Econometric Theory, **33**, 2017, 1081–1120.
doi:10.1017/S0266466616000293

LOCAL PARTITIONED QUANTILE REGRESSION

ZHENGYU ZHANG
Shanghai University of Finance and Economics

In this paper, we consider the nonparametric estimation of a broad class of quantile regression models, in which the partially linear, additive, and varying coefficient models are nested. We propose for the model a two-stage kernel-weighted least squares estimator by generalizing the idea of local partitioned mean regression (Christopeit and Hoderlein, 2006, *Econometrica* 74, 787–817) to a quantile regression framework. The proposed estimator is shown to have desirable asymptotic properties under standard regularity conditions. The new estimator has three advantages relative to existing methods. First, it is structurally simple and widely applicable to the general model as well as its submodels. Second, both the functional coefficients and their derivatives up to any given order can be estimated. Third, the procedure readily extends to censored data, including fixed or random censoring. A Monte Carlo experiment indicates that the proposed estimator performs well in finite samples. An empirical application is also provided.

1. INTRODUCTION

This paper discusses the nonparametric estimation of a general class of quantile regression models

$$Y = \theta(X) + m(Z) + g(X)^T W + U, \tag{1.1}$$

where Y is a real-valued dependent variable, T denotes transposition, $(X, W, Z) \in \mathbb{R}^{1+p_W+p_Z}$ is a $(1 + p_W + p_Z)$-dimensional random vector, $\theta(\cdot)$ and $m(\cdot)$ are unknown functions, $g(\cdot) = (g_1(\cdot), \ldots, g_{p_W}(\cdot))^T$ is an unknown p_W-vector valued function and U is an unobserved random variable whose τ-th quantile conditional on (X, W, Z) is zero for any given $0 < \tau < 1$. The estimation is based on an i.i.d. random sample $\{Y_i, X_i, W_i, Z_i\}$, $i = 1, \ldots, n$, of (Y, X, W, Z). Let the prime symbol denote the first order derivative. We describe an estimator of $(\theta'(\cdot), g(\cdot), g'(\cdot))$ that converges in probability pointwise at the optimal rate, regardless of the dimension of Z, implying no curse of dimensionality asymptotically. Moreover, our estimator has an oracle property. Specifically, the estimator is expected to perform asymptotically as if $m(\cdot)$ were known.

The author is grateful to the editor Peter C. B. Phillips, a co-editor and two anonymous referees for their constructive comments. Zhengyu Zhang is also affiliated with the Key Laboratory of Mathematical Economics (SUFE), Ministry of Education. The research is supported by the National Science Foundation of China (Grant No. 71501116). Zhengyu Zhang thanks Miffy Lee for her assistance in the research for this article. Address correspondence to Zhengyu Zhang, School of Economics, Shanghai University of Finance and Economics, 777 Guoding Road, 200433 Shanghai, China; e-mail: zy.zhang@mail.shufe.edu.cn.

© Cambridge University Press 2016

https://doi.org/10.1017/S0266466616000293 Published online by Cambridge University Press

In many applications, simple parametric models fit the data poorly, and thus a more flexible estimation approach is needed (Hardle, 1990; Horowitz, 1993; Horowitz and Lee, 2002). While a fully nonparametric estimation avoids the poor fit problem, it is usually unattractive in multivariate settings because the curse of dimensionality typically causes fully nonparametric estimates to be imprecise with moderate-sized samples. Hence, some structure must be placed on the regression function to make it useful for most econometric applications. The general regression model (1.1) is thus proposed based on this motivation. This model can reduce the effective dimension of the estimation problem, thereby achieving greater estimation precision than is possible with fully nonparametric methods, while also providing greater flexibility in the shape of the regression function than is possible with parametric models. Indeed, the regression function in (1.1) is sufficiently general to incorporate many econometrically interesting as well as empirically popular models as special cases. For example, it degenerates to a partially linear quantile regression model by imposing $\theta(\cdot) \equiv 0$ and $g(\cdot) \equiv \beta_0$, i.e.,

$$Y = m(Z) + W^T \beta_0 + U,$$

as studied by, for example, Lee (2003) and Chen and Khan (2001). The second submodel it nests is the additive quantile regression model

$$Y = \theta(X) + m(Z) + U, \tag{1.2}$$

by imposing $g(\cdot) \equiv 0$. Additive models are of considerable economic importance because they can be derived from separability assumptions on the utility or production function. Examples from production include the workhorses of the literature (Cobb-Douglas, Leontief). This additivity structure also arises in nonparametric control function and selection models (Lee, 2007). Let Z_l denote the l-th component of Z. (1.2) nests

$$Y = \theta(X) + \sum_{l=1}^{p_Z} m_l(Z_l) + U \tag{1.3}$$

but (1.2) is more general as it requires only that additive separability holds between X and Z, while the interaction among the other regressors than X is completely unrestricted. The estimation of (1.3) under the conditional quantile restriction has been studied by Yu and Lu (2004), De Gooijer and Zerom (2003), and more recently, by Horowitz and Lee (2005), among others. Another important case of (1.1) is the varying coefficient model, i.e.,

$$Y = \theta(X) + g(X)^T W + U, \tag{1.4}$$

by imposing $m(\cdot) \equiv 0$. The varying coefficient model can be regarded as a generalization of the classic linear regression model in which the marginal effects of W are nonconstant but vary across the covariates. Quantile regression estimation in this type of model has been explicitly considered by Fan and Zhang (1999), Cai and Xiao (2012), Kim (2007) and Wang, Zhu, and Zhou (2009).

https://doi.org/10.1017/S0266466616000293 Published online by Cambridge University Press

Examples include smoothly time varying coefficient models (see Fan and Zhang, 1999) and Deaton and Muellbauer's (1980) demand model in which demand for a commodity is modelled as linear in income W, with an income effect that varies nonparametrically with its own price X, together with the prices of other commodities (Z).

Recent work of Hoderlein and Mammen (2007) on local average structural derivative (LASD) estimation for nonseparable models is related in motivation to this paper. Their central result is that the derivative of the reduced-form quantile function (i.e., the given quantile of Y conditional on the covariates) is the best approximation (in the sense of minimizing the L_2 distance) to the average marginal effects over all individuals with the same observable data outcomes. A similar result was provided by Sasaki (2015), who showed under mild conditions, the quantile partial derivative identifies a weighted average of heterogeneous structural partial effects among the subpopulation of individuals at the conditional quantile of interest. The present paper concerns the estimation of marginal effects in regression quantiles, and therefore our procedure can be seen as a flexible reduced-form estimator for the LASD in Hoderlein and Mammen.

In particular, this paper proposes a new computationally feasible algorithm, named local partitioned quantile regression (LPQR), which can be used to estimate such a general model (1.1). Our estimation strategy is motivated by Christopeit and Hoderlein (2006, "CH" hereafter), who introduced a local partitioned regression (LPR) estimator for a broad class of mean regression models. They considered estimating the same regression model (1.1) but under the conditional mean restriction $\mathrm{E}(U|X, W, Z) = 0$. Based on an estimating equation modified from CH, our LPQR contains two stages. In the first stage, the conditional quantile functions of the dependent variable are nonparametrically fitted at a range of in-sample and out-of-sample points; the second stage is simply the kernel-weighted least squares. This simple algorithm ensures that a standard quantile regression programme is sufficient to compute our estimator, thus easing its implementation in practice for applied researchers. The proposed estimator is shown to have desirable asymptotic properties under standard regularity conditions.

Whereas numerous methods have been developed to estimate partially linear, additive, and varying coefficient quantile regression models, none of them is as general as our LPQR. Moreover, to our best knowledge, existing methods are unable to estimate (1.1) with censored data. Instead, we show that LPQR is sufficiently flexible to handle censored data, including either fixed or random censoring. The rest of the paper is structured as follows. The motivation of the estimation strategy and two-stage procedure are described in Section 2. Section 3 studies large sample properties. Section 4 extends LPQR to censored data. Section 5 reports some Monte Carlo results and Section 6 provides an empirical example. Section 7 concludes. All the technical details are presented in the Appendix.

https://doi.org/10.1017/S0266466616000293 Published online by Cambridge University Press

2. ESTIMATION PROCEDURE

Let $q_Y(x, w, z)$ denote the τ-th quantile of Y conditional on $(X, W, Z) = (x, w, z)$. Similar to CH, we focus on the marginal effects, namely $\partial q_Y(x, w, z)/\partial x = \theta'(x) + g'(x)^T w$ and $\partial q_Y(x, w, z)/\partial w = g(x)$ evaluated at $x = x_0$. Hence, we are interested in the parameters

$$\Theta_0 = \left\{\theta'(x_0), g(x_0)^T, g'(x_0)^T\right\}^T.$$

For these quantities, our objective is to find asymptotically normal estimators that behave like one-dimensional nonparametric estimates. For instance, $\theta'(x_0)$ should converge at the rate $\sqrt{nh^3}$, which is the optimal rate of convergence under certain smoothness assumptions.

To explain our identification strategy, we review CH's local partitioned regression. Consider the additive model (1.2) under the conditional mean restriction $\mathrm{E}(U|X, Z) = 0$. Denote

$$K_h(x) = \frac{1}{h} K\left(\frac{x - x_0}{h}\right),$$

where $K(\cdot)$ is the density kernel function and h is the bandwidth parameter to be specified subsequently. Then CH's LPR estimator for $\theta'(x_0)$ is given by regressing

$$K_h(X_i)Y_i - \mathrm{E}\{K_h(X_i)Y_i|K_h(X_i)Z_i\}$$

on

$$K_h(X_i)(X_i - x_0) - \mathrm{E}\{K_h(X_i)(X_i - x_0)|K_h(X_i)Z_i\},$$

which can be seen as a generalization of Robinson's (1988) partitioned regression estimator for the partially linear regression model.[1] The localizer $K_h(X_i)$ appearing in LPR restricts the partitioned regression to be implemented only for the observations within a small neighborhood centered on x_0. This idea of local partitioned regression was extended by CH to the more sophisticated model (1.1) under the conditional mean restriction $\mathrm{E}(U|X, W, Z) = 0$. Specifically, CH derived the following equation (equation (3.2) in CH):

$$\begin{aligned} &K_h(X)Y - \mathrm{E}\{K_h(X)Y|K_h(X)Z\} \\ &\quad = \theta'(x_0)(K_h(X)(X - x_0) - \mathrm{E}\{K_h(X)(X - x_0)|K_h(X)Z\}) \\ &\qquad + g(x_0)^T(K_h(X)W - \mathrm{E}\{K_h(X)W|K_h(X)Z\}) \\ &\qquad + g'(x_0)^T(K_h(X)(X - x_0)W - \mathrm{E}\{K_h(X)(X - x_0)W|K_h(X)Z\}) \\ &\qquad + \text{remainder terms.} \end{aligned} \tag{2.1}$$

Based on (2.1), LPR for Θ_0 is simply the OLS coefficient by regressing

$$K_h(X_i)Y_i - \mathrm{E}\{K_h(X_i)Y_i|K_h(X_i)Z_i\}$$

https://doi.org/10.1017/S0266466616000293 Published online by Cambridge University Press

on

$$K_h(X_i)(X_i - x_0) - \mathrm{E}\{K_h(X_i)(X_i - x_0)|K_h(X_i)Z_i\},$$
$$K_h(X_i)W_i - \mathrm{E}\{K_h(X_i)W_i|K_h(X_i)Z_i\}, \quad \text{and}$$
$$K_h(X_i)(X_i - x_0)W_i - \mathrm{E}\{K_h(X_i)(X_i - x_0)W_i|K_h(X_i)Z_i\}.^2$$

To extend CH's LPR to the quantile regression framework, we must interpret equation (2.1) in a slightly different manner. Let us view $K_h(\cdot)$ as a selection function in the sense that it selects only the observations sufficiently close to x_0 into the second stage of the regression. Informally let $J(x_0, z)$ denote such a set that $j \in J(x_0, z)$ if and only if X_j is regarded close to x_0 given $Z = z$. Under the conditional mean restriction, we can write $Y = E_Y(X, Z, W) + U$, where $E_Y(x, z, w)$ is the expectation of Y conditional on $X = x$, $Z = z$, and $W = w$. Then conditional on $Z = z$ and using the sample average to approximate the expectation operator, equation (2.1) can be reinterpreted as, for any $i \in J(x_0, z)$,

$$\begin{aligned} E_Y(X_i, W_i, z) - \frac{1}{\#J(x_0,z)} \sum_{j \in J(x_0,z)} E_Y(X_j, W_j, z) + \left\{ U_i - \frac{1}{\#J(x_0,z)} \sum_{j \in J(x_0,z)} U_j \right\} \\ = \theta'(x_0) \left\{ (X_i - x_0) - \frac{1}{\#J(x_0,z)} \sum_{j \in J(x_0,z)} (X_j - x_0) \right\} \\ + g(x_0)^T \left\{ W_i - \frac{1}{\#J(x_0,z)} \sum_{j \in J(x_0,z)} W_j \right\} \\ + g'(x_0)^T \left\{ (X_i - x_0)W_i - \frac{1}{\#J(x_0,z)} \sum_{j \in J(x_0,z)} (X_j - x_0)W_j \right\} + \text{residual terms}, \end{aligned}$$

where $\#J(x_0, z)$ is the number of elements in $J(x_0, z)$. Multiplying both sides of the above equation with $\#J(x_0, z)$ gives

$$\begin{aligned} \sum_{j \in J(x_0,z)} \{E_Y(X_i, W_i, z) - E_Y(X_j, W_j, z)\} = \theta'(x_0) \sum_{j \in J(x_0,z)} \{X_i - X_j\} \\ + g(x_0)^T \sum_{j \in J(x_0,z)} \{W_i - W_j\} \\ + g'(x_0)^T \sum_{j \in J(x_0,z)} \{W_i(X_i - x_0) - W_j(X_j - x_0)\} \\ + \text{residual terms}. \end{aligned}$$

To apply the above equation in the quantile regression context, we simply need to replace $E_Y(x, w, z)$ with $q_Y(x, w, z)$, the τth quantile function of Y conditional on $X = x$, $W = w$, and $Z = z$, i.e.,

$$\Pr\{Y \le q_Y(x, w, z)|X = x, W = w, Z = z\} = \tau,$$

https://doi.org/10.1017/S0266466616000293 Published online by Cambridge University Press

thus we arrive at

$$\begin{aligned}\sum_{j\in J(x_0,z)}\{q_Y(X_i,W_i,z)-q_Y(X_j,W_j,z)\} &= \theta'(x_0)\sum_{j\in J(x_0,z)}\{X_i-X_j\}\\ &\quad+g(x_0)^T\sum_{j\in J(x_0,z)}\{W_i-W_j\}\\ &\quad+g'(x_0)^T\sum_{j\in J(x_0,z)}\{W_i(X_i-x_0)-W_j(X_j-x_0)\}\\ &\quad+\text{residual terms}\end{aligned}\tag{2.2}$$

for any $i\in J(x_0,z)$. The implication of equation (2.2) is clear: given $Z=z$ and provided that X_i and X_j lie close to x_0, then for any pair of indices $i\neq j$, $q_Y(X_i,W_i,z)-q_Y(X_j,W_j,z)$ can be approximately represented as a linear combination of three independent variables: X_i-X_j, W_i-W_j, and $W_i(X_i-x_0)-W_j(X_j-x_0)$. An alternative way in which to derive the same implication as in equation (2.2) is to linearly expand $q_Y(X_i,W_i,z)-q_Y(X_j,W_j,z)$, bearing in mind that both X_i and X_j lie close to x_0. Then we arrive at

$$\begin{aligned}q_Y(X_i,W_i,z)-q_Y(X_j,W_j,z) &= \theta(X_i)-\theta(X_j)+g(X_i)^TW_i-g(X_j)^TW_j\\ &=\theta(X_i)-\theta(X_j)+g(x_0)^T\{W_i-W_j\}\\ &\quad+\{g(X_i)-g(x_0)\}^TW_i-\{g(X_j)-g(x_0)\}^TW_j\\ &=\{X_i-X_j\}\theta'(x_0)+g(x_0)^T\{W_i-W_j\}\\ &\quad+g'(x_0)^T\{(X_i-x_0)W_i-(X_j-x_0)W_j\}+r_{ij},\end{aligned}\tag{2.3}$$

where the remainder term

$$r_{ij}=\frac{1}{2}\theta''(\overline{x}_0)(X_i-X_j)^2+\frac{1}{2}g''(\overline{x}_0)^T\left\{(X_i-x_0)^2W_i-(X_j-x_0)^2W_j\right\},$$

with $\overline{x}_0$ lying between X_i and X_j. The above reasoning implies that Θ_0 is identified if the following $(2p_W+1)\times(2p_W+1)$ matrix

$$V_1=\mathrm{E}\left\{\begin{bmatrix}X_1-X_2\\W_1-W_2\\(X_1-x_0)W_1-(X_2-x_0)W_2\end{bmatrix}\begin{bmatrix}X_1-X_2\\W_1-W_2\\(X_1-x_0)W_1-(X_2-x_0)W_2\end{bmatrix}^T K_h(X_1)K_h(X_2)\right\}$$

has full rank. From some straightforward computation, we can show

$$V_1=\begin{pmatrix}h&0&0\\0&I_{p_W}&0\\0&0&h\times I_{p_W}\end{pmatrix}\times V_2\times\begin{pmatrix}h&0&0\\0&I_{p_W}&0\\0&0&h\times I_{p_W}\end{pmatrix},$$

where

$$V_2=\begin{pmatrix}V_{\theta'\theta'}&0&V_{\theta'g'}^T\\0&V_{gg}&0\\V_{\theta'g'}&0&V_{g'g'}\end{pmatrix},$$

https://doi.org/10.1017/S0266466616000293 Published online by Cambridge University Press

with

$$\kappa_2 = \int K(s)s^2 ds, \quad V_{\theta'\theta'} = \kappa_2 f_X^2(x_0),$$
$$V_{gg} = \left\{\mathrm{E}\left(WW^T|X=x_0\right) - \mathrm{E}\left(W|X=x_0\right)\mathrm{E}\left(W^T|X=x_0\right)\right\} f_X^2(x_0),$$
$$V_{g'g'} = \kappa_2 \mathrm{E}\left(WW^T|X=x_0\right) f_X^2(x_0), \quad V_{\theta'g'} = \kappa_2 \mathrm{E}\left(W|X=x_0\right) f_X^2(x_0).$$

We summarize this identification condition in the following assumption.

