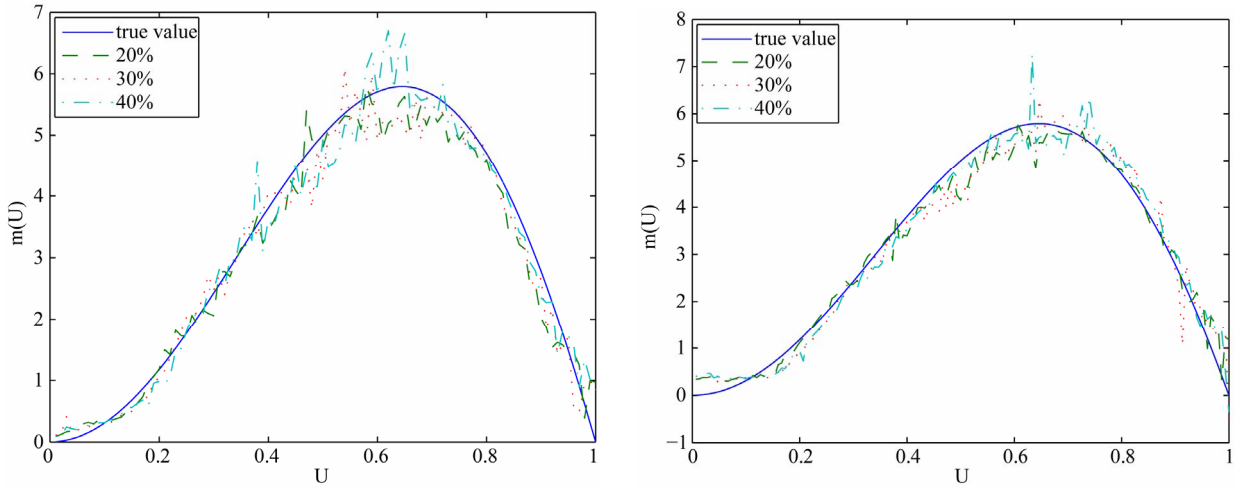

Figure 1. Curve estimates of $\hat{m}(\cdot)$ for Example 1. (a) n = 100; (b) n = 200.

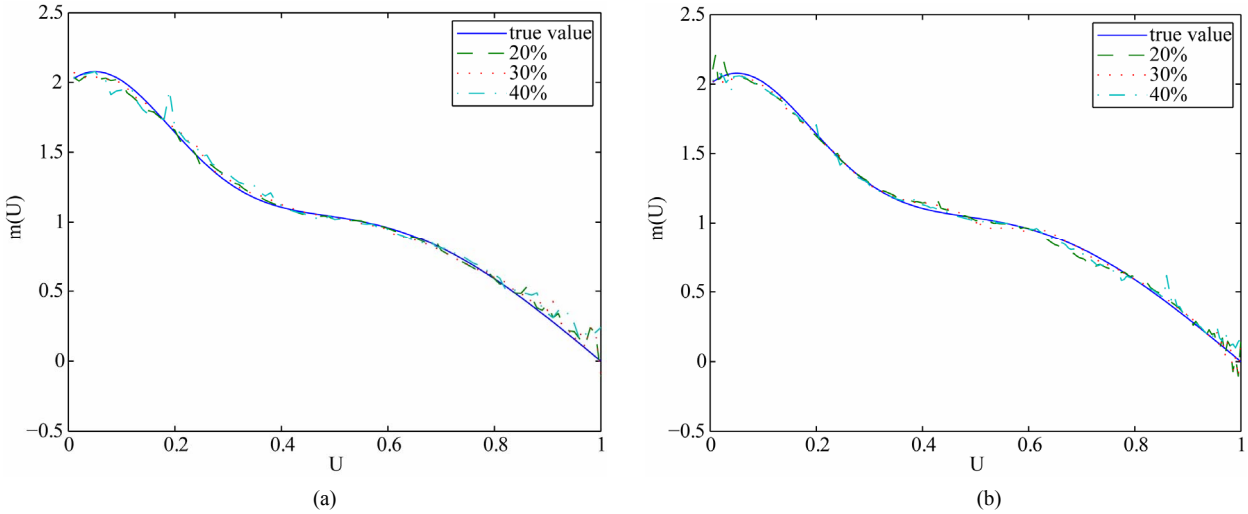

Figure 2. Curve estimates of $\hat{m}(\cdot)$ for Example 2. (a) n = 100; (b) n = 200.

