Econometric Theory, **37**, 2021, 613–631.
doi:10.1017/S0266466620000146

LOCAL COMPOSITE QUANTILE REGRESSION SMOOTHING: A FLEXIBLE DATA STRUCTURE AND CROSS-VALIDATION*

XIAO HUANG
Kennesaw State University
ZHONGJIAN LIN
Emory University

In this paper, we study the local composite quantile regression estimator for mixed categorical and continuous data. The local composite quantile estimator is an efficient and safe alternative to the local polynomial method and has been well-studied for continuous covariates. Generalization of the local composite quantile regression estimator to a flexible data structure is appealing to practitioners as empirical studies often encounter categorical data. Furthermore, we study the theoretical properties of the cross-validated bandwidth selection for the local composite quantile estimator. Under mild conditions, we derive the rates of convergence of the cross-validated smoothing parameters to their optimal benchmark values for both categorical and continuous covariates. Monte Carlo experiments show that the proposed estimator may have large efficiency gains compared with the local linear estimator. Furthermore, we illustrate the robustness of the local composite quantile estimator using the Boston housing dataset.

1. INTRODUCTION

The conditional quantile method receives tremendous attention because of its capability to investigate research questions beyond the conditional mean. Since its inception in Koenker and Bassett (1978), there have been many methodological developments of quantile methods in econometrics and statistics. For more details, see Koenker (2005, 2017) and Koenker et al. (2017). The conditional quantile method provides an additional dimension in data analysis and this feature motivates the averaging of quantiles for efficiency improvement.

The composite quantile regression (CQR) method was introduced to improve efficiency and gain robustness in parametric regressions (Zou and Yuan, 2008).

*We are grateful to Co-Editor, Liangjun Su, and three anonymous reviewers for their insightful comments and suggestions, which helped improve the paper. We also thank the KSU High Performance Computing Cluster for providing computing support as well as the audience at the 2018 Econometric Society China meeting for their helpful comments. All errors are our own. Address correspondence to Xiao Huang, Department of Economics, Finance and Quantitative Analysis, Kennesaw State University, 560 Parliament Garden Way NW, Kennesaw, GA 30144, USA; e-mail: xhuang3@kennesaw.edu

© Cambridge University Press 2020

https://doi.org/10.1017/S0266466620000146 Published online by Cambridge University Press

The CQR method introduces the quantile dimension and takes the average of a sequence of symmetric quantiles to tackle the problems raised by erratic error terms. In a nonparametric analysis, the local CQR smoothing technique has been shown to be an efficient and safe alternative to the local polynomial method. Kai, Li, and Zou (2010) derive the asymptotic bias, variance, and normality of the local CQR estimator for continuous covariates and document the relative efficiency compared with the local polynomial method for nonnormal errors.

This paper contributes to the literature in two aspects. First, we study the local linear CQR method for mixed categorical and continuous data. The local composite quantile estimator is an efficient and safe alternative to the local polynomial method and has been well-studied for continuous covariates. The generalization of the local CQR estimator to a flexible data structure is appealing to practitioners as empirical studies often encounter categorical data. Second, we provide a data-driven bandwidth selection for the local CQR smoother. The data-driven bandwidth selection is empirically oriented and handles the undesirably large bias or variance of theoretically optimal smoothing parameters.

The local CQR method has been well studied in nonparametric and semiparametric models with continuous variables; for example, Kai et al. (2011) for the varying coefficient partially linear model; Jiang et al. (2012) and Fan et al. (2013) for the single-index model; Li and Li (2016) for the local CQR smoother in Harris recurrent Markov processes; Zhao and Xiao (2014) for a weighted version of local CQR smoothing and the Cramér-Rao lower bound; and Kong and Xia (2014) for a high-dimensional analysis, to name only a few. Economic analyses often encounter categorical data, such as age, gender, race, and so forth, which limits the use of nonparametric techniques in empirical applications. Racine and Li (2004) and Li and Racine (2004) generalize local linear smoothing to allow for both categorical and continuous covariates. To take advantage of the efficiency gain and robustness of the local CQR method, we generalize the local CQR smoother to allow for a flexible data structure with both categorical and continuous covariates. We derive the asymptotic bias, variance, and normality of the local CQR smoother for the mixed categorical and continuous covariates.

It is well-known that nonparametric kernel estimators are sensitive to the choice of the smoothing parameters. In particular, undersmoothing or oversmoothing leads to a large variance or bias; see Pagan and Ullah (1999) and Li and Racine (2007). We propose a least-squares cross-validation to select optimal bandwidths for both categorical and continuous covariates in the local CQR smoother. We derive the rate of convergence of the cross-validated bandwidths to their optimal benchmark values. Unlike the local constant estimator in Racine and Li (2004) or other local polynomial estimators, a closed-form expression is unavailable for the local CQR estimator, making it difficult to derive the rate result for bandwidths. The analytic methods in Kai et al. (2010) are used to obtain the asymptotic results for the bandwidths.

We illustrate the local CQR smoother for mixed data and the data-driven bandwidth selection using Monte Carlo experiments as well as an empirical

https://doi.org/10.1017/S0266466620000146 Published online by Cambridge University Press

application to the Boston housing market to determine the impact of location factors. The Monte Carlo results show efficiency gains of the local CQR estimator for a variety of nonnormal errors. Unlike the local CQR estimator in the continuous covariate case, the local CQR estimator can improve efficiency even in the normal error case under a heteroskedastic process in the simulation. Furthermore, we find that the cross-validated bandwidths for categorical covariates in the local CQR smoother tend to be larger than those selected by cross-validation in the local linear method (Li and Racine, 2004). The empirical application in the housing market documents the robustness of the local CQR estimator. When data are artificially adapted to a more extreme case, the local CQR estimate is more robust than its local polynomial counterpart.

The paper unfolds as follows. Section 2 presents the local CQR estimator for the nonparametric model with mixed categorical and continuous data. The asymptotic bias, variance, and normality of the local CQR estimator are also derived. Section3 presents the cross-validated bandwidth selection for the local CQR estimator with mixed data and the convergence rate of the selected bandwidths to their theoretical benchmark values. We illustrate the relative efficiency of the cross-validated local CQR estimator to the corresponding local linear estimation with two Monte Carlo experiments and an empirical application to the Boston housing market in Section 4. Section 5 concludes. Propositions, lemmas, and proofs are rendered in the online Supplementary Material associated with this article.

2. LOCAL CQR SMOOTHING FOR A NONPARAMETRIC REGRESSION WITH MIXED DATA

Consider the nonparametric regression model defined by

$$Y_i = g(X_i) + \sigma(X_i)\varepsilon_i,\ i = 1, \ldots, n, \tag{1}$$

where $X_i \equiv (X_i^c, X_i^d)$ contains both continuous and categorical variables, $g(\cdot)$ is the conditional mean function, $\sigma^2(\cdot)$ is the conditional variance function, and ε_i is the error term with a mean and variance equal to zero and one, respectively. X_i^d is an $r \times 1$ vector of categorical regressors and X_i^c is a $p \times 1$ vector of continuous regressors. We use $X_{t,i}^d$ to denote the tth component of X_i^d and assume that $X_{t,i}^d \in \{0, 1, \ldots, C_t - 1\}$ for $t = 1, \ldots, r$ and that $C_t \geq 2$. $g(\cdot)$ has an unknown functional form. We use $f(x) = f(x^c, x^d)$ to denote the joint density function of X_i. We define the partial derivative of $g(X)$ with respect to X^c:

$$\beta(X) \equiv \frac{\partial g(X)}{\partial X^c} \equiv g'(X), \tag{2}$$

where we use the prime to denote the partial derivative of the function with respect to a continuous regressor. We are interested in estimating the unknown function $g(x)$.

