# Censored Regression Quantiles with Endogenous Regressors

Richard Blundell* and James L. Powell†

November 2005
(Version 3.0)

**Abstract**

This paper develops a semiparametric method for estimation of the censored regression model when some of the regressors are endogenous (and continuously distributed) and instrumental variables are available for them. A "distributional exclusion" restriction is imposed on the unobservable errors, whose conditional distribution is assumed to depend on the regressors and instruments only through a lower-dimensional "control variable," here assumed to be the difference between the endogenous regressors and their conditional expectations given the instruments. This assumption, which implies a similar exclusion restriction for the conditional quantiles of the censored dependent variable, is used to motivate a two-stage estimator of the censored regression coefficients. In the first stage, the conditional quantile of the dependent variable given the instruments and the regressors is nonparametrically estimated, as are the first-stage reduced-form residuals to be used as control variables. The second-stage estimator is a weighted least squares regression of pairwise differences in the estimated quantiles on the corresponding differences in regressors, using only pairs of observations for which both estimated quantiles are positive (i.e., in the uncensored region) and the corresponding difference in estimated control variables is small. The paper gives the form of the asymptotic distribution for the proposed estimator, and discusses how it compares to similar estimators for alternative models.

**JEL: C14, C25, C35, J22.**

**Key Words:** Censored Regression, Endogeneity, Quantile Regression, Control Function Estimation

*Address for correspondence*: James L. Powell, Department of Economics, University of California at Berkeley, Berkeley, CA 94720-3880. email: powell@econ.berkeley.edu

*University College London, Department of Economics, Gower Street, London, WC1E 6BT and Institute for Fiscal Studies. r.blundell@ucl.ac.uk

†Department of Economics, University of California, Berkeley, CA 94720-3880. powell@econ.berkeley.edu

# 1. Introduction[1]

This paper proposes an extension of quantile-based semiparametric estimation methods for the coefficients of the censored regression model to models in which some of the regressors are endogenous. Identification of the regression coefficients is based upon a quantile variant of the "control function" approach to estimation with endogenous regressors, in which the error terms are assumed to be independent of the regressors after conditioning upon the residuals of a first-stage fit of the regressors on some instrumental variables. Use of first-stage residuals (or some other identified function of the regressors and instruments) to control for endogeneity of the regressors or sample selectivity is a long-standing tradition in the estimation of parametric econometric models; examples include Dhrymes (1970), Heckman (1976, 1979), and Blundell and Smith (1989, 1991). More recently, it has been adopted to identify and estimate semiparametric (Powell 1987, 2001; Ahn and Powell 1993; Chen and Khan 1999; Honoré and Powell 2004; Das, Newey, Vella 2003; Blundell and Powell 2004) and nonparametric models (e.g., Newey, Powell, and Vella 1999; Altonji and Matzkin 1997; Blundell and Powell 2001; Imbens and Newey 2002; Florens, Heckman, Meghir, and Vytlacil 2003).

For models with additive error terms, some of these papers impose a *mean exclusion* restriction that the conditional expectation of the errors given the control variable and regressors does not depend upon the latter; for models with non-additive errors, a stronger *distributional exclusion* restriction of conditional independence of the errors and regressors given the control variable is typically assumed. This paper imposes an alternative *quantile exclusion* restriction that a particular conditional quantile of the error distribution is independent of the regressors given the control variable; like the mean exclusion restriction, this quantile exclusion restriction is implied by the stronger distributional exclusion version, but it is more suitable for the censored regression model, in which the error terms are non-addively but monotonically related to the observable dependent variable.

[1] **Acknowledgements:** We are grateful to Xioahong Chen, Hidehiko Ichimura, Guido Imbens, Michael Jansson, Thomas MaCurdy, Thomas Rothenberg, Paul Ruud, Edward Vytlacil, the referees, seminar participants at Stanford, Berkeley, NYU, Caltech, UCSD, and Columbia, and participants in the Semiparametrics in Rio conference for their helpful comments. Blundell gratefully acknowledges financial support from the Leverhulme Trust.

For the censored regression model under the quantile exclusion restriction, a two-stage estimator of the unknown regression coefficients is proposed. In the first stage, nonparametric estimation methods are used to estimate the control variable (residuals from a nonparametric regression of the endogenous regressors on the instruments) and the conditional quantile of the censored dependent variable given the regressors and instruments. The second stage constructs an estimator of the finite-dimensional regression coefficient vector through a weighted least squares regression of the differences in the estimated quantiles on differences in estimated regressors, with weights that are nonzero only when both quantile estimates are positive (and thus linear in the covariates) and the difference in the control variables is close to zero. For a given quantile, this estimator of the underlying regression coefficients is shown to be $\sqrt{n}$-consistent and asymptotically normally distributed under appropriate regularity conditions.

In the next section, the structural equations for the censored dependent variable and endogenous regressors are defined, as are the distributional exclusion and implied quantile exclusion restrictions, which are then used to motivate the definition of the estimator of the unknown regression coefficients. Section 3 discusses the nonparametric first-stage estimation methods and their key asymptotic properties, and gives the form of the asymptotic (normal) distribution of the coefficient estimator. The final section discusses how the proposed estimation approach can be extended to similar models (including coherent simultaneous Tobit models, and censored regression with selectivity), relates the estimation approach to recently-proposed control function estimators for endogenous versions of binary response and censored selection, and compares the features of the proposed estimation method to alternative identifying assumptions and estimators for the censored endogenous regression model. Statements of regularity conditions and derivations of the theoretical results are given in a mathematical appendix.

## 2. The Model and Estimation Approach

We consider the *censored regression* model, specifically, the linear regression model with left-censoring at zero,

$$y_i = \max\{0, \mathbf{x}_i'\boldsymbol{\beta}_0 + u_i\}, \tag{2.1}$$

where the scalar dependent variable $y_i$ and the vector $\mathbf{x}_i$ of explanatory variables are assumed to be jointly i.i.d. for $i = 1, ..., n$, with $\boldsymbol{\beta}_0$ a conformable vector of unknown regression coefficients and $u_i$ an unobservable scalar error term,. Variants of this model, e.g., right-censoring at an observable censoring variable $c_i$,

$$y_i = \min\{\mathbf{x}_i'\boldsymbol{\beta}_0 + u, c_i\}, \tag{2.2}$$

can be easily accommodated, provided the censoring threshold is observable for all data points and satisfies similar conditions to those imposed on the regressors.

When the regressors are assumed to be exogenous – that is, when the error term $u_i$, or some feature of its conditional distribution, is assumed independent of the regressors $\mathbf{x}_i$ – a number of root-$n$-consistent estimators for $\boldsymbol{\beta}_0$ have been proposed in the semiparametric literature (see Powell 1994, Section 3.3, for a dated survey). When some components $\mathbf{x}_i^{(e)}$ of $\mathbf{x}_i$ are endogenous, standard practice is to posit the existence of a vector $\mathbf{z}_i$ of "instrumental variables" which are exogenous in an appropriate sense. For censored data, such a model may arise if the latent (uncensored) dependent variable

$$y_i^* \equiv \mathbf{x}_i'\boldsymbol{\beta}_0 + u_i \tag{2.3}$$

is the single equation of interest in a simultaneous-equations system in which the $L$-dimensional subvector of endogenous regressors $\mathbf{x}_i^{(e)}$ is generated by the reduced form

$$\mathbf{x}_i^{(e)} = \boldsymbol{\pi}(\mathbf{z}_i) + \mathbf{v}_i, \tag{2.4}$$

with $\boldsymbol{\pi}$ some (possibly nonparametric) function of the instruments $\mathbf{z}_i$ and $\mathbf{v}_i$ the corresponding error term.

In this setting, an assumption of independence between the error terms $(u_i, \mathbf{v}_i)$ and the instruments $\mathbf{z}_i$ would generate a number of weaker exclusion restrictions which might be used to try to identify $\boldsymbol{\beta}_0$ in the system (2.1) and (2.4). One such weaker restriction would impose only independence of $u_i$ and $\mathbf{z}_i$, which would sidestep the specification of (2.4) and the associated "control variable" $\mathbf{v}_i$. This restriction figures prominently in recent results on the estimation of coefficients for linear models with endogenous regressors, including those by Chernozhukov and Hansen (2005ab) and Sawata (2005); however, as Hong and Tamer (2003) demonstrate, identification of $\boldsymbol{\beta}_0$ under this weaker restriction involves strong conditions on the regressors and instruments (e.g., existence of a set of instrument values, with positive probability, such that $\Pr\{\mathbf{x}_i'\boldsymbol{\beta}_0 > 0|\mathbf{z}_i\} = 1$). An alternative condition implied by independence of $(u_i, \mathbf{v}_i)$ and $\mathbf{z}_i$ is the distributional exclusion restriction

$$\begin{aligned} F_u(q|\mathbf{x}_i, \mathbf{z}_i) &\equiv \Pr\{u_i \leq q|\mathbf{x}_i, \mathbf{z}_i\} \\ &= \Pr\{u_i \leq q|\mathbf{v}_i\} \equiv F_u(q|\mathbf{v}_i) \qquad w.p.1, \qquad q \in R. \end{aligned} \tag{2.5}$$

This assumption is also weaker than independence of all errors and the instruments, since it does not require independence of $\mathbf{v}_i$ and $\mathbf{z}_i$; for example, it may be that $\mathbf{v}_i = \sigma(\mathbf{z}_i)\boldsymbol{\eta}_i$, where $\sigma(\mathbf{z}_i)$ is a suitable scaling function and $u_i$ and $\boldsymbol{\eta}_i$ are jointly independent of $\mathbf{z}_i$. This example makes clear that condition (2.5) is neither stronger nor weaker than independence of $u_i$ and $\mathbf{z}_i$, since it permits $\mathbf{z}_i$ to affect $u_i$ through $\mathbf{v}_i$; nevertheless, the condition does require a complete specification of the list of the relevant instrumental variables $\mathbf{z}_i$ in (2.4), along with a stochastic restriction on the first-stage residuals $\mathbf{v}_i$ (e.g., $E[\mathbf{v}_i|\mathbf{z}_i] = 0$) which permits consistent estimation of $\boldsymbol{\pi}$.