Copyright © 2013 SciRes.

Table 5. Simulation results of $\hat{y}$ for lung cancer data.

Method	$RMADE_y$	$MADE_y$	$RMSE_y$	MSE_y
LCQR5	-	0.6877 (0.7074)	-	0.9711 (2.2640)
LCQRC5	0.9941	0.6836 (0.6788)	0.9536	0.9261 (2.0939)
NQRC0.5	1.1172	0.7683 (0.5738)	0.9454	0.9181 (1.3586)

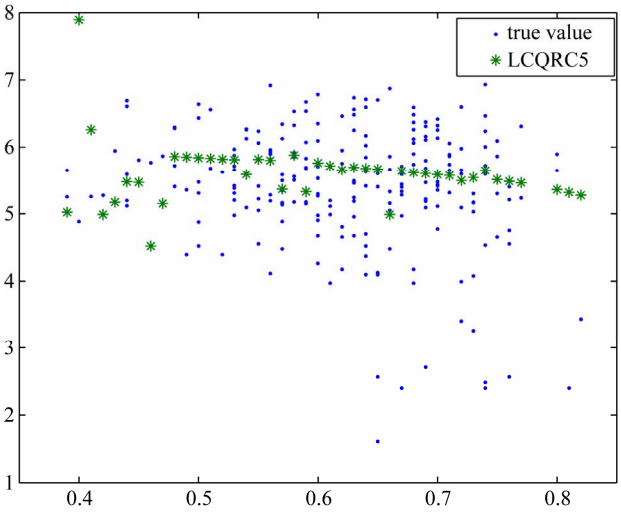

Figure 3. Curve estimates for lung cancer data.

shows that the proposal is valid.

4. Conclusion

In this work, we have focused on the LCQR for non-parametric model with censored data and its nice theoretical properties have been proven. The proposed approaches are demonstrated by simulation examples and real data applications. In addition, we believe the method can be extended to varying coefficient model (see [7]).

REFERENCES

[1] J. L. Powell, "Least Absolute Deviations Estimation for the Censored Regression Model," *Journal of Econometrics*, Vol. 25, No. 3, 1984, pp. 303-325. doi:10.1016/0304-4076(84)90004-6

[2] S. Portnoy, "Censored Regression Quantiles," *Journal of the American Statistical Association*, Vol. 98, No. 464, 2003, pp. 1001-1012. doi:10.1198/016214503000000954

[3] L. Peng and Y. Huang, "Survival Analysis with Quantile Regression Models," *Journal of the American Statistical Association*, Vol. 103, No. 482, 2008, pp. 637-649. doi:10.1198/016214508000000355

[4] H. J. Wang and L. Wang, "Locally Weighted Censored Quantile Regression," *Journal of the American Statistical Association*, Vol. 104, No. 478, 2009, pp. 1117-1128. doi:10.1198/jasa.2009.tm08230

[5] H. Zou and M. Yuan, "Composite Quantile Regression and the Oracle Model Selection Theory," *Annals of Statistics*, Vol. 36, No. 3, 2008, pp. 1108-1126. doi:10.1214/07-AOS507

[6] B. Kai, R. Li and H. Zou, "Local Composite Quantile Regression Smoothing: An Efficient and Safe Alternative to Local Polynomial Regression," *Journal of the Royal Statistical Society*, *Series B*, Vol. 72, No. 1, 2010, pp. 49-69. doi:10.1111/j.1467-9868.2009.00725.x

[7] B. Kai, R. Li and H. Zou, "New Efficient Estimation and Variable Selection Methods for Semiparametric Varying-Coefficient Partially Linear Models," *Annals of Statistics*, Vol. 39, No. 1, 2011, pp. 305-332. doi:10.1214/10-AOS842

[8] R. Jiang, Z. G. Zhou, W. M. Qian and W. Q. Shao, "Single-Index Composite Quantile Regression," *Journal of the Korean Statistical Society*, Vol. 3, No. 3, 2012, pp. 323-332. doi:10.1016/j.jkss.2011.11.001