Assumption 2.1. The matrix V_2 is positive definite.

Since equation (2.3) holds for any given z, it should also hold for each Z_k, $k = 1, \ldots, n$, Thus, we can estimate Θ_0 by minimizing the following criterion function:

$$\sum_{k=1}^{n} \sum_{1 \le i < j \le n} \left\{ q_Y(X_i, W_i, Z_k) - q_Y(X_j, W_j, Z_k) - (X_i - X_j)\beta_{\theta'} - (W_i - W_j)^T \beta_g \right.$$
$$\left. - \left(W_i(X_i - x_0) - W_j(X_j - x_0)\right)^T \beta_{g'} \right\}^2 K_h(X_i) K_h(X_j)$$

with respect to $(\beta_{\theta'}, \beta_g, \beta_{g'})$. Moreover, minimizing the above criterion function amounts to regressing

$$\{q_Y(X_i, W_i, Z_k) - q_Y(X_j, W_j, Z_k)\} K_h(X_i) K_h(X_j)$$

on

$$\{X_i - X_j\} K_h(X_i) K_h(X_j), \quad \{W_i - W_j\} K_h(X_i) K_h(X_j)$$

and

$$\{(X_i - x_0)W_i - (X_j - x_0)W_j\} K_h(X_i) K_h(X_j),$$

for $i \neq j$, $1 \le i, j, k \le n$, which provides a more intuitive connection between our estimation strategy and that of CH.

Remark 2.1. The estimation procedure described above is sufficiently flexible to estimate the other high-order derivatives of $\theta(\cdot)$ and $g(\cdot)$ in which a researcher might be interested. To see this, expand $q_Y(X_i, W_i, z) - q_Y(X_j, W_j, z)$ until the remainder term is a polynomial of order three with respect to $(X_i - x_0)^3$ and $(X_j - x_0)^3$, i.e.,

$$\begin{aligned} & q_Y(X_i, W_i, z) - q_Y(X_j, W_j, z) \\ & = \{X_i - X_j\}\theta'(x_0) + \frac{1}{2}\left\{(X_i - x_0)^2 - (X_j - x_0)^2\right\}\theta''(x_0) \\ & \quad + g(x_0)^T\{W_i - W_j\} + g'(x_0)^T\{(X_i - x_0)W_i - (X_j - x_0)W_j\} \\ & \quad + \frac{g''(x_0)^T}{2}\left\{(X_i - x_0)^2 W_i - (X_j - x_0)^2 W_j\right\} + \text{residual terms}. \end{aligned}$$

https://doi.org/10.1017/S0266466616000293 Published online by Cambridge University Press

Then one may estimate

$$\{\theta'(x_0), \theta''(x_0), g(x_0), g'(x_0), g''(x_0)\}$$

by regressing

$$\{q_Y(X_i, W_i, Z_k) - q_Y(X_j, W_j, Z_k)\}K_h(X_i)K_h(X_j)$$

on

$$\{X_i - X_j\}K_h(X_i)K_h(X_j), \quad \frac{1}{2}\{(X_i - x_0)^2 - (X_j - x_0)^2\}K_h(X_i)K_h(X_j),$$
$$\{W_i - W_j\}K_h(X_i)K_h(X_j), \quad \{(X_i - x_0)W_i - (X_j - x_0)W_j\}K_h(X_i)K_h(X_j),$$

and

$$\frac{1}{2}\{(X_i - x_0)^2 W_i - (X_j - x_0)^2 W_j\}K_h(X_i)K_h(X_j)$$

for $i \neq j$, $1 \leq i, j, k \leq n$.

Remark 2.2. LPQR can be simplified to estimate the model (1.2). In this case,

$$q_Y(X_i, z) - q_Y(X_j, z) = \theta(X_i) - \theta(X_j) = (X_i - X_j)\theta'(x_0) + \text{residual term}.$$

Thus one may estimate $\theta'(x_0)$, the marginal effect of X on Y evaluated at $X = x_0$, simply by regressing

$$\{q_Y(X_i, Z_k) - q_Y(X_j, Z_k)\}K_h(X_i)K_h(X_j) \quad \text{on} \quad \{X_i - X_j\}K_h(X_i)K_h(X_j)$$

for $1 \leq i, j, k \leq n$ and $i \neq j$. Given the estimate of $\theta'(x)$, it is straightforward to recover the other objects of potential interest. For example, $\theta(x_0)$ can be estimated up to a location shift by

$$\widehat{\theta}_n(x_0) = \int_0^{x_0} \widehat{\theta}'_n(s)ds,$$

and the average derivatives $\vartheta = \mathrm{E}\theta'(X)$ can be estimated by

$$\widehat{\vartheta}_n = \frac{1}{n}\sum_{i=1}^{n} \widehat{\theta}'_n(X_i).$$

Remark 2.3. For the varying coefficient model (1.4), LPQR also has a simplified form. Since

$$\begin{aligned} q_Y(X_i, W_i) - q_Y(X_j, W_j) = {} & \{X_i - X_j\}\theta'(x_0) + g(x_0)^T\{W_i - W_j\} \\ & + g'(x_0)^T\{(X_i - x_0)W_i - (X_j - x_0)W_j\} \\ & + \text{residual terms}, \end{aligned}$$

$\{\theta'(x_0), g(x_0), g'(x_0)\}$ can be estimated by regressing

$$\{q_Y(X_i, W_i) - q_Y(X_j, W_j)\}K_h(X_i)K_h(X_j)$$

on

$$\{X_i - X_j\}K_h(X_i)K_h(X_j), \quad \{W_i - W_j\}K_h(X_i)K_h(X_j)$$

https://doi.org/10.1017/S0266466616000293 Published online by Cambridge University Press

and

$$\{(X_i - x_0)W_i - (X_j - x_0)W_j\}K_h(X_i)K_h(X_j),$$

for $1 \le i, j \le n$ and $i \ne j$. For (1.4), LPQR has two advantages relative to existing algorithms (e.g., Kim, 2007; Wang et al., 2009; Cai and Xiao, 2012). First, it estimates not only the slope coefficient function $g(\cdot)$ but also its derivative $g'(\cdot)$. As commented in Remark 2.1, if desired, it can thus estimate the other high-order derivatives of $g(\cdot)$. Second, existing methods cannot estimate the varying coefficient model with a censored dependent variable, while LPQR can, as demonstrated in Section 4.

Remark 2.4. Our discussion focuses on the case with univariate X, which is sufficient for most empirical studies. The identification and estimation principle described above can easily be extended to the multivariate X case. To illustrate, let us consider the case where the dimension of X is two: $X = (X^1, X^2)$. Similar to (2.3), we can write

$$\begin{aligned}
&q_Y\left(X_i^1, X_i^2, W_i, z\right) - q_Y\left(X_j^1, X_j^2, W_j, z\right)\\
&\quad = \theta\left(X_i^1, X_i^2\right) - \theta\left(X_j^1, X_j^2\right) + g\left(X_i^1, X_i^2\right)^T W_i - g\left(X_j^1, X_j^2\right)^T W_j\\
&\quad = \theta\left(X_i^1, X_i^2\right) - \theta\left(X_j^1, X_j^2\right) + g\left(x_0^1, x_0^2\right)^T \{W_i - W_j\}\\
&\qquad + \left\{g\left(X_i^1, X_i^2\right) - g\left(x_0^1, x_0^2\right)\right\}^T W_i - \left\{g\left(X_j^1, X_j^2\right) - g\left(x_0^1, x_0^2\right)\right\}^T W_j\\
&\quad = \left\{X_i^1 - X_j^1\right\}\theta_1'\left(x_0^1, x_0^2\right) + \left\{X_i^2 - X_j^2\right\}\theta_2'\left(x_0^1, x_0^2\right) + g\left(x_0^1, x_0^2\right)^T \{W_i - W_j\}\\
&\qquad + g_1'\left(x_0^1, x_0^2\right)^T \left\{\left(X_i^1 - x_0^1\right)W_i - \left(X_j^1 - x_0^1\right)W_j\right\}\\
&\qquad + g_2'\left(x_0^1, x_0^2\right)^T \left\{\left(X_i^2 - x_0^2\right)W_i - \left(X_j^2 - x_0^2\right)W_j\right\} + \text{residual terms},
\end{aligned}$$

where $\theta_1'(\cdot)$ is the partial derivative of $\theta(\cdot)$ with respect to the first component of X, and $\theta_2'(\cdot)$, $g_1'(\cdot)$, $g_2'(\cdot)$ can be similarly defined. Thus $\left\{\theta_1'\left(x_0^1, x_0^2\right), \theta_2'\left(x_0^1, x_0^2\right), g\left(x_0^1, x_0^2\right), g_1'\left(x_0^1, x_0^2\right), g_2'\left(x_0^1, x_0^2\right)\right\}$ can be estimated by regressing

$$\left\{q_Y\left(X_i^1, X_i^2, W_i, Z_k\right) - q_Y\left(X_j^1, X_j^2, W_j, Z_k\right)\right\}K_h\left(X_i^1\right)K_h\left(X_j^1\right)K_h\left(X_i^2\right)K_h\left(X_j^2\right)$$

on

$$\begin{aligned}
&\left\{X_i^1 - X_j^1\right\}K_h\left(X_i^1\right)K_h\left(X_j^1\right)K_h\left(X_i^2\right)K_h\left(X_j^2\right),\\
&\left\{X_i^2 - X_j^2\right\}K_h\left(X_i^1\right)K_h\left(X_j^1\right)K_h\left(X_i^2\right)K_h\left(X_j^2\right),\\
&\left\{W_i - W_j\right\}K_h\left(X_i^1\right)K_h\left(X_j^1\right)K_h\left(X_i^2\right)K_h\left(X_j^2\right),\\
&\left\{\left(X_i^1 - x_0^1\right)W_i - \left(X_j^1 - x_0^1\right)W_j\right\}K_h\left(X_i^1\right)K_h\left(X_j^1\right)K_h\left(X_i^2\right)K_h\left(X_j^2\right)
\end{aligned}$$

and

$$\left\{\left(X_i^2 - x_0^2\right)W_i - \left(X_j^2 - x_0^2\right)W_j\right\}K_h\left(X_i^1\right)K_h\left(X_j^1\right)K_h\left(X_i^2\right)K_h\left(X_j^2\right).$$

https://doi.org/10.1017/S0266466616000293 Published online by Cambridge University Press

To operationalize the estimation strategy described above, $q_Y(X_i, W_i, Z_k)$'s and $q_Y(X_j, W_j, Z_k)$'s should be replaced by their nonparametric estimates. Thus the first stage of LPQR involves nonparametrically fitting the conditional quantiles of Y given the regressors (X, W, Z). While numerous methods are available to estimate the conditional regression quantiles in the statistics and econometrics literature, we use the local polynomial regression estimator studied by Chaudhuri (1991a, 1991b) and Guerrea and Sabbah (2012). This estimation procedure is computationally simple as it minimizes the globally convex objective function that can be handled by using linear programming methods. Moreover, it can control the order of the bias by appropriately selecting the polynomial order. Without loss of generality, we assume that Z contains only continuous components.

Let $C(x, w, z, \delta)$ denote the $p_W + p_Z + 1$ dimensional cube centered at (x, w, z) with side length 2δ. Let d_θ denote the order of differentiability of $\theta(\cdot)$ with respect to X, d_m the order of differentiability of $m(\cdot)$ with respect to Z, and d_g the order of differentiability of $g(\cdot)$ with respect to X. Let $A(d_m)$ signify the following set of p_Z-dimensional vector of non-negative integers

$$\left\{ \mathbf{a} = (a_1, \ldots, a_{p_Z}) \Big| \sum_{l=1}^{p_Z} |a_l| \leq d_m \right\}.$$

Let $\#A(d_m)$ be the cardinality of $A(d_m)$. Moreover, for convenience, let the first element in $A(d_m)$ be $(0, \ldots, 0)$. Then the local polynomial fitting of $q_Y(x, w, z)$ amounts to solving the following minimization problem

$$\begin{aligned}
&\left(\widehat{b}_n^X, \widehat{b}_n^Z, \widehat{b}_n^W, \widehat{b}_n^{XW}\right) \\
&= \arg \min_{(b^X, b^W, b^Z, b^{XW})} \sum_{k=1}^{n} I\{(X_k, W_k, Z_k) \in C(x, w, z, \delta)\} \rho_\tau \left\{ Y_k - \sum_{l=1}^{d_\theta} b_l^X (X_k - x)^l \right. \\
&\quad \left. - \sum_{l=1}^{\#A(d_m)} b_l^Z (Z_k - z)^{\alpha_l} - \sum_{p=1}^{p_W} b_p^W (W_{pk} - w_p) - \sum_{p=1}^{p_W} \sum_{l=1}^{d_g} b_{pl}^{XW} (X_k - x)^l (W_{pk} - w_p) \right\},
\end{aligned}$$

where W_{pk} and w_p are the p-th elements of W_k and w, respectively: $b^X = \left(b_1^X, \ldots, b_{d_\theta}^X\right)$, $b^Z = \left(b_1^Z, \ldots, b_{\#A(d_m)}^Z\right)$, $b^W = \left(b_1^W, \ldots, b_{p_W}^W\right)$ and $b^{XW} = \left(b_{11}^{XW}, \ldots, b_{1d_g}^{XW}, \ldots, b_{p_W d_g}^{XW}\right)$; $(Z_k - z)^{\alpha_l}$ is a shorthand notation for the product of each component of $(Z_k - z)$ raised to the corresponding component of $\alpha_l \in A(d_m)$, $\rho_\tau(u) = u(\tau - I\{u \leq 0\})$. Among the $[d_\theta + p_W(d_g + 1) + \#A(d_m)]$ components of $(\widehat{b}_n^X, \widehat{b}_n^Z, \widehat{b}_n^W, \widehat{b}_n^{XW})$, it is the coefficient corresponding to the constant term $\widehat{b}_1^Z$, i.e., the first component of $\widehat{b}_n^Z$ that estimates the conditional quantile function $q_Y(x, w, z)$ while the others are regarded as nuisance parameters. Two features of (1.1) have been exploited by the objective function of local polynomial regression. First, owing to the additivity structure, there are no interactive terms between X and Z or between Z and W. Second, the objective function is linear in W, i.e., no high-order polynomials with respect to W are incorporated, thus giving a computational advantage by reducing the dimensionality of the minimization problem.

https://doi.org/10.1017/S0266466616000293 Published online by Cambridge University Press

In the second stage, we estimate Θ_0 by

$$\begin{aligned}\widehat{\Theta}_n &= \left\{\widehat{\theta}_n'(x_0), \widehat{g}_n(x_0)^T, \widehat{g}_n'(x_0)^T\right\}^T \\ &= \arg\min_{\beta_{\theta'},\beta_g,\beta_{g'}} \sum_{k=1}^{n} \sum_{1\le i<j\le n} K_h(X_i)K_h(X_j)\big\{\widehat{q}_Y(X_i,W_i,Z_k) - \widehat{q}_Y(X_j,W_j,Z_k) \\ &\quad - (X_i - X_j)\beta_{\theta'} - (W_i - W_j)^T\beta_g - (W_i(X_i - x_0) - W_j(X_j - x_0))^T\beta_{g'}\big\}^2,\end{aligned}$$

where $\widehat{q}_Y(X_i, W_i, Z_k)$ and $\widehat{q}_Y(X_j, W_j, Z_k)$ are the local polynomial quantile regression estimates of $q_Y(X_i, W_i, Z_k)$ and $q_Y(X_j, W_j, Z_k)$, respectively. The above estimator can be equivalently expressed as

$$\widehat{\Theta}_n = S_{1n}^{-1} S_{2n}, \tag{2.4}$$

where

$$S_{1n} = \sum_{k=1}^{n} \sum_{1\le i<j\le n} \mathbf{X}_{ij}\mathbf{X}_{ij}^T K_h(X_i)K_h(X_j),$$

$$S_{2n} = \sum_{k=1}^{n} \sum_{1\le i<j\le n} \mathbf{X}_{ij}\widehat{\mathbf{Y}}_{ijk} K_h(X_i)K_h(X_j),$$

$$\mathbf{X}_{ij} = \left\{X_i - X_j, (W_i - W_j)^T, (X_i - x_0)W_i^T - (X_j - x_0)W_j^T\right\}^T,$$

and $\widehat{\mathbf{Y}}_{ijk} = \widehat{q}_Y(X_i, W_i, Z_k) - \widehat{q}_Y(X_j, W_j, Z_k)$.

According to Remark 2.3, LPQR can be simplified for (1.4), i.e.,

$$\widehat{\Theta}_n = S_{3n}^{-1} S_{4n}, \tag{2.5}$$

where

$$S_{3n} = \sum_{1\le i<j\le n} \mathbf{X}_{ij}\mathbf{X}_{ij}^T K_h(X_i)K_h(X_j),$$

$$S_{4n} = \sum_{1\le i<j\le n} \mathbf{X}_{ij}\widehat{\mathbf{Y}}_{ij} K_h(X_i)K_h(X_j),$$

and $\widehat{\mathbf{Y}}_{ij} = \widehat{q}_Y(X_i, W_i) - \widehat{q}_Y(X_j, W_j)$.

Remark 2.5. As a two-stage kernel-weighted LS estimator, LPQR shares some similarities with two-stage estimation procedures that use local polynomial quantile regression as their first-stage (e.g., Lee, 2003; Chen and Khan, 2001, 2008). Among them, Lee (2003) and Chen and Khan (2001) estimated a semiparametric partially linear regression model under the conditional quantile restriction, which is nested by our general model (1.1) as a special case.

https://doi.org/10.1017/S0266466616000293 Published online by Cambridge University Press

3. LARGE SAMPLE PROPERTIES

Let $\mathcal{X}$, $\mathcal{W}$, and $\mathcal{Z}$ signify the support of X, W, and Z, respectively. For any matrix A, denote $\|A\| = \sqrt{\text{trace}(AA^T)}$. For some $\varrho \in (0, 1]$, and any function $m(\cdot)$ defined on a set $\mathcal{D}$, we mean by $m(\cdot) \in \mathcal{C}^{\varrho}(\mathcal{D})$ that there exists a positive constant C such that

$$|m(x_1) - m(x_2)| \leq C\|x_1 - x_2\|^{\varrho}$$

for any $x_1, x_2 \in \mathcal{D}$.