The distributional exclusion restriction (2.5) is equivalent to a restriction that all of the conditional quantiles of $u_i$ given $\mathbf{x}_i$ and $\mathbf{z}_i$ are functions only of the control variable $\mathbf{v}_i \equiv \mathbf{x}_i - \boldsymbol{\pi}(\mathbf{z}_i)$:

$$Q_\alpha[u_i|\mathbf{x}_i, \mathbf{z}_i] = Q_\alpha[u_i|\mathbf{v}_i] \qquad w.p.1, \qquad \alpha \in (0,1), \tag{2.6}$$

where the conditional $\alpha^{th}$ quantile $Q_\alpha$ of $u_i$ given $\mathbf{x}_i$ and $\mathbf{z}_i$ is defined as

$$\begin{aligned} Q_\alpha[u_i|\mathbf{x}_i, \mathbf{z}_i] &\equiv F^{-1}_{u|\mathbf{x},\mathbf{z}}(\alpha|\mathbf{x}_i, \mathbf{z}_i) \\ &\equiv \inf\{q : \Pr\{u \leq q|\mathbf{x}_i, \mathbf{z}_i\} \geq \alpha\} &\quad (2.7) \\ &= \arg\min_q E\left[\rho_\alpha (u - q) |\mathbf{x}_i, \mathbf{z}_i\right], &\quad (2.8) \end{aligned}$$

where $\alpha \in (0, 1)$ and $\rho_\alpha$ is the "check function"

$$\rho_\alpha(u) = |u| \cdot |\alpha - 1\{u \leq 0\}|. \tag{2.9}$$

Of course, a weaker restriction would impose (2.6) only for a particular value of $\alpha$ (e.g., $\alpha = .5$, a median restriction). Such a quantile restriction on the errors is useful for models in which the dependent variable is monotonically related to the error term, as for the censored regression model here. Since quantiles are equivariant with respect to nondecreasing transformations, the restriction (2.6) yields a corresponding restriction on the conditional quantile of the censored dependent variable $y_i$ given the regressors $\mathbf{x}_i$ and instruments $\mathbf{z}_i$ :

$$\begin{aligned} q_i &\equiv Q_\alpha[y_i|\mathbf{x}_i, \mathbf{z}_i] \equiv q_i(\alpha) \\ &= Q_\alpha[\max\{0, \mathbf{x}_i'\boldsymbol{\beta}_0 + u_i\}|\mathbf{x}_i, \mathbf{z}_i] \\ &= \max\{0, \mathbf{x}_i'\boldsymbol{\beta}_0 + Q_\alpha[u_i|\mathbf{x}_i, \mathbf{z}_i]\} \\ &= \max\{0, \mathbf{x}_i'\boldsymbol{\beta}_0 + \lambda_\alpha(\mathbf{v}_i)\}, \end{aligned} \tag{2.10}$$

where

$$\lambda_\alpha(\mathbf{v}_i) \equiv Q_\alpha[u_i|\mathbf{v}_i]. \tag{2.11}$$

When all components of $\mathbf{x}_i$ are exogenous (with $\mathbf{x}_i = \mathbf{z}_i$), the "control function" $\lambda_\alpha(\mathbf{v}_i)$ is a constant for each $\alpha$, a result that forms the basis for the large literature on quantile estimation of censored regression, a literature which includes Powell (1984, 1986), Nawata (1992), Buchinsky and Hahn (1998), Khan and Powell (2001), Chen and Khan (2001), Chernozhukov and Hong (2002), and Hong and Tamer (2003), among others. When the regressors are endogenous and (2.6) applies, the conditional quantile of $y_i$ becomes a left-censored version of a partially-linear regression function. As discussed below, estimation of this semilinear censored regression model was considered by Chen and Khan (2001), under the assumption that the control variable $\mathbf{v}_i$ is known; the modifications of their approach in this paper are made to accommodate nonparametric estimation of the control variable.

The random variable $q_i$ in (2.10), being the conditional quantile for observable random variables, is identified; sufficient conditions for identification of $\boldsymbol{\beta}_0$ can then be derived from a similar "pairwise differencing" argument to that used by Powell (1987, 2001) for identification of the regression coefficients in a censored selection model. Specifically, for a pair of observations with both conditional quantiles being positive, the difference in the quantiles is the difference in the regression functions plus the difference in the control functions. By restricting attention to pairs of observations with identical control variables $\mathbf{v}_i$, differences in the quantiles only involve differences in the regression function, which identifies $\boldsymbol{\beta}_0$ if the regressors are sufficiently variable conditional on the control variable. Algebraically, for a pair of observations $i$ and $j$ with

$$q_i > 0,\ q_j > 0, \qquad \text{and} \qquad \mathbf{v}_i = \mathbf{v}_j, \tag{2.12}$$

it follows that

$$q_i - q_j = (\mathbf{x}_i - \mathbf{x}_j)'\boldsymbol{\beta}_0, \tag{2.13}$$

which identifies $\boldsymbol{\beta}_0$ if the regressors are sufficiently variable given the condition (2.12). When $\mathbf{v}_i$ is defined by the additive reduced form (2.4), the relation (2.13) is equivalent to

$$q_i - q_j = (\boldsymbol{\pi}(\mathbf{z}_i) - \boldsymbol{\pi}(\mathbf{z}_j))'\boldsymbol{\beta}_0, \tag{2.14}$$

so identification of $\boldsymbol{\beta}_0$ follows if the first-stage fitted values $\boldsymbol{\pi}(\mathbf{z}_i)$ have a full rank distribution given (2.12).

Again following Powell (1987, 2001) and Ahn and Powell (1993), this identification strategy suggests a two-stage approach to estimation of $\beta_0$. In the first stage, nonparametric estimates of the conditional quantile $q_i \equiv Q_\alpha[y_i|\mathbf{x}_i, \mathbf{z}_i]$ and the control variable $\mathbf{v}_i \equiv \mathbf{x}_i^{(e)} - \boldsymbol{\pi}(\mathbf{z}_i)$ are obtained using suitable methods. The second stage estimates $\boldsymbol{\beta}_0$ using a weighted least-squares regression of all pairs of differences $\hat{q}_i - \hat{q}_j$ in estimated quantiles on the corresponding differences $\mathbf{x}_i - \mathbf{x}_j$ in regressors, using weights which are nonzero only if the quantile estimates are positive, the nonparametric estimates are sufficiently precise, and the estimated control variables are sufficiently close to each other. That is, the estimator $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}_0$ is defined as

$$\begin{aligned}\hat{\boldsymbol{\beta}} &= \left[\sum_{\iota<j} K_{\mathbf{v}}\left(\frac{\hat{\mathbf{v}}_i - \hat{\mathbf{v}}_j}{h_n}\right)\hat{t}_i\hat{t}_j(\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)'\right]^{-1} \\ &\quad \cdot \sum_{\iota<j} K_{\mathbf{v}}\left(\frac{\hat{\mathbf{v}}_i - \hat{\mathbf{v}}_j}{h_n}\right)\hat{t}_i\hat{t}_j(\mathbf{x}_i - \mathbf{x}_j)(\hat{q}_i - \hat{q}_j),\end{aligned} \tag{2.15}$$

where $K_{\mathbf{v}}(\cdot)$ is a kernel function (which integrates to one, vanishes outside a compact set, and satisfies other regularity conditions), $h_n$ is a sequence of scalar bandwidth terms which tend to zero with the sample size at an appropriate rate, and $\hat{t}_i$ is a "trimming" term, constructed so that $\hat{t}_i = 0$ unless the estimated quantiles $\hat{q}_i > 0$ and $\mathbf{x}_i$ and $\mathbf{v}_i$ fall in some compact set $\mathbf{S}$ (within which the first-stage nonparametric estimators are uniformly consistent at a suitable rate), i.e.,

$$\hat{t}_i = \omega(\hat{q}_i) \cdot 1\{(\mathbf{x}_i, \mathbf{z}_i) \in \mathbf{S}\}, \tag{2.16}$$

where $\omega(q)$ is chosen so that $\omega(q) = 0$ unless $q > \varepsilon$ for some $\varepsilon > 0$. The trimming term $\hat{t}_i$ serves two purposes: restricting the regressors to the (known) set $\mathbf{S}$ ensures that the first-stage estimator $\hat{q}_i$ is uniformly consistent for all terms appearing in the double summation, while the weighting term $\omega(\hat{q})$ asymptotically eliminates observations for which the true value of $q = Q_\alpha[y_i|\mathbf{x}_i, \mathbf{z}_i]$ equals the censoring value, zero, rather than $\mathbf{x}_i'\boldsymbol{\beta}_0 + \lambda_\alpha(\mathbf{v}_i)$. While setting a fixed, positive trimming threshold $\varepsilon$ for positivity of the estimated quantile $\hat{q}_i$ makes derivation of the large-sample properties of $\hat{\boldsymbol{\beta}}$ simpler, a more sophisticated argument would permit it to tend to zero at a slower rate than the rate of convergence of $\hat{q}_i$ to $q_i$ (specified in (3.4) below), and in practice $\varepsilon$ might be taken to be zero, so that all pairs of observations with positivity of both estimated quantiles are included in the relevant summations. Similarly, restriction of the observations to the compact set $\mathbf{S}$ simplifies the asymptotic arguments, but may not be necessary for the theoretical results or in practice.

