Semiparametric Estimation of a Partially Linear Censored Regression Model
Author(s): Songnian Chen and Shakeeb Khan
Source: *Econometric Theory*, Vol. 17, No. 3 (Jun., 2001), pp. 567-590
Published by: Cambridge University Press
Stable URL: http://www.jstor.org/stable/3533117
Accessed: 09/07/2009 00:47

Econometric Theory, **17**, 2001, 567–590. Printed in the United States of America.

SEMIPARAMETRIC ESTIMATION OF A PARTIALLY LINEAR CENSORED REGRESSION MODEL

SONGNIAN CHEN
Hong Kong University of Science and Technology

SHAKEEB KHAN
University of Rochester

In this paper we propose an estimation procedure for a censored regression model where the latent regression function has a partially linear form. Based on a conditional quantile restriction, we estimate the model by a two stage procedure. The first stage nonparametrically estimates the conditional quantile function at in-sample and appropriate out-of-sample points, and the second stage involves a simple weighted least squares procedure. The proposed procedure is shown to have desirable asymptotic properties under regularity conditions that are standard in the literature. A small scale simulation study indicates that the estimator performs well in moderately sized samples.

1. INTRODUCTION AND MOTIVATION

The partially linear regression model[1] in its simplest form can be expressed as

$$y_i = x_i'\beta_0 + \phi(z_i) + \epsilon_i, \tag{1.1}$$

where y_i is an observed scalar dependent variable, (x_i, z_i) is a d-dimensional vector of observed covariates, and ϵ_i is an unobserved random variable reflecting unaccountable heterogeneity. The d_x-dimensional vector β_0 is the unknown structural parameter of interest, and the function $\phi(\cdot)$ is also unknown, representing the nonparametric component of the model.

This model has received a great deal of attention in both the applied and theoretical statistics and econometrics literature. Its popularity stems from its flexible specification, which allows for some variables to be linearly related to the response variable without imposing stringent restrictions on variables whose relationship to the response variable may be difficult to parameterize. This allows for a more general specification than the standard linear regression model,

We are grateful to J. Powell, B. Honoré, L.-F. Lee, two anonymous referees, and the co-editor Joel Horowitz for helpful comments. Songnian Chen gratefully acknowledges the financial support of grant HKUST6175/98H from the Research Grants Council of Hong Kong. An earlier version of this paper was presented at the 1998 CEME Conference at the University of Pittsburgh. Address correspondence to: Shakeeb Khan, Department of Economics, University of Rochester, Rochester, NY 14627, USA; e-mail: skhan@troi.cc.rochester.edu.

 0266-4666/01 $9.50

yet it is easier to interpret and less prone to the "dimensionality" problem that arises from adopting a fully nonparametric approach.

In the economics literature the nonparametric component $\phi(\cdot)$ has two interpretations. One is that this function represents a complicated relationship between the explanatory and response variables.[2] Alternatively, the function $\phi(\cdot)$ may be the result of sample selection (see, e.g., Powell, 1989).

In the econometrics and statistics literature there are several papers that analyze the asymptotic properties of various estimators for β_0 and/or $\phi(\cdot)$. Of particular interest is the effect of the presence of the nonparametric component on the rate of convergence of estimators for the parameters β_0. Various estimation procedures and their asymptotic properties have been established. Examples include Wahba (1984), Rice (1986), Robinson (1988), Speckman (1988), Chen (1988), and He and Shi (1996).

The purpose of this paper is to estimate the partially linear regression model when the data are censored. In many microeconometric applications, data are censored as a result of nonnegativity constraints or top coding. Unfortunately, none of the estimation procedures referred to will yield consistent estimates in these situations.

To model censored data, we consider the following partially linear latent regression framework:

$$y_i^* = x_i'\beta_0 + \phi(z_i) + \epsilon_i,$$

$$y_i = \max(y_i^*,0),$$

where y_i^* represents an unobserved latent response variable, which is only equal to the observed response variable when it exceeds the censoring value 0. Restrictions on ϵ_i need to be imposed for this model to be identified. For the linear censored regression model, Powell (1984, 1986) showed that a conditional quantile restriction on ϵ_i is sufficient for identification.

In this paper we identify and estimate the parametric component of the model under the same type of restriction. The quantile restriction we impose exhibits advantages over existing procedures introduced in the literature. For example, an estimator for a similar model proposed by Honoré and Powell (1997) is based on the assumption of independence between ϵ_i and (x_i, z_i) and thus is inconsistent in the presence of conditional heteroskedasticity. Ai and McFadden (1997) consider estimation of a wide class of latent partially linear models that includes the censored regression model, but they impose a parametric form on the distribution of ϵ_i, which results in inconsistent estimates if the distribution is misspecified.

The paper is organized as follows. The next section describes the two stage estimation procedure we adopt for the parametric component of the model. Section 3 lists sufficient regularity conditions and details the asymptotic properties of the estimator. Section 4 explores the finite sample properties of the estimator through a small scale simulation study. Section 5 provides some concluding remarks and discusses extensions of some of the ideas developed in the paper. An Appendix provides a detailed proof of the main theorem.

2. MODEL IDENTIFICATION AND DESCRIPTION OF THE PROPOSED ESTIMATOR

The model we wish to estimate can be characterized by the three equations

$$y_i^* = x_i'\beta_0 + \phi(z_i) + \epsilon_i, \tag{2.1}$$

$$y_i = \max(y_i^*,0), \tag{2.2}$$

$$P(\epsilon_i \leq 0 | x_i, z_i) = \alpha. \tag{2.3}$$

The first equation describes the partially linear relationship between an unobserved latent response variable and the observed regressors. The parametric component of the model specifies a linear relationship between the latent response and a subset of the regressors. The slope coefficients[3] β_0, a vector of dimension d_x, is the parameter of interest.

The second equation characterizes the type of censoring in the data we allow for. This equation describes a constant (known) censoring value assumed without loss of generality to be 0 and left censoring, but we can easily allow for right censoring and/or a censoring value that may vary across observations. We only require that the censoring values are known for observations that are not censored.

The third equation reflects the assumption that ϵ_i satisfies the conditional quantile restriction that its αth quantile is equal to 0 for all values of the regressors, for some fixed, known[4] $\alpha \in (0,1)$. Further restrictions on the distribution of ϵ_i discussed in the next section ensure that this conditional quantile is unique.

The equivariance property of conditional quantiles(see Powell, 1986) is the basis of our estimation procedure. It implies that the αth conditional quantile of the observed response variable y_i, which we denote by $q^\alpha(\cdot)$, is

$$q^\alpha(x_i, z_i) = \max(x_i'\beta_0 + \phi(z_i),0). \tag{2.4}$$

Equation (2.4) is the basis for the estimation procedure we introduce in this paper. The procedure involves two stages, and the following sections detail each of the steps involved.

2.1. First Stage: Local Partially Linear Polynomial Estimation

In the first stage we estimate the value of the conditional quantile function at various points. The next section discusses the in-sample and out-of-sample points at which to estimate the function to estimate β_0. Here, we describe the nonparametric procedure employed.

Nonparametric estimation of quantile functions has recently received a great deal of attention in the statistics and econometrics literature. New estimators and their asymptotic properties have been developed in Stute (1986), Bhattacharya and Gangopadhyay (1990), Chaudhuri (1991a, 1991b), Koenker, Portnoy, and Ng (1992), and Koenker, Ng, and Portnoy (1994), among others. Our approach in this paper is to extend the local polynomial estimator of the conditional quantile function introduced in Chaudhuri (1991a, 1991b) in a way that exploits the partially linear form of the model.

A description of the implementation of this stage is facilitated by introducing new notation, and the notation adopted here has been chosen deliberately to be as close as possible to that used in Chaudhuri (1991a, 1991b).

Assuming that the regressor vector has components that are either continuously or discretely distributed, we partition it as $(x_i^{(ds)}, x_i^{(c)}, z_i^{(ds)}, z_i^{(c)})$, where the superscripts $(ds),(c)$ denote discrete and continuous components, respectively. We let $d_{x_{ds}}, d_{x_c}, d_{z_{ds}}, d_{z_c}$ denote the respective dimensions of the components in the partition and set $d_{ds} = d_{x_{ds}} + d_{z_{ds}}$ and $d_c = d_{x_c} + d_{z_c}$. To characterize the distribution of the regressors we let $f_{X^{(c)},Z^{(c)}|X^{(ds)},Z^{(ds)}}(x^{(c)}, z^{(c)} \,|\, x^{(ds)}, z^{(ds)})$ and $f_{X^{(ds)},Z^{(ds)}}(x^{(ds)}, z^{(ds)})$ denote the conditional density function of $(x_i^{(c)}, z_i^{(c)})$ given $(x_i^{(ds)}, z_i^{(ds)}) = (x^{(ds)}, z^{(ds)})$ and the mass function of $(x_i^{(ds)}, z_i^{(ds)})$, respectively. Joint and marginal distributions are denoted by $f_{X,Z}(x,z)$ and $f_X(x), f_Z(z)$, respectively.

We let $C_n(x,z)$ denote the "bin" of the point x, z at which the quantile function is to be estimated and let h_n denote the sequence of "bandwidths" that governs the size of the bin. For some observation j we interpret $x_j, z_j \in C_n(x,z)$ to mean that $x_j^{(ds)} = x^{(ds)}$, $z_j^{(ds)} = z^{(ds)}$, and $x_j^{(c)}, z_j^{(c)}$ lies in the d_c-dimensional cube centered at $x^{(c)}, z^{(c)}$ with side length $2h_n$.