Assumption 3.1. The data $\{Y_i, X_i, W_i, Z_i\}$, $i = 1, \ldots, n$ is a random sample of (Y, X, W, Z) generated by (1.1). The τ-th quantile of U conditional on (X, W, Z) is zero.

Assumption 3.2. The distribution of (X, W, Z) is absolutely continuous with respect to Lebesgue measure. The probability density function of (X, W, Z), denoted as $f_{XWZ}(x, w, z)$ is continuous in (x, w, z), uniformly bounded away from 0 and belongs to $\mathcal{C}^{\varrho}(\mathcal{X} \times \mathcal{W} \times \mathcal{Z})$.

Assumption 3.3. The distribution of U conditional on (X, W, Z) has the probability density $f_{U|XWZ}(u|x, w, z)$, which is positive at $u = 0$ and continuous in a neighborhood of $u = 0$ for all $(x, w, z) \in \mathcal{X} \times \mathcal{W} \times \mathcal{Z}$; $f_{U|XWZ}(0|x, w, z)$ belongs to $\mathcal{C}^{\varrho}(\mathcal{X} \times \mathcal{W} \times \mathcal{Z})$.

Assumption 3.4. $\theta(x)$ is continuously differentiable of order $d_\theta \geq 4$ with respect to x and all its derivatives up to the d_θ-th belong to $\mathcal{C}^{\varrho}(\mathcal{X})$; $m(z)$ is continuously differentiable of order d_m with respect to z and all its derivatives up to the d_m-th belong to $\mathcal{C}^{\varrho}(\mathcal{Z})$; $g(x)$ is continuously differentiable of order $d_g \geq 4$ with respect to x and all its derivatives up to the d_g-th belong to $\mathcal{C}^{\varrho}(\mathcal{X})$;

Assumption 3.5.

(i) $K(\cdot)$ is a symmetric kernel with bounded first derivative, satisfying

$$\int K(s)ds = 1 \quad \text{and} \quad \kappa_4 = \int s^4 K(s)ds < \infty.$$

(ii) $\delta = O(n^{-\eta})$, with

$$\eta \in (1/(2\varrho + 2\min\{d_\theta, d_m, d_g\}), 1/(3 + 3p_Z + 3p_W)).$$

(iii) $nh^3 \to \infty$, $nh^5\delta^{2(p_Z+p_W+1)} \to \infty$, $\delta^{\varrho}/h^{5/2} \to 0$.

Assumption 3.1 describes the data-generating process. Assumptions 3.2 and 3.3 parallel Assumptions 1–2 in CH to impose regularity conditions on f_{XWZ} and $f_{U|XWZ}$. The smoothness conditions imposed on $\theta(\cdot)$, $m(\cdot)$, and $g(\cdot)$ in Assumption 3.4 are reasonable for a wide variety of applications and incorporate most parametric specifications as special cases. Assumption 3.5 concerns

https://doi.org/10.1017/S0266466616000293 Published online by Cambridge University Press

the choice of the kernel function and bandwidths involved in the estimation procedure. Among them, Assumption 3.5(ii) is borrowed from Lemma 4.1 in Chaudhuri, Doksum, and Samarov (1997); similar conditions can be found in the literature on semiparametric methods using local polynomial regression as their first stage estimator, such as Assumption 6 in Lee (2003), Assumption BC in Chen and Khan (2000) and Assumption 3.6 in Chen and Khan (2001). In general, these bandwidth conditions are required to control the order of the remainder term resulting from the first stage to be sufficiently small so that plugging them into the second stage will not affect the optimal convergence rate of the final estimator. Assumption 3.5(iii) parallels CH's Assumption 5 and turns out to be sufficient for our general case. However, a slight modification of these bandwidth conditions seems necessary if one focuses on a certain specialized model.

THEOREM 3.1. *Under Assumptions 2.1, 3.1–3.5, the estimator* (2.4) *has the limiting distribution*

$$\begin{pmatrix} \sqrt{nh^3} \\ \sqrt{nh} \\ \sqrt{nh^3} \end{pmatrix} \left(\widehat{\Theta}_n - \Theta_0 - h^2 V_2^{-1} \begin{pmatrix} b_{\theta'} \\ b_g \\ b_{g'} \end{pmatrix} \right) \to^d \mathcal{N}\left(0, V_2^{-1} \Omega_1 V_2^{-1}\right),$$

where

$$\begin{aligned} b_{\theta'} &= f_X^2(x_0) \frac{(\kappa_4 + 3\kappa_2^2)\theta'''(x_0)}{6} + \mathrm{E}\left(W^T \middle| X = x_0\right) f_X^2(x_0) \frac{\kappa_4 g'''(x_0)}{6} \\ &\quad + \int w^T \frac{\partial f_{XW}(x_0, w)}{\partial X} dw \cdot f_X(x_0) \cdot \frac{\kappa_4 g''(x_0)}{2} \\ &\quad - \int \frac{\partial f_{XW}(x, w)}{\partial X} dw \cdot \int w^T f_{XW}(x_0, w) dw \cdot \frac{\kappa_2^2 g''(x_0)}{2}, \end{aligned}$$

$$b_g = \left\{ \mathrm{E}\left(WW^T \middle| X = x_0\right) - \mathrm{E}\left(W \middle| X = x_0\right) \mathrm{E}\left(W^T \middle| X = x_0\right) \right\} f_X^2(x_0) \frac{\kappa_2 g''(x_0)}{2},$$

$$\begin{aligned} b_{g'} &= \mathrm{E}\left(W \middle| X = x_0\right) f_X^2(x_0) \frac{(\kappa_4 + 3\kappa_2^2)\theta'''(x_0)}{6} + \mathrm{E}\left(WW^T \middle| X = x_0\right) f_X^2(x_0) \frac{\kappa_4 g'''(x_0)}{6} \\ &\quad + \int w w^T \frac{\partial f_{XW}(x_0, w)}{\partial X} dw \cdot f_X(x_0) \frac{\kappa_4 g''(x_0)}{2} \\ &\quad - \int w \frac{\partial f_{XW}(x_0, w)}{\partial X} dw \int w^T f_{XW}(x_0, w) dw \frac{\kappa_2^2 g''(x_0)}{2}, \end{aligned}$$

and $\Omega_1 = \mathrm{E}\left(\psi_i \psi_i^T\right)$, *with* $\psi_i = \left(\psi_{\theta' i}, \psi_{gi}^T, \psi_{g' i}^T\right)^T$,

$$\begin{aligned} \psi_{\theta' i} &= \frac{1}{\sqrt{h}} \left(\tau - I\{U_i \le 0\}\right) \left(\frac{X_i - x_0}{h}\right) K\left(\frac{X_i - x_0}{h}\right) f_{UXWZ}^{-1}(0, X_i, W_i, Z_i) \\ &\quad \times f_{XW}(X_i, W_i) f_Z(Z_i) f_X(x_0), \end{aligned}$$

https://doi.org/10.1017/S0266466616000293 Published online by Cambridge University Press

$$\psi_{gi} = \frac{1}{\sqrt{h}}(\tau - I\{U_i \le 0\})\{W_i - \mathrm{E}(W|X = x_0)\}K\left(\frac{X_i - x_0}{h}\right) \times f_{UXWZ}^{-1}(0, X_i, W_i, Z_i) f_{XW}(X_i, W_i) f_Z(Z_i) f_X(x_0),$$

and

$$\psi_{g'i} = \frac{1}{\sqrt{h}}(\tau - I\{U_i \le 0\})\left(\frac{X_i - x_0}{h}\right) W_i K\left(\frac{X_i - x_0}{h}\right) f_{UXWZ}^{-1}(0, X_i, W_i, Z_i) \times f_{XW}(X_i, W_i) f_Z(Z_i) f_X(x_0).$$

COROLLARY 3.2. *Under the assumptions of Theorem 3.1,* $\left(\widehat{\theta}_n'(x_0), \widehat{g}_n'(x_0)\right)$ *is asymptotically independent of* $\widehat{g}_n(x_0)$.

Since our estimator is motivated by CH's LPR, it is not surprising that both share common theoretical properties. First, both estimators achieve the optimal speed of convergence, i.e., $\left(nh^3\right)^{-1/2}$ for the estimate of $\left(\theta'(\cdot), g'(\cdot)\right)$ and $(nh)^{-1/2}$ for $g(\cdot)$. Second, the leading bias term for both is $O_p\left(h^2\right)$. Third, both are oracle-efficient in the sense that their limiting distributions do not rely on the nuisance part $m(\cdot)$, that is, the estimators are expected to perform asymptotically the same as if $m(\cdot)$ were known. Fourth, $\left(\widehat{\theta}_n'(x_0), \widehat{g}_n'(x_0)\right)$ is shown to be asymptotically independent of $\widehat{g}_n(x_0)$. Therefore, they can be treated separately. The limiting distribution of (2.5) is summarized in the next corollary, because the varying coefficient model (1.4) is nested by (1.1) as an important special case and its asymptotic variance formula is used to generate the confidence interval in the empirical example.

COROLLARY 3.3. *Suppose Assumptions 2.1 and 3.1–3.5*(i) *hold. Moreover, assume* $\delta = O(n^{-\eta})$, *with*

$$\eta \in (1/(2\varrho + 2\min\{d_\theta, d_g\}), 1/(3 + 3p_W)),$$

and $nh^3 \to \infty$, $nh^5\delta^{2(p_W+1)} \to \infty$ *and* $\delta^\varrho / h^{5/2} \to 0$; *then* (i) *the estimator* (2.5) *has the limiting distribution:*

$$\begin{pmatrix} \sqrt{nh^3} & & \\ & \sqrt{nh} & \\ & & \sqrt{nh^3} \end{pmatrix} \left(\widehat{\Theta}_n - \Theta_0 - h^2 V_2^{-1} \begin{pmatrix} b_{\theta'} \\ b_g \\ b_{g'} \end{pmatrix} \right) \to^d \mathcal{N}\left(0, V_2^{-1}\Omega_2 V_2^{-1}\right),$$

where $\Omega_2 = \mathrm{E}\left(\psi_i \psi_i^T\right)$, *with* $\psi_i = \left(\psi_{\theta' i}, \psi_{gi}^T, \psi_{g'i}^T\right)^T$,

$$\psi_{\theta' i} = \frac{1}{\sqrt{h}}(\tau - I\{U_i \le 0\})\left(\frac{X_i - x_0}{h}\right) K\left(\frac{X_i - x_0}{h}\right) f_{UXW}^{-1}(0, X_i, W_i) \times f_{XW}(X_i, W_i) f_X(x_0),$$

$$\psi_{gi} = \frac{1}{\sqrt{h}}(\tau - I\{U_i \le 0\})\{W_i - \mathrm{E}(W|X = x_0)\}K\left(\frac{X_i - x_0}{h}\right) \times f_{UXW}^{-1}(0, X_i, W_i) f_{XW}(X_i, W_i) f_X(x_0),$$

https://doi.org/10.1017/S0266466616000293 Published online by Cambridge University Press

and

$$\psi_{g'i} = \frac{1}{\sqrt{h}}\left(\tau - I\{U_i \le 0\}\right)\left(\frac{X_i - x_0}{h}\right) W_i K\left(\frac{X_i - x_0}{h}\right) f_{UXW}^{-1}(0, X_i, W_i) \times f_{XW}(X_i, W_i)\, f_X(x_0).$$

(*ii*)

$$\sqrt{nh}\left(\widehat{g}_n(x_0) - g(x_0) - \frac{\kappa_2 g''(x_0)h^2}{2}\right) \to^d \mathcal{N}\left(0, \Sigma_g\right),$$

where

$$\Sigma_g = \frac{\tau(1-\tau)\kappa_2}{\left[f_{U|X}\left(0|x_0\right)\right]^2 f_X(x_0)} \times \left\{\mathrm{E}\left(WW^T|X = x_0\right) - \mathrm{E}\left(W|X = x_0\right)\mathrm{E}\left(W^T|X = x_0\right)\right\}^{-1}.$$

(*iii*) $\left(\widehat{\theta}_n'(x_0), \widehat{g}_n'(x_0)\right)$ *is asymptotically independent of* $\widehat{g}_n(x_0)$.

We now discuss bandwidth choice. Generally speaking, selecting optimal bandwidths in a two-step procedure is difficult. Since only h affects the convergence rate between the bandwidth parameters h and δ, the estimator is likely to be more sensitive to h than to δ. In practice, we suggest selecting δ by using a simple rule-of-thumb and then determining h based on a plug-in criterion. While selecting δ by using this rule-of-thumb is likely to result in undersmoothing, research indicates that a two-stage procedure with undersmoothing in the first stage still performs well in simulation studies (see Horowitz, 1992; Buchinsky and Hahn, 1998; Horowitz and Lee, 2005). For the first-stage local polynomial quantile regression, a simple, informal rule is used to select the bandwidth δ. Let $\delta = \widehat{A} \times n^{-\eta}$, where $\eta > 0$ is a positive number assumed to be consistent with the asymptotic order required by Assumption 3.5. To determine $\widehat{A}$, we use the rule-of-thumb bandwidth discussed by Fan and Gijbels (1996). Such a selection rule was also used by Lee (2003) and Chen and Khan (2000), among others. To determine h, researchers must consider whether the focus of interest centres on the function $g(\cdot)$ or on the derivatives $g'(\cdot)$ and $\theta'(\cdot)$. Below we provide a detailed selection procedure only for $g(\cdot)$. A similar procedure can be devised for $g'(\cdot)$ and $\theta'(\cdot)$. Let $\mathrm{Bias}_g(x)$ and $\mathrm{Var}_g(x)$ denote the bias and asymptotic variance of $\widehat{g}_n(x)$, respectively. Specifically, we choose for $g(\cdot)$ the h that minimizes

$$\widehat{\mathrm{MSE}}_g = \frac{1}{n}\sum_{i=1}^{n}\left\{\widehat{\mathrm{Bias}}_g^2(X_i) + \widehat{\mathrm{Var}}_g(X_i)\right\}. \tag{3.1}$$

We show below that (3.1) can be simplified markedly using the asymptotic distribution given by Theorem 3.1. From the specific form of V_2 in Assumption 2.1 and Lemma 4, we know the $(2,2)$-th element of V_2^{-1} is

$$V^{gg} = V_{gg}^{-1} = \left\{\left\{\mathrm{E}\left(WW^T|X = x_0\right) - \mathrm{E}\left(W|X = x_0\right)\mathrm{E}\left(W^T|X = x_0\right)\right\} f_X^2(x_0)\right\}^{-1}.$$

https://doi.org/10.1017/S0266466616000293 Published online by Cambridge University Press

Thus

$$\text{Bias}_g(x) = V_{gg}^{-1} b_g = \frac{\kappa_2 g''(x) h^2}{2}.$$

Moreover, let $\Omega_{gg} = \text{E}(\psi_{gi}\psi_{gi}^T)$, $\Omega_{\theta'\theta'} = \text{E}(\psi_{\theta' i}\psi_{\theta' i}^T)$, $\Omega_{\theta' g'} = \text{E}(\psi_{\theta' i}\psi_{g' i}^T)$. Then, from Corollary 3.2, we know that Ω_1 can be expressed as

$$\Omega_1 = \begin{pmatrix} \Omega_{\theta'\theta'} & 0 & \Omega_{\theta' g'}^T \\ 0 & \Omega_{gg} & 0 \\ \Omega_{\theta' g'} & 0 & \Omega_{g' g'} \end{pmatrix}.$$

Again, it follows from Lemma 4 that

$$\text{Var}_g(x) = (nh)^{-1} V^{gg}(x) \Omega_{gg}(x) V^{gg}(x).$$

We estimate $\text{Bias}_g(x)$ by

$$\widehat{\text{Bias}}_g(x) = \frac{\kappa_2 \widehat{g}_n''(x) h^2}{2},$$

where $\widehat{g}_n''(x)$ denotes an estimator for the second derivative of $g(\cdot)$ evaluated at $X = x$, which can be obtained by using the regression described in Remark 2.1. Finally, we estimate $\text{Var}_g(x)$ by

$$\widehat{\text{Var}}_g(x) = (nh)^{-1} \widehat{V}^{gg}(x) \widehat{\Omega}_{gg}(x) \widehat{V}^{gg}(x),$$

where

$$\widehat{\Omega}_{gg}(x) = \frac{1}{n} \sum_{i=1}^{n} \psi_{gi} \psi_{gi}^T,$$

$$\widehat{V}_{gg}(x) = \widehat{\varphi}(x) \widehat{f}_X^2(x),$$

$$\varphi(x) = \text{E}\big(W W^T \big| X = x\big) - \text{E}\big(W \big| X = x\big)\text{E}\big(W^T \big| X = x\big).$$

4. CENSORED QUANTILE REGRESSION

LPQR has the advantage that it can handle censored data. Applications require two forms of censoring, namely fixed and random censoring. The former often results from the fact that many observed economic variables are censored from above or below because of non-negativity constraints and/or top coding; in the latter case, the censoring variable is a nondegenerate random variable. To our best knowledge, the existing methods cited above are unsuitable for estimating the additive model (1.2) or the varying coefficient model (1.4) with censored dependent variables. Our proposed procedure is thus novel in this regard.