[9] R. Jiang, W. M. Qian and Z. G. Zhou, "Variable Selection and Coefficient Estimation via Composite Quantile Regression with Randomly Censored Data," *Statistics & Probability Letters*, Vol. 2, No. 2, 2012, pp. 308-317. doi:10.1016/j.spl.2011.10.017

[10] A. Gannoun, J. Saracco, A. Yuan and G. Bonney, "Non-Parametric Quantile Regression with Censored Data," *Scandinavian Journal of Statistics*, Vol. 32, No. 4, 2005, pp. 527-550. doi:10.1111/j.1467-9469.2005.00456.x

[11] C. L. Loprinzi, *et al*., "Prospective Evaluation of Prognostic Variables from Patient-Completed Questionnaires. North Central Cancer Treatment Group," *Journal of Clinical Oncology*, Vol. 12, No. 3, 1994, pp. 601-607.

[12] W. Gonzalez-Manteiga and C. Cadarso-Suarez, "Asymptotic Properties of a Generalized Kaplan-Meier Estimator with Some Applications," *Journal of Nonparametric Statistics*, Vol. 4, No. 1, 1994, pp. 65-78. doi:10.1080/10485259408832601

[13] K. Knight, "Limiting Distributions for L_1 Regression Estimators under General Conditions," *Annals of Statistics*, Vol. 26, No. 2, 1998, pp. 755-770. doi:10.1214/aos/1028144858

Copyright © 2013 SciRes.

Appendix

Lemma 1. Assume assumption A1 hold. Then

$$\left\|\hat{F}-F_0\right\| = \sup_t \sup_U \left|\hat{F}\left(t|U\right)-F_0\left(t|U\right)\right| = O_p\left(\left(\log n\right)^{1/2} n^{-1/4-\gamma_0/2}\right),$$

where $0<\gamma_0<1/4$.

Proof. This follows directly from theorem 2.1 of [12].

Proof of Theorem 1 Let

$$\sqrt{nh}\left(a_k - m\left(u_0\right) - \sigma\left(u_0\right)c_k\right) = \eta_k,$$
$$h\sqrt{nh}\left(b - m'\left(u_0\right)\right) = v, \quad \Delta_{i,k} = \left(\eta_k + vx_i\right)\Big/\sqrt{nh},$$
$$x_i = \left(U_i - u_0\right)/h,$$
$$d_i\left(u_0\right) = m\left(U_i\right) - m\left(u_0\right) - m'\left(u_0\right)\left(U_i - u_0\right) + c_k\left(\sigma\left(U_i\right) - \sigma\left(u_0\right)\right)$$

Then $\left(\eta_1,\cdots,\eta_q,v\right)$ is the minimizer of the following criterion:

$$L_n = \sum_{k=1}^{q}\sum_{i=1}^{n} K_i\Big[\omega_i\left(\hat{F},\tau_k\right)\left\{\rho_{\tau_k}\left(A_i - \sigma\left(U_i\right)c_k + d_i\left(u_0\right) - \Delta_{i,k}\right) - \rho_{\tau_k}\left(A_i - \sigma\left(U_i\right)c_k + d_i\left(u_0\right)\right)\right\} + \left(1-\omega_i\left(\hat{F},\tau_k\right)\right)\left\{\rho_{\tau_k}\left(B_i - \sigma\left(U_i\right)c_k + d_i\left(u_0\right) - \Delta_{i,k}\right) - \rho_{\tau_k}\left(B_i - \sigma\left(U_i\right)c_k + d_i\left(u_0\right)\right)\right\}\Big],$$

where $A_i = Y_i - m\left(U_i\right), B_i = Y_i^{+\infty} - m\left(U_i\right)$ and $K_i = K\left\{\left(U_i - u_0\right)/h\right\}$. To apply the identity ([13])

$$\rho_\tau\left(x-y\right) - \rho_\tau\left(x\right) = y\left\{I\left(x\le 0\right) - \tau\right\} + \int_0^y \left\{I\left(x\le z\right) - I\left(x\le 0\right)\right\}\mathrm{d}z,$$