We define $d_{ix} = \sum_{t=1}^{r} 1\{X_{t,i}^d \neq x_t^d\}$, which equals the number of disagreeing components between X_i^d and x^d. We use the following kernel function for categorical variables with bandwidth λ:

https://doi.org/10.1017/S0266466620000146 Published online by Cambridge University Press

$L_\lambda(X_i^d, x^d) = \lambda^{d_{ix}}$.

Denote $w(\cdot)$ as the kernel function associated with the continuous variables in X^c and let h be the corresponding bandwidth. Denote $W(x^c) = \Pi_{k=1}^p w(x_k^c)$. The kernel function for the mixed regressors becomes

$$K_{ix} \equiv W_{ix} L_{\lambda,ix} = W\Big(\frac{X_i^c - x^c}{h}\Big) L_\lambda\Big(X_i^d, x^d\Big).$$

Let the local CQR loss function be

$$\sum_{k=1}^{q}\sum_{i=1}^{n} \rho_{\tau_k}\Big(Y_i - a_k - b^T(X_i^c - x^c)\Big) K_{ix}, \tag{3}$$

where $\rho_{\tau_k}(u) \equiv u\big(\tau_k - 1\{u < 0\}\big)$ is the check function in the quantile regression and $\tau_k = \frac{k}{q+1}$ for $k = 1, \ldots, q$. Denote the minimizer of equation (3) with respect to $(a_1, \ldots, a_q, b^T)$ as $(\hat{a}_1, \ldots, \hat{a}_q, \hat{b}^T)$. The local linear CQR estimators of $g(x)$ and $\beta(x)$ are given by

$$\hat{g}(x) = \frac{1}{q}\sum_{k=1}^{q} \hat{a}_k, \quad \hat{\beta}(x) = \hat{b}.$$

Note that equation (3) treats the continuous regressor X^c in a local linear fashion and the categorical regressor X^d in a local constant fashion. This paper focuses on the estimation of $g(\cdot)$. Adding the local linear structure helps reduce bias, similar to the result in the local least-squares (LLS) method. To estimate the derivative, we can use the local quadratic CQR method. The number of quantiles, q, is essentially another tuning parameter and is chosen by the researcher. We show in the simulation that there can be considerable gain by choosing a relatively small q.

2.1. Asymptotic Theory

This section investigates the asymptotic properties of the proposed local linear CQR smoother for the flexible data structure. To derive the asymptotic results, we make the following assumptions.

Assumption 1. (i) The support of X, $\mathcal{X} \equiv \mathcal{X}^c \times \mathcal{X}^d$, is compact; (ii) $g(X)$ is twice differentiable with respect to X^c in $\mathcal{X}^c$; (iii) $(X_i, Y_i), i = 1, 2, \ldots, n$ are i.i.d. and $\mathbb{E}(Y_i|X_i) = g(X_i)$ almost surely; and (iv) Let $g''(X)$ denote the Hessian of g with respect to X^c. Then $\int g''(X)^2 f(X) dX > 0$.

Assumption 2. (i) $f(X)$ is twice differentiable with respect to X^c and $E\big[f^4(X)\big] < \infty$; and (ii) the density of X, $f(X)$, is positive.

Assumption 3. The bandwidth satisfies $h \to 0, \lambda \to 0$, and $nh^p \to \infty$ as $n \to \infty$.

Denote

$$\mu_j = \int u^j w(u) du \qquad \text{and} \qquad \nu_j = \int u^j w^2(u) du, \qquad j = 0, 1, \ldots. \tag{4}$$

https://doi.org/10.1017/S0266466620000146 Published online by Cambridge University Press

Assumption 4. The kernel function $W(x^c)$ has compact support and is bounded and symmetric such that $\int vv^T W(v)dv = \mu_2 I_p < \infty$ and $\int vv^T W^2(v)dv = \nu_2 I_p < \infty$, where I_p is a $p \times p$ identity matrix. Furthermore, $w(v)$ is twice differentiable and $\int v^3 w'(v)dv < \infty$.

Assumption 5. (i) $\sigma^2(\cdot)$ is twice differentiable with respect to X^c; and (ii) $\mathbb{E}[\sigma^4(X)] < \infty$.

Assumption 6. ε_i has a symmetric distribution with a continuous and positive density. $E(\varepsilon_i|X_i) = 0$ and $Var(\varepsilon_i|X_i) = 1$.

Remark 1. Assumption 1(i) and (ii) as well as Assumption 2 are standard conditions used in the nonparametric literature. Assumption 1(iii) is essentially an exogenous condition. The i.i.d. assumption can be relaxed to allow for weak dependence; see Su, Chen, and Ullah (2009). Assumption 1(iv) indicates that all covariates are relevant. In the nonparametric literature, there is a technique of smoothing out irrelevant variables by data-driven cross-validation bandwidth selections; that is, $\hat{h} \to \infty$ and $\hat{\lambda} = 1$. See Hall, Racine, and Li (2004) and Hall, Li, and Racine (2007) for details. In writing equation (4), we also implicitly assume that both μ_j and ν_j exist and they can be viewed as moments of the individual kernels $w(u)$ and $w^2(u)$. Assumption 4 further assumes that kernel moments for different elements of u are zero. This assumption simplifies the presentation of the results and can be commonly found in the nonparametric literature; for example, Ruppert and Wand (1994). We assume that $\sigma^2(X)$ is twice differentiable in X^c as a second-order Taylor series expansion is needed to obtain the leading terms of σ at the point of interest. The symmetry requirement in Assumption 6 is used in the proof of Theorem 1 to obtain the explicit bias expression and it ensures that the leading term in the bias is $O(h^2 + \lambda)$, which vanishes as $n \to \infty$. An additional remark on error symmetry is given after Theorem 1.

