Econometrica, Vol. 75, No. 4 (July, 2007), 1191–1208

NONPARAMETRIC INSTRUMENTAL VARIABLES ESTIMATION OF A QUANTILE REGRESSION MODEL

BY JOEL L. HOROWITZ AND SOKBAE LEE[1]

We consider nonparametric estimation of a regression function that is identified by requiring a specified quantile of the regression "error" conditional on an instrumental variable to be zero. The resulting estimating equation is a nonlinear integral equation of the first kind, which generates an ill-posed inverse problem. The integral operator and distribution of the instrumental variable are unknown and must be estimated nonparametrically. We show that the estimator is mean-square consistent, derive its rate of convergence in probability, and give conditions under which this rate is optimal in a minimax sense. The results of Monte Carlo experiments show that the estimator behaves well in finite samples.

KEYWORDS: Statistical inverse, endogenous variable, instrumental variable, optimal rate, nonlinear integral equation, nonparametric regression.

1. INTRODUCTION

QUANTILE REGRESSION MODELS are increasingly important in applied econometrics. This paper is concerned with nonparametric estimation of a quantile regression model that has a possibly endogenous explanatory variable and is identified through an instrumental variable. Specifically, we estimate the function g in the model

$$Y = g(X) + U, \tag{1.1}$$

$$\mathbf{P}(U \leq 0 | W = w) = q, \tag{1.2}$$

where Y is the dependent variable, X is an explanatory variable, W is an instrument for X, U is an unobserved random variable, and q is a known constant satisfying $0 < q < 1$. We do not assume that $\mathbf{P}(U \leq 0 | X = x)$ is independent of x. Therefore, the explanatory variable X may be endogenous in the quantile regression model (1.1)–(1.2). The function g is assumed to satisfy regularity conditions, but is otherwise unknown. In particular, it does not belong to a known, finite-dimensional parametric family. The data are an independent and identically distributed random sample, $\{Y_i, X_i, W_i : i = 1, \ldots, n\}$, of (Y, X, W). We present an estimator of g, derive its L_2 rate of convergence in probability, and provide conditions under which this rate is the fastest possible in a minimax sense.

[1]The research of Joel L. Horowitz was supported in part by NSF Grant 00352675. The work of both authors was supported in part by ESRC Grant RES-000-22-0704 and by the Leverhulme Trust through its funding of the Centre for Microdata Methods and Practice and the research program "Evidence, Inference, and Inquiry." We thank three anonymous referees for many helpful comments.

Model (1.1)–(1.2) subsumes the nonseparable quantile regression model of Chernozhukov, Imbens, and Newey (2007) (hereinafter CIN). CIN were concerned with identification and estimation of the function H in the model

$$Y = H(X, V), \tag{1.3}$$

where V is an unobserved, continuously distributed random variable that is independent of an instrument, W, H is strictly increasing in its second argument, and V has positive density on its support. It can be assumed without loss of generality that $V \sim U[0,1]$. Under these conditions, $P[Y - H(X,q) \leq 0|W = w] = q$ for any w in the support of W and any q such that $0 < q < 1$. Define $U = Y - H(X,q)$. Then (1.3) becomes

$$Y = H(X,q) + U; \quad P(U \leq 0|W = w) = q$$

for any $q \in (0,1)$. Thus, CIN's model is equivalent to model (1.1)–(1.2) with $g(X) = H(X,q)$.

Estimators of linear quantile regression models with endogenous right-hand side variables have been described by Amemiya (1982), Powell (1983), Chen and Portnoy (1996), Chernozhukov and Hansen (2006), Ma and Koenker (2006), Blundell and Powell (2007), Lee (2007), and Sakata (2007). Chernozhukov and Hansen (2004) and Januszewski (2002) used such models in economic applications. Research on nonparametric estimation of quantile regression models is relatively recent. Chesher (2003) investigated nonparametric identification of derivatives of the unknown functions in a triangular array structure. Chernozhukov and Hansen (2005) gave conditions under which g in (1.1)–(1.2) is identified. CIN gave conditions for consistency of a series estimator of g in (1.3). The estimator's rate of convergence is unknown.