The estimator $\hat{\boldsymbol{\beta}}$ is a variant of the estimator proposed by Ahn and Powell (1993) for the censored selection model; it differs only in the control variable used (first-stage residuals rather than the conditional probability of selection), form of the trimming term $\hat{t}_i$, and, most importantly, replacement of the dependent variable $y_i$ with its estimated conditional quantile $\hat{q}_i$.

## 3. Large Sample Properties

To derive the asymptotic distribution of the estimator $\hat{\boldsymbol{\beta}}$, we use a combination of regularity conditions and derivations taken from Ahn and Powell (1993), Khan and Powell (2001), and Chen and Khan (2001); the statement of these conditions are given in a mathematical appendix. Primitive conditions on the components of the model include the assumption that the conditional c.d.f of the error terms $u_i$ given the control variable $\mathbf{v}_i$ is smooth, with conditional density function bounded away from zero when $\mathbf{x}_i$ and $\mathbf{z}_i$ are in the compact set $\mathbf{S}$, ensuring uniqueness of the conditional quantile $\lambda_\alpha(\mathbf{v}_i) = Q_\alpha[u_i|\mathbf{v}_i] = Q_\alpha[u_i|\mathbf{x}_i, \mathbf{z}_i]$. Model regularity also involves assumptions which ensure that (i) the differences in regressors $\mathbf{x}_i - \mathbf{x}_j$ have full-dimensional support given positivity of $q_i$ and $q_j$ and equality of $\mathbf{v}_i$ and $\mathbf{v}_j$, (ii) the conditional quantile $q_i$ is continously distributed (possibly with zero density) in a neighborhood of zero, the censoring value, (iii) the conditional expectation function $\boldsymbol{\pi}(\mathbf{z}_i) \equiv E[\mathbf{x}_i^{(e)}|\mathbf{z}_i]$ is a smooth function of any continuously-distributed components of $\mathbf{z}_i$, and (iv) the control variable $\mathbf{v}_i$ is continuously distributed on the compact set $\mathbf{V} \equiv \{\mathbf{v} = \mathbf{x} - \boldsymbol{\pi}(\mathbf{z}) : (\mathbf{x}, \mathbf{z}) \in \mathbf{S}\}$. The restrictions on the joint support of $\mathbf{x}_i$ and $\mathbf{v}_i$ are somewhat weaker than imposed by Blundell and Powell (2001) for nonparametric identification of the average structural function (ASF), due to the assumed "single-index" structure of the underlying regression function for the censored latent variable.

One intermediate result following from these regularity conditions is that, if the nonparametric estimators $\hat{q}_i$ and $\hat{\mathbf{v}}_i$ were replaced by their true values $q_i$ and $\mathbf{v}_i$ (and analogously for the trimming terms $\hat{t}_l$ in (2.16)), the corresponding estimator

$$\begin{aligned} \widetilde{\boldsymbol{\beta}} &= \left[\sum_{\iota<j} K_{\mathbf{v}}\left(\frac{\mathbf{v}_i - \mathbf{v}_j}{h_n}\right) t_i t_j (\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)'\right]^{-1} \\ &\quad \cdot \sum_{\iota<j} K_{\mathbf{v}}\left(\frac{\mathbf{v}_i - \mathbf{v}_j}{h_n}\right) t_i t_j (\mathbf{x}_i - \mathbf{x}_j)(q_i - q_j) \end{aligned} \tag{3.1}$$

would converge to the true value $\boldsymbol{\beta}_0$ at a rate faster than the square root of the sample size, i.e.,

$$\sqrt{n}(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) = o_p(1). \tag{3.2}$$

Thus, the asymptotic variance of the limiting (normal) distribution of $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ will depend exclusively on the large-sample behavior of the first-stage nonparametric estimators of $q_i$ and $\mathbf{v}_i$. For the preliminary nonparametric estimator of the conditional quantile $q_i$, we adopt the "local polynomial quantile regression" estimator proposed by Chaudhuri (1991a,b). The implementation of this estimator is as described in, say, Khan and Powell (2001); in short, this estimator is the intercept term $\hat{q}_i$ in the minimization problem

$$\begin{pmatrix} \hat{q}_i \\ \hat{\boldsymbol{\gamma}}_i \end{pmatrix} = \arg\min_{q,\boldsymbol{\gamma}} \sum_{l=1}^{n} K_{\mathbf{w}} \left( \frac{\mathbf{w}_i - \mathbf{w}_l}{\delta_n} \right) \rho_\alpha(y_l - q - g(\mathbf{w}_i - \mathbf{w}_l; p, \boldsymbol{\gamma})), \tag{3.3}$$

where $\mathbf{w}_i$ is the $r$-dimensional vector of distinct components of $\mathbf{x}_i$ and $\mathbf{z}_i$, $K_w(\cdot)$ is an $r$-dimensional product kernel of $Uniform(-1/2, 1/2)$ densities (i.e., an indicator for inclusion in the unit hypercube centered at zero), $\delta_n$ is a bandwidth sequence which converges to zero as $n \to \infty$ at a particular rate (specified in, $\rho_\alpha(\cdot)$ is the check function defined in (2.9), and $g(\mathbf{w};p, \boldsymbol{\gamma})$ is a $p^{\text{th}}$-order multivariate polynomial in the vector argument $\mathbf{w}$, with zero intercept and remaining coefficients $\boldsymbol{\gamma}$. Heuristically, this estimator $\hat{q}_i$ is the fitted value at observation $i$ of a polynomial quantile regression estimator of $y$ on $\mathbf{w}$, using only those observations $l$ for which $\mathbf{w}_l$ is within $\delta_n/2$ of $\mathbf{w}_i$ (component-by-component).

Under the conditions imposed in the appendix – which include infinite smoothness of the unknown density and uncensored quantile function and a sufficiently-large polynomial degree $p$ – Lemma 4.1 of Chaudhuri, Doksum, and Samarov (1997) implies that the local polynomial quantile estimator $\hat{q}_i$ is uniformly consistent for $q_i$ for the untrimmed observations (using the true value $t_i \equiv \omega(q_i) \cdot 1\{(\mathbf{x}_i, \mathbf{z}_i) \in \mathbf{S}\}$ of the trimming term) at a rate slightly less than $\sqrt{n}$: specifically,

$$\max_i t_i |\hat{q}_i - q_i| = o_p(n^{-3/8}). \tag{3.4}$$

Furthermore, the difference between the estimator and its true value has an asymptotically linear representation

$$\hat{q}_i - q_i = \frac{1}{n\delta_n^s} \sum_{l=1}^{n} K_w \left( \frac{\mathbf{w}_i - \mathbf{w}_l}{\delta_n} \right) [f_{\mathbf{w}}(\mathbf{w}_i)]^{-1} \cdot \psi_l + R_{in}, \tag{3.5}$$

where $f_{\mathbf{w}}$ is the joint (discrete and continuous) density of $\mathbf{w}_i$, $s$ is the number of continuously-distributed components of $w_i$ (with remaining components assumed to have finite support), $\psi_l$ is the usual influence function term

$$\psi_l \equiv \left[ f_{y|\mathbf{x}.\mathbf{z}}(q_l) \right]^{-1} \cdot (1\{y_l \le q_l\} - \alpha) \tag{3.6}$$

for quantile regression, and the remainder term $R_{in}$ is negligible in its effect on the asymptotic distribution of the second-step estimator $\hat{\boldsymbol{\beta}}$. (More precisely, when expression (3.5) is inserted into the definition of $\hat{\boldsymbol{\beta}}$, the terms involving $R_{in}$ will converge to zero when normalized by $\sqrt{n}$.) By construction, the influence function term $\psi_l$ has conditional expectation zero given $\mathbf{x}_i$ and $\mathbf{z}_i$; it is the source of the contribution to the asymptotic variance of $\hat{\boldsymbol{\beta}}$ due to the preliminary nonparametric estimation of the conditional quantile $q_l$.

The other contribution to the asymptotic variance of the coefficient estimator $\hat{\boldsymbol{\beta}}$ comes from the preliminary nonparametric estimation of the control variable $\mathbf{v}_i$. Rather than impose a particular form of the estimator – e.g., kernel versus local polynomial regression – we instead impose a "high level" assumption on the estimator $\hat{\mathbf{v}}_i$, requiring it to have analogous properties to the local quantile regression estimator $\hat{q}_i$. Specifically, we assume it is also uniformly consistent for the untrimmed observations,

$$\max_i t_i ||\hat{\mathbf{v}}_i - \mathbf{v}_i|| = o_p(n^{-3/8}), \tag{3.7}$$

and has a similar asymptotic linear representation

$$\hat{\mathbf{v}}_i - \mathbf{v}_i = \frac{1}{nb_n^k} \sum_{l=1}^{n} K_{\mathbf{z}} \left( \frac{\mathbf{z}_i - \mathbf{z}_l}{b_n} \right) [f_{\mathbf{z}}(\mathbf{z}_i)]^{-1} \cdot \boldsymbol{\zeta}_l + \mathbf{r}_{in}, \tag{3.8}$$

where $\boldsymbol{\zeta}_j \equiv \boldsymbol{\zeta}(\mathbf{x}_j, \mathbf{z}_j)$ is an influence function with conditional mean zero given $\mathbf{z}_j$, $f_{\mathbf{z}}$ is the (mixed discrete and continuous) density of $\mathbf{z}_i$ (assumed to have $k$ continuously components and $\dim\{\mathbf{z}_i\} - k$ discrete components), $K_{\mathbf{z}}(\cdot)$ is a suitable kernel function, and $r_{in}$ is remainder term which is negligible in the asymptotic distribution of $\hat{\boldsymbol{\beta}}$. These conditions will hold if $\mathbf{v}_i$ is defined as $\mathbf{v}_i \equiv \mathbf{x}_i^{(e)} - E[\mathbf{x}_i^{(e)}|\mathbf{z}_i]$ and $\boldsymbol{\pi}(\mathbf{z}_i) \equiv E[\mathbf{x}_i^{(e)}|\mathbf{z}_i]$ is estimated either by kernel regression (with a higher-order bias-reducing kernel for $K_{\mathbf{z}}$) or local polynomial (mean) regression (with polynomial order $p$ sufficiently large), in which case the influence function term $\boldsymbol{\zeta}_i$ reduces to the negative of the control variable itself, i.e., $\boldsymbol{\zeta}_i = \boldsymbol{\pi}(\mathbf{z}_i) - \mathbf{x}_i^{(e)} = -\mathbf{v}_i$. The condition (3.8) might also be shown to hold for nonparametric estimators of the alternative versions of the control variable $\mathbf{v}_i$ discussed in the following section.