Let $F_\varepsilon(\cdot)$ and $f_\varepsilon(\cdot)$ be the CDF and PDF of the error term ε, respectively. Define

$$R_1(q) = \frac{1}{q^2} \sum_{k=1}^{q} \sum_{k'=1}^{q} \frac{\tau_{kk'}}{f_\varepsilon(c_k) f_\varepsilon(c_{k'})}, \tag{5}$$

where $c_k = F_\varepsilon^{-1}(\tau_k)$ and $\tau_{kk'} = \tau_k \wedge \tau_{k'} - \tau_k \tau_{k'}$. Define, for $k = 1, \ldots, q$,

$$u_k \equiv \sqrt{nh^p}\{a_k - g(x) - \sigma(x)c_k\}, \quad v \equiv \sqrt{nh^{p+2}}[b - \beta(x)],$$
$$\Delta_{i,k} \equiv \frac{u_k}{\sqrt{nh^p}} + \frac{v^T(X_i^c - x^c)}{\sqrt{nh^{p+2}}}, \quad r_i \equiv g(X_i) - g(x) - \beta(x)(X_i^c - x^c),$$
$$d_{i,k} \equiv c_k[\sigma(X_i) - \sigma(x)] + r_i, \quad \eta_{i,k}^* \equiv 1\{\varepsilon_i \le c_k - d_{i,k}/\sigma(X_i)\} - \tau_k.$$

Denote $\theta = (u_1, \ldots, u_q, v^T)^T$. Let $S_{n,11}$ be a $q \times q$ diagonal matrix with diagonal elements $\frac{f_\varepsilon(c_k)}{nh^p} \sum_{i=1}^n \frac{K_{ix}}{\sigma(X_i)}$, $S_{n,12}$ be a $q \times p$ matrix with the kth row

https://doi.org/10.1017/S0266466620000146 Published online by Cambridge University Press

$\frac{f_\varepsilon(c_k)}{nh^{p+1}}\sum_{i=1}^n \frac{K_{ix}(X_i^c-x^c)^T}{\sigma(X_i)}$, and $S_{n,22}$ be a $p\times p$ matrix $\frac{1}{nh^{p+2}}\sum_{k=1}^q f_\varepsilon(c_k)\sum_{i=1}^n \frac{K_{ix}(X_i^c-x^c)(X_i^c-x^c)^T}{\sigma(X_i)}$. Denote

$$S_n=\begin{pmatrix} S_{n,11} & S_{n,12}\\ S_{n,12}^T & S_{n,22}\end{pmatrix}.$$

Let $W_n^*=(w_{11}^*,\ldots,w_{1q}^*;w_{21}^{*T})^T$ with

$$w_{1k}^*=\frac{1}{\sqrt{nh^p}}\sum_{i=1}^n K_{ix}\eta_{i,k}^* \quad\text{and}\quad w_{21}^*=\frac{1}{\sqrt{nh^{p+2}}}\sum_{k=1}^q\sum_{i=1}^n K_{ix}\eta_{i,k}^*(X_i^c-x^c).$$

Let

$$S=\begin{pmatrix} f(c_1) & 0 & \cdots & 0 & 0\\ 0 & f(c_2) & \cdots & 0 & 0\\ \vdots & \vdots & \ddots & \vdots & \vdots\\ 0 & 0 & \cdots & f(c_q) & 0\\ 0 & 0 & \cdots & 0 & \mu_2 I_p\sum_{k=1}^q f(c_k)\end{pmatrix}.$$

Similarly, let Σ_{11} be a $q\times q$ matrix with (k,k')-element $\nu_0\tau_{kk'}$ and let Σ_{22} be $(\sum_{k,k'=1}^q \tau_{kk'})I_p\nu_2$. Define

$$\Sigma=\begin{pmatrix}\Sigma_{11} & 0\\ 0 & \Sigma_{22}\end{pmatrix}.$$

Denote $\eta_{i,k}=1\{\varepsilon_i\le c_k\}-\tau_k$ and $W_n=(w_{11},\ldots,w_{1q},w_{21}^T)^T$ with

$$w_{1k}=\frac{1}{\sqrt{nh^p}}\sum_{i=1}^n K_{ix}\eta_{i,k} \quad\text{and}\quad w_{21}=\frac{1}{\sqrt{nh^{p+2}}}\sum_{k=1}^q\sum_{i=1}^n K_{ix}(X_i^x-x^c)\eta_{i,k}.$$

Proposition 3 in the online supplement gives the asymptotic property of $\hat\theta$, which is used to derive the asymptotic properties of $\hat g(X)$ in Theorem 1. The bias and variance in Proposition 3 can be used to directly construct the bias and variance in Theorem 1. The notation $d_{\tilde x,x}=1$ in Theorem 1 is defined similarly to d_{ix} below equation (2) and it measures the number of disagreeing components between $\tilde x$ and x.

THEOREM 1. *Under Assumptions 1 to 6, the asymptotic conditional bias and variance of the local linear CQR estimator $\hat g(x)$ are given by*

https://doi.org/10.1017/S0266466620000146 Published online by Cambridge University Press

$$Bias(\hat{g}(x)|X) = \left[h^2 \frac{tr[\beta'(x)]\mu_2}{2} + \lambda \right.$$
$$\left. \times \sum_{\tilde{x}^d, d_{\tilde{x},x}=1} \frac{\sigma(x) f(x^c, \tilde{x}^d)[g(x^c, \tilde{x}^d) - g(x)]}{f(x)\sigma(x^c, \tilde{x}^d)} \right] (1 + o_p(1)),$$
$$Var(\hat{g}(x)|X) = \frac{1}{nh^p} \frac{\sigma^2(x)}{f(x)} \nu_0 R_1(q)(1 + o_p(1)).$$

Furthermore, conditioning on X, we have

$$\sqrt{nh^p} \left[\hat{g}(x) - g(x) - Bias(\hat{g}(x)|X) \right] \xrightarrow{d} N\left(0, \frac{\sigma^2(x)}{f(x)} \nu_0 R_1(q)\right). \tag{6}$$

See the online supplement for the proof.

Remark 2. From Theorem 1, the mean squared error (MSE), omitting small order terms, is given by

$$\text{MSE}(x) = (C_1 h^2 + C_2 \lambda)^2 + \frac{C_3}{nh^p}, \tag{7}$$

where $C_1, C_2,$ and C_3 are corresponding coefficients in the bias and variance expressions. Minimizing equation (7) with respect to both h and λ will not give the optimal bandwidth h as we can choose λ such that the first term on the right-hand side of equation (7) is zero with no solution for h. Adding higher order terms such as $O(\lambda^2)$ and $O(\lambda h)$ in the bias formula may partially solve this problem by expressing the optimal λ as a function of h and solve a higher order polynomial in h in the MSE. However, a closed-form solution for optimal h is difficult to obtain. Instead, minimizing the leading terms of the cross-validation objective function gives closed-form solutions to h and λ and allows us to study the convergence rate of bandwidths from the cross-validation exercise.

Remark 3. The first term in the bias expression is a standard result from the local linear regression. This part of bias essentially measures the impact of curvature in $g(x)$ due to the continuous variables in x. Hence, more curvature will lead to larger diagonal elements in $\beta'(x)$, a larger trace of the Hessian matrix of $g(x)$, and a larger bias. The second term is the additional bias due to the presence of categorical regressors. Categorical variables cause jumps in $g(x)$. The term $g(x^c, \tilde{x}^d) - g(x)$ and the summation over $\tilde{x}^d$ suggest that more jumps necessarily lead to a larger bias. This jump-related bias is also documented in Thm. 1.1 in Racine and Li (2004). Finally, we notice that there is no derivative of $f(x)$ in the bias expression, which implies that despite the presence of categorical variables, the local CQR estimator preserves the design-adaptive property of the LLS estimator.

Remark 4. The error ε_i is assumed to have a symmetric distribution and the immediate consequence is $\sum_{k=1}^{q} c_k = 0$. In the proof of Theorem 1, this helps remove the term $\frac{1}{q}\sum_{k=1}^{q} \sigma(x) c_k$ from the bias expression. Without the symmetric

https://doi.org/10.1017/S0266466620000146 Published online by Cambridge University Press

assumption, $Bias(\hat{g}(x)|X)$ will have an extra term, which does not affect the asymptotic variance result, and asymptotic normality still holds. In practice, if asymmetry in the error distribution is a concern, one can do a bias correction by removing an estimate of $\frac{1}{q}\sum_{k=1}^{q}\sigma(x)c_k$.