There has also been research on nonparametric estimation of g in the model

$$Y = g(X) + U; \quad \mathbf{E}(U|W = w) = 0. \tag{1.4}$$

Here, as in (1.1)–(1.2), X is a possibly endogenous explanatory variable and W is an instrument for X, but the quantile restriction (1.2) is replaced by the condition $\mathbf{E}(U|W = w) = 0$. Blundell and Powell (2003), Darolles, Florens, and Renault (2006), Florens (2003), Newey and Powell (2003), Newey, Powell, and Vella (1999), Carrasco, Florens, and Renault (2005), Hall and Horowitz (2005), and Horowitz (2007) discussed estimators of g in (1.4). Our estimator of g in (1.1)–(1.2) is related to Hall and Horowitz's (2005) estimator of g in (1.4), but for reasons that will now be explained, estimating g in (1.1)–(1.2) presents problems that are different from those of estimating g in (1.4).

In both (1.1)–(1.2) and (1.4), the relation that identifies g is an operator equation,

$$\mathcal{T}g = \theta, \tag{1.5}$$

say, where $\mathcal{T}$ is a nonsingular integral operator and θ is a function. $\mathcal{T}$ and θ are unknown, but can be estimated consistently without difficulty. However, $\mathcal{T}^{-1}$ is discontinuous in both (1.1)–(1.2) and (1.4), so g cannot be estimated consistently by replacing $\mathcal{T}$ and θ in (1.5) with estimators. This "ill-posed inverse problem" is familiar in the literature on integral equations. See, for example, Groetsch (1984), Engl, Hanke, and Neubauer (1996), and Kress (1999). It is dealt with by regularizing (that is, modifying) $\mathcal{T}$ to make the resulting inverse operator continuous. As in Darolles, Florens, and Renault (2006) and Hall and Horowitz (2005), we use Tikhonov (1963a, 1963b) regularization. This consists of choosing the estimator $\hat{g}$ to solve

$$\text{(1.6)} \qquad \underset{\varphi\in\mathcal{G}}{\text{minimize}}: \|\hat{\mathcal{T}}\varphi - \hat{\theta}\|^2 + a_n\|\varphi\|^2,$$

where $\mathcal{G}$ is a set of functions; $\hat{\mathcal{T}}$ and $\hat{\theta}$, respectively, are consistent estimators of $\mathcal{T}$ and θ; $\{a_n\}$ is a sequence of nonnegative constants that converges to 0 as $n \to \infty$; and

$$\|\nu\|^2 = \int \nu(x)^2\,dx$$

for any square integrable function ν.[2] In model (1.4), $\mathcal{T}$ and $\hat{\mathcal{T}}$ are linear operators, and the first-order condition for (1.6) is a linear integral equation (a Fredholm equation of the first kind). In (1.1)–(1.2), however, $\mathcal{T}$ and $\hat{\mathcal{T}}$ are nonlinear operators, and the first-order condition for (1.6) is a nonlinear integral equation.

The nonlinearity of $\mathcal{T}$ in (1.1)–(1.2) complicates the task of deriving the rate of convergence of $\hat{g}$ to g. In contrast to many other nonlinear estimation problems, using a Taylor series expansion to make a linear approximation to the first-order condition is unattractive because, as a consequence of the ill-posed inverse problem, the approximation error dominates other sources of estimation error and controls the rate of convergence of $\hat{g}$ unless very strong assumptions are made about the probability density function of (Y, X, W). To avoid such assumptions, we use a modified version of a method developed by Engl, Hanke, and Neubauer (1996, Theorem 10.7) to derive the rate of convergence of $\hat{g}$. This method works directly from the objective function in (1.6), rather than the first-order condition.

[2]More generally, the penalization term could be $a_n\|\varphi - \varphi^*\|$ for some function $\varphi^* \in \mathcal{G}$. This does not affect identification of g or the consistency and rate of convergence of our estimator under our assumptions. It might affect the finite-sample performance of the estimator if one knows a φ^* that is close to the unknown function g, but other estimation approaches that may have better properties are also available if one has such information. Here, we assume that such information is not available.