Next, we let k denote the order of differentiability of $\phi(z)$ with respect to $z^{(c)}$, and we let A denote the set of all d_{z_c}-dimensional vectors of non-negative integers $\{b_l\}$ where the sum of the components of b_l, which we denote by $[b_l]$, is less than or equal to k. We "naturally" order this set so its first element corresponds to $[b_l] = 0$ and let $s(A)$ denote the number of elements in this set. For any $s(A)$-dimensional vector θ, we denote its lth component by $\theta_{(l)}$, and we let β_c denote a d_{x_c}-dimensional vector.

Our first stage estimator estimates the $d_{x_c} + s(A)$-dimensional vector of parameters at any point by minimizing the following objective function:[5]

$$\hat{\beta}_c, \hat{\theta} = \operatorname{argmin}_{\beta_c, \theta} \sum_{x_j, z_j \in C_n(x,z)} \rho_\alpha\left(y_j - (x_j^{(c)} - x^{(c)})'\beta_c - \sum_{l=1}^{s(A)} \theta_{(l)} (z_j^{(c)} - z^{(c)})^{b_l} \right), \tag{2.5}$$

where $\rho_\alpha(\cdot) \equiv \alpha|\cdot| + (2\alpha - 1)(\cdot)I[\cdot < 0]$ is the loss function associated with a quantile restriction (see Koenker and Bassett, 1978), and for the two d_{z_c}-dimensional vectors $(z_j^{(c)} - z^{(c)})$ and b_l, the value $(z_j^{(c)} - z^{(c)})^{b_l}$ is shorthand notation for the product of each component of $(z_j^{(c)} - z^{(c)})$ raised to the corresponding component of b_l.

As discussed in Buchinsky and Hahn (1998), the minimizer of this type of objective function is a solution to a linear programming problem. Efficient algorithms, such as that proposed by Barrodale and Roberts (1973), converge to a global minimizer in a finite number of simplex iterations. The value $\hat{\theta}_{(1)}$ estimates $q^{\alpha}(x, z)$. The other parameters estimated are simply "nuisance" parameters in this context and are estimated only to improve the performance of the estimators of the parameters of interest.

Remark 1. This local polynomial estimator is different from that adopted in Chaudhuri (1991a, 1991b) and Chaudhuri, Doksum, and Samarov (1997). Specifically, we exploit the partially linear form of the model. This is reflected in the fact that we only adopt a linear expansion with respect to x_i and do not include interaction terms between x_i and z_i in the objective function. This will have the computational advantage of reducing the dimensionality of the minimization problem.

2.2. Second Stage: Weighted Least Squares

The previous stage estimation procedure provided estimates of the quantile function at any point. In this section, we illustrate how estimators at both in-sample and out-of-sample points can be used to construct an estimator for the parameter of interest β_0. We note that for an in-sample observation (x_i, z_i) such that $q^{\alpha}(x_i, z_i)$ is positive, we have

$$q^{\alpha}(x_i, z_i) = x_i' \beta_0 + \phi(z_i). \tag{2.6}$$

We also note that if for some $x_j \neq x_i$, it is also the case that the quantile function evaluated at the out-of-sample point (x_j, z_i) is positive, then

$$q^{\alpha}(x_j, z_i) = x_j' \beta_0 + \phi(z_i). \tag{2.7}$$

Equations (2.6) and (2.7) imply that the nonparametric component of the quantile function could be "differenced out":

$$q^{\alpha}(x_j, z_i) - q^{\alpha}(x_i, z_i) = (x_j - x_i)' \beta_0. \tag{2.8}$$

This suggests a least squares type estimator of β_0, using differenced values of $q^{\alpha}(\cdot)$ (as long as both are positive) estimated in the first stage as dependent variables and differenced values of x_i as independent variables.

One practical issue concerning the implementation of this estimation procedure is the selection of the out-of-sample points (x_j, z_i). We propose letting the values of z_i in the sample govern the selection of the out-of-sample values. Specifically, we let $\ell(z_j, z_i)$ denote a "selection function," and for the in-sample observation (x_i, z_i) we select all out-of-sample observations (x_j, z_i) such that the in-sample observation (x_j, z_j) satisfies the condition that $\ell(z_j, z_i)$ is positive.

One example of a selection function would be $\ell(z_j, z_i) \equiv 1$, in which case the quantile function would be estimated at all $n(n-1)$ pairs x_j, z_i. Another example of the selection function would be $\ell(z_j, z_i) = I[\|z_i - z_j\| \leq \delta]$ where $\|\cdot\|$ denotes the Euclidean norm and δ is a small positive constant. In this case, only observations where z_j is close to z_i would be selected, so the quantile function would have to be estimated at far fewer out-of-sample points. As we discuss later on, one motivation for this out-of-sample selection procedure is that particular choices of the selection function correspond to versions of "partial mean" and "kernel weighted" estimators previously proposed in the literature. For now, we simply note that computational issues motivate selection functions resembling the second example and asymptotic efficiency may motivate the selection function of the first example.

Before defining our estimator, one other implementation issue is selecting which observations to keep based on their estimated quantile function value. Instead of the one-zero rule $I[\hat{q}^\alpha(x_i, z_i) > 0]$, where $I[\cdot]$ denotes the usual indicator function and $\hat{q}^\alpha(x_i, z_i)$ denotes the first stage estimator, we propose a smooth (i.e., continuously differentiable with bounded derivative) weighting function $\omega(\cdot)$, giving greater weight to observations that have a larger quantile function value. For technical reasons, we bound the support of ω away from 0, giving positive weight only to observations where the estimated quantile function value exceeds a small positive constant c. This type of weighting function was considered in Buchinsky and Hahn (1998).

We formally define our estimator as the minimizer of the following weighted least squares type objective function:

$$\hat{\beta} = \operatorname{argmin}_\beta \frac{1}{n(n-1)} \sum_{i \neq j} \hat{\omega}_{ii} \hat{\omega}_{ji} \tau_{ii} \tau_{ji} \ell(z_j, z_i)(\Delta \hat{q}^\alpha_{ji} - \Delta x'_{ji}\beta)^2, \tag{2.9}$$

where

$\hat{\omega}_{ii}\hat{\omega}_{ji} \equiv \omega(\hat{q}^\alpha(x_i, z_i))\omega(\hat{q}^\alpha(x_j, z_i))$;

$\tau_{ii}\tau_{ji} \equiv \tau(x_i, z_i)\tau(x_j, z_i)$ is the product of a "trimming function" evaluated at the in-sample and selected out-of-sample points. The support of $\tau(\cdot)$ is denoted by $\mathcal{W} = \mathcal{X} \times \mathcal{Z}$, where $\mathcal{X}, \mathcal{Z}$ are compact subsets of the supports of x_i, z_i, respectively;

$\Delta \hat{q}^\alpha_{ji}$ denotes $\hat{q}^\alpha(x_j, z_i) - \hat{q}^\alpha(x_i, z_i)$;

Δx_{ji} denotes $x_j - x_i$.

Remark 2.

(i) This stage of the estimation procedure is as simple to compute as weighted least squares and involves no optimization routines to carry out.

(ii) The "trimming" functions incorporated in the objective function serve to bound the density of the regressors away from 0. This is to alleviate the "denominator" problem that arises when a preliminary nonparametric estimator is used in a second stage estimator.

(iii) It is worth pointing out how different selection functions relate to different estimators of the (uncensored) partially linear model. If $\ell(z_j, z_i) \equiv 1$, so all pairs are considered, the estimator appears similar to the partial mean approach adopted in Newey (1994). In the context of our estimator, selecting all pairs has two disadvantages. The first is that the nonparametric estimator of the quantile function at the out-of-sample point may be imprecise if z_i and z_j are far apart. The second is that the procedure could be quite computationally expensive as it would involve $O(n^2)$ minimizations in the first stage.

At the other extreme, if we view $\ell(z_j, z_i)$ as depending on the distance between z_i and z_j and the sample size, the estimator can resemble the kernel weighted least squares estimator in Powell (1989) if the distance goes to 0 as the sample size increases. We point out that in contrast to his approach, the "selection distance" need not change with the sample size for consistency.

The next section discusses the asymptotic properties of this estimation procedure.

3. ASYMPTOTIC PROPERTIES OF THE ESTIMATOR

Regularity conditions will first be outlined before proceeding to the main theorem; specific assumptions are imposed on the parameter space, the distributions of ϵ_i and the regressors, the order of smoothness of the function $\phi(\cdot)$, and the bandwidth sequence h_n.

Assumption (3.1) (Full Rank Condition). Denoting $\omega_{ii}\omega_{ji} \equiv \omega(q^\alpha(x_i, z_i)) \times \omega(q^\alpha(x_j, z_i))$, the $d_x \times d_x$ matrix V, defined as

$$E[\omega_{ii}\omega_{ji}\tau_{ii}\tau_{ji}\ell(z_j, z_i)\Delta x_{ji}\Delta x_{ji}'],$$

is full rank.

Assumption (3.2) (Random Sampling). The sequence of $d+1$-dimensional vectors (ϵ_i, x_i, z_i) is independent and identically distributed.

Assumption (3.3) (Regressor Distribution).