Remark 5. Zhao and Xiao (2014) propose a weighted quantile estimator where the weights are obtained from minimizing the part of the asymptotic variance that contains the weights—a quadratic form that gives a closed-form solution. A similar exercise can also be done for the local CQR approach here. Based on the asymptotic variance expression in Theorem 1, we can obtain a closed-form solution for the weights. An iterative procedure can be used by adopting some starting values for the weights to do cross-validation and obtain the bandwidths. We also use the estimated bandwidths to recalculate the weights. These steps can be repeated until certain convergence criterion is met. The computation cost will be high as the iterative procedure calls for multiple rounds of cross-validation.

Remark 6. The uniform inference of the nonparametric quantile function of mixed categorical and continuous data is challenging and under development. Guerre and Sabbah (2012) provide a uniform bias study and a Bahadur representation for local polynomial estimators for continuous data. In this paper, we do not aim to provide a thorough uniform analysis of the local CQR smoother over quantiles, covariates, and the smoothing parameter.

Remark 7. With an explicit formula for the asymptotic bias, we can replace all terms in $Bias(\hat{g}(x)|X)$ with their sample analogs to correct bias. The procedure may require the local quadratic CQR to obtain a robust estimator for the derivative. We can estimate the density functions with the standard local linear estimator. Similarly, we replace those terms in the conditional variance with their consistent estimators to estimate the asymptotic variance.

3. CROSS-VALIDATION BANDWIDTH SELECTION

In this section, we study the selection of bandwidth based on a least-squares cross-validation. The general procedure for bandwidth selection with mixed regressors is discussed in Racine and Li (2004) and Li and Racine (2004) for a nonparametric regression and in Li, Lin, and Racine (2013) for a nonparametric quantile regression. We extend the work in Racine and Li (2004) and Li and Racine (2004) to the case of the local CQR. The form of the local linear CQR estimator differs from the local constant/linear estimator in Racine and Li (2004) and Li and Racine (2004), leading to some differences in the proof, although the convergence results for the bandwidths are similar.

From the proof of Proposition 3 in the online supplement, we have

$$\begin{pmatrix}\hat{u}_1\\ \vdots\\ \hat{u}_q\end{pmatrix} = -\frac{\sigma(x)}{f(x)}\left(S^{-1}\right)_{11} W^*_{1n}, \tag{8}$$

https://doi.org/10.1017/S0266466620000146 Published online by Cambridge University Press

where $\left(S^{-1}\right)_{11}$ is a diagonal matrix; that is, the upper-left block matrix of S^{-1}. W_{1n}^* are the first q elements of W_n^*. In equation (8), we also omit the $o_p(1)$ term in equation (A.6) in the online supplement as it will not affect the analysis of the cross-validation. See Lemma 8 for a proof.

Let ι_q be a $q \times 1$ vector of ones. Substitute equation (8) into $\hat{g}(x)$ to obtain

$$\hat{g}(x) = -\frac{1}{q\sqrt{nh^p}} \frac{\sigma(x)}{f(x)} \iota_q^T \left(S^{-1}\right)_{11} W_{1n}^* + g(x) = -\frac{1}{nh^p q} \frac{\sigma(x)}{f(x)} \sum_{k=1}^{q} \sum_{i=1}^{n} \frac{K_{ix} \eta_{i,k}^*}{f(c_k)} + g(x). \tag{9}$$

Recall $\hat{g}(x) = \sum_{k=1}^{q} \hat{a}_k / q$. Combining the definition of u_k in Section 2 and equation (8) gives the result in equation (9). Writing $\hat{g}(x)$ and also $\hat{g}_{-i}(x)$ in equation (10) in this format helps us derive the asymptotic results for $\hat{h}$ and $\hat{\lambda}$.

Define the leave-one-out estimator $\hat{g}_{-i}$:

$$\hat{g}_{-i} = \hat{g}_{-i}(x) = -\frac{1}{nh^p q} \frac{\sigma(x)}{f(x)} \sum_{k=1}^{q} \sum_{\substack{j=1 \\ j \neq i}}^{n} \frac{K_{ji} \eta_{j,k}^*}{f(c_k)} + g(x), \tag{10}$$

where

$$K_{ji} = W_{ji} L_{ji}, \quad W_{ji} = W\left(\frac{X_j^c - X_i^c}{h}\right), \quad L_{ji} = L_\lambda\left(X_j^d, X_i^d\right),$$

and $\eta_{j,k}^*$ is defined similarly to $\eta_{i,k}^*$ in Section 2.

The loss function for the cross-validation exercise is given by

$$CV(h, \lambda) = n^{-1} \sum_{i=1}^{n} \left[Y_i - \hat{g}_{-i}(X_i)\right]^2. \tag{11}$$

Remark 8. Our local composite quantile method provides an efficient and safe alternative to estimate the conditional mean function. Therefore, we adopt the least-squares cross-validation function in equation (11). It is worth pointing out that the cross-validation bandwidth selection for nonparametric quantile regression is underdeveloped. If the objective is to obtain a quantile estimator, we may use a quantile loss function in cross-validation for the nonparametric quantile regression. To do that, it is possible to rewrite a quantile loss function in a quadratic form similar to that in Proposition 2 in the online supplement. We can then mimic the steps in this section and the proof of Theorem 2 to derive the convergence rates for the bandwidths. It will be interesting to compare the convergence rates for the bandwidths between the two types of cross-validation functions. Such an inquiry is beyond the scope of this current paper and we defer it to future research.

https://doi.org/10.1017/S0266466620000146 Published online by Cambridge University Press

Let $\hat{h}$ and $\hat{\lambda}$ be the cross-validation solution of h and λ, respectively, in equation (11). Using equation (1) and expanding the quadratic term in equation (11) gives us

$$CV(h,\lambda) = n^{-1}\sum_{i=1}^{n}(g_i - \hat{g}_{-i})^2 + 2n^{-1}\sum_{i=1}^{n}\sigma(X_i)\varepsilon_i(g_i - \hat{g}_{-i}) + n^{-1}\sum_{i=1}^{n}\sigma(X_i)^2\varepsilon_i^2$$
$$= CV_1(h,\lambda) + n^{-1}\sum_{i=1}^{n}\sigma(X_i)^2\varepsilon_i^2,$$

where

$$CV_1(h,\lambda) = n^{-1}\sum_{i=1}^{n}(g_i - \hat{g}_{-i})^2 + 2n^{-1}\sum_{i=1}^{n}\sigma(X_i)\varepsilon_i(g_i - \hat{g}_{-i}). \tag{12}$$