Substitution of the asymptotic linear representations (3.5) and (3.8) into expression (2.15) for the estimator $\hat{\boldsymbol{\beta}}$, along with the usual projection arguments for the resulting second-order U-statistic, yields an asymptotic linear representation for $\hat{\boldsymbol{\beta}}$ of the form

$$\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 = \boldsymbol{\Sigma}_{xx}^{-1} \cdot \frac{1}{n} \sum_{\iota=1}^{n} t_i f_{\mathbf{v}}(\mathbf{v}_i) \left\{ \psi_i + \frac{\partial \lambda_\alpha(\mathbf{v}_i)}{\partial \mathbf{v}'} \boldsymbol{\zeta}_i \right\} [\tau_i \mathbf{x}_i - \boldsymbol{\mu}_i] + o_p(n^{-1/2}), \tag{3.9}$$

where

$$\begin{aligned} t_i &\equiv \omega(q_i) \cdot 1\{(\mathbf{x}_i, \mathbf{z}_i) \in \mathbf{S}\}, & (3.10) \\ \tau_i &\equiv E[t_i|\mathbf{v}_i], & (3.11) \\ \boldsymbol{\mu}_i &\equiv E[t_i \mathbf{x}_i|\mathbf{v}_i], & (3.12) \end{aligned}$$

and

$$\boldsymbol{\Sigma}_{xx} \equiv E\left[t_i f_v(\mathbf{v}_i)(\tau_i \mathbf{x}_i - \boldsymbol{\mu}_i)(\tau_i \mathbf{x}_i - \boldsymbol{\mu}_i)'\right]. \tag{3.13}$$

Thus $\hat{\beta}$ will be asymptotically normal,

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \xrightarrow{d} N(0, \boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Omega}\boldsymbol{\Sigma}_{xx}^{-1}), \tag{3.14}$$

for

$$\begin{aligned} \boldsymbol{\Omega} &\equiv Var\left[t_i f_{\mathbf{v}}(\mathbf{v}_i)\left\{\psi_i + \frac{\partial \lambda(\mathbf{v}_i)}{\partial \mathbf{v}'}\boldsymbol{\zeta}_i\right\}[\tau_i \mathbf{x}_i - \boldsymbol{\mu}_i]\right] & (3.15) \\ &\equiv Var\left[\boldsymbol{\eta}_i\right], \end{aligned}$$

where

$$\boldsymbol{\eta}_i \equiv t_i f_{\mathbf{v}}(\mathbf{v}_i)\left\{\psi_i + \frac{\partial \lambda_\alpha(\mathbf{v}_i)}{\partial \mathbf{v}'}\boldsymbol{\zeta}_i\right\}[\tau_i \mathbf{x}_i - \boldsymbol{\mu}_i]. \tag{3.16}$$

An estimator of the asymptotic covariance matrix $\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Omega}\boldsymbol{\Sigma}_{xx}^{-1}$, useful for large-sample inference on $\boldsymbol{\beta}_0$, would straightforward to construct, but its consistency would require more structure on the nonparametric estimation of the control variable $\mathbf{v}_i$. A consistent estimator of the matrix $\boldsymbol{\Sigma}_{xx}^{-1}$ is a suitably-normalized version of the first matrix in the definition of $\hat{\boldsymbol{\beta}}$ in (2.15),

$$\hat{\boldsymbol{\Sigma}}_{xx}^{-1} \equiv \left[\binom{n}{2}^{-1} h_n^{-L} \sum_{\iota<j} K_{\mathbf{v}}\left(\frac{\hat{\mathbf{v}}_i - \hat{\mathbf{v}}_j}{h_n}\right) \hat{t}_i \hat{t}_j (\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)'\right]^{-1}; \tag{3.17}$$

demonstration of its consistency is a step in the derivation of the asymptotic linearity result (3.9). It is consistent estimation of the "middle matrix" $\Omega = Var[\boldsymbol{\eta}_i]$ that is more problematic, requiring an estimator of the influence function term $\boldsymbol{\eta}_i$. Given suitably consistent estimators $\hat{\psi}_i$ and $\hat{\boldsymbol{\zeta}}_i$ of the influence function terms $\psi_i$ and $\boldsymbol{\zeta}_i$ – that is, estimators satisfying

$$\frac{1}{n}\sum_{i=1}^{n} t_i \left\|\mathbf{x}_i\right\| \left[\left(\hat{\psi}_i - \psi_i\right)^2 + \left\|\hat{\boldsymbol{\zeta}}_i - \boldsymbol{\zeta}_i\right\|^2\right] \xrightarrow{p} 0 \tag{3.18}$$

as $n \to \infty$ – and defining

$$D_{\mathbf{v}}(\mathbf{u}) \equiv \frac{\partial K_{\mathbf{v}}(\mathbf{u})}{\partial \mathbf{u}'}, \tag{3.19}$$

a consistent estimator of $\boldsymbol{\Omega}$ is

$$\hat{\boldsymbol{\Omega}} \equiv \frac{1}{n} \sum_{i=1}^{n} \hat{\boldsymbol{\eta}}_i \hat{\boldsymbol{\eta}}_i', \tag{3.20}$$

where

$$\hat{\boldsymbol{\eta}}_i \equiv \frac{2}{n-1} \sum_{j \neq i, j=1}^{n} h_n^{-L} \left\{ K_{\mathbf{v}} \left( \frac{\hat{\mathbf{v}}_i - \hat{\mathbf{v}}_j}{h_n} \right) \left( \hat{\psi}_i - \hat{\psi}_j \right) + h_n^{-1} D_{\mathbf{v}} \left( \frac{\hat{\mathbf{v}}_i - \hat{\mathbf{v}}_j}{h_n} \right) \left( \hat{\boldsymbol{\zeta}}_i - \hat{\boldsymbol{\zeta}}_j \right) \right\} \hat{t}_i \hat{t}_j (\mathbf{x}_i - \mathbf{x}_j) \tag{3.21}$$

is the estimator of the influence function for $\hat{\boldsymbol{\beta}}$. Demonstration of the consistency of $\hat{\boldsymbol{\Omega}}$ under condition (3.18) would follow from arguments similar to those for, e.g., Theorem 6.1 of Powell (1987, 2001). However, this leaves open the question of construction of estimators $\hat{\psi}_i$ and $\hat{\boldsymbol{\zeta}}_i$ satisfying condition (3.18), which in turn would involve nonparametric estimation of the conditional density function $f_{y|\mathbf{x}.\mathbf{z}}(q_l)$ of the uncensored values of $y$ given $\mathbf{x}$ and $\mathbf{z}$ (a component of $\psi_i$ in (3.6)), as well as specification of a particular form for the control variable $\mathbf{v}_i$ (e.g., the first stage residuals in (2.4), for which $\hat{\boldsymbol{\zeta}}_i = \hat{\boldsymbol{\pi}}(\mathbf{z}_i) - \mathbf{x}_i^{(e)} = -\hat{\mathbf{v}}_i$).

## 4. Alternative Models and Estimation Methods

As noted above, the estimation approach adopted here is very similar to that proposed by Chen and Khan (2003) for the partially-linear censored regression model, i.e., the model

$$y_i = \max\{0, \mathbf{x}_i'\boldsymbol{\beta}_0 + \lambda(\mathbf{v}_i) + \varepsilon_i\}, \tag{4.1}$$

where the regressors $\mathbf{v}_i$ in the nonparametric component $\lambda(\mathbf{v}_i)$ are assumed known for all $i$, the nonparametric function $\lambda(\cdot)$ is smooth, and the error terms $\varepsilon_i$ have conditional $\alpha^{\text{th}}$ quantile equal to zero given $\mathbf{x}_i$ and $\mathbf{v}_i$ and have uniformly positive conditional densities at this quantile. Since the "control variable" $\mathbf{v}_i$ needs not be estimated in their setup, they sidestep the need for a "kernel matching" term $K_{\mathbf{v}}\left(h_n^{-1}(\mathbf{v}_i - \mathbf{v}_j)\right)$, instead using local polynomial quantile regression of $y_i$ on $\mathbf{x}_i$ and $\mathbf{v}_i$ to "difference off" the nonparametric components. That is, for the model (4.1), for a pair of observations $i$ and $j$, if the conditional quantile function of $y_i$ given $\mathbf{x}_i = \mathbf{x}$ and $\mathbf{v}_i = \mathbf{v}$ are evaluated at the two different $\mathbf{x}$ values but the same $\mathbf{v}$ value, and if

$$Q_\alpha[y|\mathbf{x}_i, \mathbf{v}_i] > 0,\ Q_\alpha[y|\mathbf{x}_j, \mathbf{v}_i] > 0, \tag{4.2}$$

then since $Q_\alpha[y|\mathbf{x}, \mathbf{v}] = \mathbf{x}'\boldsymbol{\beta}_0 + \lambda(\mathbf{v})$ when $Q_\alpha[y|\mathbf{x}, \mathbf{v}] > 0$, it follows that