we have

$$\begin{aligned} L_n &= \sum_{k=1}^{q}\sum_{i=1}^{n} K_i\Delta_{i,k}\omega_i\left(\hat{F},\tau_k\right)\left[I\left(A_i < \sigma\left(U_i\right)c_k - d_i\left(u_0\right)\right) - \tau_k\right] \\ &\quad + \sum_{k=1}^{q}\sum_{i=1}^{n} K_i\int_0^{\Delta_{i,k}} \omega_i\left(\hat{F},\tau_k\right)\left[I\left(A_i < \sigma\left(U_i\right)c_k - d_i\left(u_0\right) + t\right) - I\left(A_i < \sigma\left(U_i\right)c_k - d_i\left(u_0\right)\right)\right]\mathrm{d}t \\ &\quad + \sum_{k=1}^{q}\sum_{i=1}^{n} K_i\Delta_{i,k}\left(1-\omega_i\left(\hat{F},\tau_k\right)\right)\left[I\left(B_i < \sigma\left(U_i\right)c_k - d_i\left(u_0\right)\right) - \tau_k\right] \\ &\quad + \sum_{k=1}^{q}\sum_{i=1}^{n} K_i\int_0^{\Delta_{i,k}} \left(1-\omega_i\left(\hat{F},\tau_k\right)\right)\left[I\left(B_i < \sigma\left(U_i\right)c_k - d_i\left(u_0\right) + t\right) - I\left(B_i < \sigma\left(U_i\right)c_k - d_i\left(u_0\right)\right)\right]\mathrm{d}t \\ &\triangleq L_{1n} + L_{2n} + L_{3n} + L_{4n}. \end{aligned}$$

Since $Y^{+\infty}$ is any value sufficiently large to exceed all $m\left(U_i\right)$, $L_{3n} = -\sum_{k=1}^{q}\sum_{i=1}^{n} K_i\Delta_{i,k}\left(1-\omega_i\left(\hat{F},\tau_k\right)\right)\tau_k$ and $L_{4n} = 0$, then $L_{1n} + L_{3n} = \sum_{k=1}^{q}\sum_{i=1}^{n} K_i\Delta_{i,k}\left(\omega_i\left(\hat{F},\tau_k\right)I\left(A_i < \sigma\left(U_i\right)c_k - d_i\left(u_0\right)\right) - \tau_k\right)$.

Denote $L_{2n} = \sum_{k=1}^{q} L_{2n}^{(k)}$, where

$$L_{2n}^{(k)} = \sum_{i=1}^{n} K_i\int_0^{\Delta_{i,k}} \omega_i\left(\hat{F},\tau_k\right)\left[I\left(A_i < \sigma\left(U_i\right)c_k - d_i\left(u_0\right) + t\right) - I\left(A_i < \sigma\left(U_i\right)c_k - d_i\left(u_0\right)\right)\right]\mathrm{d}t.$$

$$\begin{aligned} &\omega_i\left(F,\tau_k\right)I\left(A_i < \sigma\left(U_i\right)c_k - d_i\left(u_0\right)\right) \\ &= I\left(C_i > m\left(U_i\right) + \sigma\left(U_i\right)c_k - d_i\left(u_0\right), T_i \le m\left(U_i\right) + \sigma\left(U_i\right)c_k - d_i\left(u_0\right)\right) + I\left(C_i \le m\left(U_i\right) + \sigma\left(U_i\right)c_k - d_i\left(u_0\right), T_i \le C_i\right) \\ &\quad + I\left(C_i \le m\left(U_i\right) + \sigma\left(U_i\right)c_k - d_i\left(u_0\right), T_i > C_i\right)\times\left(1 - \frac{1-\tau_k}{1-F_i\left(C_i\right)}I\left(F_i\left(C_i\right) < \tau_k\right)\right). \end{aligned}$$

By the conditional independence of T and C given U, we have

Copyright © 2013 SciRes.

$$E\left[I\left(C>t,T<t\right)\middle|U\right]=P\left(C>t\middle|U\right)P\left(T<t\middle|U\right)=\left[1-G\left(t\middle|U\right)\right]F_0\left(t\middle|U\right),$$

$$E\left[I\left(C<t,T<C\right)\middle|U\right]=E_{C|U}\left[I\left(C<t\right)P\left(T<C\middle|C,U\right)\right]=\int_{-\infty}^{t}F_0\left(s\middle|U\right)g\left(s\middle|U\right)\mathrm{d}s.$$