(3.3a) $f_{X^{(c)},Z^{(c)}|X^{(ds)},Z^{(ds)}}(x^{(c)}, z^{(c)}|x^{(ds)}, z^{(ds)})$ is bounded away from 0 and ∞ on $\mathcal{W}$.
(3.3b) $f_{X^{(ds)},Z^{(ds)}}(x^{(ds)}, z^{(ds)})$ has a finite number of mass points on $\mathcal{W}$.

Assumption (3.4) (Residual Distribution). For all $x_i, z_i \in \mathcal{W}$, the conditional distribution of ϵ_i given x_i, z_i has a density function denoted by $f_{\epsilon|X,Z}(e|x_i, z_i)$, which is positive at 0 and continuous at all values in a neighborhood of 0.

Assumption (3.5) (Order of Smoothness). For some $\varrho \in (0,1]$, and any function f, and set $\mathcal{D}$, we adopt the notation $f \in \mathcal{C}^{\varrho}(\mathcal{D})$ to mean there exists a positive constant K such that

$$\|f(\aleph_1) - f(\aleph_2)\| \le K\|\aleph_1 - \aleph_2\|^{\varrho}$$

for all $\aleph_1, \aleph_2 \in \mathcal{D}$. Letting $\mathcal{W}^{(c)} = \mathcal{X}^{(c)} \times \mathcal{Z}^{(c)}$ denote the subset of $\mathcal{W}$ corresponding to the continuous components, we impose the following smoothness conditions.

(3.5a) $f_{X,Z}(x_i, z_i), f_{\epsilon,X,Z}(0, x_i, z_i), \tau(x_i, z_i) \in \mathcal{C}^{\varrho}(\mathcal{W}^{(c)}) \quad \forall\, x_i^{(ds)}, z_i^{(ds)}$

$\ell(z_i, z_j) \in \mathcal{C}^{\varrho}(\mathcal{Z}^{(c)} \times \mathcal{Z}^{(c)}) \quad \forall\, z_i^{(ds)}, z_j^{(ds)}$

(3.5b) $\phi(z_i)$ is continuously differentiable in $z_i^{(c)}$ of order k, with kth order derivatives $\in \mathcal{C}^{\varrho}(\mathcal{Z}^{(c)})$. We let $p = k + \varrho$ denote the order of smoothness of this function.

Assumption (3.6) (Bandwidth Conditions). The bandwidth sequence used in the first stage is of the form

$$h_n = c^* n^{-\kappa},$$

where c^* is some constant and $\kappa \in \big((1/2p), (1/3d_c)\big)$.

Remark 3. These assumptions are quite standard when compared to existing conditions in the semiparametric literature. We only comment on some important features.

(i) The full rank condition in Assumption 3.1 reflects the necessary conditions for identification in both the partially linear and censored quantile regression models. To illustrate this point, we consider the case when the trimming and selection functions are set to 1. By the law of iterated expectations, the matrix V can be expressed as $E[E[\omega_{ii}\omega_{ji}(x_j - x_i)(x_j - x_i)' | z_i = z_j]]$. Now it becomes clear that the full rank condition fails if z_i determines x_i. This is to be expected because the full rank condition for the uncensored partially linear model fails also in this case, as discussed in Robinson (1988). Furthermore, the condition also fails if $\omega_{ii} = 0$ with high probability. This is analogous to the full rank condition discussed for the linear censored regression model. As discussed in Powell (1984, 1986), information for β_0 is only available from observations for which the conditional quantile function exceeds the censoring point. Thus the full rank condition is satisfied when there is sufficient variability in x_i conditional on z_i for the observations where the quantile function is positive.

(ii) Assumption 3.3 allows for both discretely and continuously distributed regressors. This is an important condition for most data sets in economics, where the effects of categorical variables, such as race and gender, are often of interest.

(iii) Assumption 3.4, which requires a positive conditional residual density function in a neighborhood of 0, is also necessary for identification, as it ensures uniqueness of the conditional quantile function.

(iv) Assumption 3.6, which allows for a wide range of bandwidth sequences, excludes the optimal bandwidth sequence derived in Chaudhuri (1991b). As is often the case for semiparametric estimators, "undersmoothing" is necessary to attain convergence at the parametric rate. It should also be mentioned that the lower bound on the bandwidth exponent, $1/2p$, is only due to the presence of nonparametric function $\phi(\cdot)$. Thus in principle, one could use different bandwidths for x_i and z_i, though this will only have a second order effect on the asymptotics.

These conditions enable us to characterize the asymptotic properties of the estimator. The main theorem of this paper, whose proof is left to the Appendix, establishes that $\hat{\beta}$ converges at the parametric rate and has an asymptotically normal distribution. Before stating the theorem, we define the following functions that characterize the influence function in the linear representation of $\hat{\beta}$.

$$\bar{\ell}(x,z) = E[\ell(z_i,z)|x_i = x], \tag{3.1}$$

$$\bar{\nabla}_1(z) = E[\omega_{ii}\tau_{ii}x_i|z_i = z], \tag{3.2}$$

$$\bar{\nabla}_2(z) = E[\omega_{ii}\tau_{ii}|z_i = z], \tag{3.3}$$

$$\Xi_1(z_i,z) = E[\omega(q^\alpha(x_i,z))\tau(x_i,z)x_i|z_i]\ell(z_i,z), \tag{3.4}$$

$$\Xi_2(z_i,z) = E[\omega(q^\alpha(x_i,z))\tau(x_i,z)|z_i]\ell(z_i,z), \tag{3.5}$$

$$\bar{\Xi}_1(z) = E[\Xi_1(z_i,z)], \tag{3.6}$$

$$\bar{\Xi}_2(z) = E[\Xi_2(z_i,z)]. \tag{3.7}$$

We can now state the main theorem of this paper. Its proof is left to the Appendix.

THEOREM 1. *Let $\bar{\Delta}_i$ denote the $d_x \times 1$ vector*

$$\omega_{ii}\tau_{ii}f_{\epsilon,X,Z}^{-1}(0,x_i,z_i)(\psi_1(x_i,z_i) + \psi_2(x_i,z_i))(\alpha - I[y_i \le q^\alpha(x_i,z_i)]),$$

where

$$\psi_1(x_i,z_i) = (f_X(x_i)f_Z(z_i)\bar{\ell}(x_i,z_i))(x_i\bar{\nabla}_2(z_i) - \bar{\nabla}_1(z_i)),$$

$$\psi_2(x_i,z_i) = f_{X,Z}(x_i,z_i)(x_i\bar{\Xi}_2(z_i) - \bar{\Xi}_1(z_i)).$$

Then under Assumptions 3.1–3.6

$$\sqrt{n}(\hat{\beta} - \beta_0) \Rightarrow N(0, V^{-1}\Omega V^{-1}), \tag{3.8}$$

where $\Omega = E[\bar{\Delta}_i\bar{\Delta}_i']$.

We conclude this section with a brief discussion on how this estimator can be used to construct an estimator of the nonparametric component $\phi(\cdot)$. For estimation of the function at an arbitrary point z, we let $S_n(z)$ denote the index set

$$\{i : 1 \le i \le n, \|z_i - z\| \le \hbar\},$$

where $\hbar$ is a fixed small positive constant. Given that the relationship

$$\phi(z) = q^\alpha(x_i,z) - x_i'\beta_0$$

holds when $q^\alpha(x_i,z) > 0$, we propose a weighted average estimator of $\phi(z)$ by replacing the quantile function with its nonparametric estimate, replacing β_0 with $\hat{\beta}$, and averaging over observations in $S_n(z)$:

$$\hat{\phi}(z) = \frac{\sum_{i \in S_n(z)} \omega(\hat{q}^\alpha(x_i, z))(\hat{q}^\alpha(x_i, z) - x_i'\hat{\beta})}{\sum_{i \in S_n(z)} \omega(\hat{q}^\alpha(x_i, z))}.$$

Asymptotically, this estimator will be equivalent to the infeasible estimator that subtracts the index $x_i'\beta_0$ from the quantile function and will converge at a nonparametric rate. The rate will be slower than the optimal rate established in Chaudhuri (1991b) because of the undersmoothing required in the bandwidth conditions.

4. FINITE SAMPLE PROPERTIES

In this section, the finite sample properties of the proposed estimation procedure are examined by a simulation study. We simulated from designs for which the latent equation followed designs considered in Robinson (1988). These were of the form

$$y_i^* = \zeta + x_i\beta_0 + \phi(z_i) + \epsilon_i,$$

where the covariates x_i, z_i were drawn from a bivariate standard normal distribution, with correlation 0.5, and ϵ_i was drawn from a standard normal distribution. The simulation study considered four designs, corresponding to the following specifications of $\phi(\cdot)$:

1. $\phi(z_i) = z_i$
2. $\phi(z_i) = z_i^2$
3. $\phi(z_i) = \sin(\pi z_i)$
4. $\phi(z_i) = \sinh(z_i)$

The latent dependent variable was censored from below at 0; the value of the parameter of interest, β_0, was set to 1; and the parameter ζ was varied across designs to keep the degree of censoring constant at 30%.