$$Q_\alpha[y|\mathbf{x}_i, \mathbf{v}_i] - Q_\alpha[y|\mathbf{x}_j, \mathbf{v}_i] = (\mathbf{x}_i - \mathbf{x}_j)'\boldsymbol{\beta}_0 \tag{4.3}$$

when condition (4.2) is satisfied. As argued above, this implies that $\boldsymbol{\beta}_0$ will be identified if $\mathbf{x}_i - \mathbf{x}_j$ is sufficiently variable given $\mathbf{v}_i = \mathbf{v}_j$. In the first stage of their two-step procedure, Chen and Khan use a variant of the local polynomial quantile regression estimator defined in (3.3) (which substitutes $\mathbf{v}_i$ for $\mathbf{z}_i$ and exploiting the linearity of the conditional quantile as a function of $\mathbf{x}$) to construct estimators $\hat{Q}_{ii}$ and $\hat{Q}_{ji}$ of $Q_\alpha[y|x_i, v_i]$ and $Q_\alpha[y|x_j, v_i]$, respectively; their second step obtains a $\sqrt{n}$-consistent estimator of $\boldsymbol{\beta}_0$ from a least squares regression of the differences $\hat{Q}_{ii} - \hat{Q}_{ji}$ on $(\mathbf{x}_i - \mathbf{x}_j)$ across all pairs of observations,

$$\tilde{\boldsymbol{\beta}} \equiv \arg\min_{\boldsymbol{\beta}} \sum_{i,j} \hat{t}_{ij} \left(\hat{Q}_{ii} - \hat{Q}_{ji} - (\mathbf{x}_i - \mathbf{x}_j)'\boldsymbol{\beta}\right)^2, \tag{4.4}$$

where the "trimming terms" $\hat{t}_{ij}$ equal zero unless $\hat{Q}_{ii} > 0$, $\hat{Q}_{ij} > 0$, and $x_i$, $x_j$, and $v_i$ are in a compact set. The asymptotic distribution of this estimator has a similar form to the estimator $\hat{\boldsymbol{\beta}}$ proposed above, except for the absence of the $\boldsymbol{\zeta}_i$ component (since no first-stage estimation of $\mathbf{v}_i$ is needed), and a similar absence of the density function term $f_{\mathbf{v}}(\mathbf{v}_i)$ of the control variable (which arises for $\hat{\boldsymbol{\beta}}$ because of the kernel term $K_{\mathbf{v}}\left(h_n^{-1}(\mathbf{v}_i - \mathbf{v}_j)\right)$). To extend Chen and Khan's approach to the present context, the Chaudhuri's (1991a,b) large-sample results for the local polynomial quantile regression estimator would have to be extended to accommodate preliminary nonparametric estimation of the conditioning variables $\hat{\mathbf{v}}_i$, which seems a daunting theoretical challenge. In contrast, the estimation approach proposed here, though it requires nonparametric estimation of a higher-dimensional function (that is, the conditional quantile of $y_i$ given $\mathbf{x}_i$ and $\mathbf{z}_i$ rather than given $\mathbf{v}_i$ and a linear form in $\mathbf{x}_i$), can use Chaudhuri's results "off the shelf" in deriving the form of the second-stage estimator of $\beta_0$.

Another closely-related estimation approach is the pairwise-differencing estimator of the partially-linear censored regression model proposed by Honoré and Powell (2004). Under the distributional exclusion restriction (2.5), it is possible to construct functions of pairs of observations and the unknown regression coefficients that have conditional mean zero given $\mathbf{v}_i = \mathbf{v}_j$ and the regressors $\mathbf{x}_i$ and $\mathbf{x}_j$ when evaluated at the true parameter value $\boldsymbol{\beta}_0$, but have nonzero conditional mean otherwise. The form of this function was given by Honoré (1992) for estimation of censored regression models with panel data and fixed effects; in the present setting, for the model (2.1) under the distributional exclusion restriction, if $\mathbf{v}_i = \mathbf{v}_j$ for a pair of observations $i$ and $j$, and if

$$\varepsilon_{ij}(\boldsymbol{\beta}) \equiv \max\{y_i - \mathbf{x}_i'\boldsymbol{\beta}, -\mathbf{x}_j'\boldsymbol{\beta}\}, \tag{4.5}$$

then

$$\varepsilon_{ij}(\boldsymbol{\beta}_0) = \max\{u_i, -\mathbf{x}_i'\boldsymbol{\beta}_0, -\mathbf{x}_j'\boldsymbol{\beta}_0\}$$

has the same conditional distribution as

$$\varepsilon_{ji}(\boldsymbol{\beta}_0) = \max\{u_j, -\mathbf{x}_i'\boldsymbol{\beta}_0, -\mathbf{x}_j'\boldsymbol{\beta}_0\}, \tag{4.6}$$

so the difference $\varepsilon_{ij}(\boldsymbol{\beta}_0) - \varepsilon_{ji}(\boldsymbol{\beta}_0)$ is symmetrically distributed around zero given $\mathbf{v}_i = \mathbf{v}_j$ and $\mathbf{x}_i$, $\mathbf{x}_j$, $\mathbf{z}_i$, and $\mathbf{z}_j$. This generates conditional moment restrictions of the form

$$E[m\left(\varepsilon_{ij}(\boldsymbol{\beta}_0) - \varepsilon_{ji}(\boldsymbol{\beta}_0)\right)|\mathbf{v}_i = \mathbf{v}_j, \mathbf{x}_i, \mathbf{x}_j, \mathbf{z}_i, \mathbf{z}_j] = 0 \tag{4.7}$$

if $m(\cdot)$ is an odd function of its argument (and this expectation exists). To exploit this moment condition, Honoré and Powell propose estimation of $\boldsymbol{\beta}_0$ by

$$\tilde{\beta} = \arg\min_{\boldsymbol{\beta}} \sum_{\iota<j} K_{\mathbf{v}}\left(\frac{\hat{\mathbf{v}}_i - \hat{\mathbf{v}}_j}{h_n}\right) s(y_i, y_j, (\mathbf{x}_i - \mathbf{x}_j)'\boldsymbol{\beta}), \tag{4.8}$$

where $s(\cdot)$ is the criterion function from Honoré's (1992) panel-data censored regression estimator, $K_{\mathbf{v}}(\cdot)$ is a kernel (integrating to one), and $\hat{\mathbf{v}}_i$ is a consistent estimator of the control variable estimator of $\mathbf{v}_i$. Honoré and Powell assume this latter estimator is linear in a $\sqrt{n}$-consistent estimator, i.e., $\hat{\mathbf{v}}_i = \mathbf{x}_i - \hat{\mathbf{\Pi}}\mathbf{z}_i$ with $\hat{\mathbf{\Pi}} = \mathbf{\Pi}_0 + O_p(n^{-1/2})$ for some $\mathbf{\Pi}_0$, but extension of the analysis to permit a nonparametric first-stage estimation method should be feasible, following Ahn and Powell's (1993) similar extension of the censored selection estimator proposed by Powell (1987, 2001). This estimation approach would avoid the preliminary nonparametric estimation of the conditional quantile $q_i$ of $y_i$ given $\mathbf{x}_i$ and $\mathbf{z}_i$, but, unlike either the Chen and Khan (2003) estimator or the approach proposed in this paper, the second-step estimator $\tilde{\boldsymbol{\beta}}$ in (4.8) is not in "closed form," requiring minimization of a non-quadratic objective, which may make it more computationally burdensome for large sample sizes.

Yet another related estimation approach was proposed by Das, Newey, and Vella (2003) for the censored selection model with endogenous regressors. For this model, the latent dependent variable $y_i^* = g(\mathbf{x}_i)+u_i$ is observed only if some other latent variable $d_i^* = \tau(\mathbf{z}_i)+\varepsilon_i$ is positive, and the regressors $\mathbf{x}_i$ satisfy the same reduced form relation (2.4); the system of equations generating the observable variables $y_i$, $d_i$, and $\mathbf{x}_i$ from the instruments $\mathbf{z}_i$ and the unobservable errors thus takes the triangular form

$$\begin{aligned} y_i &= d_i \cdot (g(\mathbf{x}_i) + u_i)\,, \\ d_i &= 1\{\tau(\mathbf{z}_i) + \varepsilon_i \geq 0\}, \\ \mathbf{x}_i &= \boldsymbol{\pi}(\mathbf{z}_i) + \mathbf{v}_i, \end{aligned} \tag{4.9}$$

with $d_i$ and $\mathbf{x}_i$ first determined by $\mathbf{z}_i$ and the error terms $\varepsilon_i$ and $\mathbf{v}_i$, and $y_i$ then determined by $d_i$, $\mathbf{x}_i$, and the remaining error term $u_i$. This model clearly includes the censored regression model as a special case, taking $g(\mathbf{x}_i)$ to have a linear form $g(\mathbf{x}_i) = \mathbf{x}_i'\beta_0$, with $\boldsymbol{\tau}(\mathbf{z}_i) \equiv \boldsymbol{\pi}(\mathbf{z}_i)'\beta_0$ and $\varepsilon_i = u_i + \mathbf{v}_i'\beta_0$. For their general model, Das, Newey, and Vella note that the assumption that the errors $(u_i, \varepsilon_i, \mathbf{v}_i)$ are independent of the instruments $\mathbf{z}_i$ yields an additive form for the conditional mean of the uncensored values of the dependent variable $y_i$ :

$$E[y_i|d_i = 1, \mathbf{x}_i, \mathbf{z}_i] = g(\mathbf{x}_i) + \lambda(p_i, \mathbf{v}_i), \tag{4.10}$$

where $p_i = p(\mathbf{z}_i) = \Pr\{d_i = 1|\mathbf{z}_i\}$ is the propensity score,

$$p_i = p(\mathbf{z}_i) \equiv E[d_i|\mathbf{z}_i]. \tag{4.11}$$

Indeed, the additive form in (4.10) would follow from a weaker distributional exclusion restriction for the error terms $u_i$ and $\varepsilon_i$ given the first-stage error $\mathbf{v}_i$,

$$\Pr\{u_i \leq u, \varepsilon_i \leq e|\mathbf{v}_i, \mathbf{z}_i\} = \Pr\{u_i \leq u, \varepsilon_i \leq e|\mathbf{v}_i\}\,, \tag{4.12}$$

which would permit, say, conditional heteroskedasticity of $\mathbf{v}_i$ given $\mathbf{z}_i$ but still yield the mean exclusion restriction

$$E[u_i|d_i = 1, \mathbf{x}_i, \mathbf{z}_i] = \lambda(p_i, \mathbf{v}_i) \tag{4.13}$$

for the errors $u_i$ in the equation of interest.