4.1. Fixed Censoring

Without loss of generality, we assume that the dependent variable Y in (1.1) is subject to the following two-sided fixed censoring:

https://doi.org/10.1017/S0266466616000293 Published online by Cambridge University Press

$$Y^* = \underline{c} I\{Y < \underline{c}\} + Y I\{\underline{c} \le Y \le \overline{c}\} + \overline{c} I\{Y > \overline{c}\}, \tag{4.1}$$

with $\underline{c} < \overline{c}$ the two fixed constants.[3] Under fixed censoring, only the censored dependent variable Y^* instead of Y is observed, thus $q_Y(\cdot)$. In other words, the conditional quantile of Y in the second stage is not identified straightforwardly as in the uncensored case. This difficulty can be overcome by invoking the well-known equivariance property of conditional quantiles, exploited by Powell (1984, 1986) to deal with censored linear quantile regression models. The property states that the τ-th conditional quantile of the observed response variable Y^* relates to that of the latent response variable Y through

$$q_{Y^*}(\cdot) = \min\{\max\{q_Y(\cdot), \underline{c}\}, \overline{c}\}. \tag{4.2}$$

(4.2) implies that $q_{Y^*}(\cdot)$ is as informative as $q_Y(\cdot)$ as long as $\underline{c} < q_{Y^*}(\cdot) < \overline{c}$ and in this case $q_Y(\cdot) = q_{Y^*}(\cdot)$. Based on this observation, the estimator (2.4) changes into

$$\widehat{\Theta}_n = S_{5n}^{-1} S_{6n}, \tag{4.3}$$

where

$$S_{5n} = \sum_{k=1}^{n} \sum_{1 \le i < j \le n} \mathbf{X}_{ij} \mathbf{X}_{ij}^T K_h(X_i) K_h(X_j) t\{\widehat{q}_{Y^*}(X_i, W_i, Z_k)\} t\{\widehat{q}_{Y^*}(X_j, W_j, Z_k)\},$$

$$S_{6n} = \sum_{k=1}^{n} \sum_{1 \le i < j \le n} \mathbf{X}_{ij} \widehat{\mathbf{Y}}_{ijk}^* K_h(X_i) K_h(X_j) t\{\widehat{q}_{Y^*}(X_i, W_i, Z_k)\} t\{\widehat{q}_{Y^*}(X_j, W_j, Z_k)\},$$

$$\widehat{\mathbf{Y}}_{ijk}^* = \widehat{q}_{Y^*}(X_i, W_i, Z_k) - \widehat{q}_{Y^*}(X_j, W_j, Z_k), \quad \text{and}$$

$$t(q) = I\{\underline{c} < q < \overline{c}\}.$$

Compared with (2.4), the estimator (4.3) incorporates the trimming device $t(\cdot)$, ensuring that the individual summands for which

$$q_{Y^*}(X_i, W_i, Z_k) \notin [\underline{c}, \overline{c}] \text{ or } q_{Y^*}(X_j, W_j, Z_k) \notin [\underline{c}, \overline{c}]$$

will not contribute to the second-stage weighted least squares. As expected, the estimator (4.3) has the same convergence rate as (2.4) and its limiting distribution can be derived in a similar manner as in Theorem 3.1.

4.2. Random Censoring

To illustrate this case, we focus on random right censoring. The observations consist of a random sample of (Y^*, D, X, W, Z), where D is a binary variable indicating whether the dependent variable is censored:

$$D = I\{Y < C\}, \quad Y^* = \min\{Y, C\},$$

where the latent dependent variable Y is generated by (1.1) and the random variable C is assumed to be independent of Y conditional on (X, W, Z). Let $S(\cdot|x, w, z)$ signify the survival function of C conditional on $(X, W, Z) = (x, w, z)$.

https://doi.org/10.1017/S0266466616000293 Published online by Cambridge University Press

In the presence of random censoring, the identification strategy introduced in Section 2 still works. One complication brought about by random censoring is that in the first-stage local polynomial fitting of conditional quantile functions, $q_{Y^*}(\cdot)$ is the τ-th quantile of Y if and only if $D = 1$. However, the naive modification of the local polynomial regression objective function by restricting the summation to be across those k's with $D_k = 1$ results in a biased estimator. We can tackle this problem in at least three possible ways. One is by replacing $\rho_\tau(\cdots)$ in the local polynomial regression objective function with its conditional expectation given (Y^*, X, W, Z, D) instead of (Y, X, W, Z) (Honore, Khan, and Powell, 2002). The second is to apply the redistribution-of-mass idea of Efron (1967); see also Portnoy (2003), Peng and Huang (2008), and Wang and Wang (2009). The third strategy, which we consider below, is based on the observation that $\mathrm{E}[D/S(Y^*|X, W, Z)] = 1$ (Bang and Tsiatis, 2002; Kong, Linton, and Xia, 2013). Adopting the same notation as in Section 2 leads to the following minimization problem:

$$\begin{aligned}
&\left(\widehat{b}_n^X, \widehat{b}_n^Z, \widehat{b}_n^W, \widehat{b}_n^{XW}\right) \\
&\quad = \arg \min_{(b^X, b^W, b^Z, b^{XW})} \sum_{k=1}^{n} \frac{D_k}{S(Y_k^*|X_k, W_k, Z_k)} \\
&\qquad \times I\{(X_k, W_k, Z_k) \in C(x, w, z, \delta)\}\rho_\tau \left\{ Y_k - \sum_{l=1}^{d_\theta} b_l^X (X_k - x)^l \right. \\
&\qquad \left. - \sum_{l=1}^{\#A(d_m)} b_l^Z (Z_k - z)^{\alpha_l} - \sum_{p=1}^{p_W} b_p^W (W_{pk} - w_p) - \sum_{p=1}^{p_W} \sum_{l=1}^{d_g} b_{pl}^{XW} (X_k - x)^l (W_{pk} - w_p) \right\}
\end{aligned}$$

in the first stage. Likewise, among all the $[d_\theta + p_W(d_g + 1) + \#A(d_m)]$ components of $\left(\widehat{b}_n^X, \widehat{b}_n^Z, \widehat{b}_n^W, \widehat{b}_n^{XW}\right)$, the first component of $\widehat{b}_n^Z$ estimates the conditional quantile function $q_Y(x, w, z)$. In practice, $S(\cdot|x, w, z)$ is unknown and should be replaced by its empirical counterpart $\widehat{S}_n(\cdot|x, w, z)$ through, for example, the local Kaplan–Meier estimator (Gonzalez-Manteiga and Cadarso-Suarez, 1994). Once these $\widehat{q}_Y(\cdot)$'s are appropriately estimated, the resulting LPQR is the same as in (2.4).

5. MONTE CARLO SIMULATION

In this section, the finite sample performance of LPQR is assessed by a small-scale simulation study. Without loss of generality, we focus on median regression only, namely $\tau = 0.5$. As theory (Theorem 3.1) predicts that the limiting distribution of LPQR does not rely on any specific functional form of the nuisance part $m(\cdot)$, we empirically evaluate this oracle property by comparing the feasible two-stage estimator with an infeasible oracle estimator in which the functional form $m(\cdot)$ is assumed to be known. We find that the performance of the feasible LPQR gradually closes to that of the oracle as the sample size grows. The data-generating

https://doi.org/10.1017/S0266466616000293 Published online by Cambridge University Press

process is specified as

$$Y = \underbrace{X + X^3/3 + 2\ln\left((X+1)^2+1\right)}_{\theta(X)} + \underbrace{\sin Z}_{m(Z)} + \underbrace{\left(X^2+1\right)}_{g(X)} W + U,$$

where $U = 0.5 \times \left(|X| + |Z|\right) \times \epsilon$, the components of (X, Z, W, ϵ) are mutually independent, with X and Z uniformly distributed on $[-3, 3]$ and W and ϵ distributed as $\mathcal{N}(0, 1)$. The function $\theta(x)$ is created so that

$$\theta'(x) = 1 + x^2 + 4(x+1)\left\{(x+1)^2+1\right\}^{-1}$$

displays complex local behaviour around $x = -1$ whereas $g(x) = x^2 + 1$ and $g'(x) = 2x$ have a globally simple structure. The sample size is $n \in \{500, 1000, 2000\}$.

To implement LPQR, we must choose $K(\cdot)$ as well as the bandwidths h and δ. The uniform kernel $K(s) = I\{-0.5 \le s \le 0.5\}$ is used throughout the simulation. Let $\delta = \widehat{A} \times n^{-1/9}$, where the rate $n^{-1/9}$ is consistent with the guidelines in Assumption 3.5 with $\varrho = 1$ and $p_W = p_Z = 1$. $q_Y(\cdot)$'s are fitted by a polynomial of order three. To obtain a rough range of $\widehat{A}$, we use the rule-of-thumb advocated by Fan and Gijbels (1996, p. 202), which amounts to setting

$$\widehat{A} = 2.813 \times \left\{ \frac{\tau(1-\tau)\left\{\widehat{f}_U\left(\widehat{q}_U(\tau)\right)\right\}^{-2} \int \omega_0(x)dx}{\sum_{i=1}^{n} \left(\widehat{\theta}_n^{(4)}(X_i) + \widehat{g}_n^{(4)}(X_i)^T W_i\right)^2 \omega_0(X_i)} \right\}^{1/9}. \tag{5.1}$$

In (5.1), $\omega_0(x) = I\{|x| \le 2.5\}$; $\widehat{f}_U(\cdot)$ is the kernel density estimate of the residuals of a global polynomial fit, and $\widehat{q}_U(\cdot)$ is the τ-th sample quantile of the residuals. In addition, $\theta^{(4)}(\cdot)$ and $g^{(4)}(\cdot)$ denote the fourth derivatives of $\theta(\cdot)$ and $g(\cdot)$, respectively. The global polynomial fit is obtained by carrying out a median regression of Y on a constant term, $X, \ldots, X^4$, $W, XW, \ldots, X^4W$, and $Z, \ldots, Z^4$. For our data generating process, preliminary simulations yield rule-of-thumb constants ranging from 0.98 to 1.41. Based on these results, $\widehat{A}$ varies from 0.5 to 2 in order to explore sensitivity to bandwidth choice. Given δ, we determine h by adopting the procedure outlined at the end of Section 3.

To understand the quality of the actual fit, Figure 1 shows the true functions and their derivatives as solid lines and displays the 25% and 75% empirical quantiles of 1,000 realizations of the estimators with $n = 1,000$. This figure shows that LPQR yields satisfactory results for $g(\cdot)$. Moreover, in spite of its significantly slower convergence rate $(nh^3)^{-1/2}$, LPQR yields reasonable results for $\theta'(\cdot)$ and $g'(\cdot)$, as the general shape of the functions is still captured well. To quantify these estimators, the average squared error (ASE) for $g(\cdot)$ and $\theta'(\cdot)$ are defined as

$$\mathrm{ASE}_g = \frac{1}{n}\sum_{i=1}^{n} \left(\widehat{g}_n(X_i) - g(X_i)\right)^2$$

https://doi.org/10.1017/S0266466616000293 Published online by Cambridge University Press

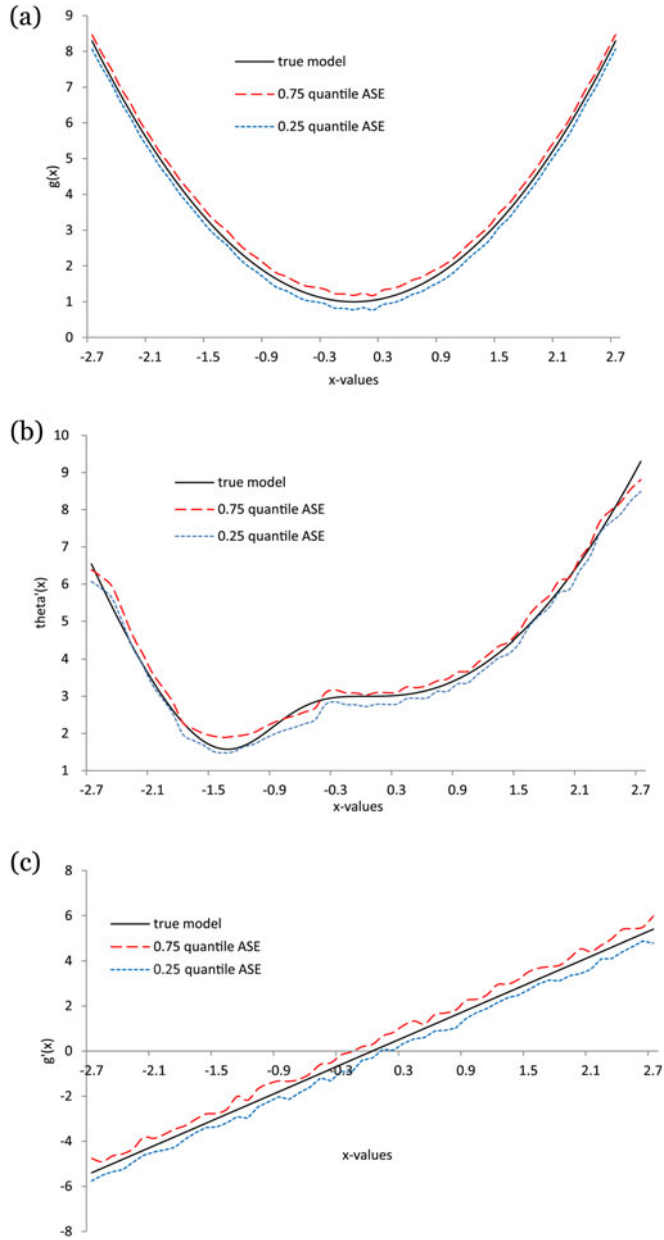


FIGURE 1. (a) Realizations of LPQR estimators for g(x) for $n = 1,000$, ordered according to quantiles of ASE. (b) Realizations of LPQR estimators for theta'(x) for $n = 1,000$, ordered according to quantiles of ASE. (c) Realizations of LPQR estimators for g'(x) for $n = 1,000$, ordered according to quantiles of ASE.

https://doi.org/10.1017/S0266466616000293 Published online by Cambridge University Press

and

$$\mathrm{ASE}_{\theta'} = \frac{1}{n}\sum_{i=1}^{n}\left(\widehat{\theta}'_n(X_i) - \theta'(X_i)\right)^2,$$

respectively.[4] To understand the oracle property predicted by Theorem 3.1, we also compute an infeasible oracle estimator (i.e., we compute LPQR by assuming that $m(\cdot)$ is known). Although fewer effective covariates may change the optimal selection of δ, thereby making a sequence of order $O\left(n^{-1/9}\right)$ too large for the current case, we still use the same bandwidth as before when computing the oracle estimator by way of a comparison.

Table 1 summarizes the empirical median and standard deviation (in parentheses) of 1,000 realizations of ASE for feasible LPQR and its corresponding infeasible oracle estimator with different h's. These results show that the LPQR estimators are slightly biased, but that this bias declines with the sample size. In particular, the estimate of g is accurate and provides acceptable precision even for $n = 500$. However, compared with $\widehat{g}_n$, $\widehat{\theta}'_n$ shows much larger bias and dispersion, because it has a slower convergence rate than $\widehat{g}_n$. Moreover, these estimators are somewhat insensitive to the choice of $\widehat{A}$ as long as it is not too small. The infeasible oracle

TABLE 1. Median and standard deviation (in parentheses) of 1,000 ASE values for median regression

	ASE_g		$\mathrm{ASE}_{\theta'}$	
$\widehat{A}$	Feasible	Oracle	Feasible	Oracle
		$n = 500$		
0.5	0.0539 (0.0693)	0.0398 (0.0511)	0.9399 (1.6305)	0.6926 (1.2022)
1.0	0.0533 (0.0681)	0.0393 (0.0502)	0.9242 (1.6177)	0.6814 (1.1827)
1.5	0.0546 (0.0658)	0.0402 (0.0485)	0.9638 (1.5892)	0.7106 (1.0717)
2.0	0.0562 (0.0650)	0.0414 (0.0479)	0.9875 (1.5542)	0.7280 (1.0459)
		$n = 1,000$		
0.5	0.0363 (0.0572)	0.0297 (0.0475)	0.6942 (1.3462)	0.5737 (1.0856)
1.0	0.0358 (0.0563)	0.0295 (0.0463)	0.6812 (1.3356)	0.5625 (1.0623)
1.5	0.0369 (0.0544)	0.0310 (0.0460)	0.7139 (1.3121)	0.5900 (0.9844)
2.0	0.0382 (0.0537)	0.0324 (0.0446)	0.7335 (1.2832)	0.6135 (0.9521)
		$n = 2,000$		
0.5	0.0195 (0.0465)	0.0177 (0.0414)	0.4643 (1.0945)	0.4134 (0.9746)
1.0	0.0191 (0.0457)	0.0171 (0.0407)	0.4538 (1.0859)	0.4041 (0.9459)
1.5	0.0202 (0.0442)	0.0181 (0.0394)	0.4804 (1.0667)	0.4277 (0.9498)
2.0	0.0211 (0.0436)	0.0189 (0.0388)	0.4963 (1.0433)	0.4419 (0.9290)

https://doi.org/10.1017/S0266466616000293 Published online by Cambridge University Press

estimator generally outperforms the feasible one in terms of both median ASE and standard deviation. When $n = 500$, the magnitude of ASE for feasible LPQR is generally 35–40% larger than that for the oracle. When $n = 1,000$, the efficiency loss ratio diminishes to about 22–25%. When $n = 2,000$, the ratio is only 10–12%. These results indicate that the performance of the feasible LPQR estimator becomes gradually closer to that of the oracle one as the sample size rises.