The term $n^{-1}\sum_{i=1}^{n}\sigma(X_i)^2\varepsilon_i^2$ can be omitted in optimization as it is independent of both h and λ. Define

$$\gamma_i = \sigma(X_i)/f(X_i), \quad \delta_i = \sigma(X_i)^2/f(X_i).$$

Substitute equation (10) into the first term on the r.h.s. of equation (12) to obtain

$$\frac{1}{n}\sum_{i=1}^{n}(g_i - \hat{g}_{-i})^2 \quad = \frac{1}{q^2}\sum_{k=1}^{q}\sum_{m=1}^{q}\frac{1}{f(c_k)f(c_m)}\frac{1}{n^3h^{2p}}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\sum_{\substack{l=1\\ l\neq i}}^{n}\gamma_i^2 K_{ji}K_{li}\eta_{j,k}^*\eta_{l,m}^*. \tag{13}$$

Substitute equation (10) into the second term on the r.h.s. of equation (12) to obtain

$$\frac{2}{n}\sum_{i=1}^{n}\sigma(X_i)\varepsilon_i(g_i - \hat{g}_{-i}) = \frac{2}{q}\sum_{k=1}^{q}\frac{1}{f(c_k)}\frac{1}{n^2h^p}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\delta_i K_{ji}\eta_{j,k}^*\varepsilon_i. \tag{14}$$

Hence, $CV_1(h,\lambda)$ can be written as

$$CV_1(h,\lambda) = \frac{1}{q^2}\sum_{k=1}^{q}\sum_{m=1}^{q}\frac{1}{f(c_k)f(c_m)}S_{1,km} + \frac{2}{q}\sum_{k=1}^{q}\frac{1}{f(c_k)}S_{2,k}, \tag{15}$$

where

$$S_{1,km} = \frac{1}{n^3h^{2p}}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\sum_{\substack{l=1\\ l\neq i}}^{n}\gamma_i^2 K_{ji}K_{li}\eta_{j,k}^*\eta_{l,m}^*, \quad S_{2,k} = \frac{1}{n^2h^p}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\delta_i K_{ji}\eta_{j,k}^*\varepsilon_i.$$

Lemmas 6 and 7 in the online supplement show that the leading term in the cross-validation can be obtained in the following form:

$$CV_0 = A_1h^4 - A_2h^2\lambda + A_3\lambda^2 + A_4(nh^p)^{-1}, \tag{16}$$

where the coefficients A_1 to A_4 are constants defined in the proof of Theorem 2. Let h_0 and λ_0 be the values of h and λ, respectively, which minimizes equation (16). It is

https://doi.org/10.1017/S0266466620000146 Published online by Cambridge University Press

shown in the proof of Theorem 2 that h_0 and λ_0 are $O(n^{-1/(p+4)})$ and $O(n^{-2/(p+4)})$, respectively. The next theorem establishes the rates of convergence of $(\hat{h}-h_0)/h_0$ and $(\hat{\lambda}-\lambda_0)/\lambda_0$.

THEOREM 2. *Under Assumptions 1 to 6, we have*

$$\left(\hat{h}-h_0\right)/h_0 = O_p(n^{-p/[2(p+4)]}) \text{ and } \left(\hat{\lambda}-\lambda_0\right)/\lambda_0 = O_p(n^{-1/2}).$$

See the online supplement for the proof.

Remark 9. In collecting the leading terms of $S_{1,km}$ in Lemma 6, the dimension parameter p appears in the exponent with the form of $-p$ and does not appear in the leading terms in $S_{2,k}$. This simplifies our analysis as compared with Thm. 1.2 in Racine and Li (2004) since there is no need to separately discuss the results for $p \leq 3$ and $p \geq 4$.

Remark 10. Similar to the results in Racine and Li (2004), the categorical regressor X_i^d may also include variables with a natural ordering such as self-reported financial status (low, middle, or high income groups). The distance function can be modified as $d_{ix} = \sum_{t=1}^{r} |X_{t,i}^d - x_t^d|$. The convergence results for $\hat{h}$ and $\hat{\lambda}$ can be obtained similarly to those in Theorem 2.

Remark 11. Based on equations (A.11) and (A.13) in the online supplement, we know that both $\hat{h}$ and $\hat{\lambda}$ are linear functions of B_1 and B_2, respectively. Both B_1 and B_2 are summations of zero-mean random variables with finite variance. Hence, both B_1 and B_2 have an asymptotic normal distribution, which can be used to establish the asymptotic normality of $n^{p/[2(p+4)]}(\hat{h}-h)$ and $\sqrt{n}(\hat{\lambda}-\lambda)$.

Remark 12. A more sophisticated result can be derived by assuming different bandwidths for different components in both X_i^c and X_i^d. See Li and Racine (2004) for the rates of convergence of the smoothing parameters in this case. The results in Theorem 2 can also be extended to this case.

Remark 13. Kai et al. (2010) estimate the optimal bandwidth based on the relationship between the optimal CQR bandwidth and the optimal LLS bandwidth. They also propose a "pilot" selector method to calculate the optimal CQR bandwidth. Using a cross-validation for bandwidth selection complements the study in Kai et al. (2010).

Denote $\hat{g}_{\hat{h},\hat{\lambda}}(x)$ as the local CQR estimator with cross-validation selected bandwidths. We restate the asymptotic result for $\hat{g}_{\hat{h},\hat{\lambda}}(x)$ in Theorem 1 in terms of the cross-validated bandwidths.

Corollary 1. *Under Assumptions 1 to 6, the asymptotic conditional bias and variance of the local linear CQR estimator* $\hat{g}_{\hat{h},\hat{\lambda}}(x)$ *are given by*

https://doi.org/10.1017/S0266466620000146 Published online by Cambridge University Press

$$Bias(\hat{g}_{\hat{h},\hat{\lambda}}(x)|X) = \left[\hat{h}^2 \frac{tr[\beta'(x)]\mu_2}{2} + \hat{\lambda} \times \sum_{\tilde{x}^d, d_{\tilde{x},x}=1} \frac{\sigma(x) f(x^c, \tilde{x}^d)[g(x^c, \tilde{x}^d) - g(x)]}{f(x)\sigma(x^c, \tilde{x}^d)}\right](1 + o_p(1)),$$

$$Var(\hat{g}_{\hat{h},\hat{\lambda}}(x)|X) = \frac{1}{n\hat{h}^p} \frac{\sigma^2(x)}{f(x)} \nu_0 R_1(q)(1 + o_p(1)).$$

Furthermore, conditioning on X, we have

$$\sqrt{n\hat{h}^p}\left[\hat{g}_{\hat{h},\hat{\lambda}}(x) - g(x) - Bias(\hat{g}_{\hat{h},\hat{\lambda}}(x)|X)\right] \xrightarrow{d} N\left(0, \frac{\sigma^2(x)}{f(x)} \nu_0 R_1(q)\right). \quad \textbf{(17)}$$

See the online supplement for the proof.

4. MONTE CARLO EXPERIMENTS AND AN EMPIRICAL ILLUSTRATION

In this section, we provide simulation results and an empirical illustration for the robustness of the local CQR estimator.

4.1. Monte Carlo Experiments

We use Monte Carlo simulations to illustrate the finite sample performance of the local CQR with data-driven bandwidth selection. We compare our results to the LLS estimator in Racine and Li (2004) under several scenarios. Although we do not have an analytic result for relative efficiency between the local CQR and LLS estimators, the Monte Carlo results document large efficiency gains of the local CQR smoother compared with its LLS counterpart in a heteroskedastic process.