While it is tempting to adopt this estimation approach for the special case of the censored selection model considered in this paper – specifically, replacing the nonparametric regression function $g(\mathbf{x}_i)$ in (4.10) with the linear form $\mathbf{x}_i'\boldsymbol{\beta}_0$ and fitting that equation via a least squares procedure – the relation (4.10) will not suffice to identify the entire $\boldsymbol{\beta}_0$ vector for the censored regression model (2.1) with reduced form (2.4) if the error terms $u_i$ and $\mathbf{v}_i$ are assumed independent of $\mathbf{z}_i$ (with invertible c.d.f.). As Robinson (1988) notes, a requirement for identification of $\boldsymbol{\beta}_0$ for estimation of the semilinear model (4.10) would be that the deviation of the regressors $\mathbf{x}_i$ from their conditional expectations given the control variables $p_i$ and $\mathbf{v}_i$, i.e.,

$$\mathbf{x}_i^* \equiv \mathbf{x}_i - E[\mathbf{x}_i|p_i, \mathbf{v}_i], \tag{4.14}$$

have a full dimensional distribution, so that, say, $E[\mathbf{x}_i^*\mathbf{x}_i^{*\prime}]$ would have full rank. However, since $\mathbf{v}_i = \mathbf{x}_i - \boldsymbol{\pi}(\mathbf{z}_i)$ from the reduced form,

$$\mathbf{x}_i^* \equiv \boldsymbol{\pi}(\mathbf{z}_i) - E[\boldsymbol{\pi}(\mathbf{z}_i)|p_i, \mathbf{v}_i]; \tag{4.15}$$

furthermore, for the censored regression model

$$\begin{aligned} p_i &= \Pr\{\mathbf{x}_i'\boldsymbol{\beta}_0 + u_i > 0|\mathbf{z}_i\} \\ &= \Pr\{\boldsymbol{\pi}(\mathbf{z}_i)'\boldsymbol{\beta}_0 + u_i + \mathbf{v}_i'\boldsymbol{\beta}_0 > 0|\mathbf{z}_i\} \\ &= \gamma(\boldsymbol{\pi}(\mathbf{z}_i)'\boldsymbol{\beta}_0) \end{aligned} \tag{4.16}$$

for an invertible function $\gamma(\cdot)$ if the composite error terms $u_i + \mathbf{v}_i'\boldsymbol{\beta}_0$ is independent of $\mathbf{z}_i$. This would imply that the deviations $\mathbf{x}_i^*$ would have a singular distribution with $Var(\mathbf{x}_i^{*\prime}\boldsymbol{\beta}_0) = 0$. Thus $\boldsymbol{\beta}_0$ would at best be identified up to a normalization unless $u_i + \mathbf{v}_i'\boldsymbol{\beta}_0$ is assumed *not* to be independent of $\mathbf{z}_i$, requiring non-independence of $\mathbf{v}_i$ and $\mathbf{z}_i$ if the distributional exclusion restriction (2.5) is imposed, and ruling out the leading case of independence of errors and instruments.

For similar reasons, the estimation approach for the endogenous censored regression model considered in the previous sections could not easily be generalized to permit the regressors $\mathbf{x}_i$ to be observed only when the censored dependent variable $y_i$ is positive, i.e., if the reduced form specification (2.4) is replaced with

$$\mathbf{x}_i = 1\{y_i > 0\} \cdot (\boldsymbol{\pi}(\mathbf{z}_i) + \mathbf{v}_i) \tag{4.17}$$

(as in the "Type 3 Tobit" model, in Amemiya's (1984) terminology). Estimation of the reduced-form regression function $\boldsymbol{\pi}(\mathbf{z}_i)$ (and thus the control variable $\mathbf{v}_i$) is problematic when the errors are independent of the instruments, since the propensity score (needed to control for selection in (4.17) would be a function of a linear combination of the same $\boldsymbol{\pi}(\mathbf{z}_i)$ that is the object of estimation, yielding a multicollinearity problem for the conditional mean of the uncensored $\mathbf{x}_i$ terms. Thus, identification of $\boldsymbol{\beta}_0$ when the regressors are only observed when $y_i > 0$ may require alternative assumptions to the distributional exclusion restrictions imposed here, e.g., joint symmetry of $u_i$ and $\mathbf{v}_i$ around zero given $\mathbf{z}_i$, as imposed by Honoré, Kyriazidou, and Udry (1997) for semiparametric estimation of the Type 3 Tobit model.

Though conditioning on positivity of the censored dependent variable $y_i$ causes difficulties for the censored regression model considered here, the present estimation approach can be adapted to permit $y_i$ and $\mathbf{x}_i$ to be subject to additional sample selection. Following Das, Newey, and Vella (2003), if

$$\begin{aligned} y_i &= d_i \cdot \max\{0, \mathbf{x}_i'\boldsymbol{\beta}_0 + u_i\}, \\ \mathbf{x}_i &= d_i \cdot (\boldsymbol{\pi}(\mathbf{z}_i) + \mathbf{v}_i), \\ d_i &= 1\{\tau(\mathbf{z}_i) + \varepsilon_i \geq 0\}, \end{aligned} \tag{4.18}$$

and the error terms are independent of $\mathbf{z}_i$, then the conditional quantile of $y_i$ given $\mathbf{x}_i$, $\mathbf{z}_i$, and $d_i = 1$ will take the form

$$Q_\alpha[y_i | d_i = 1, \mathbf{x}_i, \mathbf{z}_i] = \max\{0, \mathbf{x}_i'\boldsymbol{\beta}_0 + \lambda_\alpha(p_i, \mathbf{v}_i)\}, \tag{4.19}$$

and a generalization of the estimator in (2.15) would treat both the propensity score $p_i$ and the first-stage errors $\mathbf{v}_i$ as control variables, with a kernel weight depending upon both. Estimation of the control variable $\mathbf{v}_i$ using the selected observations on the endogenous variable $\mathbf{x}_i$ is the nonparametric censored selection problem treated by Das, Newey, and Vella.

The estimation approach considered here can also be extended to other structural forms for the control variable $\mathbf{v}_i$. For example, a special case of the "coherent simultaneous Tobit" model considered by Blundell and Smith (1989, 1994) has a pair of scalar dependent variables $y_1$ and $y_2$ determined by

$$\begin{aligned} y_1 &= \max\{0, \alpha_0 y_1 + \gamma_0 y_2 + \mathbf{z}_1'\boldsymbol{\beta}_0 + u\} \\ y_2 &= \delta_0 y_1 + \mathbf{z}_2'\boldsymbol{\pi}_0 + v, \end{aligned} \tag{4.20}$$

which requires the parametric restriction

$$\alpha_0 + \gamma_0\delta_0 = 0$$

for the model to be coherent, i.e., the model is assumed to take the form

$$\begin{aligned} y_1 &= \max\{0, \gamma_0(y_2 - \delta_0 y_1) + \mathbf{z}_1'\boldsymbol{\beta}_0 + u\} \\ y_2 &= \delta_0 y_1 + \mathbf{z}_2'\boldsymbol{\pi}_0 + v. \end{aligned} \tag{4.21}$$

For this model, the control variable $v_i$ would be the difference of a linear combination $y_2 - \delta_0 y_1$ from its conditional mean given the instruments $\mathbf{z}_i$, and instrumental variables estimation methods would be needed to estimate $\delta_0$ and $\boldsymbol{\pi}_0$ (and thus $v_i$) in the first stage.

Another example of an alternative control variable comes from the "nonadditive triangular model" setup of Imbens and Newey (2002), in which a single component $x_{1i}$ of the regressors $\mathbf{x}_i$ is assumed to be endogenous, with a nonseparable reduced form

$$x_{1i} = \phi(\mathbf{z}_i, v_i) \tag{4.22}$$

which is assumed invertible in $v_i$ for all $\mathbf{z}_i$ (but otherwise unknown). Imbens and Newey show how $\phi$ can be normalized so that $v_i$ is a "generalized propensity score,"

$$v_i = F_{x_1|\mathbf{z}}(x_{1i}|\mathbf{z}_i), \tag{4.23}$$

where $F_{x_1|\mathbf{z}}$ is the conditional c.d.f. of $x_{1i}$ given $\mathbf{z}_i$; thus the estimation of the control variable $v_i$ in the first stage of our approach would involve nonparametric estimation of the conditional cumulative distribution function of the endogenous regressor given the instruments $\mathbf{z}_i$.

Finally, though the estimator $\hat{\boldsymbol{\beta}} \equiv \hat{\boldsymbol{\beta}}_\alpha$ was defined in (2.15) only for a particular choice of $\alpha$, the distributional exclusion restriction (2.5) implies that estimators based upon different $\alpha$ values will be jointly $\sqrt{n}$-consistent and asymptotically normal for the same vector $\boldsymbol{\beta}_0$ of slope coefficients, so the distributional exclusion restriction can be tested using contrasts of $\hat{\boldsymbol{\beta}}_\alpha$ estimates across different values of $\alpha \in (0,1)$. Also, under the distributional exclusion restriction, a more efficient estimator of $\boldsymbol{\beta}_0$ can be obtained from an optimal linear combination of $\hat{\boldsymbol{\beta}}_\alpha$ estimators across different values of $\alpha$.