Section 2 presents the estimator for the case in which X and W are scalars. Section 3 gives conditions under which the estimator is consistent and has the fastest possible rate of convergence in a minimax sense. Section 4 extends these results to a multivariate model in which X is a vector that may have some exogenous components. Section 5 summarizes a Monte Carlo investigation of the estimator's finite-sample performance. Section 6 gives concluding comments. The proofs of theorems are in the electronic supplement to this article (Horowitz and Lee (2007)).

2. THE ESTIMATOR

This section describes our estimator of g in (1.1)–(1.2) when X and W are scalars. Let $F_{Y|XW}$ denote the distribution function of Y conditional on (X, W). We assume that the conditional distribution of Y has a probability density function, $f_{Y|XW}$, with respect to Lebesgue measure. We also assume that (X, W) has a probability density function with respect to Lebesgue measure, f_{XW}. Let f_W denote the marginal density of W. Define $F_{YXW} = F_{Y|XW} f_{XW}$ and $f_{YXW} = f_{Y|XW} f_{XW}$. Assume without loss of generality that the support of (X, W) is contained in $[0, 1]^2$.

It follows from (1.1)–(1.2) that

$$\int_0^1 F_{YXW}[g(x), x, w]\, dx = q f_W(w) \tag{2.1}$$

for almost every w. We assume that (2.1) uniquely identifies g up to a set of x values whose Lebesgue measure is 0. Chernozhukov and Hansen (2005, Theorem 4) gave sufficient conditions for this assumption to hold.

Now define the operator $\mathcal{T}$ on $L_2[0, 1]$ by

$$(\mathcal{T}\varphi)(w) = \int_0^1 F_{YXW}[\varphi(x), x, w]\, dx, \tag{2.2}$$

where φ is any function in $L_2[0, 1]$. Then (2.1) is equivalent to the operator equation

$$\mathcal{T} g = q f_W.$$

Identifiability of g is equivalent to assuming that $\mathcal{T}$ is invertible. Thus,

$$g = q\mathcal{T}^{-1} f_W. \tag{2.3}$$

However, $\mathcal{T}^{-1}$ is discontinuous because the Fréchet derivative of $\mathcal{T}$ is a compact operator and, therefore, has an unbounded inverse if f_{YXW} is "well behaved." Consequently, g cannot be estimated consistently by replacing $\mathcal{T}$ and

f_W with consistent estimators in (2.3). As was explained in Section 1, we use Tikhonov regularization to deal with this problem.

We now describe the version of problem (1.6) that we solve to obtain the regularized estimator of g. We assume that f_{YXW} has $r_{yx} > 0$ continuous derivatives with respect to any combination of its first two arguments and r_w continuous derivatives with respect to its third argument. Our results hold even if f_{YXW} or its derivatives are discontinuous at one or more boundaries of the support of (Y, X, W) if the kernel function K that we now define is replaced by a boundary kernel. Let K denote a continuously differentiable kernel function whose support is $[-1, 1]$, that is symmetrical about 0, and that satisfies

$$\int_{-1}^{1} \nu^j K(\nu)\, d\nu = \begin{cases} 1, & \text{if } j = 0, \\ 0, & \text{if } 1 \le j \le \max(1, r-1), \end{cases} \tag{2.4}$$

with $r = r_{yx}$ or r_w. Let $h_{yx}, h_w > 0$ denote bandwidth parameters. Define $K_h(\nu) = K(\nu/h)$ for $h = h_{yx}$ or h_w and any $\nu \in [-1, 1]$. We estimate f_W, f_{YXW}, and F_{YXW}, respectively, by

$$\hat{f}_W(w) = \frac{1}{nh_w} \sum_{i=1}^{n} K_{h_w}(w - W_i),$$

$$\hat{f}_{YXW}(y, x, w) = \frac{1}{nh_{yx}^2 h_w} \sum_{i=1}^{n} K_{h_{yx}}(y - Y_i) K_{h_{yx}}(x - X_i) K_{h_w}(w - W_i),$$

and

$$\hat{F}_{YXW}(y, x, w) = \int_{-\infty}^{y} \hat{f}_{YXW}(\nu, x, w)\, d\nu.$$

Define the operator $\hat{\mathcal{T}}$ by

$$(\hat{\mathcal{T}}\varphi)(w) = \int_0^1 \hat{F}_{YXW}[\varphi(x), x, w]\, dx$$

for any $\varphi \in L_2[0, 1]$. Let $\|\varphi\|$ denote the L_2 norm of φ. Define $\mathcal{G} = \{\varphi \in L_2[0, 1] : \|\varphi\|^2 \le M\}$ for some constant $M < \infty$. Our estimator of g is any solution to the problem

$$\hat{g} = \arg\min_{\varphi \in \mathcal{G}} S_n(\varphi) \equiv \|\hat{\mathcal{T}}\varphi - q\hat{f}_W\|^2 + a_n \|\varphi\|^2. \tag{2.5}$$