6. AN EMPIRICAL EXAMPLE

The empirical example presented in this section illustrates the benefits of the proposed estimator. Li, Huang, Li, and Fu (2002) used a sample of Chinese firms in the nonmetal mineral manufacturing industry to examine whether intermediate production and management expense affects output neutrally. The data are drawn from the Third Industrial Census of China conducted by the National Statistics Bureau of China in 1995. The sample consists of 1406 nonmetal mineral manufacturing firms (Code 31 in the survey). In the production function, the dependent variable Y is logged output (the logged value added in thousand RMB). The two inputs are the logged value of capital assets in thousand RMB (W_K) and logged average number of employees (W_L). The logged intermediate production and management expense (X) is included as a proxy for R&D to examine the possible non-neutral efficiency effect in production. The intermediate production and management expense consists of the R&D expense, expenses for upgrading existing equipment and training employees, and other expenses indirectly related to the production of output. Note that although R&D is only a part of the intermediate production and management expense, R&D itself is not available in this dataset. Summary statistics for these variables can be found in Li et al. (2002). The semiparametric model used by Li et al. (2002) is the following varying coefficient model:

$$Y = \theta(X) + W_K g_K(X) + W_L g_L(X) + U, \tag{6.1}$$

where the coefficients of W_K and W_L vary with X without specifying the functional forms $g_K(\cdot)$ and $g_L(\cdot)$. Based on a series of preliminary regressions and tests, Li et al. (2002) argued that (6.1) significantly outperforms a log-linear Cobb-Douglas model, a partially linear model, and any other benchmark parametric model when describing the relationship between output and the explanatory variables. This greater level of performance occurs because (6.1) allows the marginal productivity of labour and capital to depend on the firm's R&D value, while other models (e.g., the partially linear model) assume that R&D can only shift the level of the production frontier (intercept), thus placing the restriction that R&D affects the stochastic frontier neutrally. While the empirical results of Li et al. (2002) were based on analysing (6.1) under the mean restriction, we supplement their analysis by running a quantile regression. Compared with the mean regression in which the marginal productivity of labour and capital depends only on X, our quantile regression implicitly assumes that $g_K(\cdot)$ and $g_L(\cdot)$ not only depend on X

https://doi.org/10.1017/S0266466616000293 Published online by Cambridge University Press

but also vary with τ (ranking of unobserved factors that affect a firm's productivity such as entrepreneurial or managerial skills):

$$q_Y(\tau, X, W_K, W_L) = \theta(\tau, X) + W_K g_K(\tau, X) + W_L g_L(\tau, X). \quad \textbf{(6.2)}$$

We estimate $g_K(\cdot)$ and $g_L(\cdot)$ by using the proposed LPQR method. The kernel function and bandwidths are selected by following the same procedure as shown in Section 5. In particular, we use the uniform kernel $\delta = \widehat{A} \times n^{-1/9}$, where $\widehat{A}$ is determined by (5.1), and then solve h by minimizing (3.1). For robustness,

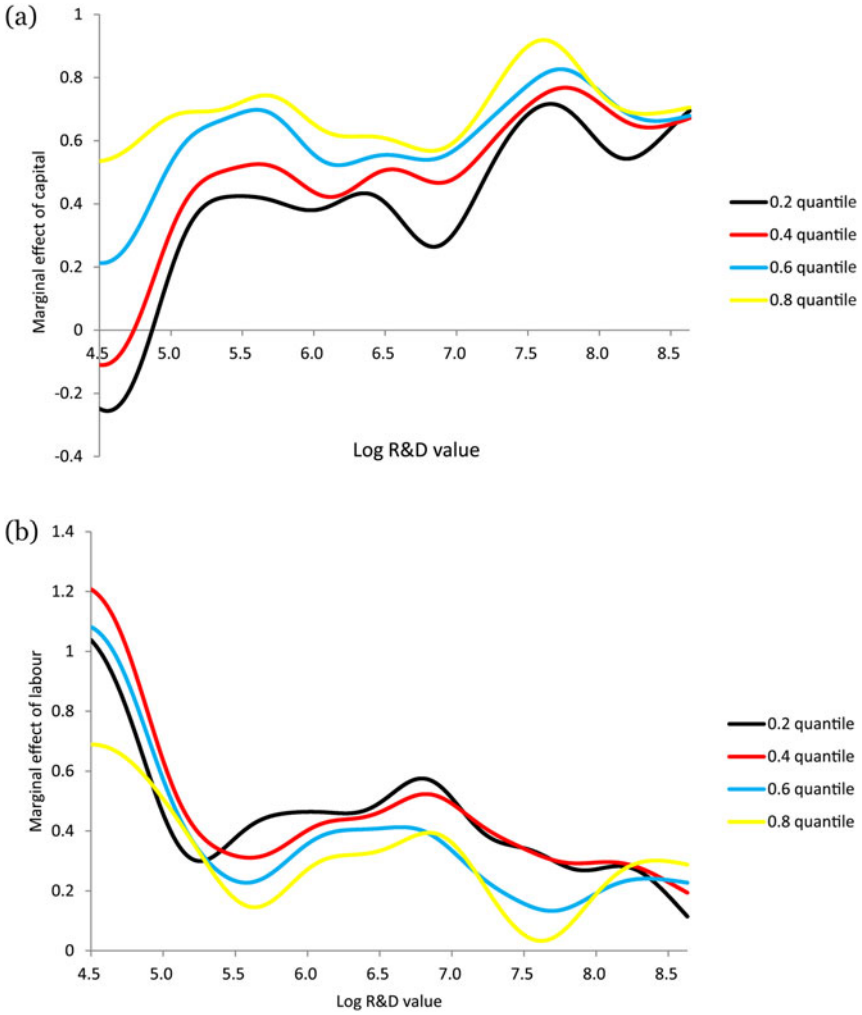


FIGURE 2. (a) Functional coefficient on capital. (b) Functional coefficient on labour.

https://doi.org/10.1017/S0266466616000293 Published online by Cambridge University Press

we try different bandwidths that are close to the one determined above. We find that different bandwidth choices do not change the overall shape of the functional coefficients or alter the fundamental empirical observations.

Figure 2 plots the functional coefficients $g_K(\cdot)$ and $g_L(\cdot)$ estimated at $\tau = 0.2$, 0.4, 0.6, and 0.8. For any given τ, the coefficient of capital shows an upward trend as X increases, while the coefficient of labour shows a downward trend. The former implies that higher intermediate expenses increase firm efficiency in utilizing capital; however, the latter is puzzling. Nonetheless, given that most sample firms are state-owned, the decreasing coefficient of labour indicates a surplus of workers in these firms, pointing to the so-called "concealed unemployment" effect, as the government did not allow these firms to lay off extra employees in order to avoid social unrest. In fact, the quantile regression reveals more than the above. Examining these coefficients from the perspective of different τ's shows that the coefficient of capital is increasing in τ for most values of X and that the coefficient of labour is decreasing in τ for each of the middle values of X. While X is viewed as an observed proxy for R&D-related expenses, τ represents the ranking of unobserved factors that may affect firm productivity but are indirectly observed inputs of the production process such as entrepreneurial or managerial skills. These quantile regression results thus show that unobserved factors affecting firm productivity such as entrepreneurial or managerial skills may have played a similar role as the observed intermediate production and management expense in the production process.

7. CONCLUSION

This paper introduces a new two-stage local partitioned quantile regression for a broad class of quantile regression models that nest partially linear, additive, and varying coefficient models. The new estimator is structurally simple and widely applicable to the general specification as well as its submodels. Moreover, it estimates not only the functional coefficient but also its derivatives up to any given order. The procedure readily extends to censored data, including fixed or random censoring. Asymptotic properties of the proposed estimator are formally established. Future research could use the presented results to test restrictions against more general alternatives. Moreover, since our estimators are robust against censoring, tests constructed from these estimators would be similarly robust.

NOTES

1. For the partially linear model $Y = X^T\beta_0 + m(Z) + U$ under $\mathrm{E}(U|X, Z) = 0$, Robinson (1988) proposed estimating β_0 by regressing $Y_i - \mathrm{E}(Y_i|Z_i)$ on $X_i - \mathrm{E}(X_i|Z_i)$.

2. In practice, one should first nonparametrically estimate the conditional mean functions in the regressand and regressors before implementing OLS. Thus, LPR is a two-stage estimation method.

3. This general design incorporates the case of no censoring by artificially setting $\overline{c}$ sufficiently large and $\underline{c}$ sufficiently small.

4. The results concerning $\mathrm{ASE}_{g'}$ are similar to $\mathrm{ASE}_{\theta'}$, thus not reported here.

https://doi.org/10.1017/S0266466616000293 Published online by Cambridge University Press

REFERENCES

Bang, H. & A.A. Tsiatis (2002) Median regression with censored cost data. *Biometrics* 58, 643–649.
Buchinsky, M. & J. Hahn (1998) An alternative estimator for the censored quantile regression model. *Econometrica* 66, 627–651.
Cai, Z. & Z. Xiao (2012) Semiparametric quantile regression estimation in dynamic models with partially varying coefficients. *Journal of Econometrics* 167, 413–425.
Chaudhuri, P. (1991a) Nonparametric estimates of regression quantiles and their local bahadur representation. *Annals of Statistics* 19, 760–777.
Chaudhuri, P. (1991b) Global nonparametric estimation of conditional quantiles and their derivatives. *Journal of Multivariate Analysis* 39, 246–269.
Chaudhuri, P., K. Doksum, & A. Samarov (1997) On average derivative quantile regression. *Annals of Statistics* 25, 715–744.
Chen S. & S. Khan (2000) Estimating censored regression models in the presence of nonparametric multiplicative heteroscedasticity. *Journal of Econometrics* 98, 283–316.
Chen S. & S. Khan (2001) Semiparametric estimation of a partially linear censored regression model. *Econometric Theory* 17, 567–590.
Chen, S. & S. Khan (2008) Semiparametric estimation of nonstationary censored panel data models with time varying factor loads. *Econometric Theory* 24, 1149–1173.
Christopeit, N. & S. Hoderlein (2006) Local partitioned regression. *Econometrica* 74, 787–817.
De Gooijer, J.G. & D. Zerom (2003) On additive conditional quantiles with high-dimensional covariates. *Journal of the American Statistical Association* 98, 135–146.
Deaton, A. & J. Muellbauer (1980) An almost ideal demand system. *American Economic Review* 70, 312–326.
Efron, B. (1967) The two-sample problem with censored data. In L. Le Cam & J. Neyman (eds.), *Proceedings of the Fifth Berkeley Symposium in Mathematical Statistics, IV*, pp. 831–853. Prentice Hall.
Fan, J. & I. Gijbels (1996) *Local Polynomial Modelling and its Applications*. Chapman and Hall.
Fan, J. & W. Zhang (1999) Statistical estimation in varying coefficient models. *The Annals of Statistics* 27, 1491–1518.
Gonzalez-Manteiga, W. & C. Cadarso-Suarez (1994) Asymptotic properties of a generalized Kaplan-Meier estimator with some applications. *Journal of Nonparametric Statistics* 4, 65–78.
Guerre, E. & C. Sabbah (2012) Uniform bias study and bahadur representation for local polynomial estimators for the conditional quantile function. *Econometric Theory* 28, 87–129.
Hardle, W. (1990) *Applied Nonparametric Regression*. Cambridge University Press.
Hoderlein, S. & E. Mammen (2007) Identification of marginal effects in nonseparable models without monotonicity. *Econometrica* 75, 1513–1518.
Honore, B., S. Khan, & J.L. Powell (2002) Quantile regression under random censoring. *Journal of Econometrics* 109, 212–221.
Horowitz, J.L. (1992) A smoothed maximum score estimator for the binary response model. *Econometrica* 60, 505–531.
Horowitz, J.L. (1993) Semiparametric estimation of a work-trip mode choice model. *Journal of Econometrics* 58, 49–70.
Horowitz, J.L. & S. Lee (2002) Semiparametric methods in applied econometrics: Do the models fit the data? *Statistical Modeling* 2, 3–22.
Horowitz, J.L. & S. Lee (2005) Nonparametric estimation of an additive quantile regression model. *Journal of the American Statistical Association* 100, 1238–1249.
Kim, M.O. (2007) Quantile regression with varying coefficients. *The Annals of Statistics* 35, 92–108.
Kong, E., O. Linton, & Y. Xia (2013) Global bahadur representation for nonparametric censored regression quantiles and its application. *Econometric Theory* 29, 941–968.
Lee, S. (2003) Efficient semiparametric estimation of partially linear quantile regression model. *Econometric Theory* 19, 1–31.

https://doi.org/10.1017/S0266466616000293 Published online by Cambridge University Press

Lee, S. (2007) Endogeneity in quantile regression models: A control function approach. *Journal of Econometrics* 141, 1131–1158.
Li, Q., C.J., Huang, D. Li, & T.T. Fu (2002) Semiparametric smooth coefficient models. *Journal of Business and Economic Statistics* 20, 412–422.
Peng, L. & Y. Huang (2008) Survival analysis with quantile regression models. *Journal of the American Statistical Association* 103, 637–649.
Portnoy, S. (2003) Censored regression quantiles. *Journal of the American Statistical Association* 98, 1001–1012.
Powell, J.L. (1984) Least absolute deviations estimation for the censored regression model. *Journal of Econometrics* 25, 303–325.
Powell, J.L. (1986) Censored regression quantiles. *Journal of Econometrics* 32, 143–155.
Powell, J.L., J.H. Stock, & T.M. Stoker (1989) Semiparametric estimation of index coefficients. *Econometrica* 57, 1404–1430.
Robinson, P. (1988) Root-n-consistent semiparametric regression. *Econometrica* 56, 931–954.
Sasaki, Y. (2015) What do quantile regressions identify for general structural functions. *Econometric Theory* 31, 1102–1116.
Wang, H. & L. Wang (2009) Locally weighted censored quantile regression. *Journal of the American Statistical Association* 104, 1117–1128.
Wang, H.J., Z. Zhu, & J. Zhou (2009) Quantile regression in partially linear varying coefficient models. *Annals of Statistics* 37, 3841–3866.
Yu, K. & Z. Lu (2004) Local linear additive quantile regression. *Scandinavian Journal of Statistics* 31, 333–346.

APPENDIX: Proofs

The following notations are maintained throughout the proof. $q_{iik} = q_Y(X_i, W_i, Z_k)$, $q_{jjk} = q_Y(X_j, W_j, Z_k)$, $\widehat{q}_{iik} = \widehat{q}_Y(X_i, W_i, Z_k)$, $\widehat{q}_{jjk} = \widehat{q}_Y(X_j, W_j, Z_k)$;

$$D_1 = \begin{pmatrix} \sqrt{nh^3} & 0 & 0 \\ 0 & \sqrt{nh} \times I_{p_W} & 0 \\ 0 & 0 & \sqrt{nh^3} \times I_{p_W} \end{pmatrix}, \quad D_2 = \begin{pmatrix} h & 0 & 0 \\ 0 & I_{p_W} & 0 \\ 0 & 0 & h \times I_{p_W} \end{pmatrix},$$

$$D_3 = \begin{pmatrix} \sqrt{nh} & 0 & 0 \\ 0 & \sqrt{nh} \times I_{p_W} & 0 \\ 0 & 0 & \sqrt{nh} \times I_{p_W} \end{pmatrix}.$$

LEMMA A.1.

$$\frac{1}{n^2(n-1)} D_2^{-1} S_{1n} D_2^{-1} \to^p V_2,$$

where V_2 is defined in Assumption 2.1.

Proof. Note that V_2 is a $(2p_W + 1) \times (2p_W + 1)$ matrix. We verify the result element-by-element. For the $(1, 1)$-th element, we need to show

$$\begin{aligned} \dagger_1 &= \frac{1}{n^2(n-1)h^2} \sum_{k=1}^{n} \sum_{1 \le i < j \le n} (X_i - X_j)^2 K_h(X_i) K_h(X_j) \\ &= \frac{1}{n(n-1)h^2} \sum_{1 \le i < j \le n} (X_i - X_j)^2 K_h(X_i) K_h(X_j) = \kappa_2 f_X^2(x_0) + o_p(1). \end{aligned}$$

https://doi.org/10.1017/S0266466616000293 Published online by Cambridge University Press

We prove this by verifying that $\mathrm{E}(\dagger_1) = \kappa_2 f_X^2(x_0) + o(1)$ and $\mathrm{Var}(\dagger_1) = o(1)$. Direct computation gives

$$\begin{aligned}
\mathrm{E}(\dagger_1) &= \frac{1}{2h^2}\int (x_1 - x_2)^2 K_h(x_1) K_h(x_2) f_X(x_1) f_X(x_2) dx_1 dx_2 \\
&= \frac{1}{2}\int (u-v)^2 K(u) K(v) f_X(x_0+uh) f_X(x_0+vh) du dv \\
&= \frac{1}{2}\left\{\int u^2 K(u) f_X(x_0+uh) du \int f_X(x_0+vh) K(v) dv \right. \\
&\quad + \int v^2 K(v) f_X(x_0+vh) dv \int f_X(x_0+uh) K(u) du \\
&\quad \left. + 2\int u K(u) f_X(x_0+uh) du \int v K(v) f_X(x_0+vh) dv \right\} \\
&= \frac{1}{2}\left\{\int u^2 K(u) f_X(x_0) du \int f_X(x_0) K(v) dv \right. \\
&\quad + \int v^2 K(v) f_X(x_0) dv \int f_X(x_0) K(u) du \\
&\quad \left. + 2\int u K(u) f_X(x_0) du \int v K(v) f_X(x_0) dv \right\} + O(h) \\
&= \kappa_2 f_X^2(x_0) + o(1).
\end{aligned}$$

As $\dagger_1$ is a second order U-statistic with the kernel

$$\xi(x_1, x_2) = \frac{1}{h^2}(x_1 - x_2)^2 K_h(x_1) K_h(x_2),$$

it follows from the Hoeffding theorem for U-statistics that the leading term of $\mathrm{Var}(\dagger_1)$ is given by

$$\frac{1}{n}\mathrm{Cov}(\xi(x_1, x_2), \xi(x_1, x_3)) = \frac{1}{n}\mathrm{E}(\xi(x_1, x_2)\xi(x_1, x_3)) - \frac{1}{n}\mathrm{E}^2(\xi(x_1, x_2)).$$