To implement the estimation procedure, we fixed α at 0.5 and set the order of the polynomial in z to 2. To select the bandwidth, we treated the nonparametric procedure as a one-dimensional problem. This is consistent with the theory because the asymptotic arguments were mainly governed by the rate at which the bandwidth of z_i converges, as alluded to in Remark 3(iv). Selecting the bandwidth in two step estimators is a difficult problem, but there are procedures that incorporate the undersmoothing prescribed by the theory. Examples include the procedures used in Horowitz (1992) and Buchinsky and Hahn (1998), which both perform well in the simulation studies they consider. For our study, we considered bandwidths that decreased to 0 at the rate $n^{-2/7}$, as this rate is consistent with the guidelines in Assumption 3.6 when $p = 2$ and $d = 1$. To select the constant of the bandwidth, we first considered a modified version of the "rule of thumb" bandwidth discussed on page 202 of Fan and Gijbels (1996). This was of the form

$$\left(\frac{\alpha(1-\alpha)\phi_Z(0)^{-2}}{\frac{1}{n}\sum_{i=1}^{n} \hat{q}_{gl}^{\alpha'''}(x_i, z_i)^2} \right)^{2/7},$$

where $\phi_z(\cdot)$ denotes the probability density function (p.d.f.) of the standard normal distribution and $\hat{q}_{gl}^{\alpha'''}$ denotes an estimator of the third derivative of the quantile function obtained from a global cubic fit. For these designs, preliminary simulations yielded average rule of thumb constants ranging from 2.4 to 2.75. Based on this result, we considered bandwidth constants from 1.75 to 3.50 with interval lengths of 0.25 to explore sensitivity to bandwidth choice.

For the weighting function $\omega(\cdot)$, we used the same function adopted in Buchinsky and Hahn (1998):

$$\omega(q^\alpha) = \left(\frac{e^{q^\alpha - 2c}}{1 + e^{q^\alpha - 2c}} - \frac{e^{-c}}{1 + e^{-c}} \right) \left(\frac{2 + e^c + e^{-c}}{e^c - e^{-c}} \right) I[c < q^\alpha < 3c] + I[q^\alpha > 3c]$$

and set $c = 0.1$. Preliminary studies showed results to be insensitive to the choice of weighting function and c. This is consistent with the results found in Buchinsky and Hahn (1998).

As a final implementation procedure, we set $\ell(z_j, z_i) = I[|z_i - z_j| \le \delta]$. In light of Remark 2(iii), a reasonable choice in practice would be to set $\delta = c_\delta \hat{\sigma}_z$, where c_δ is a small constant, say, 5%, and $\hat{\sigma}_z$ is the sample standard deviation of z_i. We first considered constants c_δ ranging from 0.005 to 2. Because preliminary results were insensitive to the choice of this constant, to save on computation time we reported results for $\delta = 0.005\hat{\sigma}_z$ for the complete study.

Each Monte Carlo experiment involved 801 replications for sample sizes of 100, 200, 400, and 800. Tables 1–4 report four statistics for the estimation of β_0 (mean, median, root mean square error [RMSE] and mean absolute deviation [MAD]) for the eight bandwidth constants. The simulation study was performed mostly in Gauss, with the first stage values tabulated using Fortran 77. For the four sample sizes considered, average times per replication for the first design and smallest bandwidth constant were 0.137, 0.389, 1.317, and 5.845 seconds on a Pentium II 400 MHz PC.

Qualitatively, the results are similar for all the designs considered. Except for the case when the bandwidth constant is set to 1.75, the behavior of the estimator seems to be in accordance with the asymptotic theory. The values of the bias and the RMSE consistently shrink at a rate of the square root of the sample size. When the constant is set to 1.75, it appears larger sample sizes than those considered in the study are necessary for the asymptotic theory to be reflected. Other than that, results are quite insensitive to values of the constant in the neighborhood of the rule of thumb choice, with the best performance corresponding to constants of 2.25 or 2.50 for designs I, III, and IV. For design II a bandwidth constant of 3.00 achieved the best results, which was consistent with its rule of thumb value being larger than that for the other designs.

TABLE 1. Monte Carlo simulation: Design 1

	$\phi(z_i) = z_i$							
	$c^* = 1.75$	$c^* = 2.00$	$c^* = 2.25$	$c^* = 2.50$	$c^* = 2.75$	$c^* = 3.00$	$c^* = 3.25$	$c^* = 3.50$
$n = 100$								
Mean bias	−0.2527	−0.2087	−0.2264	−0.2195	−0.2249	−0.2397	−0.2545	−0.2683
Median bias	−0.2167	−0.2268	−0.2319	−0.2252	−0.2257	−0.2437	−0.2572	-0.2747
RMSE	1.0764	0.5422	0.4741	0.4157	0.3965	0.3821	0.3729	0.3681
MAD	0.5679	0.4131	0.3666	0.3309	0.3163	0.3104	0.3085	0.3095
$n = 200$								
Mean bias	−0.1485	−0.1608	−0.1518	−0.1591	−0.1676	−0.1773	−0.1892	−0.2019
Median bias	−0.1742	−0.1729	−0.1669	−0.1693	−0.1753	−0.1821	−0.1934	−0.2067
RMSE	1.0540	0.4584	0.3303	0.2882	0.2698	0.2610	0.2592	0.2610
MAD	0.6418	0.3507	0.2669	0.2341	0.2186	0.2138	0.2154	0.2202
$n = 400$								
Mean bias	−0.1852	−0.1282	−0.1151	−0.1170	−0.1213	−0.1275	−0.1353	−0.1447
Median bias	−0.1451	−0.1318	−0.1087	−0.1138	−0.1209	−0.1207	−0.1285	−0.1407
RMSE	0.9184	0.3332	0.2424	0.2103	0.1946	0.1879	0.1863	0.1879
MAD	0.5796	0.2631	0.1912	0.1675	0.1565	0.1526	0.1533	0.1573
$n = 800$								
Mean bias	−0.1615	−0.0814	−0.0777	−0.0806	−0.0844	−0.0892	−0.0949	−0.1013
Median bias	−0.1116	−0.0724	−0.0741	−0.0746	−0.0790	−0.0844	−0.0926	−0.1013
RMSE	0.9899	0.2454	0.1810	0.1577	0.1461	0.1400	0.1377	0.1381
MAD	0.4731	0.1930	0.1430	0.1253	0.1166	0.1130	0.1127	0.1145

TABLE 2. Monte Carlo simulation: Design 2

	$\phi(z_i) = z_i^2$							
	$c^* = 1.75$	$c^* = 2.00$	$c^* = 2.25$	$c^* = 2.50$	$c^* = 2.75$	$c^* = 3.00$	$c^* = 3.25$	$c^* = 3.50$
$n = 100$								
Mean bias	−0.3359	−0.3419	−0.3429	−0.3241	−0.3227	−0.3361	−0.3480	−0.3600
Median bias	−0.3432	−0.3650	−0.3653	−0.3310	−0.3339	−0.3464	−0.3559	−0.3715
RMSE	1.1385	0.7065	0.6028	0.5363	0.4879	0.4606	0.4468	0.4434
MAD	0.6978	0.5495	0.4775	0.4320	0.4037	0.3888	0.3836	0.3847
$n = 200$								
Mean bias	−0.2474	−0.2650	−0.2626	−0.2631	−0.2615	−0.2667	−0.2771	−0.2871
Median bias	−0.2312	−0.2886	−0.2808	−0.2733	−0.2681	−0.2647	−0.2775	−0.2870
RMSE	1.2586	0.6190	0.4548	0.3810	0.3491	0.3338	0.3301	0.3306
MAD	0.7218	0.4718	0.3702	0.3183	0.2934	0.2850	0.2867	0.2926
$n = 400$								
Mean bias	−0.2797	−0.2275	−0.2114	−0.2041	−0.1987	−0.1983	−0.2030	−0.2107
Median bias	−0.0869	−0.2656	−0.2315	−0.2081	−0.1973	−0.2008	−0.2095	−0.2139
RMSE	1.0933	0.5645	0.3425	0.2927	0.2656	0.2515	0.2463	0.2465
MAD	0.6670	0.4192	0.2783	0.2420	0.2235	0.2152	0.2141	0.2182
$n = 800$								
Mean bias	−0.1384	−0.1283	−0.1212	−0.1170	−0.1157	−0.1167	−0.1192	−0.1229
Median bias	−0.1369	−0.1284	−0.1201	−0.1165	−0.1147	−0.1157	−0.1201	−0.1238
RMSE	0.1733	0.1610	0.1522	0.1462	0.1432	0.1424	0.1430	0.1446
MAD	0.1463	0.1367	0.1293	0.1245	0.1224	0.1223	0.1237	0.1264