Under our assumptions, a function $\hat{g}$ that minimizes S_n always exists, though it may not be unique (Bissantz, Hohage, and Munk (2004)).

3. THEORETICAL PROPERTIES

This section gives conditions under which the estimator $\hat{g}$ of Section 2 is consistent and under which $\|\hat{g}-g\|^2 \to^p 0$ at the fastest possible rate in a minimax sense.

3.1. *Consistency*

This section gives conditions under which $\mathbf{E}\|\hat{g}-g\|^2 \to 0$ as $n \to \infty$.

ASSUMPTION 1: (a) *The function g is identified. That is,* (2.1) *specifies $g(x)$ uniquely up to a set of x values whose Lebesgue measure is* 0. (b) $\|g\|^2 \le M$ *for some constant $M < \infty$.*

ASSUMPTION 2: *f_{YXW} has $r_{yx} > 0$ continuous derivatives with respect to any combination of its first two arguments and $r_w > 0$ continuous derivatives with respect to its third argument. The mixed partial derivatives $D_y^{r_1} D_x^{r_2} D_w^{r_3} f_{YXW}$ exist whenever $r_1 + r_2 < r_{yx}$ and $r_3 < r_w$. The derivatives and f_{YXW} are bounded in absolute value by M.*

Define $\delta_n = h_{yx}^{2r_{yx}} + h_w^{2r_w} + 1/(nh_w)$.

ASSUMPTION 3: *As $n \to \infty$, $a_n \to 0$, $\delta_n \to 0$, and $\delta_n/a_n \to 0$.*

ASSUMPTION 4: *The kernel function K is supported on* $[-1, 1]$, *is continuously differentiable, is symmetrical about* 0, *and satisfies* (2.4).

In Assumption 2, r_{yx} and r_w need not be integers. If r_{yx} or r_w is not an integer, then the assumption means that

$$\begin{aligned}&\left|D^{[r]}[f_{YXW}(y_1, x_1, w_1) - f_{YXW}(y_2, x_2, w_2)]\right|\\&\quad \le M\|(y_1, x_1, w_1) - (y_2, x_2, w_2)\|^{r-[r]},\end{aligned}$$

where $r = r_{yx}$ or r_w, $[r]$ is the integer part of r, $D^{[r]} f_{YXW}$ denotes an order $[r_{yx}]$ derivative of f_{YXW} with respect to its first two arguments or the $[r_w]$ derivative of f_{YXW} with respect to its third argument, and $\|z_1 - z_2\|$ is the Euclidean distance between the points z_1 and z_2.

Assumptions 1(b) and 2 are standard in nonparametric estimation. Assumptions 3 and 4 are satisfied by a wide range of choices of h_{yx}, h_w, a_n, and K. The choice of h_w in applications is discussed in Section 3.2.

THEOREM 1: *Let Assumptions* 1–4 *hold. Then*

$$\lim_{n\to\infty} \mathbf{E}\|\hat{g}-g\|^2 = 0. \tag{3.1}$$

Result (3.1) implies that $\|\hat{g}-g\|^2$ converges in probability to 0 as $n \to \infty$. The next section obtains the rate of convergence.