# 5. Appendix: Regularity Conditions

The following regularity conditions on the model and estimation procedure are sufficient (but by no means necessary) for the derivation of the asymptotically normal distribution of $\hat{\beta}$.

*Assumption M (Random Sampling and Model):* The random vectors $\{(u_i, \mathbf{x}_i', \mathbf{z}_i')\}$ are independently and identically distributed, and $y_i$ is generated from the censored regression model (2.1).

*Assumption R (Distribution of Regressors):* The vector $\mathbf{w}_i$ of distinct components of $\mathbf{x}_i$ and $\mathbf{z}_i$ is composed of two subvectors $\mathbf{w}_i^{(d)}$ and $\mathbf{w}_i^{(c)}$, where the the subvector $\mathbf{w}_i^{(d)}$ of discrete components has finite support (with marginal density $f_d(\mathbf{w}^{(d)})$) and the continuous components $\mathbf{w}_i^{(c)}$ have a conditional density function $f_c(\mathbf{w}^{(c)}|\mathbf{w}^{(d)})$ that is absolutely continuous and bounded away from zero on a known compact subset $\mathbf{S}$ of possible values of $\mathbf{w}_i$.

*Assumption E (Distribution of Errors):* The error terms $\{u_i\}$ satisfy the conditional quantile restriction (2.6), and for all $\mathbf{w}_i$ in $\mathbf{S}$, the conditional distribution of $u_i$ given $\mathbf{w}_i$ has a density function $f_{u|w}(u|\mathbf{w}_i) = f_{u|v}(u|\mathbf{v}_i)$ which is uniformly (in $\mathbf{v}_i$) positive when evaluated at its conditional $\alpha^{th}$ quantile $\lambda_\alpha(\mathbf{v}_i) \equiv Q_\alpha[u_i|\mathbf{v}_i]$.

*Assumption C (Control Variable):* The $L$-dimensional control variable $\mathbf{v}_i = \mathbf{x}_i^{(e)} - E[\mathbf{x}_i^{(e)}|\mathbf{z}_i]$ is continuously distributed with density function $f_{\mathbf{v}}(\mathbf{v})$ which is uniformly positive.

*Assumption S (Smoothness):* For all $\mathbf{w}_i$ in $\mathbf{S}$, the conditional densities $f_c$ and $f_{u|v}$ and the conditional quantile function $\lambda_\alpha$ defined in assumptions $R$ and $E$, as well as the density $f_{\mathbf{v}}$ of the control variable $\mathbf{v}_i$, are infinitely smooth in their arguments, i.e., continuously differentiable to arbitrary order.

*Assumption QE (Nonparametric Quantile Estimation):* The preliminary nonparametric estimator $\hat{q}_i$ of the conditional quantile

$$\begin{aligned} q_i &\equiv Q_\alpha[y_i|\mathbf{x}_i, \mathbf{z}_i] \\ &= \max\{0, \mathbf{x}_i'\boldsymbol{\beta}_0 + \lambda_\alpha(\mathbf{v}_i)\}, \end{aligned}$$

of $y_i$ given $\mathbf{x}_i$ and $\mathbf{z}_i$ is a local polynomial quantile regression estimator, as defined in (3.3), with polynomial order $p$ satisfying

$$p > 3s/2$$

where $s = \dim\{\mathbf{w}_i^{(c)}\}$, and the first-step bandwidth term $\delta_n$ satisfies

$$\delta_n = c^* n^{-\kappa}$$

for some $c^* > 0$ and $\kappa \in (1/(2p+s), 1/(4s))$.

*Assumption CE (Control Variable Estimation):* The preliminary nonparametric estimator $\hat{\mathbf{v}}_i$ of $\mathbf{v}_i \equiv E[\mathbf{x}_i|\mathbf{z}_i = \mathbf{z}]$ satisfies

$$\max_i t_i ||\mathbf{v}_i - \mathbf{v}_i|| = o_p(n^{-3/8}),$$

and has the asymptotically linear representation (3.8), with bandwidth $b_n$ satisfying

$$b_n^{-1} = o(n^{1/(4k)})$$

(for $k$ being the number of continuous components of $\mathbf{z}_i$), and with remainder term $\mathbf{r}_{in}$ satisfying

$$\max_i t_i ||\mathbf{r}_{in}|| = o_p(n^{-1/2})$$

and influence function $\boldsymbol{\zeta}_i$ having $E||\boldsymbol{\zeta}_i||^2 < \infty$.

*Assumption K (Second-Step Kernel):* The second-step kernel function $K_{\mathbf{v}}$ appearing in the definition (2.15) of $\hat{\boldsymbol{\beta}}$ is an $M^{th}$-order kernel of the form

$$K_{\mathbf{v}}(\mathbf{v}) = \sum_{j=1}^{M/2} a_j b_j^L \rho(b_j \mathbf{v}),$$

where

(i) the (even) integer $M$ has $M > 4(L+2)$, where $L = \dim(\mathbf{v}_i)$;

(ii) the function $\rho(\mathbf{v})$ is an infinitely smooth density function (nonnegative and integrating to one) which is symmetric about zero (i.e., $\rho(\mathbf{v}) = \rho(||\mathbf{v}||)$ ) and equals zero for $||\mathbf{v}|| \geq 1$;

(iii) the constants $b_1, ..., b_{M/2}$ are positive and distinct (but otherwise arbitrary); and

(iv) the constants $a_1, ..., a_{M/2}$ are chosen to satisfy the linear restrictions

$$\begin{aligned} \sum_j a_j &= 1, \\ \sum_j a_j b_j^{-2m} &= 0 \qquad \text{for} \qquad m = 1, ..., (M/2) - 1. \end{aligned}$$

*Assumption B2 (Second-Step Bandwidth):* The bandwidth sequence $h_n$ appearing in definition (2.15) of $\hat{\boldsymbol{\beta}}$ is of the form

$$h_n = c_n \cdot n^{-\gamma},$$

with $c_0 < c_n < c_0^{-1}$ for some $c_0 > 0$ and some $\gamma \in (1/(2M), (1-2\kappa s)/(4L+8))$, where $M$ is the order of the kernel given in assumption K, $\kappa$ is the first-step bandwidth rate given in assumption QE, $s = \dim\{\mathbf{w}_i^{(c)}\}$, and $L = \dim(\mathbf{v}_i)$.

*Assumption T (Trimming Function):* The weighting function $\omega(q)$ used to construct the trimming variable $\hat{t}_i$ in (2.16) is twice continuously differentiable (with bounded derivatives) and has $0 \leq \omega(q) \leq 1$, with $\omega(q) = 0$ if $q < \varepsilon$ for some $\varepsilon > 0$.

*Assumption I (Identification):* The matrix $\boldsymbol{\Sigma}_{xx}$ defined in (3.13) is positive definite.

Given this lengthy list of regularity conditions, the $\sqrt{n}$-consistency and asymptotic normality of $\hat{\boldsymbol{\beta}}$ indeed follows from the same arguments given in Ahn and Powell (1993) and Chen and Khan (2001). Rather than duplicate those arguments in detail here, we discuss how the imposed conditions yield the key steps in the derivation of the asymptotic distribution of the estimator.

Since the estimator $\hat{\boldsymbol{\beta}}$ involves three nonparametrically estimated components – the conditional quantile estimator $\hat{q}_i$ for $y_i$ given $\mathbf{x}_i$ and $\mathbf{z}_i$, the control variable $\hat{\mathbf{v}}_i$, and an implicit estimator of the control function $\lambda_\alpha(\mathbf{v}_i)$ in the second-step – their rates of convergence interact in the asymptotic behavior of $\hat{\boldsymbol{\beta}}$, and the convergence rates for their respective bandwidths are also interdependent. The strong smoothness assumptions on the unknown distribution functions imply that the orders of the kernel function and local polynomial can be taken to be large enough to ensure a fast $n^{-3/8}$ rate of convergence for all of the nonparametric estimators even though the corresponding bandwidths converge to zero slowly in $n$.

This uniform convergence rate is faster than the usual $n^{-1/4}$ rate in the previous papers. Writing

$$\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\Sigma}}_{\mathbf{xx}}^{-1}\hat{\boldsymbol{\Sigma}}_{\mathbf{x}q}, \tag{5.1}$$

with

$$\hat{\boldsymbol{\Sigma}}_{\mathbf{xx}} \equiv \binom{n}{2}^{-1} h_n^{-L} \sum_{\iota<j} K_{\mathbf{v}}\left(\frac{\hat{\mathbf{v}}_i - \hat{\mathbf{v}}_j}{h_n}\right) \hat{t}_i\hat{t}_j(\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)' \tag{5.2}$$

and

$$\hat{\boldsymbol{\Sigma}}_{\mathbf{x}q} \equiv \binom{n}{2}^{-1} h_n^{-L} \sum_{\iota<j} K_{\mathbf{v}}\left(\frac{\hat{\mathbf{v}}_i - \hat{\mathbf{v}}_j}{h_n}\right) \hat{t}_i\hat{t}_j(\mathbf{x}_i - \mathbf{x}_j)(\hat{q}_i - \hat{q}_j), \tag{5.3}$$

the presence of the term $h_n^{-L} = o(n^{1/8})$ makes the faster $n^{-3/8}$ rate (midway between $n^{-1/4}$ and $n^{-1/2}$) convenient when showing that any (quadratic) remainder terms are negligible when multiplied by $\sqrt{n}$.