Table 3. Monte Carlo simulation: Design 3

	$\phi(z_i) = \sin(\pi z_i)$							
	$c^* = 1.75$	$c^* = 2.00$	$c^* = 2.25$	$c^* = 2.50$	$c^* = 2.75$	$c^* = 3.00$	$c^* = 3.25$	$c^* = 3.50$
$n = 100$								
Mean bias	−0.2361	−0.2349	−0.2274	−0.2267	−0.2353	−0.2463	−0.2587	−0.2731
Median bias	−0.2256	−0.2353	−0.2269	−0.2348	−0.2323	−0.2483	−0.2662	−0.2846
RMSE	0.8917	0.6060	0.5173	0.4274	0.4039	0.3860	0.3762	0.3726
MAD	0.5796	0.4425	0.3874	0.3433	0.3260	0.3174	0.3123	0.3133
$n = 200$								
Mean bias	−0.1264	−0.1481	−0.1380	−0.1520	−0.1643	−0.1770	−0.1904	−0.2031
Median bias	−0.1763	−0.1645	−0.1524	−0.1559	−0.1595	−0.1783	−0.1952	−0.2071
RMSE	1.5432	0.4719	0.3414	0.2894	0.2672	0.2599	0.2594	0.2616
MAD	0.6938	0.3696	0.2728	0.2350	0.2190	0.2150	0.2168	0.2227
$n = 400$								
Mean bias	−0.2252	−0.1344	−0.1183	−0.1202	−0.1266	−0.1353	−0.1448	−0.1553
Median bias	−0.1661	−0.1394	−0.1056	−0.1105	−0.1251	−0.1333	−0.1452	−0.1577
RMSE	0.9255	0.3941	0.2711	0.2291	0.2104	0.2021	0.1995	0.2013
MAD	0.6207	0.3079	0.2104	0.1815	0.1687	0.1651	0.1656	0.1695
$n = 800$								
Mean bias	−0.1538	−0.0679	−0.0654	−0.0712	−0.0778	−0.0855	−0.0936	−0.1021
Median bias	−0.0985	−0.0583	−0.0598	−0.0669	−0.0749	−0.0826	−0.0925	−0.1011
RMSE	0.9532	0.2638	0.1887	0.1613	0.1474	0.1411	0.1393	0.1403
MAD	0.5026	0.2068	0.1481	0.1265	0.1169	0.1135	0.1134	0.1158

Table 4. Monte Carlo simulation: Design 4

	$\phi(z_i) = \sinh(z_i)$							
	$c^* = 1.75$	$c^* = 2.00$	$c^* = 2.25$	$c^* = 2.50$	$c^* = 2.75$	$c^* = 3.00$	$c^* = 3.25$	$c^* = 3.50$
$n = 100$								
Mean bias	−0.2631	−0.2027	−0.2185	−0.2187	−0.2246	−0.2391	−0.2538	−0.2682
Median bias	−0.2029	−0.2243	−0.2130	−0.2211	−0.2291	−0.2393	−0.2536	−0.2689
RMSE	1.3424	0.5311	0.4640	0.4087	0.3909	0.3783	0.3702	0.3656
MAD	0.5845	0.4037	0.3582	0.3261	0.3122	0.3069	0.3054	0.3073
$n = 200$								
Mean bias	−0.1455	−0.1566	−0.1461	−0.1537	−0.1636	−0.1744	−0.1871	−0.2004
Median bias	−0.1589	−0.1594	−0.1588	−0.1603	−0.1699	−0.1778	−0.1882	−0.2032
RMSE	0.9972	0.4525	0.3256	0.2828	0.2652	0.2571	0.2558	0.2585
MAD	0.6303	0.3449	0.2626	0.2294	0.2148	0.2106	0.2124	0.2181
$n = 400$								
Mean bias	−0.1737	−0.1224	−0.1108	−0.1135	−0.1187	−0.1253	−0.1335	−0.1434
Median bias	−0.1451	−0.1251	−0.1000	−0.1078	−0.1156	−0.1196	−0.1258	−0.1370
RMSE	0.9108	0.3304	0.2387	0.2066	0.1914	0.1851	0.1839	0.1859
MAD	0.5818	0.2601	0.1878	0.1643	0.1537	0.1499	0.1510	0.1555
$n = 800$								
Mean bias	−0.1284	−0.0707	−0.0712	−0.0736	−0.0776	−0.0828	−0.0889	−0.0957
Median bias	−0.1269	−0.0719	−0.0786	−0.0740	−0.0769	−0.0842	−0.0922	−0.0981
RMSE	0.6695	0.2460	0.1758	0.1518	0.1400	0.1337	0.1316	0.1319
MAD	0.4714	0.1933	0.1418	0.1216	0.1126	0.1084	0.1079	0.1096

Though the rates of convergence of the bias and RMSE agreed with the asymptotic theory, the estimator exhibited significant mean and median biases for all designs in sample sizes of 100 and 200. This is not unusual for two step estimators with preliminary nonparametric estimators for such sample sizes. It should also be pointed out that the finite sample performance would be expected to deteriorate for a given sample size if the number of regressors increased, as a result of a second order "curse of dimensionality." In the context of our estimator, it is the dimension of z_i that should be of the biggest concern.

Overall, the simulation results indicate that our estimation procedure performs well enough in moderately sized samples to be used in practice. We advise caution in its use if the sample size is less than 100 or when the dimensionality of z_i is high.

5. SUMMARY AND CONCLUDING REMARKS

This paper introduces an estimation procedure for estimating the partially linear regression model in the presence of censored data. The estimator is shown to have favorable asymptotic properties. The results of a small scale simulation study indicate that the procedure performs reasonably well in finite samples. The main advantages of this procedure are that the resulting estimator is simple to compute and that it is "robust" to very general forms of conditional heteroskedasticity. This is in contrast to the estimation procedure proposed in Honoré and Powell (1997). However it should be noted that their procedure covers a wide range of nonlinear models, whereas ours is designed specifically for the censored regression model.

The results of this paper suggest areas for further research. Specifically, it would be interesting to compare the results of this paper, which adopted a local approach to estimating the model, to one that adopted a global approach. He and Shi (1996) propose a global (B-spline) quantile estimator of the uncensored partially linear regression model. Their results are not directly comparable to ours because they do not allow for censoring and consider only homoskedastic models, but it may be possible to extend their results to allow for censoring and conditional heteroskedasticity in a fashion analogous to approaches taken in Powell (1986) or Buchinsky and Hahn (1998) for the linear censored model.

NOTES

1. The model is also referred to as the semilinear regression and semiparametric regression model in the literature.
2. See Engle et al. (1986) and Stock (1989) for important empirical examples.
3. Note that an intercept term is not identified for this model because of its nonparametric component.
4. In practice, the median function, which corresponds to $\alpha = 0.5$, is usually considered as a result of its "central location" interpretation.
5. For technical reasons, we actually require the assumption that this minimization occur over a compact subset of $\Re^{d_{x_c}+s(A)}$.

REFERENCES

Ai, C. & D. McFadden (1997) Estimation of some partially specified nonlinear models. *Journal of Econometrics* 76, 1–37.

Barrodale, I. & F. Roberts (1973) An improved algorithm for discrete L_1 linear approximation. *SIAM Journal of Numerical Analysis* 10, 839–848.

Bhattacharya, P.K. & A.K. Gangopadhyay (1990) Kernel and nearest neighbor estimation of a conditional quantile. *Annals of Statistics* 18, 1400–1415.

Buchinsky, M. & J. Hahn (1998) An alternative estimator for the censored quantile regression model. *Econometrica* 66, 627–651.

Chaudhuri, P. (1991a) Nonparametric quantile regression. *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, H. (1988) Convergence rates for parametric components in a partly linear model. *Annals of Statistics* 16, 136–146.

Engle, R.F., C.W.J. Granger, J. Rice, & A. Weiss (1986) Semiparametric estimates of the relation between weather and electricity demand. *Journal of the American Statistical Association* 76, 817–823.

Fan, J. & I. Gijbels (1996) *Local Polynomial Modelling and Its Applications*. New York: Chapman and Hall.

He, X. & P. Shi (1996) Bivariate tensor-product *B*-splines in a partly linear model. *Journal of Multivariate Analysis* 58, 162–181.

Honoré, B.E. & J.L. Powell (1997) Pairwise Difference Estimators for Non-linear Models. Working paper.

Horowitz, J.L. (1992) A smoothed maximum score estimator for the binary response model. *Econometrica* 60, 505–531.

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

Koenker, R., S. Portnoy, & P. Ng (1992) Nonparametric estimation of conditional quantile function. In Y. Dodge (ed.), *Proceedings of the Conference on L_1-Statistical Analysis and Related Methods*, pp. 217–229. Amsterdam: Elsevier.

Koenker, R., P. Ng, & S. Portnoy (1994) Quantile smoothing splines. *Biometrika* 81, 673–680.

Newey, W.K. (1994) Kernel estimation of partial means and a general variance estimator. *Econometric Theory* 10, 233–253.

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. (1989) Semiparametric Estimation of Censored Selection Models. Manuscript.

Powell, J.L., J.H. Stock, & T.M. Stoker (1989) Semiparametric estimation of index coefficients. *Econometrica* 57, 1404–1430.

Rice, J. (1986) Convergence rates for partially splined models. *Statistics and Probability Letters* 4, 203–208.

Robinson, P.M. (1988) Root-*N*-consistent semiparametric regression. *Econometrica* 56, 931–954.

Serfling, R.J. (1980) *Approximation Theorems of Mathematical Statistics*. New York: Wiley.

Speckman, P. (1988) Kernel smoothing in partial linear models. *Journal of the Royal Statistical Society, Series B* 50, 413–436.

Stock, J.H. (1989) Nonparametric policy analysis. *Journal of the American Statistical Association* 84, 1461–1481.

Stute, W. (1986) Conditional empirical processes. *Annals of Statistics* 14, 638–647.

Wahba, G. (1984) Partial spline models for the semiparametric estimation of functions of several variables, pp. 319–329. In *Statistical Analysis of Time Series*. Tokyo: Institute of Statistical Mathematics.