For both local CQR and LLS estimates, cross-validation is used to select the optimal bandwidth. The optimal bandwidth is then used to evaluate the mean function, $g(X)$, at 200 equally spaced grid points for X^c on $[-1.5, 1.5]$ and 200 simulated *Bernoulli*(0.5) random variables for X^d. We assess the performance of each estimator by calculating average squared errors (ASE) defined as

$$\text{ASE}(\hat{g}) = \frac{1}{200} \sum_{i=1}^{200} \left[\hat{g}(X_i) - g(X_i)\right]^2,$$

where $\hat{g}(X_i)$ is the estimate for $g(X_i)$.

We use a sample size of 400 in each cross-validation exercise, select the optimal bandwidths, and calculate the ASE for the local CQR and LLS methods. This process is repeated 400 times. Two measures are used to compare the performance between the LLS and local CQR methods. Let $j = 1, \ldots, 400$ index the jth replication. The mean average squared errors (MASE) and the ratio of average

https://doi.org/10.1017/S0266466620000146 Published online by Cambridge University Press

TABLE 1. Simulation results for DGP 1

	MASE	RASE		Ratio of bandwidths	
		Mean	Std. dev.	h	λ
Standard normal					
CQR_5	0.934	1.041	0.812	1.154	2.828
CQR_9	0.938	1.046	0.775	1.136	2.548
CQR_{19}	0.933	1.047	0.770	1.138	2.390
Laplace					
CQR_5	1.005	1.048	0.310	1.118	3.055
CQR_9	1.007	1.044	0.307	1.106	1.654
CQR_{19}	0.981	1.004	0.242	1.112	2.963
t_3					
CQR_5	1.025	1.131	0.588	1.105	2.431
CQR_9	1.006	1.101	0.459	1.102	2.510
CQR_{19}	1.003	1.094	0.494	1.097	2.843
$0.95\mathcal{N}(0,1)+0.05\mathcal{N}(0,9)$					
CQR_5	0.883	0.981	0.444	1.164	2.935
CQR_9	0.882	0.987	0.492	1.163	2.252
CQR_{19}	0.879	0.989	0.482	1.149	2.558
$0.95\mathcal{N}(0,1)+0.05\mathcal{N}(0,100)$					
CQR_5	0.922	1.024	0.470	1.136	2.437
CQR_9	0.917	1.012	0.475	1.150	2.620
CQR_{19}	0.918	1.014	0.469	1.131	2.329

squared errors (RASE) are given by

$$\mathrm{MASE}(\hat{g}) = \frac{\frac{1}{400}\sum_{j=1}^{400}\mathrm{ASE}(\hat{g}_{\mathrm{LLS}})_j}{\frac{1}{400}\sum_{j=1}^{400}\mathrm{ASE}(\hat{g}_{\mathrm{CQR}})_j}, \quad \mathrm{RASE}(\hat{g}) = \frac{1}{400}\sum_{j=1}^{400}\frac{\mathrm{ASE}(\hat{g}_{\mathrm{LLS}})_j}{\mathrm{ASE}(\hat{g}_{\mathrm{CQR}})_j},$$

where $\hat{g}_{LLS}$ and $\hat{g}_{CQR}$ are LLS and local CQR estimates, respectively, while the subscript j indicates the ASE obtained from the jth replication. Tables 1 and 2 report MASE, RASE, and their associated standard deviations. In addition, the two tables also report the average ratios of optimal bandwidths selected in the local CQR and LLS methods over 400 replications. A larger-than-one ratio for either h or λ suggests that, on average, the local CQR uses a larger bandwidth than LLS does.

We consider two data-generating processes (DGPs): a homoskedastic error case and a heteroskedastic error case. The first DGP is given by

$$Y_i = \sin(2X_i^c) + 2\exp(-16(X_i^c)^2) + 0.5X_i^d + X_i^c X_i^d + 0.5\varepsilon_i,$$

https://doi.org/10.1017/S0266466620000146 Published online by Cambridge University Press

TABLE 2. Simulation results for DGP 2

	MASE	RASE		Ratio of bandwidths	
		Mean	Std. dev.	h	λ
Standard normal					
CQR_5	1.098	1.152	0.408	0.997	1.381
CQR_9	1.012	1.048	0.241	0.991	1.673
CQR_{19}	0.992	1.017	0.150	0.993	1.479
Laplace					
CQR_5	1.224	1.331	0.473	1.019	1.644
CQR_9	1.162	1.229	0.338	1.036	2.052
CQR_{19}	1.140	1.191	0.466	1.026	1.584
t_3					
CQR_5	1.498	1.604	0.771	0.994	1.581
CQR_9	1.424	1.492	0.801	0.996	1.753
CQR_{19}	1.270	1.339	0.793	0.986	1.230
$0.95\mathcal{N}(0,1)+0.05\mathcal{N}(0,9)$					
CQR_5	1.064	1.104	0.406	1.002	1.365
CQR_9	1.031	1.079	0.272	0.992	1.553
CQR_{19}	1.000	1.028	0.190	0.994	1.693
$0.95\mathcal{N}(0,1)+0.05\mathcal{N}(0,100)$					
CQR_5	1.259	1.324	0.533	0.976	1.434
CQR_9	1.211	1.254	0.429	0.986	1.425
CQR_{19}	1.118	1.155	0.270	0.993	1.423

where $X_i^c \sim N(0,1)$ and $X_i^d \sim Bernoulli(0.5)$. The five error distributions for ε_i include: standard normal, Laplace distribution with a mean of 0 and a standard deviation of 1, t distribution with 3 degrees of freedom, and mixture normal distributions, $0.95N(0,1)+0.05N(0,\sigma^2)$, with $\sigma = 3$ or 10. We consider three choices of q in the local CQR estimation: $q = 5, 9, 19$, which are referred to as CQR_5, CQR_9, and CQR_{19}, respectively, in the tables. This is the same DGP used in Ex. 1 in Kai et al. (2010).

In the second DGP, the categorical variable appears in both the conditional mean and conditional variance, resembling the specification in equation (1).

$$Y_i = X_i^c \sin(2\pi X_i^c) + 0.5X_i^d + X_i^c X_i^d + 10^{-1}(2+\cos(2\pi X_i^c) + X_i^c X_i^d)\varepsilon_i,$$

where $X_i \sim U(0,1)$ and other specifications are the same as those in DGP 1. DGP 2 introduces a significant departure from DGP 1 in that it incorporates not only heteroskedasticity but also categorical variables in the conditional variance.

In Table 1, under the standard normal error distribution, most of the MASEs are slightly less than one and most of the RASEs are greater than one. Overall, both

https://doi.org/10.1017/S0266466620000146 Published online by Cambridge University Press

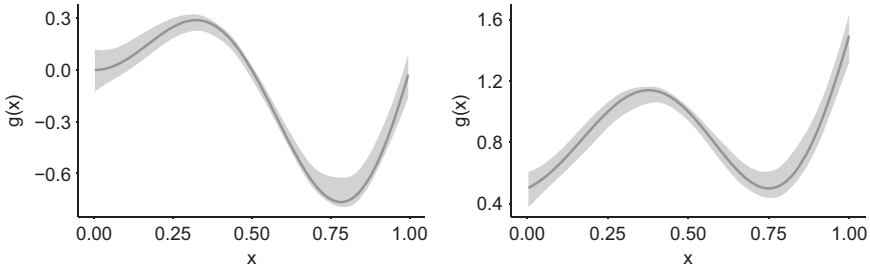


FIGURE 1. DGP 2 normal distribution coverage results. Shaded interval: asymptotic 95% confidence interval averaged over 1,000 replications; solid line: $g(x)$. $X^d = 0$ in the left panel and $X^d = 1$ in the right panel.

measures are close to one, suggesting that the local CQR and LLS yield similar efficiency results. We also notice that the selected bandwidth for CQR is larger than that in LLS for both h and λ. The difference is more notable for λ.