Notice that $\mathrm{E}(\xi(x_1, x_2)) = O(1)$, together with

$$\begin{aligned}
&\mathrm{E}(\xi(x_1, x_2)\xi(x_1, x_3)) \\
&\quad = \frac{1}{h^4}\int (x_1 - x_2)^2 (x_1 - x_3)^2 K_h^2(x_1) K_h(x_2) K_h(x_3) f_X(x_1) f_X(x_2) f_X(x_3) dx_1 dx_2 dx_3 \\
&\quad = \frac{1}{h}\int (u-v)^2 (u-s)^2 K^2(u) K(v) K(s) f_X(x_0+uh) f_X(x_0+vh) \\
&\qquad \times f_X(x_0+sh) du dv ds = O\left(\frac{1}{h}\right),
\end{aligned}$$

implies

$$\mathrm{Var}(\dagger_1) = O\left(\frac{1}{nh}\right) = o(1).$$

Applying a similar argument to the $(1,2)$-th element of V_2, we obtain

https://doi.org/10.1017/S0266466616000293 Published online by Cambridge University Press

$$\frac{1}{n^2(n-1)h}\sum_{k=1}^{n}\sum_{1\le i<j\le n}(X_i-X_j)(W_i-W_j)^T K_h(X_i)K_h(X_j)$$
$$=\underbrace{\mathrm{E}\left(\frac{1}{n^2(n-1)h}\sum_{k=1}^{n}\sum_{1\le i<j\le n}(X_i-X_j)(W_i-W_j)^T K_h(X_i)K_h(X_j)\right)}_{\dagger_2}+o_p(1),$$

where

$$\begin{aligned}\dagger_2 &= \frac{1}{2h}\int (x_1-x_2)(w_1-w_2)^T K_h(x_1)K_h(x_2)f_{XW}(x_1,w_1)f_{XW}(x_2,w_2)dx_1dx_2dw_1dw_2\\ &= \frac{1}{2}\int (u-v)(w_1-w_2)^T K(u)K(v)f_{XW}(x_0+uh,w_1)f_{XW}(x_0+vh,w_2)dudvdw_1dw_2\\ &= \frac{1}{2}\int\left\{\int uK(u)f_{XW}(x_0+uh,w_1)du\int K(v)f_{XW}(x_0+vh,w_2)dv\right\}(w_1-w_2)^T dw_1dw_2\\ &\quad -\frac{1}{2}\int\left\{\int vK(v)f_{XW}(x_0+vh,w_2)dv\int K(u)f_{XW}(x_0+uh,w_1)du\right\}(w_1-w_2)^T dw_1dw_2\\ &= O(h)=o(1).\end{aligned}$$

Addressing the remaining elements of V_2 in a similar way, we obtain

$$\begin{aligned}&\frac{1}{n^2(n-1)h^2}\sum_{k=1}^{n}\sum_{1\le i<j\le n}(X_i-X_j)\{(X_i-x_0)W_i-(X_j-x_0)W_j\}^T K_h(X_i)K_h(X_j)\\ &= \frac{1}{2h^2}\int (x_1-x_2)\{(x_1-x_0)w_1-(x_2-x_0)w_2\}^T K_h(x_1)K_h(x_2)f_{XW}(x_1,w_1)\\ &\quad\times f_{XW}(x_2,w_2)dx_1dx_2dw_1dw_2+o_p(1)\\ &= \frac{1}{2}\int (u-v)(uw_1-vw_2)^T K(u)K(v)f_{XW}(x_0+uh,w_1)\\ &\quad\times f_{XW}(x_0+vh,w_2)dudvdw_1dw_2+o_p(1)\\ &= \kappa_2\mathrm{E}(W|X=x_0)f_X^2(x_0)+o_p(1),\end{aligned}$$

$$\begin{aligned}&\frac{1}{n^2(n-1)}\sum_{k=1}^{n}\sum_{1\le i<j\le n}(W_i-W_j)(W_i-W_j)^T K_h(X_i)K_h(X_j)\\ &= \frac{1}{2}\int (w_1-w_2)(w_1-w_2)^T K_h(x_1)K_h(x_2)f_{XW}(x_1,w_1)\\ &\quad\times f_{XW}(x_2,w_2)dx_1dx_2dw_1dw_2+o_p(1)\\ &= \left\{\mathrm{E}\left(WW^T|X=x_0\right)-\mathrm{E}(W|X=x_0)\mathrm{E}\left(W^T|X=x_0\right)\right\}f_X^2(x_0)+o_p(1),\end{aligned}$$

$$\begin{aligned}&\frac{1}{n^2(n-1)h}\sum_{k=1}^{n}\sum_{1\le i<j\le n}(W_i-W_j)\{(X_i-x_0)W_i-(X_j-x_0)W_j\}^T K_h(X_i)K_h(X_j)\\ &= \frac{1}{2h}\int (w_1-w_2)\{(x_1-x_0)w_1-(x_2-x_0)w_2\}^T K_h(x_1)K_h(x_2)\end{aligned}$$

https://doi.org/10.1017/S0266466616000293 Published online by Cambridge University Press

$$\times f_{XW}(x_1, w_1) f_{XW}(x_2, w_2) dx_1 dx_2 dw_1 dw_2 + o_p(1)$$
$$= \frac{1}{2} \int (w_1 - w_2)(uw_1 - vw_2)^T K(u) K(v) f_{XW}(x_0 + uh, w_1)$$
$$\times f_{XW}(x_0 + vh, w_2) du dv dw_1 dw_2 + o_p(1)$$
$$= O(h) + o_p(1),$$

and

$$\frac{1}{n^2(n-1)h^2} \sum_{k=1}^{n} \sum_{1 \le i < j \le n} \{(X_i - x_0)W_i - (X_j - x_0)W_j\}\{(X_i - x_0)W_i - (X_j - x_0)W_j\}^T$$
$$\times K_h(X_i) K_h(X_j) = \kappa_2 \mathrm{E}\left(WW^T \middle| X = x_0\right) f_X^2(x_0) + o_p(1),$$

which completes the proof. ■

LEMMA A.2.

$$\ddagger_3 = \frac{1}{n^2(n-1)} D_3 D_2^{-1} \sum_{k=1}^{n} \sum_{1 \le i < j \le n} \mathbf{X}_{ij} \left(\mathbf{Y}_{ijk} - \mathbf{X}_{ij}^T \Theta_0\right) K_h(X_i) K_h(X_j)$$
$$= D_1 \left(h^2 \begin{pmatrix} b_{\theta'} \\ b_g \\ b_{g'} \end{pmatrix} + o_p(h^2) \right).$$

Proof. Decompose $\ddagger_3$ into $\dagger_3 + \dagger_4$, where

$$\dagger_3 = \frac{D_3 D_2^{-1}}{n^2(n-1)} \sum_{k=1}^{n} \sum_{1 \le i < j \le n} \mathbf{X}_{ij} \left\{\theta(X_i) - \theta(X_j) - \left(X_i - X_j\right)\theta'(x_0)\right\} K_h(X_i) K_h(X_j)$$

and

$$\dagger_4 = \frac{D_3 D_2^{-1}}{n^2(n-1)} \sum_{k=1}^{n} \sum_{1 \le i < j \le n} \mathbf{X}_{ij} K_h(X_i) K_h(X_j)$$
$$\times \left\{ \left(g(X_i) - g(x_0)\right)^T W_i - \left(g(X_j) - g(x_0)\right)^T W_j - g'(x_0)^T \left((X_i - x_0)W_i - (X_j - x_0)W_j\right) \right\}.$$

Partition $\dagger_3$ into

$$\left\{ \dagger_3^1, \left(\dagger_3^2\right)^T, \left(\dagger_3^3\right)^T \right\}^T,$$

where

$$\dagger_3^1 = \frac{\sqrt{nh^3}}{n^2(n-1)h^2} \sum_{k=1}^{n} \sum_{1 \le i < j \le n} (X_i - X_j)\{\theta(X_i) - \theta(X_j) - (X_i - X_j)\theta'(x_0)\} K_h(X_i) K_h(X_j),$$

$$\dagger_3^2 = \frac{\sqrt{nh}}{n^2(n-1)} \sum_{k=1}^{n} \sum_{1 \le i < j \le n} (W_i - W_j)\{\theta(X_i) - \theta(X_j) - (X_i - X_j)\theta'(x_0)\} K_h(X_i) K_h(X_j),$$

https://doi.org/10.1017/S0266466616000293 Published online by Cambridge University Press

and

$$\dagger_3^3 = \frac{\sqrt{nh^3}}{n^2(n-1)h^2}\sum_{k=1}^{n}\sum_{1\le i<j\le n}\left\{\left(X_i-x_0\right)W_i-\left(X_j-x_0\right)W_j\right\}$$
$$\times\left\{\theta(X_i)-\theta(X_j)-\left(X_i-X_j\right)\theta'(x_0)\right\}K_h(X_i)K_h(X_j).$$

Notice that

$$\theta(X_i)-\theta(X_j)-\theta'(x_0)(X_i-X_j)=\frac{1}{2}\theta''(x_0)(X_i-X_j)^2+\frac{1}{6}\theta'''(x_0)(X_i-X_j)^3$$
$$+\frac{1}{24}\theta''''(\overline{x}_0)(X_i-X_j)^4$$

with $\overline{x}_0$ lying between X_i and X_j. For $\dagger_3^1$, direct computation gives

$$\begin{aligned}
\mathrm{E}\left(\dagger_3^1\right) &= \frac{\sqrt{nh^3}}{2h^2}\int\left\{\frac{1}{2}\theta''(x_0)\,(x_1-x_2)^3+\frac{1}{6}\theta'''(x_0)\,(x_1-x_2)^4+\frac{1}{24}\theta''''(\overline{x}_0)\,(x_1-x_2)^5\right\}\\
&\quad\times K_h(x_1)K_h(x_2)f_X(x_1)f_X(x_2)dx_1dx_2\\
&= \frac{\sqrt{nh^3}}{2}\int\left\{\frac{h}{2}\theta''(x_0)\,(u-v)^3+\frac{h^2}{6}\theta'''(x_0)\,(u-v)^4+\frac{h^3}{24}\theta''''(\overline{x}_0)\,(u-v)^5\right\}\\
&\quad\times K(u)K(v)f_X(x_0+uh)f_X(x_0+vh)dudv\\
&= \frac{\sqrt{nh^3}}{2}\left\{\int\frac{h}{2}\theta''(x_0)\,(u-v)^3\,K(u)K(v)f_X(x_0+uh)f_X(x_0+vh)dudv\right.\\
&\quad\left.+\int\frac{h^2}{6}\theta'''(x_0)\,(u-v)^4\,K(u)K(v)f_X(x_0+uh)f_X(x_0+vh)dudv+O\left(h^3\right)\right\}\\
&= \sqrt{nh^3}\left\{h^2\left(f_X^2(x_0)\frac{(\kappa_4+3\kappa_2^2)\theta'''(x_0)}{6}\right)+o\left(h^2\right)\right\},
\end{aligned}$$

using

$$\int(u-v)^3K(u)K(v)dudv=0,\quad\int(u-v)^4K(u)K(v)dudv=2\kappa_4+6\kappa_2^2,$$

$$\int(u-v)^3uK(u)K(v)dudv=-\int(u-v)^3vK(u)K(v)dudv.$$

Then to conclude that

$$\dagger_3^1=\sqrt{nh^3}\left\{h^2f_X^2(x_0)\frac{(\kappa_4+3\kappa_2^2)\theta'''(x_0)}{6}+o_p(h^2)\right\},$$

it suffices to show that $\mathrm{Var}(\dagger_3^1)=o(nh^3h^4)=o(nh^7)$. As $\dagger_3^1$ is a second order U-statistic with the kernel

$$\xi(x_1,x_2)=\frac{\sqrt{nh^3}}{h^2}(x_1-x_2)\left\{\theta(x_1)-\theta(x_2)-(x_1-x_2)\theta'(x_0)\right\}K_h(x_1)K_h(x_2),$$

it follows from the Hoeffding theorem for U-statistics that the leading term of $\mathrm{Var}(\dagger_3^1)$ is given by

$$\frac{1}{n}\mathrm{Cov}\left(\xi(x_1,x_2),\xi(x_1,x_3)\right)=\frac{1}{n}\mathrm{E}\left(\xi(x_1,x_2)\xi(x_1,x_3)\right)-\frac{1}{n}\mathrm{E}^2\left(\xi(x_1,x_2)\right).$$

https://doi.org/10.1017/S0266466616000293 Published online by Cambridge University Press

As

$$\mathrm{E}\left(\xi(x_1,x_2)\right)=O\left(\sqrt{nh^3}h^2\right),$$

and

$$\begin{aligned}
&\mathrm{E}\left(\xi(x_1,x_2)\xi(x_1,x_3)\right)\\
&\quad=\frac{nh^3}{h^4}\int(x_1-x_2)(x_1-x_3)\left\{\theta(x_1)-\theta(x_2)-(x_1-x_2)\theta'(x_0)\right\}\\
&\qquad\times\left\{\theta(x_1)-\theta(x_3)-(x_1-x_3)\theta'(x_0)\right\}K_h^2(x_1)K_h(x_2)K_h(x_3)\\
&\qquad\times f_X(x_1)f_X(x_2)f_X(x_3)dx_1dx_2dx_3\\
&\quad=nh^4\int(u-v)(u-s)\left\{\frac{1}{2}\theta''(x_0)(u-v)^2+\frac{h}{6}\theta'''(\overline{x}_0)(u-v)^3\right\}\\
&\qquad\times\left\{\frac{1}{2}\theta''(x_0)(u-s)^2+\frac{h}{6}\theta'''(\overline{x}_0)(u-s)^3\right\}K^2(u)K(v)K(s)\\
&\qquad\times f_X(x_0+uh)f_X(x_0+vh)f_X(x_0+sh)dudvds\\
&\quad=O\left(nh^4\right),
\end{aligned}$$

we have

$$\mathrm{Var}(\dagger_3^1)=O\left(h^7\right)+O\left(h^4\right)=O\left(h^4\right)=o\left(nh^7\right)$$

as $nh^3\to\infty$. Similarly, for $\dagger_3^2$, direct computation gives

$$\begin{aligned}
\mathrm{E}\left(\dagger_3^2\right)&=\frac{\sqrt{nh}}{2}\int(w_1-w_2)\left\{\frac{1}{2}\theta''(x_0)(x_1-x_2)^2+\frac{1}{6}\theta'''(\overline{x}_0)(x_1-x_2)^3\right\}\\
&\qquad\times K_h(x_1)K_h(x_2)f_{XW}(x_1,w_1)f_{XW}(x_2,w_2)dx_1dx_2dw_1dw_2\\
&=\frac{\sqrt{nh}}{2}\int(w_1-w_2)\left\{\frac{h^2}{2}\theta''(x_0)(u-v)^2+\frac{h^3}{6}\theta'''(\overline{x}_0)(u-v)^3\right\}K(u)K(v)\\
&\qquad\times f_{XW}(x_0+uh,w_1)f_{XW}(x_0+vh,w_2)dudvdw_1dw_2=\sqrt{nh}\cdot o\left(h^2\right)
\end{aligned}$$

using

$$\begin{aligned}
&\int w_1(u-v)^2K(u)K(v)f_{XW}(x_0,w_1)f_{XW}(x_0,w_2)dudvdw_1dw_2\\
&\quad=\int w_2(u-v)^2K(u)K(v)f_{XW}(x_0,w_1)f_{XW}(x_0,w_2)dudvdw_1dw_2.
\end{aligned}$$

Then to conclude that

$$\dagger_3^2=\sqrt{nh}\cdot o_p\left(h^2\right),$$

it suffices to show that $\mathrm{Var}(\dagger_3^2)=o(nhh^4)=o(nh^5)$. As $\dagger_3^2$ is a second order U-statistic with the kernel

$$\xi(x_1,x_2)=\sqrt{nh}(w_1-w_2)\{\theta(x_1)-\theta(x_2)-(x_1-x_2)\theta'(x_0)\}K_h(x_1)K_h(x_2),$$

https://doi.org/10.1017/S0266466616000293 Published online by Cambridge University Press

it follows from the Hoeffding theorem for U-statistics that the leading term of $\mathrm{Var}(\dagger_3^2)$ is given by

$$\frac{1}{n}\mathrm{Cov}\left(\xi\left(x_1,x_2\right),\xi\left(x_1,x_3\right)\right)=\frac{1}{n}\mathrm{E}\left(\xi(x_1,x_2)\xi(x_1,x_3)\right)-\frac{1}{n}\mathrm{E}^2\left(\xi\left(x_1,x_2\right)\right).$$

As

$$\mathrm{E}\left(\xi(x_1,x_2)\right)=O\left(\sqrt{nh}h^3\right),$$

and

$$\begin{aligned}
&\mathrm{E}\left(\xi(x_1,x_2)\xi(x_1,x_3)\right)\\
&\quad=nh\int\left(w_1-w_2\right)\left\{\frac{1}{2}\theta''(x_0)\left(x_1-x_2\right)^2+\frac{1}{6}\theta'''(\overline{x}_0)\left(x_1-x_2\right)^3\right\}\\
&\qquad\times\left\{\frac{1}{2}\theta''(x_0)\left(x_1-x_3\right)^2+\frac{1}{6}\theta'''(\overline{x}_0)\left(x_1-x_3\right)^3\right\}\left(w_1-w_3\right)K_h^2(x_1)K_h(x_2)K_h(x_3)\\
&\qquad\times f_{XW}\left(x_1,w_1\right)f_{XW}\left(x_2,w_2\right)f_{XW}\left(x_3,w_3\right)dx_1dx_2dx_3dw_1dw_2dw_3\\
&\quad=nh^4\int\left(w_1-w_2\right)\left\{\frac{1}{2}\theta''(x_0)\left(u-v\right)^2+\frac{h}{6}\theta'''(\overline{x}_0)\left(u-v\right)^3\right\}\left(w_1-w_3\right)\\
&\qquad\times\left\{\frac{1}{2}\theta''(x_0)\left(u-s\right)^2+\frac{h}{6}\theta'''(\overline{x}_0)\left(u-s\right)^3\right\}K^2(u)K(v)K(s)f_{XW}\left(x_0+uh,w_1\right)\\
&\qquad\times f_{XW}\left(x_0+vh,w_2\right)f_{XW}\left(x_0+sh,w_3\right)dudvdsdw_1dw_2dw_3=O(nh^4),
\end{aligned}$$

we have

$$\mathrm{Var}(\dagger_3^2)=O(h^7)+O(h^4)=O(h^4)=o(nh^5)$$

as $nh\to\infty$. For $\dagger_3^3$,

$$\begin{aligned}
\mathrm{E}(\dagger_3^3)&=\frac{\sqrt{nh^3}}{2h^2}\int\left\{\frac{1}{2}\theta''(x_0)\left(x_1-x_2\right)^2+\frac{1}{6}\theta'''(x_0)\left(x_1-x_2\right)^3+\frac{1}{24}\theta''''(\overline{x}_0)\left(x_1-x_2\right)^4\right\}\\
&\qquad\times\left\{\left(x_1-x_0\right)w_1-\left(x_2-x_0\right)w_2\right\}K_h(x_1)K_h(x_2)f_{XW}\left(x_1,w_1\right)\\
&\qquad\times f_{XW}\left(x_2,w_2\right)dx_1dx_2dw_1dw_2\\
&=\frac{\sqrt{nh^3}}{2}\int\left\{\frac{h}{2}\theta''(x_0)\left(u-v\right)^2\left(uw_1-vw_2\right)K(u)K(v)f_{XW}\left(x_0+uh,w_1\right)\right.\\
&\qquad\times f_{XW}\left(x_0+vh,w_2\right)dudvdw_1dw_2\\
&\qquad+\int\frac{h^2}{6}\theta'''(x_0)\left(u-v\right)^3\left(uw_1-vw_2\right)K(u)K(v)f_{XW}\left(x_0+uh,w_1\right)\\
&\qquad\left.\times f_{XW}\left(x_0+vh,w_2\right)dudvdw_1dw_2+O(h^3)\right\}\\
&=\sqrt{nh^3}\left\{h^2\mathrm{E}\left(W\middle|X=x_0\right)f_X^2(x_0)\frac{(\kappa_4+3\kappa_2^2)\theta'''(x_0)}{6}+o(h^2)\right\}.
\end{aligned}$$