APPENDIX

To keep expressions notationally simple, in this section we let $q_{ii}^{\alpha}, q_{ji}^{\alpha}$ denote $q^{\alpha}(x_i, z_i)$ and $q^{\alpha}(x_j, z_i)$, respectively. We denote estimated values by $\hat{q}_{ii}^{\alpha}, \hat{q}_{ji}^{\alpha}$; also, we let C_{ni} denote $C_n(x_i, z_i)$ and let $N_n(x_i, z_i) = \sum_{j \neq i} I[(x_j, z_j) \in C_{ni}]$.

The proof involves establishing a linear representation for the estimator $\hat{\beta}$. We work with the relationship

$$\hat{\beta} - \beta_0 = \hat{S}_{xx}^{-1} \hat{S}_{xy}, \tag{A.1}$$

where

$$\hat{S}_{xx} = \frac{1}{n(n-1)} \sum_{i \neq j} \hat{\omega}_{ii} \hat{\omega}_{ji} \tau_{ii} \tau_{ji} \ell(z_j, z_i) \Delta x_{ji} \Delta x_{ji}' \tag{A.2}$$

and

$$\hat{S}_{xy} = \frac{1}{n(n-1)} \sum_{i \neq j} \hat{\omega}_{ii} \hat{\omega}_{ji} \tau_{ii} \tau_{ji} \ell(z_j, z_i) \Delta x_{ji} (\hat{q}_{ji}^{\alpha} - \hat{q}_{ii}^{\alpha} - \Delta x_{ji}' \beta_0). \tag{A.3}$$

Our strategy in proving the theorem is to evaluate the probability limit of $\hat{S}_{xx}$ and a linear representation for $\hat{S}_{xy}$. We begin by establishing the following two lemmas, which correspond to two uniform convergence results for the nonparametric estimator of the conditional quantile function. The lemmas are proven for the in-sample observations only, as we note that identical arguments can be used for the out-of-sample points. The first lemma establishes a rate uniform over points where the quantile function is bounded away from the censoring point. The result follows directly from the uniform rates derived in Chaudhuri (1991b) and Chaudhuri et al. (1997). (These uniform rates were based on the assumption that the regressors were continuously distributed. As mentioned on page 252 of Chaudhuri (1991b), this was only assumed to ensure that $N_n(x_i, z_i)$ increases at the appropriate rate. This rate will be satisfied under Assumption 3.3.)

LEMMA 1. (From Chaudhuri et al. Lemma 4.3a). *Under Assumptions 3.2–3.6,*

$$\max_{1 \le i \le n, q_{ii}^{\alpha} \ge c/2} |\hat{q}_{ii}^{\alpha} - q_{ii}^{\alpha}| = o_p(n^{-1/4}).$$

The second uniform result involves an exponential bound for points in a neighborhood of the censoring point.

LEMMA 2. *Under Assumptions 3.2–3.6, let $\mathcal{W}_c$ denote the set*

$$\{(x_i, z_i) \in \mathcal{W}, q^{\alpha}(x_i, z_i) \le c/2\}$$

and let A_n denote the event

$$\{\hat{q}^{\alpha}(x_i, z_i) \ge c \quad \text{for all } (x_i, z_i) \in \mathcal{W}_c\}.$$

Then there exist constants C_1, C_2 such that

$$P(A_n) \le C_1 e^{-C_2 n h_n^{d_c}}.$$

Proof. We first derive a similar result for the nonparametric estimator that fits a polynomial of degree 0 and denote this by $\bar{q}_{ii}^{\alpha}$. We consider attaining an exponential bound for the probability of the event

$$\bar{A}_n = \{\bar{q}_{ii}^{\alpha} \geq 3c/4 \text{ for all } (x_i, z_i) \in \mathcal{W}_c\}.$$

For a pair of positive constants $c_1 < c_2$ we define the event E_n as

$$\{c_1 n h_n^{d_c} \leq N_n(x_i, z_i) \leq c_2 n h_n^{d_c} \text{ for all } (x_i, z_i) \in \mathcal{W}_c\}.$$

By Theorem 3.1(i) in Chaudhuri (1991b), we can choose c_1, c_2 such that

$$P(E_n^c) \leq c_3 e^{-c_4 n h_n^{d_c}}, \tag{A.4}$$

where c_3, c_4 are positive constants and E_n^c denotes the complement of the event E_n. Thus it will suffice to derive an exponential rate for the probability of the event $\bar{A}_n \cap E_n$. We note that $\bar{A}_n$ implies the event

$$\frac{1}{N_n(x_i, z_i)} \sum_{(x_j, z_j) \in C_{ni}} I[y_j \geq 3c/4] \geq (1-\alpha) \quad \text{for all } (x_i, z_i) \in \mathcal{W}_c. \tag{A.5}$$

Now, by the continuity of q_{ii}^{α} and the compactness of $\mathcal{W}$, we have for n larger than some N_0, $q_{jj}^{\alpha} < 2c/3$ if $(x_j, z_j) \in C_{ni}$ for all $(x_i, z_i) \in \mathcal{W}_c$. Thus by Assumption 3.4 there is a positive constant λ_1 such that for $(x_j, z_j) \in C_{ni}$

$$P(y_j \geq 3c/4 | x_j, z_j) \leq P(\epsilon_j \geq c/12 | x_j, z_j) \leq (1-\alpha) - \lambda_1,$$

and the probability of the event $\bar{A}_n \cap E_n$ is bounded above by

$$P\left(\sum_{x_j, z_j \in C_{ni}} I[y_j \geq 3c/4] - E[I[y_j \geq 3c/4] | x_j, z_j] \geq c_1 n h^{d_c} \lambda_1 \cap E_n\right) \leq e^{-2\lambda_1^2 c_1^2 n h_n^{d_c}}, \tag{A.6}$$

where the exponential bound follows by Hoeffding's inequality. Thus the exponential bound follows for $\bar{q}_{ii}^{\alpha}$ by picking C_1 and C_2 such that the bounds in (A.4) and (A.6) are satisfied. Finally, the conclusion of the lemma follows for $\hat{q}_{ii}^{\alpha}$ by showing that

$$|\hat{q}_{ii}^{\alpha} - \bar{q}_{ii}^{\alpha}| < c/4 \quad \text{for all } (x_i, z_i) \in \mathcal{W}. \tag{A.7}$$

This result becomes apparent by expressing the local polynomial estimator $\hat{q}_{ii}^{\alpha}$ as an estimator with polynomial of degree 0 and dependent variable equal to

$$y_j^* = y_j - (x_j^{(c)} - x_i^{(c)})'\hat{\beta}_c - \sum_{l=2}^{s(A)} \hat{\theta}_{(l)} (z_j^{(c)} - z_i^{(c)})^{b_l},$$

where $\hat{\beta}_c$ and $\hat{\theta}_{(l)}$, $l \neq 1$ (which are constrained to lie in a compact set by construction) are the local polynomial estimators of the coefficient on $x_i^{(c)}$ and the derivatives of $\phi(\cdot)$, respectively. Thus $|y_j^* - y_j| < c/4$ for all $(x_i, z_i) \in \mathcal{W}$ for all n larger than some N_1, as the local polynomial estimators are restricted to lie in a bounded set. This establishes (A.7). ■

We now proceed to the expression in (A.1). The following lemma establishes the probability limit of $\hat{S}_{xx}$.

LEMMA 3.

$$\hat{S}_{xx} \xrightarrow{p} V. \tag{A.8}$$

Proof. A mean value expansion of the function $\hat{\omega}(\cdot)$ around the true value of the quantile function yields

$$\hat{S}_{xx} = \frac{1}{n(n-1)} \sum_{i \neq j} \omega_{ii} \omega_{ji} \tau_{ii} \tau_{ji} \ell(z_j, z_i) \Delta x_{ji} \Delta x'_{ji} + R_{1n} + R_{2n}, \tag{A.9}$$

where

$$R_{1n} = \frac{1}{n(n-1)} \sum_{i \neq j} \omega'^{*}_{ii} \omega^{*}_{ji} \tau_{ii} \tau_{ji} \ell(z_j, z_i)(\hat{q}^{\alpha}_{ii} - q^{\alpha}_{ii}) \Delta x_{ji} \Delta x'_{ji},$$

$$R_{2n} = \frac{1}{n(n-1)} \sum_{i \neq j} \omega^{*}_{ii} \omega'^{*}_{ji} \tau_{ii} \tau_{ji} \ell(z_j, z_i)(\hat{q}^{\alpha}_{ji} - q^{\alpha}_{ji}) \Delta x_{ji} \Delta x'_{ji},$$

where * denotes intermediate values. We establish that $R_{1n} = o_p(1)$. We note that it can be decomposed as the sum of

$$\frac{1}{n(n-1)} \sum_{i \neq j} \omega'^{*}_{ii} \omega^{*}_{ji} \tau_{ii} \tau_{ji} \ell(z_j, z_i)(\hat{q}^{\alpha}_{ii} - q^{\alpha}_{ii}) \Delta x_{ji} \Delta x'_{ji} I[q^{\alpha}_{ii} \geq c/2],$$

$$\frac{1}{n(n-1)} \sum_{i \neq j} \omega'^{*}_{ii} \omega^{*}_{ji} \tau_{ii} \tau_{ji} \ell(z_j, z_i)(\hat{q}^{\alpha}_{ii} - q^{\alpha}_{ii}) \Delta x_{ji} \Delta x'_{ji} I[q^{\alpha}_{ii} < c/2].$$

The first term is $o_p(1)$ by Lemma 1. For the second term, we note that if $\omega'^{*}_{ii} \omega^{*}_{ji}$ is positive, then $\hat{q}^{\alpha}_{ii} \geq c$, so this second term is bounded above in absolute value by

$$\frac{1}{n(n-1)} \sum_{i \neq j} C |(\hat{q}^{\alpha}_{ii} - q^{\alpha}_{ii})| I[q^{\alpha}_{ii} < c/2, \hat{q}^{\alpha}_{ii} \geq c],$$

where C is some constant. This term is $o_p(1)$ by Lemma 2.