Therefore,

$$\begin{aligned}&E\left[\omega_i\left(F_0,\tau_k\right)I\left(A_i\le\sigma\left(U_i\right)c_k-d_i\left(u_0\right)\middle|U\right)\right]\\&=\left\{1-G\left(m\left(U\right)+\sigma\left(U_i\right)c_k-d\middle|U\right)\right\}F_0\left(m\left(U\right)+\sigma\left(U_i\right)c_k-d\middle|U\right)+\int_{-\infty}^{m(U)+\sigma(U_i)c_k-d}F_0\left(s\middle|U\right)g\left(s\middle|U\right)\mathrm{d}s\\&\quad+\int_{-\infty}^{m(U)+\sigma(U_i)c_k-d}\left\{1-F_0\left(s\middle|U\right)\right\}\times\left[1-\frac{1-\tau_k}{1-F_0\left(s\middle|U\right)}I\left(F_0\left(s\middle|U\right)<\tau_k\right)\right]g\left(s\middle|U\right)\mathrm{d}s.\end{aligned}$$

By Lemma 1, we have

$$\begin{aligned}&\left|\sum_{i=1}^{n}K_i\int_0^{\Delta_{i,k}}E\left\{\left[\omega_i\left(\hat{F},\tau_k\right)-\omega_i\left(F_0,\tau_k\right)\right]\left[I\left(A_i\le\sigma\left(U_i\right)c_k+t\right)-I\left(A_i\le\sigma\left(U_i\right)c_k\right)\right]\middle|U\right\}\mathrm{d}t\right|\\&\le\sum_{i=1}^{n}K_i\int_0^{\Delta_{i,k}}E\left\{\left|\omega_i\left(\hat{F},\tau_k\right)-\omega_i\left(F_0,\tau_k\right)\right|\left|I\left(A_i\le\sigma\left(U_i\right)c_k+t\right)-I\left(A_i\le\sigma\left(U_i\right)c_k\right)\right|\middle|U\right\}\mathrm{d}t\\&=\sum_{i=1}^{n}K_i\int_0^{\Delta_{i,k}}E\left\{O\left(\left|\hat{F}-F_0\right|\right)\left|I\left(A_i\le\sigma\left(U_i\right)c_k+t\right)-I\left(A_i\le\sigma\left(U_i\right)c_k\right)\right|\middle|U\right\}\mathrm{d}t\\&=\sum_{i=1}^{n}K_i\int_0^{\Delta_{i,k}}E\left\{\left|I\left(A_i\le\sigma\left(U_i\right)c_k+t\right)-I\left(A_i\le\sigma\left(U_i\right)c_k\right)\right|\middle|U\right\}\mathrm{d}t\cdot o_p\left(1\right)\to 0.\end{aligned}$$

Then, we can obtain

$$\begin{aligned}&E\left[L_{2n}^{(k)}\middle|U\right]\\&=\sum_{i=1}^{n}K_i\int_0^{\Delta_{i,k}}E\left\{\left(\omega_i\left(\hat{F},\tau_k\right)\right)\left[I\left(A_i<\sigma\left(U_i\right)c_k-d_i\left(u_0\right)+t\right)-I\left(A_i<\sigma\left(U_i\right)c_k-d_i\left(u_0\right)\right)\right]\middle|U\right\}\mathrm{d}t\\&=\sum_{i=1}^{n}K_i\int_0^{\Delta_{i,k}}E\left(\omega_i\left(F_0,\tau_k\right)\left[I\left(A_i<\sigma\left(U_i\right)c_k-d_i\left(u_0\right)+t\right)-I\left(A_i<\sigma\left(U_i\right)c_k-d_i\left(u_0\right)\right)\right]\middle|U\right)\mathrm{d}t+o_p\left(1\right)\\&\to\frac{1}{2}f_U\left(u_0\right)\left(\eta_k,v\right)\begin{pmatrix}\Gamma_k & 0\\0 & \mu_2\Gamma_k\end{pmatrix}\left(\eta_k,v\right)^{\mathrm{T}}.\end{aligned}$$