Consistency of $\hat{\boldsymbol{\Sigma}}_{\mathbf{xx}}$ for $\boldsymbol{\Sigma}_{\mathbf{xx}}$ is a straightforward combination of the arguments in Lemma 3 of Chen and Khan (2001) and part (i) of Theorem 3.1 of Ahn and Powell (1993). Then, decomposing the $\hat{\mathbf{\Sigma}}_{\mathbf{x}q}$ term as

$$\begin{aligned}
\hat{\mathbf{\Sigma}}_{\mathbf{x}q} &\equiv \binom{n}{2}^{-1} h_n^{-L} \sum_{\iota<j} K_{\mathbf{v}}\left(\frac{\hat{\mathbf{v}}_i-\hat{\mathbf{v}}_j}{h_n}\right)[\hat{t}_i\hat{t}_j - t_it_j](\mathbf{x}_i-\mathbf{x}_j)(\hat{q}_i-\hat{q}_j) \\
&\quad + \binom{n}{2}^{-1} h_n^{-L} \sum_{\iota<j} K_{\mathbf{v}}\left(\frac{\hat{\mathbf{v}}_i-\hat{\mathbf{v}}_j}{h_n}\right) t_it_j(\mathbf{x}_i-\mathbf{x}_j)(q_i-q_j) \qquad (5.4)\\
&\quad + \binom{n}{2}^{-1} h_n^{-L} \sum_{\iota<j} K_{\mathbf{v}}\left(\frac{\mathbf{v}_i-\mathbf{v}_j}{h_n}\right) t_it_j(\mathbf{x}_i-\mathbf{x}_j)((\hat{q}_i-q_i)-(\hat{q}_j-q_j)) \\
&\quad + \mathbf{r}_n \qquad (5.5)\\
&\equiv \mathbf{S}_1+\mathbf{S}_2+\mathbf{S}_3+\mathbf{r}_n \qquad (5.6)
\end{aligned}$$

the remainder term $\mathbf{r}_n$, which involves quadratic terms in the differences between the nonparametric estimators and their estimators, is $o_p(n^{-1/2})$ for the reasons given in the previous paragraph. The term $\mathbf{S}_1$ can also be shown to be $o_p(n^{-1/2})$ using the same argument as in the proof of condition (A.10) in Chen and Khan (2001). Apart from the possibility that the dimension $L$ of the vector of control variables $\mathbf{v}_i$ may exceed one, the term $\mathbf{S}_2$ is identical to the term $\hat{\mathbf{S}}_{\mathbf{zv}}$ that appears in the proof of Theorem 3.1 of Ahn and Powell (1993); extension of those arguments to account for a preliminary $L$-dimensional nonparametric estimator yields

$$\sqrt{n}\mathbf{S}_2 = \frac{1}{\sqrt{n}}\sum_{\iota=1}^{n} t_i f_{\mathbf{v}}(\mathbf{v}_i)\left(\frac{\partial\lambda_\alpha(\mathbf{v}_i)}{\partial\mathbf{v}'}\right)\boldsymbol{\zeta}_i\left[\tau_i\mathbf{x}_i-\boldsymbol{\mu}_i\right]+o_p(1). \qquad (5.7)$$

Finally, adaptation of the proof of Lemmae 5 and 6 of Chen and Khan (2001) – replacing their fixed "selection function" $l(\mathbf{v}_i,\mathbf{v}_j)$ with the kernel weight $h_n^{-L}K_{\mathbf{v}}\left(h_n^{-1}(\mathbf{v}_i-\mathbf{v}_j)\right)$, which depends upon the sample size $n$ – implies that

$$\sqrt{n}\mathbf{S}_3 = \frac{1}{\sqrt{n}}\sum_{\iota=1}^{n} t_i f_{\mathbf{v}}(\mathbf{v}_i)\psi_i\left[\tau_i\mathbf{x}_i-\boldsymbol{\mu}_i\right]+o_p(1), \qquad (5.8)$$

where $\psi_i$ is defined in (3.6). Combining these results yields the asymptotically linear representation (3.9) for $\hat{\boldsymbol{\beta}}$, from which its asymptotic normal distribution follows.

## 6. References


Ahn, H. (1995), "Non-parametric Two Stage Estimation of Conditional Choice Probabilities in a Binary Choice Model under Uncertainty", *Journal of Econometrics*, 67, 337-378.

Ahn, H. and J.L. Powell (1993), "Semiparametric Estimation of Censored Selection Models with a Nonparametric Selection Mechanism," *Journal of Econometrics*, 58: 3-29.

Amemiya, T. (1984), *Advanced Econometrics*, Harvard University Press.

Altonji, J. and R.l. Matzkin (1997), "Panel Data Estimators for Nonseparable Models with Endogenous Regressors," manuscript, Northwestern University.

Blundell, R. and J.L. Powell (2001), "Endogeneity in Nonparametric and Semiparametric Regression Models," in Turnovsky, S. and L.P. Hansen, eds., *Advances in Economics and Econometrics: Proceedings from the Eighth World Congress*, Vol. 2, Cambridge University Press.

Blundell, R. and J.L. Powell (2004), "Endogeneity in Semiparametric Binary Response Models," *Review of Economic Studies*, forthcoming.

Blundell, R.W. and R.J. Smith (1989), "Estimation in a Class of Simultaneous Equation Limited Dependent Variable Models," *Review of Economic Studies*, 56: 37-58.

Blundell, R.W. and R.J. Smith (1994), "Coherency and Estimation in Simultaneous Models with Censored or Qualitative Dependent Variables," *Journal of Econometrics*, 64: 355-373.

Buchinsky, M. and J. Hahn (1998), "An Alternative Estimator for the Quantile Regression Model," *Econometrica*, 66: 653-672.

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., Doksum, K., and A. Samarov (1997), "On Average Derivative Quantile Regression," *Annals of Statistics*, 25: 715-744.

Chen, S. and S. Khan (1999), "$\sqrt{N}$-Consistent Estimation of Heteroskedastic Sample Selection Models," manuscript.

Chen, S. and S. Khan (2001), "Semiparametric Estimation of a Partially Linear Censored Regression Model," *Econometric Theory*, 17:567-590.

Chernozhukov, V. and C. Hansen (2005a), "An IV Model of Quantile Treatment Effects" *Econometrica*, forthcoming.

Chernozhukov, V. and C. Hansen (2005b), "Inference on the Instrumental Quantile Regression Process for Structural and Treatment Effect Models," *Journal of Econometrics*, forthcoming.

Chernozhukov, V. and H. Hong (2002), "Three-Step Censored Quantile Regression" *Journal of American Statistical Association*, 97: 872–882.

Das, M., W.K. Newey, and F. Vella (2003), "Nonparametric Estimation of Sample Selection Models," *Review of Economic Studies*, 70: 33-58.

Dhrymes, P.J (1970), *Econometrics: Statistical Foundations and Applications*, Springer-Verlag.

Florens, J.P., J.J. Heckman, C. Meghir, and E.J. Vytlacil (2003), "Instrumental Variables, Local Instrumental Variables, and Control Functions," working paper.

Heckman, J.J. (1976), "The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models," *Annals of Economic and Social Measurement*, 5: 475-492.

Heckman, J.J. (1979), "Sample Selection Bias as a Specification Error," *Econometrica*, 47: 153-162.

Hong, H. and E. Tamer (2003), "Inference in Censored Models with Endogenous Regressors," *Econometrica*, 71: 905–932.

Honoré, B.E (1992), "Trimmed LAD and Least Squares Estimation of Truncated and Censored Regression Models with Fixed Effects," *Econometrica*, 60: 533-565.

Honoré, B.E., E. Kyriazidou, and C. Udry (1997), "Estimation of Type 3 Tobit Models Using Symmetric Trimming and Pairwise Comparisons," *Journal of Econometrics*, 76: 107-128.

Honoré, B.E. and J.L. Powell (2004), "Pairwise Difference Estimators for Nonlinear Models," in Andrews, D.W.K. and J.H. Stock, eds., *Identification and Inference in Econometric Models*, Cambridge University Press, forthcoming.

Imbens, G.W. and W.K. Newey (2002), "Nonparametric Identification of Triangular Simultaneous Equations Models Without Additivity," manuscript.

Khan, S. and J.L. Powell (2001), "Two-Step Estimation of Semiparametric Censored Regression Models," *Journal of Econometrics*, 103: 73-110.

Nawata, K. (1992), Robust Estimation Based Upon Group-Adjusted Data In Censored Regression Models," *Journal of Econometrics*, 43: 337-362.

Newey, W.K., J.L. Powell, and F. Vella (1999), "Nonparametric Estimation of Triangular Simultaneous Equations Models," *Econometrica*, 67: 565-603.

Powell, J.L. (1984), "Least Absolute Deviations Estimation of 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. (1987), "Semiparametric Estimation of Bivariate Latent Variable Models," Social Systems Research Institute, University of Wisconsin-Madison, Working Paper No. 8704.

Powell, J.L. (1994), "Estimation of Semiparametric Models," in Engle, R. and D. McFadden, eds, *Handbook of Econometrics*, Vol. 4, Elsevier.

Powell, J.L. (2001), "Semiparametric Estimation of Censored Selection Models," in C. Hsiao, K. Morimune, and J.L. Powell, eds., *Nonlinear Econometric Inference*, Cambridge University Press.

Robinson, P.M. (1988), "Root $n$-consistent semiparametric regression", *Econometrica*, 56, 931-954.

Sawata, S. (2005), "Instrumental Variable Estimation Based on Conditional Median Restriction," manuscript, University of British Columbia.

Telser, L.G. (1964), "Iterative Estimation of a Set of Linear Regression Equations," *Journal of the American Statistical Association*, 59: 845-862.