The same argument can be used to show that R_{2n} is $o_p(1)$. Thus we have established that

$$\hat{S}_{xx} = \frac{1}{n(n-1)} \sum_{i \neq j} \omega_{ii} \omega_{ji} \tau_{ii} \tau_{ji} \ell(z_j, z_i) \Delta x_{ji} \Delta x'_{ji} + o_p(1),$$

and the conclusion of the lemma thus follows from the law of large numbers for U-statistics (see, e.g., Serfling, 1980, p. 190). ■

We next establish a linear representation for $\hat{S}_{xy}$. We note that it can be decomposed as

$$\hat{S}^{(1)}_{xy} - \hat{S}^{(2)}_{xy} + R_n,$$

where

$$\hat{S}_{xy}^{(1)} = \frac{1}{n(n-1)} \sum_{i \neq j} \hat{\omega}_{ii} \hat{\omega}_{ji} \tau_{ii} \tau_{ji} \ell(z_j, z_i)(\hat{q}_{ji}^{\alpha} - q_{ji}^{\alpha}) \Delta x_{ji},$$

$$\hat{S}_{xy}^{(2)} = \frac{1}{n(n-1)} \sum_{i \neq j} \hat{\omega}_{ii} \hat{\omega}_{ji} \tau_{ii} \tau_{ji} \ell(z_j, z_i)(\hat{q}_{ii}^{\alpha} - q_{ii}^{\alpha}) \Delta x_{ji},$$

$$R_n = \frac{1}{n(n-1)} \sum_{i \neq j} \hat{\omega}_{ii} \hat{\omega}_{ji} \tau_{ii} \tau_{ji} \ell(z_j, z_i) \Delta x_{ji}(q_{ji}^{\alpha} - q_{ii}^{\alpha} - \Delta x_{ji}' \beta_0).$$

We first show that the remainder term R_n is asymptotically negligible.

LEMMA 4.

$R_n = o_p(n^{-1/2})$.

Proof. We note that

$$\frac{1}{n(n-1)} \sum_{i \neq j} \omega_{ii} \omega_{ji} \tau_{ii} \tau_{ji} \ell(z_j, z_i) \Delta x_{ji}(q_{ji}^{\alpha} - q_{ii}^{\alpha} - \Delta x_{ji}' \beta_0) = 0,$$

so we need only show that

$$\frac{1}{n(n-1)} \sum_{i \neq j} (\hat{\omega}_{ii} \hat{\omega}_{ji} - \omega_{ii} \omega_{ji}) \tau_{ii} \tau_{ji} \ell(z_j, z_i) \Delta x_{ji}(q_{ji}^{\alpha} - q_{ii}^{\alpha} - \Delta x_{ji}' \beta_0) = o_p(n^{-1/2}). \quad \textbf{(A.10)}$$

A mean value expansion of $\hat{\omega}_{ii} \hat{\omega}_{ji}$ around $\omega_{ii} \omega_{ji}$ yields the sum of the terms

$$\frac{1}{n(n-1)} \sum_{i \neq j} \omega_{ii}'^{*} \omega_{ji}^{*} \tau_{ii} \tau_{ji} \ell(z_j, z_i)(\hat{q}_{ii}^{\alpha} - q_{ii}^{\alpha}) \Delta x_{ji}(q_{ji}^{\alpha} - q_{ii}^{\alpha} - \Delta x_{ji}' \beta_0)$$

$$\frac{1}{n(n-1)} \sum_{i \neq j} \omega_{ii}^{*} \omega_{ji}'^{*} \tau_{ii} \tau_{ji} \ell(z_j, z_i)(\hat{q}_{ji}^{\alpha} - q_{ji}^{\alpha}) \Delta x_{ji}(q_{ji}^{\alpha} - q_{ii}^{\alpha} - \Delta x_{ji}' \beta_0).$$

We only show that the first of the preceding terms is asymptotically negligible, as the same arguments can be applied to the second term. Note we can multiply each term in the summation by $I[q_{ii}^{\alpha} = 0, q_{ji}^{\alpha} > 0] + I[q_{ii}^{\alpha} > 0, q_{ji}^{\alpha} = 0]$ because terms in the summation are 0 if both values of the quantile function are positive. Also, $\omega_{ii}'^{*} \omega_{ji}^{*} I[q_{ii}^{\alpha} = 0]$ and $\omega_{ii}'^{*} \omega_{ji}^{*} I[q_{ji}^{\alpha} = 0]$ are bounded above by a constant times $I[q_{ii}^{\alpha} = 0, \hat{q}_{ii}^{\alpha} \geq c]$ and a constant times $I[q_{ji}^{\alpha} = 0, \hat{q}_{ji}^{\alpha} \geq c]$, respectively. So by Lemma 2, (A.10) holds. ■

The following lemma establishes the linear representation for $\hat{S}_{xy}^{(1)}$.

LEMMA 5. *Under Assumptions 3.2–3.6,*

$$\hat{S}_{xy}^{(1)} = \frac{1}{n} \sum_{i=1}^{n} \omega_{ii} \tau_{ii} f_{\epsilon, X, Z}^{-1}(0, x_i, z_i) \psi_1(x_i, z_i)(\alpha - I[y_i \leq q_{ii}^{\alpha}]) + o_p(n^{-1/2}). \quad \textbf{(A.11)}$$

Proof. A mean value expansion of $\hat{\omega}_{ii} \hat{\omega}_{ji}$ around $\omega_{ii} \omega_{ji}$ yields

$$\hat{S}_{xy}^{(1)} = \frac{1}{n(n-1)} \sum_{i \neq j} \omega_{ii} \omega_{ji} \tau_{ii} \tau_{ji} \ell(z_j, z_i)(\hat{q}_{ji}^{\alpha} - q_{ji}^{\alpha}) \Delta x_{ji} + R_{1n} + R_{2n},$$

where now

$$R_{1n} = \frac{1}{n(n-1)} \sum_{i \neq j} \omega_{ii}^* \omega_{ji}'^* \tau_{ii} \tau_{ji} \ell(z_j, z_i)(\hat{q}_{ji}^\alpha - q_{ji}^\alpha)^2 \Delta x_{ji},$$

$$R_{2n} = \frac{1}{n(n-1)} \sum_{i \neq j} \omega_{ii}'^* \omega_{ji}^* \tau_{ii} \tau_{ji} \ell(z_j, z_i)(\hat{q}_{ji}^\alpha - q_{ji}^\alpha)(\hat{q}_{ii}^\alpha - q_{ii}^\alpha) \Delta x_{ji}.$$

We only show that $R_{1n} = o_p(n^{-1/2})$, as the same argument can be applied to R_{2n}. We decompose R_{1n} as the sum of the two components

$$\frac{1}{n(n-1)} \sum_{i \neq j} \omega_{ii}^* \omega_{ji}'^* \tau_{ii} \tau_{ji} \ell(z_j, z_i)(\hat{q}_{ji}^\alpha - q_{ji}^\alpha)^2 \Delta x_{ji} I[q_{ij}^\alpha \geq c/2],$$

$$\frac{1}{n(n-1)} \sum_{i \neq j} \omega_{ii}^* \omega_{ji}'^* \tau_{ii} \tau_{ji} \ell(z_j, z_i)(\hat{q}_{ji}^\alpha - q_{ji}^\alpha)^2 \Delta x_{ji} I[q_{ij}^\alpha < c/2].$$

By Lemma 1 the first of the preceding terms is $o_p(n^{-1/2})$, and by the same argument used in the proof of Lemma 3, using Lemma 2, the second term is also $o_p(n^{-1/2})$. We can therefore work with the expression

$$\frac{1}{n(n-1)} \sum_{i \neq j} \omega_{ii} \omega_{ji} \tau_{ii} \tau_{ji} \ell(z_j, z_i)(\hat{q}_{ji}^\alpha - q_{ji}^\alpha) \Delta x_{ji}. \tag{A.12}$$

Our next step is to insert a linear representation for $(\hat{q}_{ji}^\alpha - q_{ji}^\alpha)$. This type of representation has been well established in Chaudhuri (1991a) and Chaudhuri et al. (1997). Its introduction requires some additional notation, and we deliberately adopt the notation used in Chaudhuri et al. (1997). For estimation of q_{ii}^α using observations $j \neq i$, we let $b(h_n, x_j - x_i, z_j - z_i)$ denote the $d_{x_c} + s(A)$-dimensional vector

$$b(h_n, x_j - x_i, z_j - z_i) = (1, (x_j^{(c)} - x_i^{(c)})', \{h_n^{-[u]}(z_j^{(c)} - z_i^{(c)})^u, 1 \leq [u] \leq k\}')'. \tag{A.13}$$