Table 2 shows that there can be large efficiency gains in using local CQR when the DGP exhibits heteroskedasticity, particularly when categorical variables appear in the conditional variance. Both MASE and RASE demonstrate the efficiency gain in using local CQR. It is also interesting to observe that, unlike the results in Table 1, the selected optimal bandwidths in the LLS and local CQR are very close for the continuous variable, although they still differ significantly for the categorical variable.

Next, we use the asymptotic variance in Theorem 1 to construct confidence intervals and compare those confidence intervals with the ones obtained from the bootstrap method. Simulations in all figures are based on DGP 2.

Let $q = 5$ and the sample size be 400. The following steps are taken to study the coverage of the 95% confidence interval based on asymptotic results. Step 1 simulates a sample of 400 observations. Step 2 uses the variance expression in Theorem 1 to calculate the standard error at each of the 400 points. Step 3 replicates the process 1,000 times and calculates the mean confidence intervals. In the 1,000 replications, values of X^c and X^d are fixed but error terms vary across replications. At each of the 400 points, we average the 1,000 confidence intervals and use the critical values ± 1.96 to obtain the mean confidence interval. The figures plot the mean confidence interval for all 400 data points along with the actual value $g(X)$.

This section reports the results from the normal and t errors in Figures 1 and 2. Results for the other three error distributions can be found in Figures 1–3 in the online supplement. In general, the asymptotic confidence intervals work well. The confidence intervals are tighter in the middle range of X^c when $X^d = 0$ compared with the case when $X^d = 1$. Among the five error distributions under consideration, the confidence intervals for the Laplace error are relatively wide.

Finally, we compare confidence intervals obtained from the asymptotic result and the bootstrap method. The number of bootstraps is 1,000. For each of the

https://doi.org/10.1017/S0266466620000146 Published online by Cambridge University Press

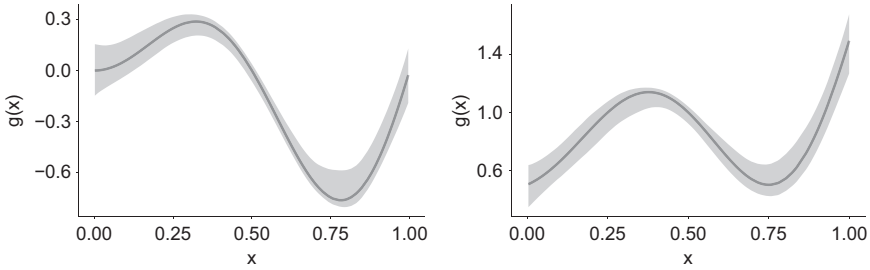


FIGURE 2. DGP 2 t distribution coverage results. Shaded interval: asymptotic 95% confidence interval averaged over 1,000 replications; solid line: $g(x)$. $X^d = 0$ in the left panel and $X^d = 1$ in the right panel.

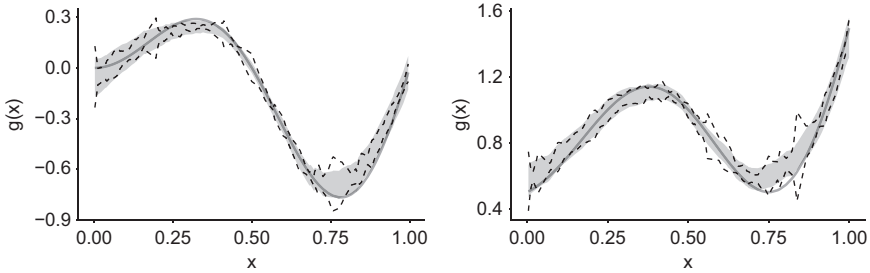


FIGURE 3. DGP 2 normal distribution results. Dashed lines: bootstrap 95% confidence interval; shaded interval: asymptotic 95% confidence interval; solid line: $g(x)$. $X^d = 0$ in the left panel and $X^d = 1$ in the right panel.

five error distributions, we simulate only one sample data of 400 observations to calculate the standard errors at each data point. To perform the bootstrap, we randomly sample 400 pairs of $\left(Y_i, X_i^c, X_i^d\right)$ from the simulated sample data to do cross-validation and use the selected bandwidths to obtain a sequence of 400 $\hat{g}$. We repeat this process 1,000 times so that there are 1,000 estimates at each of the $400(x^c, x^d)$ points. We calculate the standard deviation of the 1,000 estimates $(\hat{g})$ at each (x^c, x^d) and use it as the bootstrap standard error.

Since the discrete regressor takes the values of 0 and 1, we produce two figures for each of the five error distributions; that is, one for $X^d = 0$ and the other for $X^d = 1$. To conserve space, only the results for the normal and t errors are reported. Figures for the other three error distributions can be found in the online supplement. Asymptotic confidence intervals are constructed using the critical values ± 1.96 and are reported as the shaded areas in the figures. The bootstrap results are given by the dashed lines. The true value of $g(x^c, x^d)$ is added as a solid line.

Unlike the results in Figures 1 and 2, the (shaded) asymptotic confidence intervals in Figures 3 and 4 are obtained from a single simulation of 400 observations. In most cases, the bootstrap method provides some improvement over the asymptotic

https://doi.org/10.1017/S0266466620000146 Published online by Cambridge University Press

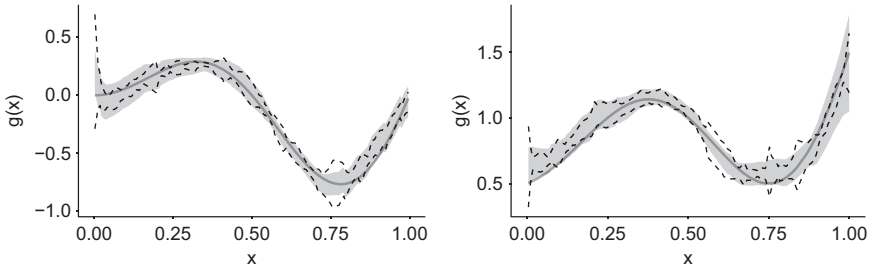


FIGURE 4. DGP 2 t distribution results. Dashed lines: bootstrap 95% confidence interval; shaded interval: asymptotic 95% confidence interval; solid line: $g(x)$. $X^d = 0$ in the left panel and $X^d = 1$ in the right panel.

results. However, the improvement is not uniform across the range of the data. For example, the shaded confidence interval is tighter around $x = 0.5$ than the dashed intervals in Figure 3. In Figure 4 in the online supplement, the bootstrap results are inferior to the asymptotic results in the case of the Laplace errors.