Applying similar arguments, we can show

$$\dagger_3^3=\sqrt{nh^3}\left\{h^2\mathrm{E}\left(W\middle|X=x_0\right)f_X^2(x_0)\frac{(\kappa_4+3\kappa_2^2)\theta'''(x_0)}{6}+o_p(h^2)\right\}.$$

https://doi.org/10.1017/S0266466616000293 Published online by Cambridge University Press

Similarly, we partition $\dagger_4$ into

$$\left\{\dagger_4^1, (\dagger_4^2)^T, (\dagger_4^3)^T\right\}^T,$$

where

$$\dagger_4^1 = \frac{\sqrt{nh^3}}{n^2(n-1)h^2}\sum_{k=1}^{n}\sum_{1\le i<j\le n}(X_i - X_j)\big\{(g(X_i)-g(x_0))^T W_i - (g(X_j)-g(x_0))^T W_j$$
$$- g'(x_0)^T((X_i - x_0)W_i - (X_j - x_0)W_j)\big\}K_h(X_i)K_h(X_j),$$

$$\dagger_4^2 = \frac{\sqrt{nh}}{n^2(n-1)}\sum_{k=1}^{n}\sum_{1\le i<j\le n}(W_i - W_j)\big\{(g(X_i)-g(x_0))^T W_i - (g(X_j)-g(x_0))^T W_j$$
$$- g'(x_0)^T((X_i - x_0)W_i - (X_j - x_0)W_j)\big\}K_h(X_i)K_h(X_j),$$

and

$$\dagger_4^3 = \frac{\sqrt{nh^3}}{n^2(n-1)h^2}\sum_{k=1}^{n}\sum_{1\le i<j\le n}\big\{(X_i - x_0)W_i - (X_j - x_0)W_j\big\}K_h(X_i)K_h(X_j)$$
$$\times\big\{(g(X_i)-g(x_0))^T W_i - (g(X_j)-g(x_0))^T W_j - g'(x_0)^T((X_i - x_0)W_i - (X_j - x_0)W_j)\big\}.$$

Notice that

$$(g(X_i)-g(x_0))^T W_i - (g(X_j)-g(x_0))^T W_j - g'(x_0)^T\{(X_i - x_0)W_i - (X_j - x_0)W_j\}$$
$$= \left\{\frac{1}{2}g''(x_0)(X_i - x_0)^2 + \frac{1}{6}g'''(x_0)(X_i - x_0)^3 + \frac{1}{24}g''''(\overline{x}_0)(X_i - x_0)^4\right\}^T W_i$$
$$- \left\{\frac{1}{2}g''(x_0)(X_j - x_0)^2 + \frac{1}{6}g'''(x_0)(X_j - x_0)^3 + \frac{1}{24}g''''(\overline{x}_0)(X_j - x_0)^4\right\}^T W_j,$$

with $\overline{x}_0$ lying between X_i and X_j. Then the desired results follow from applying the same arguments as we address $\dagger_3^1$, $\dagger_3^2$, $\dagger_3^2$, and

$$\mathrm{E}(\dagger_4^1) = \frac{\sqrt{nh^3}}{2h^2}\int (x_1 - x_2)\left\{\left(\frac{1}{2}g''(x_0)(x_1 - x_0)^2 + \frac{1}{6}g'''(x_0)(x_1 - x_0)^3\right.\right.$$
$$\left. + \frac{1}{24}g''''(\overline{x}_0)(x_1 - x_0)^4\right)^T w_1$$
$$- \left(\frac{1}{2}g''(x_0)(x_2 - x_0)^2 + \frac{1}{6}g'''(x_0)(x_2 - x_0)^3 + \frac{1}{24}g''''(\overline{x}_0)(x_2 - x_0)^4\right)^T w_2\Big\}$$
$$\times K_h(x_1)K_h(x_2)f_{XW}(x_1, w_1)\, f_{XW}(x_2, w_2)\, dx_1 dx_2 dw_1 dw_2$$
$$= \frac{\sqrt{nh^3}}{2}\int (u - v)\, K(u)K(v)f_{XW}(x_0 + uh, w_1)$$
$$\times f_{XW}(x_0 + vh, w_2)\, du dv dw_1 dw_2\left\{\left(\frac{h}{2}g''(x_0)u^2 + \frac{h^2}{6}g'''(x_0)u^3\right)^T\right.$$
$$\left.\times w_1 - \left(\frac{h}{2}g''(x_0)v^2 + \frac{h^2}{6}g'''(x_0)v^3\right)^T w_2 + O(h^3)\right\}$$

https://doi.org/10.1017/S0266466616000293 Published online by Cambridge University Press

$$= \sqrt{nh^3}\Big\{h^2\Big(\mathrm{E}(W^T|X=x_0)f_X^2(x_0)\frac{\kappa_4 g'''(x_0)}{6} + \int w^T \frac{\partial f_{XW}(x_0,w)}{\partial X}dw \cdot f_X(x_0)\cdot\frac{\kappa_4 g''(x_0)}{2} - \frac{\partial f_X(x_0)}{\partial X}\cdot\int w^T f_{XW}(x_0,w)dw\cdot\frac{\kappa_2^2 g''(x_0)}{2}\Big) + o(h^2)\Big\},$$

$$\mathrm{E}(\dagger_4^2) = \frac{\sqrt{nh}}{2}\int (w_1-w_2) f_{XW}(x_1,w_1) f_{XW}(x_2,w_2)\Big\{\Big(\frac{1}{2}g''(x_0)(x_1-x_0)^2 + \frac{1}{6}g'''(\overline{x}_0)(x_1-x_0)^3\Big)^T w_1 - \Big(\frac{1}{2}g''(x_0)(x_2-x_0)^2 + \frac{1}{6}g'''(\overline{x}_0)(x_2-x_0)^3\Big)^T w_2\Big\} \times K_h(x_1)K_h(x_2)dx_1dx_2dw_1dw_2$$

$$= \sqrt{nh}\Big\{h^2\Big\{\mathrm{E}(WW^T|X=x_0) - \mathrm{E}(W|X=x_0)\mathrm{E}(W^T|X=x_0)\Big\} \times f_X^2(x_0)\frac{\kappa_2 g''(x_0)}{2} + o(h^2)\Big\},$$

and

$$\mathrm{E}(\dagger_4^3) = \frac{\sqrt{nh^3}}{2h^2}\int \{(X_i-x_0)W_i - (X_j-x_0)W_j\} K_h(x_1)K_h(x_2) f_{XW}(x_1,w_1) f_{XW}(x_2,w_2) \times \Big\{\Big(\frac{1}{2}g''(x_0)(x_1-x_0)^2 + \frac{1}{6}g'''(x_0)(x_1-x_0)^3 + \frac{1}{24}g''''(\overline{x}_0)(x_1-x_0)^4\Big)^T w_1 - \Big(\frac{1}{2}g''(x_0)(x_2-x_0)^2 + \frac{1}{6}g'''(x_0)(x_2-x_0)^3 + \frac{1}{24}g''''(\overline{x}_0)(x_2-x_0)^4\Big)^T w_2\Big\}dx_1dx_2dw_1dw_2$$

$$= \frac{\sqrt{nh^3}}{2}\int (uw_1 - vw_2) K(u)K(v) f_{XW}(x_0+uh,w_1) f_{XW}(x_0+vh,w_2)dudvdw_1dw_2 \times \Big\{\Big(\frac{h}{2}g''(x_0)u^2 + \frac{h^2}{6}g'''(x_0)u^3\Big)^T w_1 - \Big(\frac{h}{2}g''(x_0)v^2 + \frac{h^2}{6}g'''(x_0)v^3\Big)^T w_2 + O(h^3)\Big\}$$

$$= \sqrt{nh^3}\Big\{h^2\Big(\mathrm{E}(WW^T|X=x_0)f_X^2(x_0)\frac{\kappa_4 g'''(x_0)}{6} + \int ww^T\frac{\partial f_{XW}(x_0,w)}{\partial X}dw\cdot f_X(x_0)\frac{\kappa_4 g''(x_0)}{2} - \int w\frac{\partial f_{XW}(x_0,w)}{\partial X}dw\cdot\int w^T f_{XW}(x_0,w)dw\cdot\frac{\kappa_2^2 g''(x_0)}{2}\Big) + o(h^2)\Big\}. \quad \blacksquare$$

LEMMA A.3.

$$2\ddagger_4 = \frac{2D_3D_2^{-1}}{n^2(n-1)}\sum_{k=1}^n \sum_{1\le i<j\le n} \mathbf{X}_{ij}\,(\widehat{q}_{iik} - q_{iik})\,K_h(X_i)K_h(X_j) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \psi_i + o_p(1).$$

Proof. To obtain the asymptotic distribution, we will plug in a Bahadur type expansion of $\widehat{q}_{iik} - q_{iik}$. Fortunately this type of expansion has been already well established in Chaudhuri (1991a, 1991b) and Chaudhuri et al. (1997). According to Chaudhuri (1991a), $\widehat{q}_{iik} - q_{iik}$ permits the following asymptotic linear representation:

$$\widehat{q}_{iik} - q_{iik} = \#C^{-1}(X_i,W_i,Z_k,\delta)e_1^T G^{-1}(X_i,W_i,Z_k)\sum_{l\ne i,k} L(\delta, X_l-X_i, W_l-W_i, Z_l-Z_k) \times (\tau - I\{Y_l \le q_Y(X_l,W_l,Z_l)\})\,I\{(X_l,W_l,Z_l)\in C(X_i,W_i,Z_k,\delta)\} + R_n(X_i,W_i,Z_k),$$

https://doi.org/10.1017/S0266466616000293 Published online by Cambridge University Press

where the remainder term

$$\max_{X_i\in\mathcal{X},W_i\in\mathcal{W},Z_k\in\mathcal{Z}}|R_n(X_i,W_i,Z_k)|=o(n^{-1/2})$$

by Assumption 3.5(ii) and Lemma 4.1 in Chaudhuri et al. (1997), $\#C(x,w,z,\delta)$ is the number of elements in $C(x,w,z,\delta)$, $L(\delta,X_l-x,W_l-w,Z_l-z)$ denotes the $[d_\theta+\#A(d_m)+p_W(d_g+1)t]$-dimensional vector

$$L(\delta,X_l-x,W_l-w,Z_l-z)=\left\{1,\frac{X_l-x}{\delta},\ldots,\frac{(X_l-x)^{d_\theta}}{\delta^{d_\theta}},\frac{(W_l-w)^T}{\delta},\frac{(W_l-w)^T(X_l-x)}{\delta^2},\ldots,\right.$$
$$\left.\frac{(W_l-w)^T(X_l-x)^{d_g}}{\delta^{d_g+1}},\left\{\delta^{-|\mathbf{a}|}(Z_l-z)^{\mathbf{a}},1\le|\mathbf{a}|\le d_m\right\}^T\right\}^T,$$

with $|\mathbf{a}|=\sum_{l=1}^{p_Z}a_l$, e_1 is a $[d_\theta+\#A(d_m)+p_W(d_g+1)]$-dimensional column vector with the first component being one and the remaining components being zero; $G(x,w,z)$ is the density weighted conditional expectation of the outer product of $L(\delta,X_l-x,W_l-w,Z_l-z)$ given that $(X_l,W_l,Z_l)\in C(x,w,z,\delta)$. Let $N(x,w,z)=(1/n)\mathrm{E}[\#C(x,w,z,\delta)]$. Plugging the representation for $\widehat{q}_{iik}-q_{iik}$ into $\ddagger_4$ gives

$$\begin{aligned}2\ddagger_4=&\frac{D_3D_2^{-1}}{n(n-1)(n-2)(n-3)}\sum_{i\neq j\neq k\neq l}\mathbf{X}_{ij}K_h(X_i)K_h\left(X_j\right)N^{-1}(X_i,W_i,Z_k)e_1^TG^{-1}(X_i,W_i,Z_k)\\&\times L(\delta,X_l-X_i,W_l-W_i,Z_l-Z_k)(\tau-I\{Y_l\le q_Y(X_l,W_l,Z_l)\})\\&\times I\{(X_l,W_l,Z_l)\in C(X_i,W_i,Z_k,\delta)\}+o_p(1).\end{aligned}$$

Partition $2\ddagger_4$ into

$$2\ddagger_4=\left\{2\ddagger_4^1,2\left(\ddagger_4^2\right)^T,2\left(\ddagger_4^3\right)^T\right\}^T,$$

where

$$\begin{aligned}2\ddagger_4^1=&\frac{\sqrt{h}}{\sqrt{n}(n-1)(n-2)(n-3)}\sum_{i\neq j\neq k\neq l}\left(\frac{X_i-X_j}{h}\right)K_h(X_i)K_h(X_j)N^{-1}(X_i,W_i,Z_k)\\&\times e_1^TG^{-1}(X_i,W_i,Z_k)L(\delta,X_l-X_i,W_l-W_i,Z_l-Z_k)(\tau-I\{Y_l\le q_Y(X_l,W_l,Z_l)\})\\&\times I\{(X_l,W_l,Z_l)\in C(X_i,W_i,Z_k,\delta)\}+o_p(1),\end{aligned}$$

$$\begin{aligned}2\ddagger_4^2=&\frac{\sqrt{h}}{\sqrt{n}(n-1)(n-2)(n-3)}\sum_{i\neq j\neq k\neq l}(W_i-W_j)K_h(X_i)K_h(X_j)N^{-1}(X_i,W_i,Z_k)\\&\times e_1^TG^{-1}(X_i,W_i,Z_k)L(\delta,X_l-X_i,W_l-W_i,Z_l-Z_k)(\tau-I\{Y_l\le q_Y(X_l,W_l,Z_l)\})\\&\times I\{(X_l,W_l,Z_l)\in C(X_i,W_i,Z_k,\delta)\}+o_p(1),\end{aligned}$$

and

$$\begin{aligned}2\ddagger_4^3=&\frac{\sqrt{h}}{\sqrt{n}(n-1)(n-2)(n-3)}\sum_{i\neq j\neq k\neq l}\left\{\left(\frac{X_i-x_0}{h}\right)W_i-\left(\frac{X_j-x_0}{h}\right)W_j\right\}\\&\times K_h(X_i)K_h(X_j)N^{-1}(X_i,W_i,Z_k)e_1^TG^{-1}(X_i,W_i,Z_k)\\&\times L(\delta,X_l-X_i,W_l-W_i,Z_l-Z_k)(\tau-I\{Y_l\le q_Y(X_l,W_l,Z_l)\})\\&\times I\{(X_l,W_l,Z_l)\in C(X_i,W_i,Z_k,\delta)\}+o_p(1).\end{aligned}$$

https://doi.org/10.1017/S0266466616000293 Published online by Cambridge University Press

Our next step is applying the projection theorem for U-statistic to

$$2\ddagger_4^1 = \frac{1}{\sqrt{n}(n-1)(n-2)(n-3)} \sum_{i\neq j\neq k\neq l} \mathcal{F}_1\left(\xi_i, \xi_j, \xi_k, \xi_l\right) + o_p(1),$$

where $\xi_i = (Y_i, X_i, W_i, Z_i)$ and

$$\begin{aligned}
&\mathcal{F}_1(\xi_i, \xi_j, \xi_k, \xi_l) \\
&\quad = h^{-3/2} I\left\{(X_l, W_l, Z_l) \in C(X_i, W_i, Z_k, \delta)\right\} \left(\frac{X_i - X_j}{h}\right) \\
&\qquad \times K\left(\frac{X_i - x_0}{h}\right) K\left(\frac{X_j - x_0}{h}\right) N^{-1}(X_i, W_i, Z_k) \\
&\qquad \times e_1^T G^{-1}(X_i, W_i, Z_k) L(\delta, X_l - X_i, W_l - W_i, Z_l - Z_k)\left(\tau - I\{Y_l \leq q_Y(X_l, W_l, Z_l)\}\right).
\end{aligned}$$

According to Powell, Stock, and Stoker (1989), the condition sufficient for the projection theorem is

$$\mathrm{E}\left|\mathcal{F}_1(\xi_i, \xi_j, \xi_k, \xi_l)\right|^2 = o(n). \tag{A.1}$$

Noting that the smallest eigenvalue of $G(X_i, W_i, Z_k)$ is bounded away from zero for all n, and that $I\left\{(X_l, W_l, Z_l) \in C(X_i, W_i, Z_k, \delta)\right\} L(\delta, X_l - X_i, W_l - W_i, Z_l - Z_k)$ is bounded above by one, thus the only term in $\mathcal{F}_1(\cdot, \cdot, \cdot, \cdot)$ that becomes unbounded as the sample size increases is $N^{-1}(X_i, W_i, Z_k)$. Noting that $N(X_i, W_i, Z_k)$ can be expressed as

$$\delta^{p_W + p_Z + 1} \int_{[-1,1]^{p_W}} \int_{[-1,1]^{p_Z}} \int_{[-1,1]} f_{XWZ}(X_i + \delta t_1, W_i + \delta t_2, Z_k + \delta t_3) dt_1 dt_2 dt_3.$$