Denote the $(d_{x_c} + s(A)) \times (d_{x_c} + s(A))$ matrix $G_n(x_i, z_i)$ as the density weighted conditional expectation of the outer product of $b(h_n, x_j - x_i, z_j - z_i)$ given that $x_j, z_j \in C_{ni}$. Let $e_{(1)}$ denote a $d_{x_c} + s(A)$-dimensional vector whose first component is 1 and remaining components are 0 and denote its transpose by $e_{(1)}'$. From Lemma 4.1 in Chaudhuri et al. (1997), the linear representation can be expressed as

$$\begin{aligned} \hat{q}_{ii}^\alpha - q_{ii}^\alpha &= N_n(x_i, z_i)^{-1} e_{(1)}' G_n(x_i, z_i)^{-1} \\ &\quad \times \sum_{j=1, j \neq i}^{n} b(h_n, x_j - x_i, z_j - z_i)(\alpha - I[y_j \leq q_{jj}^\alpha]) I[x_j, z_j \in C_{ni}] \\ &\quad + R_n(x_i, z_i), \end{aligned} \tag{A.14}$$

where $R_n(x_i, z_i)$ is $o_p(n^{-1/2})$ *uniformly* in $(x_i, z_i) \in \mathcal{W}$. We insert the nonnegligible component of this representation into (A.12). Following identical arguments as used in Chaudhuri et al. (1997), this yields a third order U-statistic plus a remainder term that is $o_p(n^{-1/2})$. The U-statistic is of the form

$$\frac{1}{n(n-1)(n-2)} \sum_{i \neq j \neq k} \omega_{ii}\omega_{ji}\tau_{ii}\tau_{ji}\ell(z_j,z_i)e'_{(1)}f_n(x_j,z_i)^{-1}G_n(x_j,z_i)^{-1}$$
$$\times\, b(h_n,x_k-x_j,z_k-z_i)(\alpha - I[y_k \le q^{\alpha}_{kk}])I[x_k \in C^x_{nj}]I[z_k \in C^z_{ni}]\Delta x_{ji}, \qquad \textbf{(A.15)}$$

where here $f_n(x_j,z_i) = (1/n)E[N_n(x_j,z_i)|x_j,z_i]$ and superscripts on C_{ni} denote bins with respect to the specified regressors. The preceding U-statistic has a kernel function that depends on the sample size. Projection theorems for such cases have been developed in Powell, Stock, and Stoker (1989), for example. Writing the U-statistic as

$$\frac{1}{n(n-1)(n-2)} \sum_{i \neq j \neq k} \mathcal{F}_n(\xi_i,\xi_j,\xi_k), \qquad \textbf{(A.16)}$$

where $\xi_i \equiv (y_i,x'_i,z'_i)'$, the condition sufficient for the projection theorem is

$$E[\|\mathcal{F}_n(\cdot,\cdot,\cdot)\|^2] = o(n).$$

Noting that the smallest eigenvalue of $G_n(x_j,z_i)$ is bounded away from 0 for all n, and that $b(h_n,x_k-x_j,z_k-z_j)I[x_k \in C^x_{nj}]I[z_k \in C^z_{ni}]$ is bounded above by 1, the only term in $\mathcal{F}_n(\cdot,\cdot,\cdot)$ that becomes unbounded as the sample size increases is $f_n^{-1}(x_j,z_i)$. Noting that $f_n(x_j,z_i)$ can be expressed as

$$h_n^{d_c}\int_{[-1,1]^{d_{xc}}}\int_{[-1,1]^{d_{zc}}} f_{X,Z}(x_j^{(ds)},x_j^{(c)}+h_nt,z_i^{(ds)},z_i^{(c)}+h_nu)\,du\,dt,$$

it follows that

$$E[\|\mathcal{F}_n(\xi_i,\xi_j,\xi_k)\|^2] = O(h_n^{-2d_c}) = o(n),$$

where the second equality follows from Assumption 3.6.

We can therefore work with the projection of the U-statistic. Noting that $E[\mathcal{F}_n(\xi_i,\xi_j,\xi_k)|\xi_i] = E[\mathcal{F}_n(\xi_i,\xi_j,\xi_k)|\xi_j] = 0$, the projection theorem yields

$$\frac{1}{n(n-1)(n-2)} \sum_{i \neq j \neq k} \mathcal{F}_n(\xi_i,\xi_j,\xi_k) = \frac{1}{n}\sum_{i=1}^{n} E[\mathcal{F}_n(\cdot,\cdot,\xi_i)] + o_p(n^{-1/2}),$$

where $E[\mathcal{F}_n(\cdot,\cdot,\xi_i)]$ denotes the conditional expectation of $\mathcal{F}_n(\cdot,\cdot,\cdot)$ given its third argument. Therefore, it remains only to establish that

$$\frac{1}{n}\sum_{i=1}^{n} E[\mathcal{F}_n(\cdot,\cdot,\xi_i)] - \omega_{ii}\tau_{ii}f^{-1}_{\epsilon,X,Z}(0,x_i,z_i)\psi_1(x_i,z_i)(\alpha - I[y_i \le q^{\alpha}_{ii}])$$
$$= o_p(n^{-1/2}). \qquad \textbf{(A.17)}$$

To evaluate the limit of the projection, we first note that by Lemma 4.2b of Chaudhuri et al. (1997), we have

$$h_n^{d_c}e'_{(1)}(f_n(x_i,z_i)G_n(x_i,z_i))^{-1} = f^{-1}_{\epsilon,X,Z}(0,x_i,z_i)e'_{(1)} + r_n(x_i,z_i),$$

where

$$\lim_{n\to\infty} h_n^{-\varrho}E[\|r_n(x_i,z_i)\|^2] = 0.$$

Next, by partitioning the regressor vector into its discrete and continuous components, we note that by a change of variables we can express $E[\mathcal{F}_n(\cdot,\cdot,\xi_i)]$ as

$$
\begin{aligned}
&(I[y_i \le q_{ii}^{\alpha}] - \alpha) \\
&\quad \times \int_{[-1,1]^{d_{xc}}} \int_{[-1,1]^{d_{zc}}} \int_{\mathcal{X}} \int_{\mathcal{Z}} \omega(q^{\alpha}(x^{(ds)}, x^{(c)}, z_i^{(ds)}, z_i^{(c)} + h_n t)), \\
&\qquad \omega(q^{\alpha}(x_i^{(ds)}, x_i^{(c)} + h_n u, z_i^{(ds)}, z_i^{(c)} + h_n t)) \\
&\quad \times \tau(x^{(ds)}, x^{(c)}, z_i^{(ds)}, z_i^{(c)} + h_n t)\tau(x_i^{(ds)}, x_i^{(c)} + h_n u, z_i^{(ds)}, z_i^{(c)} + h_n t) \\
&\quad \times \ell(z^{(ds)}, z^{(c)}, z_i^{(ds)}, z_i^{(c)} + h_n t) \\
&\quad \times f_{\epsilon,X,Z}^{-1}(0, x_i^{(ds)}, x_i^{(c)} + h_n u, z_i^{(ds)}, z_i^{(c)} + h_n t) f_Z(z_i^{(ds)}, z_i^{(c)} + h_n t) \\
&\quad \times f_X(x_i^{(ds)}, x_i^{(c)} + h_n u) \\
&\quad \times dF_{X|Z}(x^{(ds)}, x^{(c)} | z_i^{(ds)}, z_i^{(c)} + h_n t)\, dF_{Z|X}(z^{(ds)}, z^{(c)} | x_i^{(ds)}, x_i^{(c)} + h_n u) du dt \\
&\quad + R_n(x_i, z_i), \qquad \textbf{(A.18)}
\end{aligned}
$$

where $F_{X|Z}, F_{Z|X}$ denote conditional distribution functions and (A.18) represents the conditional expectation by a Stieltjes integral. The remainder term $R_n(x_i, z_i)$ satisfies

$$E[\|R_n(x_i, z_i)\|^2] = O(h_n^{\varrho}). \qquad \textbf{(A.19)}$$

By the dominated convergence theorem, it follows that for all $(x_i, z_i) \in \mathcal{W}$, the integral in (A.18) converges to

$$\omega_{ii}\tau_{ii} f_{\epsilon,X,Z}^{-1}(0, x_i, z_i)\psi_1(x_i, z_i).$$

Therefore, another application of the dominated convergence theorem and (A.19) imply that the variance of

$$E[\mathcal{F}_n(\cdot,\cdot,\xi_i)] - f_{\epsilon,X,Z}^{-1}(0, x_i, z_i)\psi_1(x_i, z_i)(\alpha - I[y_i \le q_{ii}^{\alpha}])$$

converges to 0. Thus equation (A.17) holds by Chebyshev's inequality. This establishes the proof of the lemma. ■

By an identical argument, we can establish a linear representation for $\hat{S}_{xy}^{(2)}$.

LEMMA 6. *Under Assumptions 3.2–3.6,*

$$\hat{S}_{xy}^{(2)} = \frac{1}{n}\sum_{i=1}^{n} \omega_{ii}\tau_{ii} f_{\epsilon,X,Z}^{-1}(0, x_i, z_i)\psi_2(x_i, z_i)(\alpha - I[y_i \le q_{ii}^{\alpha}]) + o_p(n^{-1/2}). \qquad \textbf{(A.20)}$$

Combining the three previous lemmas yields a linear representation for $\hat{S}_{xy}$. Along with Lemma 3, the theorem follows by an application of Slutsky's theorem, using Assumption 3.1.