$$\begin{aligned}&\mathrm{Var}\left[L_{2n}^{(k)}\middle|U\right]\\&=\sum_{i=1}^{n}\mathrm{Var}\left\{\left(K_i\int_0^{\Delta_{i,k}}\omega_i\left(\hat{F},\tau_k\right)\left[I\left(A_i<\sigma\left(U_i\right)c_k-d_i\left(u_0\right)+t\right)-I\left(A_i<\sigma\left(U_i\right)c_k-d_i\left(u_0\right)\right)\right]\mathrm{d}t\right)\middle|U\right\}\\&\le\sum_{i=1}^{n}E\left\{\left(K_i\int_0^{\Delta_{i,k}}\left(\omega_i\left(\hat{F},\tau_k\right)\left[I\left(A_i<\sigma\left(U_i\right)c_k-d_i\left(u_0\right)+t\right)-I\left(A_i<\sigma\left(U_i\right)c_k-d_i\left(u_0\right)\right)\right]\mathrm{d}t\right)^2\middle|U\right\}\\&=o\left(\sum_{i=1}^{n}K_i^2\Delta_{i,k}^2\right)=o_p\left(1\right).\end{aligned}$$

So, we can obtain $L_{2n}^{(k)}\to\frac{1}{2}f_U\left(u_0\right)\left(\eta_k,v\right)\begin{pmatrix}\Gamma_k & 0\\0 & \mu_2\Gamma_k\end{pmatrix}\left(\eta_k,v\right)^{\mathrm{T}}$, then

$$\hat{\eta}_k=f_U^{-1}\left(u_0\right)\Gamma_k^{-1}M_k,$$

where

Copyright © 2013 SciRes.

$$M_k(u_0)=\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}K_i\left[\tau_k-\omega_i\left(\hat{F},\tau_k\right)I\left(Y_i<m(U_i)+\sigma(U_i)c_k-d_i(u_0)\right)\right].$$

Note that the error is symmetric, thus $\sum_{k=1}^{q}c_k=0$, then it follows that

$$\sqrt{nh}\left[\hat{m}(U_0)-m(U_0)\right]=\frac{1}{q}\sum_{k=1}^{q}\hat{\eta}_k.$$

Since $E\left[\tau_k-\omega_i\left(F_0,\tau_k\right)I\left(Y_i<m(U_i)+\sigma(U_i)c_k\right)\right]=0$, then

$$\begin{aligned}
&\frac{1}{\sqrt{nh}}E\left[M_k(u_0)|U\right]\\
&=\frac{1}{nh}\sum_{i=1}^{n}K_iE\left(\left[\tau_k-\omega_i\left(\hat{F},\tau_k\right)I\left(Y_i<m(U_i)+\sigma(U_i)c_k-d_i(u_0)\right)\right]|U\right)\\
&=\frac{1}{nh}\sum_{i=1}^{n}K_iE\left(\left[\omega_i\left(F_0,\tau_k\right)\left\{I\left(Y_i<m(U_i)+\sigma(U_i)c_k\right)-I\left(Y_i<m(U_i)+\sigma(U_i)c_k-d_i(u_0)\right)\right\}\right]|U\right)+o_p(1)\\
&=\frac{1}{nh}\sum_{i=1}^{n}K_id_i(u_0)\Gamma_k\left(1+o_p(1)\right)\\
&=\frac{1}{2}f_U(u_0)\mu_2h^2m''(u_0)\Gamma_k+o_p\left(h^2\right).
\end{aligned}$$

So, we can obtain

$$\begin{aligned}
&\text{bias}\left[\hat{m}(u_0)|U\right]=\frac{f_U^{-1}(u_0)}{q\sqrt{nh}}\sum_{k=1}^{q}\Gamma_k^{-1}E\left[M_k|U\right]=\frac{1}{2}\mu_2h^2m''(u_0),\\
&\text{var}\left[\hat{m}(u_0)|U\right]=\frac{1}{nh}\Sigma+o_p\left(\frac{1}{nh}\right).
\end{aligned}$$

This completes the proof.

Copyright © 2013 SciRes.