Thus

$$\mathrm{E}\left|\mathcal{F}_1(\xi_i, \xi_j, \xi_k, \xi_l)\right|^2 = O\left(h^{-5}\delta^{-2(p_W + p_Z + 1)}\right).$$

and (A.1) holds by Assumption 3.5(iii). Moreover, by Lemma 4.2(b) in Chaudhuri et al. (1997),

$$\delta^{p_W + p_Z + 1} e_1^T \left\{N(X_i, W_i, Z_k) G(X_i, W_i, Z_k)\right\}^{-1} = f_{UXWZ}^{-1}(0, X_i, W_i, Z_k) e_1^T + r_n(X_i, W_i, Z_k),$$

where

$$\lim_{n\to\infty} \delta^{-\varrho} \mathrm{E}\left[\left\|r_n\left(X_i, W_i, Z_j\right)\right\|^2\right] = 0.$$

Now we are in position to evaluate the limit of the projection $\mathrm{E}(\mathcal{F}_1(\cdot, \cdot, \cdot, \xi_l))$. By change of variables, we can express $\mathrm{E}(\mathcal{F}_1(\cdot, \cdot, \cdot, \xi_l))$ as

$$\begin{aligned}
&\frac{1}{\sqrt{h}}\left(\tau - I\{Y_l \leq q_Y(X_l, W_l, Z_l)\}\right) \int_{-\infty}^{+\infty} \int_{[-1,1]^{1+p_W+p_Z}} \left(\frac{X_l + \delta t_1 - x_0 - hu}{h}\right) K\left(\frac{X_l + \delta t_1 - x_0}{h}\right) K(u) \\
&\qquad \times f_{UXWZ}^{-1}(0, X_l + \delta t_1, W_l + \delta t_2, Z_l + \delta t_3) f_{XW}(X_l + \delta t_1, W_l + \delta t_2) \\
&\qquad \times f_Z(Z_l + \delta t_3) f_X(x_0 + uh) dt_1 dt_2 dt_3 du \\
&\quad = {}^{\dagger_5} \frac{1}{\sqrt{h}}\left(\tau - I\{Y_l \leq q_Y(X_l, W_l, Z_l)\}\right) \int_{-\infty}^{+\infty} \int_{[-1,1]^{1+p_W+p_Z}} \left(\frac{X_l - x_0}{h}\right) K\left(\frac{X_l + \delta t_1 - x_0}{h}\right) K(u)
\end{aligned}$$

https://doi.org/10.1017/S0266466616000293 Published online by Cambridge University Press

$$\times f_{UXWZ}^{-1}(0, X_l+\delta t_1, W_l+\delta t_2, Z_l+\delta t_3) f_{XW}(X_l+\delta t_1, W_l+\delta t_2)$$
$$\times f_Z(Z_l+\delta t_3) f_X(x_0+uh) dt_1 dt_2 dt_3 du$$
$$+ \dagger_6 \frac{\delta}{h^{3/2}} (\tau - I\{Y_l \le q_Y(X_l, W_l, Z_l)\}) \int_{-\infty}^{+\infty} \int_{[-1,1]^{1+p_W+p_Z}} t_1 K\left(\frac{X_l+\delta t_1 - x_0}{h}\right) K(u)$$
$$\times f_{UXWZ}^{-1}(0, X_l+\delta t_1, W_l+\delta t_2, Z_l+\delta t_3) f_{XW}(X_l+\delta t_1, W_l+\delta t_2)$$
$$\times f_Z(Z_l+\delta t_3) f_X(x_0+uh) dt_1 dt_2 dt_3 du$$
$$+ \dagger_7 \frac{1}{\sqrt{h}} (\tau - I\{Y_l \le q_Y(X_l, W_l, Z_l)\}) \int_{-\infty}^{+\infty} \int_{[-1,1]^{1+p_W+p_Z}} u K\left(\frac{X_l+\delta t_1 - x_0}{h}\right) K(u)$$
$$\times f_{UXWZ}^{-1}(0, X_l+\delta t_1, W_l+\delta t_2, Z_l+\delta t_3) f_{XW}(X_l+\delta t_1, W_l+\delta t_2)$$
$$\times f_Z(Z_l+\delta t_3) f_X(x_0+uh) dt_1 dt_2 dt_3 du.$$

By the dominated convergence theorem,

$$\dagger_5 = \frac{1}{\sqrt{h}} (\tau - I\{Y_l \le q_Y(X_l, W_l, Z_l)\}) \left(\frac{X_l - x_0}{h}\right) K\left(\frac{X_l - x_0}{h}\right) f_{UXWZ}^{-1}(0, X_l, W_l, Z_l)$$
$$\times f_{XW}(X_l, W_l) f_Z(Z_l) f_X(x_0) + \frac{1}{h^{3/2}} O_p\left(\delta^\varrho + h^2 + \frac{\delta}{h}\right).$$

Moreover, $\dagger_6 = O_p\left(\frac{\delta}{h^{3/2}}\right)$ and $\dagger_7 = O_p\left(h^{1/2}\right)$. Thus under Assumption 3.5, we conclude

$$2\ddagger_4^1 = \frac{1}{\sqrt{nh}} \sum_{l=1}^{n} (\tau - I\{Y_l \le q_Y(X_l, W_l, Z_l)\}) \left(\frac{X_l - x_0}{h}\right) K\left(\frac{X_l - x_0}{h}\right)$$
$$\times f_{UXWZ}^{-1}(0, X_l, W_l, Z_l) f_{XW}(X_l, W_l) f_Z(Z_l) f_X(x_0) + o_p(1).$$

For $2\ddagger_4^2$, we consider

$$2\ddagger_4^2 = \frac{1}{\sqrt{n}(n-1)(n-2)(n-3)} \sum_{i \ne j \ne k \ne l} \mathcal{F}_2(\xi_i, \xi_j, \xi_k, \xi_l) + o_p(1),$$

where

$$\mathcal{F}_2(\xi_i, \xi_j, \xi_k, \xi_l)$$
$$= h^{-3/2} I\{(X_l, W_l, Z_l) \in C(X_i, W_i, Z_k, \delta)\}(W_i - W_j)$$
$$\times K\left(\frac{X_i - x_0}{h}\right) K\left(\frac{X_j - x_0}{h}\right) N^{-1}(X_i, W_i, Z_k)$$
$$\times e_1^T G^{-1}(X_i, W_i, Z_k) L(\delta, X_l - X_i, W_l - W_i, Z_l - Z_k)(\tau - I\{Y_l \le q_Y(X_l, W_l, Z_l)\}).$$

Under Assumption 3.5,

$$\mathrm{E}\left|\mathcal{F}_2(\xi_i, \xi_j, \xi_k, \xi_l)\right|^2 = O\left(h^{-3}\delta^{-2(p_W+p_Z+1)}\right) = o(n).$$

The limit of the projection $\mathrm{E}(\mathcal{F}_2(\cdot, \cdot, \cdot, \xi_l))$ can be expressed as

$$\frac{1}{\sqrt{h}} (\tau - I\{Y_l \le q_Y(X_l, W_l, Z_l)\}) \int_{\mathcal{W}} \int_{-\infty}^{+\infty} \int_{[-1,1]^{1+p_W+p_Z}} (W_l + \delta t_2 - w) K\left(\frac{X_l + \delta t_1 - x_0}{h}\right) K(u)$$
$$\times f_{UXWZ}^{-1}(0, X_l+\delta t_1, W_l+\delta t_2, Z_l+\delta t_3) f_{XW}(X_l+\delta t_1, W_l+\delta t_2)$$

https://doi.org/10.1017/S0266466616000293 Published online by Cambridge University Press

$$
\begin{aligned}
&\times f_Z(Z_l+\delta t_3) f_{XW}(x_0+uh,w)dt_1dt_2dt_3dudw\\
=\ &{}^{\dagger_8}\frac{1}{\sqrt{h}}(\tau - I\{Y_l \le q_Y(X_l,W_l,Z_l)\})\int_{\mathcal{W}}\int_{-\infty}^{+\infty}\int_{[-1,1]^{1+p_W+p_Z}}(W_l-w)K\left(\frac{X_l+\delta t_1-x_0}{h}\right)K(u)\\
&\times f^{-1}_{UXWZ}(0,X_l+\delta t_1,W_l+\delta t_2,Z_l+\delta t_3) f_{XW}(X_l+\delta t_1,W_l+\delta t_2)\\
&\times f_Z(Z_l+\delta t_3) f_{XW}(x_0+uh,w)dt_1dt_2dt_3dudw\\
&+{}^{\dagger_9}\frac{\delta}{\sqrt{h}}(\tau - I\{Y_l \le q_Y(X_l,W_l,Z_l)\})\int_{\mathcal{W}}\int_{-\infty}^{+\infty}\int_{[-1,1]^{1+p_W+p_Z}}t_2K\left(\frac{X_l+\delta t_1-x_0}{h}\right)K(u)\\
&\times f^{-1}_{UXWZ}(0,X_l+\delta t_1,W_l+\delta t_2,Z_l+\delta t_3) f_{XW}(X_l+\delta t_1,W_l+\delta t_2)\\
&\times f_Z(Z_l+\delta t_3) f_{XW}(x_0+uh,w)dt_1dt_2dt_3dudw.
\end{aligned}
$$

By the dominated convergence theorem,

$$
\begin{aligned}
{}^{\dagger_8} &= \frac{1}{\sqrt{h}}(\tau - I\{Y_l \le q_Y(X_l,W_l,Z_l)\})\int_{\mathcal{W}}(W_l-w)K\left(\frac{X_l-x_0}{h}\right)f^{-1}_{UXWZ}(0,X_l,W_l,Z_l)\\
&\quad\times f_{XW}(X_l,W_l) f_Z(Z_l) f_{XW}(x_0,w)dw+\frac{1}{\sqrt{h}}O_p\left(\delta^{\varrho}+h^2+\frac{\delta}{h}\right)\\
&= \frac{1}{\sqrt{h}}(\tau - I\{Y_l \le q_Y(X_l,W_l,Z_l)\})\{W_l-\mathrm{E}(W|X=x_0)\}K\left(\frac{X_l-x_0}{h}\right)\\
&\quad\times f^{-1}_{UXWZ}(0,X_l,W_l,Z_l) f_{XW}(X_l,W_l) f_Z(Z_l) f_X(x_0)+o_p(1).
\end{aligned}
$$

Moreover, ${}^{\dagger_9}=O\left(\frac{\delta}{h^{1/2}}\right)$. Thus under Assumption 3.5, we conclude

$$
\begin{aligned}
2\ddagger_4^2 &= \frac{1}{\sqrt{nh}}\sum_{l=1}^{n}(\tau - I\{Y_l \le q_Y(X_l,W_l,Z_l)\})\{W_l-\mathrm{E}(W|X=x_0)\}K\left(\frac{X_l-x_0}{h}\right)\\
&\quad\times f^{-1}_{UXWZ}(0,X_l,W_l,Z_l) f_{XW}(X_l,W_l) f_Z(Z_l) f_X(x_0)+o_p(1).
\end{aligned}
$$

Applying a similar argument, we obtain

$$
\begin{aligned}
2\ddagger_4^3 &= \frac{1}{\sqrt{nh}}\sum_{l=1}^{n}(\tau - I\{Y_l \le q_Y(X_l,W_l,Z_l)\})\left(\frac{X_l-x_0}{h}\right)W_lK\left(\frac{X_l-x_0}{h}\right)\\
&\quad\times f^{-1}_{UXWZ}(0,X_l,W_l,Z_l) f_{XW}(X_l,W_l) f_Z(Z_l) f_X(x_0)+o_p(1),
\end{aligned}
$$

which completes the proof. ■

LEMMA A.4. *The inverse of the following symmetric matrix*

$$
\begin{pmatrix} A & 0 & C\\ 0 & D & 0\\ C^T & 0 & F\end{pmatrix}
$$

is given by

$$
\begin{pmatrix} Q^{-1} & 0 & -Q^{-1}CF^{-1}\\ 0 & D^{-1} & 0\\ -F^{-1}C^TQ^{-1} & 0 & F^{-1}+F^{-1}C^TQ^{-1}CF^{-1}\end{pmatrix}
$$

provided that D, F, and $Q=A-CF^{-1}C^T$ are nonsingular.

https://doi.org/10.1017/S0266466616000293 Published online by Cambridge University Press

Proof. Straightforward. ■

Proof of Theorem 3.1.

$$
\begin{aligned}
D_1(\widehat{\Theta}_n - \Theta_0) &= D_1(S_{1n}^{-1}S_{2n} - \Theta_0) = D_1 S_{1n}^{-1}(S_{2n} - S_{1n}\Theta_0) \\
&= D_3 D_2 S_{1n}^{-1} D_2 D_2^{-1}(S_{2n} - S_{1n}\Theta_0) = D_2 S_{1n}^{-1} D_2 D_3 D_2^{-1}(S_{2n} - S_{1n}\Theta_0) \\
&= \underbrace{\left\{\frac{1}{n^2(n-1)} D_2^{-1} S_{1n} D_2^{-1}\right\}^{-1}}_{\ddagger_1} \underbrace{\left\{\frac{1}{n^2(n-1)} D_3 D_2^{-1}(S_{2n} - S_{1n}\Theta_0)\right\}}_{\ddagger_2}.
\end{aligned}
$$

It follows from Lemma 1 that $\ddagger_1 \to^p V_2$. The limiting distribution of $\ddagger_2$ remains to be established. Decompose $\ddagger_2$ into

$$
\begin{aligned}
\ddagger_2 &= \frac{1}{n^2(n-1)} D_3 D_2^{-1} \sum_{k=1}^{n} \sum_{1\le i<j\le n} \mathbf{X}_{ij}\left(\widehat{\mathbf{Y}}_{ijk} - \mathbf{X}_{ij}^T\Theta_0\right) K_h(X_i) K_h(X_j) \\
&= \underbrace{\frac{1}{n^2(n-1)} D_3 D_2^{-1} \sum_{k=1}^{n} \sum_{1\le i<j\le n} \mathbf{X}_{ij}\left(\mathbf{Y}_{ijk} - \mathbf{X}_{ij}^T\Theta_0\right) K_h(X_i) K_h(X_j)}_{\ddagger_3} \\
&\quad + \underbrace{\frac{1}{n^2(n-1)} D_3 D_2^{-1} \sum_{k=1}^{n} \sum_{1\le i<j\le n} \mathbf{X}_{ij}\left(\widehat{q}_{iik} - q_{iik}\right) K_h(X_i) K_h(X_j)}_{\ddagger_4} \\
&\quad - \underbrace{\frac{1}{n^2(n-1)} D_3 D_2^{-1} \sum_{k=1}^{n} \sum_{1\le i<j\le n} \mathbf{X}_{ij}\left(\widehat{q}_{jjk} - q_{jjk}\right) K_h(X_i) K_h(X_j)}_{\ddagger_5},
\end{aligned}
$$

where $\mathbf{Y}_{ijk} = q_{iik} - q_{jjk}$. We show in Lemma 2 that

$$
\ddagger_3 = D_1\left[h^2\begin{pmatrix} b_{\theta'} \\ b_g \\ b_{g'} \end{pmatrix} + o_p(h^2)\right].
$$

Notice that $\ddagger_4 - \ddagger_5 = 2\ddagger_4$. It then follows from Lemma 3 that

$$
2\ddagger_4 = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \begin{pmatrix} \psi_{\theta' i} \\ \psi_{gi} \\ \psi_{g' i} \end{pmatrix} + o_p(1).
$$

Additionally, according to Lemma 4, one important feature of V_2 defined in Assumption 2.1 is that its inverse V_2^{-1} permits the form

$$
V_2^{-1} = \begin{pmatrix} V^{\theta'\theta'} & 0 & (V^{\theta' g'})^T \\ 0 & V^{gg} & 0 \\ V^{\theta' g'} & 0 & V^{g' g'} \end{pmatrix},
$$

https://doi.org/10.1017/S0266466616000293 Published online by Cambridge University Press

which further implies that $D_1^{-1}V_2^{-1}D_1 = V_2^{-1}$. By combining these preliminary results, we have

$$D_1\left(\widehat{\Theta}_n - \Theta_0 - h^2 V_2^{-1}\begin{pmatrix} b_{\theta'} \\ b_g \\ b_{g'} \end{pmatrix}\right) = D_1\left(\widehat{\Theta}_n - \Theta_0 - h^2 D_1^{-1}V_2^{-1}D_1\begin{pmatrix} b_{\theta'} \\ b_g \\ b_{g'} \end{pmatrix}\right)$$
$$= D_1\left(\widehat{\Theta}_n - \Theta_0\right) - V_2^{-1}D_1 h^2\begin{pmatrix} b_{\theta'} \\ b_g \\ b_{g'} \end{pmatrix}$$
$$= V_2^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi_i + o_p(1) \to^d \mathcal{N}\left(0, V_2^{-1}\Omega_1 V_2^{-1}\right),$$

which completes the proof of (i). ■

Proof of Corollary 3.2. It follows from direct computation that $\mathrm{E}(\psi_{\theta' i}\psi_{gi}) = o(1)$ and $\mathrm{E}\left(\psi_{gi}\psi_{g'i}^T\right) = o(1)$ by making use of $\int sK(s)ds = 0$. In other words, the matrix Ω_1 has the form

$$\Omega_1 = \begin{pmatrix} \Omega_{\theta'\theta'} & 0 & (\Omega_{\theta' g'})^T \\ 0 & \Omega_{gg} & 0 \\ \Omega_{\theta' g'} & 0 & \Omega_{g'g'} \end{pmatrix},$$

The desired result then follows from this specific form of Ω_1 and V_2^{-1}, and from Lemma 4. ■

https://doi.org/10.1017/S0266466616000293 Published online by Cambridge University Press