4.2. Empirical Illustration

In this section, we illustrate the robustness of the local CQR estimator for mixed data and the cross-validation bandwidth selection. We investigate the impact of the location factors on the housing prices of 506 census tracts in Boston from the 1970 census. We use two important location factors: "chas" indicating whether the tract shares a boundary with the Charles River, and "dis" denoting the weighted distances to five Boston employment centers. We take the median value of owner-occupied homes as our dependent variable.

We estimate the model with both the local CQR method and the LLS approach for mixed data for different data settings. We employ three data settings to check the robustness of the local CQR method and compare it to the LLS approach: (1) Original data containing censored median value (at \$50,000) (Data 1); (2) data with all censored observations removed (Data 2); and (3) the artificially generated extreme data with all censored median house values \$50,000 replaced by \$200,000 (Data 3). The motivation behind the artificially-generated Data 2 and Data 3 is to check the robustness of the two methods in different data cases. This mimics the empirical illustration strategy in Kai et al. (2010).

We report the results for observations with tracts not sharing a boundary with the Charles River (471 out of 506) in Figure 5. The local CQR method is robust for all three data settings and the LLS approach produces a nonrobust result for the generated data. The outcomes confirm the robustness of the local CQR method for mixed data, similar to the results obtained from using only continuous variables in Kai et al. (2010).

https://doi.org/10.1017/S0266466620000146 Published online by Cambridge University Press

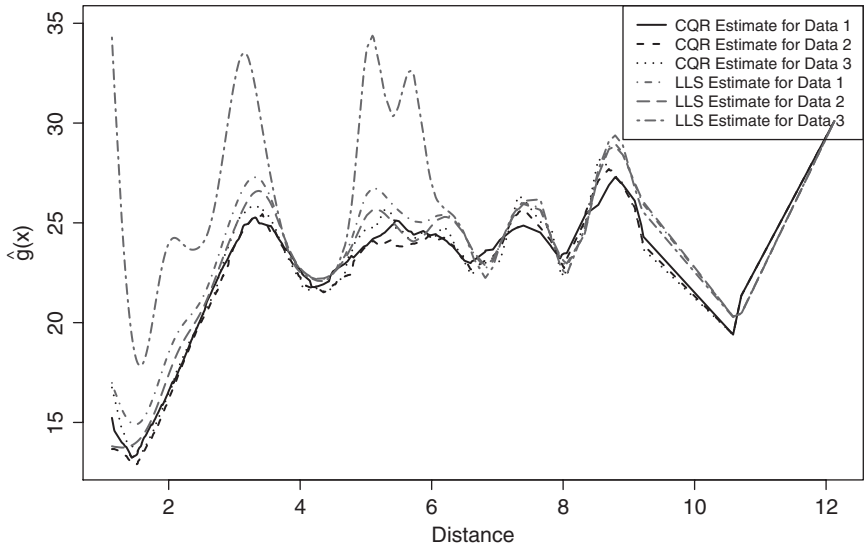


FIGURE 5. Local CQR and LLS comparison.

5. CONCLUSION

In this paper, we generalize the local CQR smoothing method to a nonparametric regression with mixed categorical and continuous regressors. The local CQR estimation is a more robust and safe alternative method to the local polynomial method with the flexible data structure. Cross-validation is used to select the optimal bandwidths for both types of regressors. We derive the asymptotic distribution of the local CQR estimator and the rates of convergence of bandwidths from the cross-validation. Our simulation study shows efficiency gains of the proposed local CQR smoother under a variety of DGPs when compared with the local linear estimator for several normal and nonnormal error distributions. We illustrate the robustness of our method using an empirical inquiry of the Boston housing market to determine the impacts of location factors, distance to business centers, and Charles River boundaries, on the median house price.

SUPPLEMENTARY MATERIAL

To view supplementary material for this article, please visit http://dx.doi.org/10.1017/S0266466620000146

REFERENCES

Fan, Y., W.K. Härdle, W. Wang & L. Zhu (2013) Composite Quantile Regression for the Single-Index Model. Working Paper.

https://doi.org/10.1017/S0266466620000146 Published online by Cambridge University Press

Guerre, E. & C. Sabbah (2012) Uniform bias study and Bahadur representation for local polynomial estimators of the conditional quantile function. *Econometric Theory* 28(1), 87–129.

Hall, P., Q. Li & J.S. Racine (2007) Nonparametric estimation of regression functions in the presence of irrelevant regressors. *The Review of Economics and Statistics* 89(4), 784–789.

Hall, P., J. Racine & Q. Li (2004) Cross-validation and the estimation of conditional probability densities. *Journal of the American Statistical Association* 99(468), 1015–1026.

Jiang, R., Z.-G. Zhou, W.-M. Qian & W.-Q. Shao (2012) Single-index composite quantile regression. *Journal of the Korean Statistical Society* 41(3), 323–332.

Kai, B., R. Li & H. Zou (2010) Local composite quantile regression smoothing: An efficient and safe alternative to local polynomial regression. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 72(1), 49–69.

Kai, B., R. Li & H. Zou (2011) New efficient estimation and variable selection methods for semiparametric varying-coefficient partially linear models. *Annals of Statistics* 39(1), 305.

Koenker, R. (2005) *Quantile Regression*. Cambridge University Press.

Koenker, R. (2017) Quantile regression: 40 years on. *Annual Review of Economics* 9, 155–176.

Koenker, R. & G. Bassett (1978) Regression quantiles. *Econometrica* 46(1), 33–50.

Koenker, R., V. Chernozhukov, X. He & L. Peng (2017) *Handbook of Quantile Regression*. CRC Press.

Kong, E. & Y. Xia (2014) An adaptive composite quantile approach to dimension reduction. *Annals of Statistics* 42(4), 1657–1688.

Li, D. & R. Li (2016) Local composite quantile regression smoothing for Harris recurrent Markov processes. *Journal of Econometrics* 194(1), 44–56.

Li, Q., J. Lin & J.S. Racine (2013) Optimal bandwidth selection for nonparametric conditional distribution and quantile functions. *Journal of Business & Economic Statistics* 31(1), 57–65.

Li, Q. & J. Racine (2004) Cross-validated local linear nonparametric regression. *Statistica Sinica* 14(2), 485–512.

Li, Q. & J.S. Racine (2007) *Nonparametric Econometrics: Theory and Practice*. Princeton University Press.

Pagan, A. & A. Ullah (1999) *Nonparametric Econometrics*. Cambridge University Press.

Racine, J. & Q. Li (2004) Nonparametric estimation of regression functions with both categorical and continuous data. *Journal of Econometrics* 119(1), 99–130.

Ruppert, D. & M.P. Wand (1994) Multivariate locally weighted least squares regression. *Annals of Statistics* 22, 1346–1370.

Su, L., Y. Chen & A. Ullah (2009) Functional coefficient estimation with both categorical and continuous data. In *Nonparametric Econometric Methods*, pp. 131–167. Emerald Group Publishing Limited.

Zhao, Z. & Z. Xiao (2014) Efficient regressions via optimally combining quantile information. *Econometric Theory* 30(6), 1272–1314.

Zou, H. & M. Yuan (2008) Composite quantile regression and the oracle model selection theory. *Annals of Statistics* 36(3), 1108–1126.

https://doi.org/10.1017/S0266466620000146 Published online by Cambridge University Press