3.2. *Rate of Convergence*

In model (1.4), where $\mathcal{T}$ is a linear operator, the source of the ill-posed inverse problem is that the sequence of singular values of $\mathcal{T}$ (or, equivalently, eigenvalues of $\mathcal{T}^*\mathcal{T}$, where $\mathcal{T}^*$ is the adjoint of $\mathcal{T}$) converges to 0. Consequently, the rate at which $\|\hat{g} - g\|^2$ converges to 0 depends on the rate of convergence of the singular values (or eigenvalues). See Hall and Horowitz (2005). In (1.1)–(1.2), where $\mathcal{T}$ is nonlinear, the source of the ill-posed inverse problem is convergence to 0 of the singular values of the Fréchet derivative of $\mathcal{T}$ at g. Denote this derivative by T_g. The rate of convergence of $\|\hat{g} - g\|^2$ in (1.1)–(1.2) depends on the rate of convergence of the singular values of T_g or, equivalently, of the eigenvalues of $T_g^* T_g$, where T_g^* is the adjoint of T_g. Accordingly, the regularity conditions for our rate of convergence result are framed in terms of the spectral representation of $T_g^* T_g$.

The Fréchet derivative of $\mathcal{T}$ at g is the operator T_g defined by

$$(T_g \varphi)(w) = \int_0^1 f_{YXW}[g(x), x, w]\varphi(x)\, dx.$$

The adjoint operator is defined by

$$(T_g^* \varphi)(w) = \int_0^1 f_{YXW}[g(w), w, x]\varphi(x)\, dx.$$

We assume that $T_g^* T_g$ is nonsingular, so its eigenvalues are strictly positive. Let $\{\lambda_j, \phi_j : j = 1, 2, \ldots\}$ denote the eigenvalues and orthonormal eigenvectors of $T_g^* T_g$ ordered so that $\lambda_1 \geq \lambda_2 \geq \cdots > 0$. Under our assumptions, $T_g^* T_g$ is a compact operator, so $\{\phi_j\}$ forms a basis for $L_2[0, 1]$. Therefore, we may write

$$g(x) = \sum_{j=1}^{\infty} b_j \phi_j(x), \tag{3.2}$$

where the Fourier coefficients b_j are given by

$$b_j = \int_0^1 g(x)\phi_j(x)\, dx.$$

We make the following additional assumptions.

ASSUMPTION 5: (a) *There are constants* $\alpha > 1$, $\beta > 1$, *and* $C_0 < \infty$ *such that* $\beta - 1/2 \leq \alpha < 2\beta - 1$, $j^{-\alpha} \leq C_0 \lambda_j$, *and* $|b_j| \leq C_0 j^{-\beta}$ *for all* $j \geq 1$. (b) *Moreover,* $r_w \geq (2\beta + \alpha - 1)/2$ *and* $r_{yx} > (2\beta + \alpha - 1)/\alpha$.

ASSUMPTION 6: *There is a finite constant* $L > 0$ *such that*

$$\|\mathcal{T}(g_1) - \mathcal{T}(g_2) - T_{g_2}(g_1 - g_2)\| \leq 0.5L\|g_1 - g_2\|^2 \tag{3.3}$$

for any $g_1, g_2 \in L_2[0,1]$ *and*

$$\sum_{j=1}^{\infty} \frac{b_j^2}{\lambda_j} < \frac{1}{L}. \tag{3.4}$$

ASSUMPTION 7: *The tuning parameters* h_{yx}, h_w, *and* a_n *satisfy* $h_{yx} = C_{h_{yx}} n^{-\gamma}$, $h_w = C_{h_w} n^{-1/(2r_w+1)}$, *and* $a_n = C_a n^{-\alpha/(2\beta+\alpha)}$, *where* $C_{h_{yx}}$, C_{h_w}, *and* C_a *are positive, finite constants, and* γ *is a constant satisfying*

$$\frac{2\beta+\alpha-1}{2r_{yx}(2\beta+\alpha)} < \gamma < \frac{\alpha}{2(2\beta+\alpha)}.$$

In Assumption 5(a), α characterizes the severity of the ill-posed inverse problem. As α increases, the problem becomes more severe and the fastest possible rate of convergence of any estimator of g decreases. The parameter β characterizes the complexity of g. If β is large, then g can be well approximated by the finite-dimensional parametric model that is obtained by truncating the series on the right-hand side of (3.2) at $j = J$ for some small integer J. A finite-dimensional g can be estimated with a $n^{-1/2}$ rate of convergence, so the rate of convergence of $\|\hat{g} - g\|^2$ increases as β increases.[3] In Assumption 7, the rate of convergence of h_w is the usual one for nonparametric density estimation. Accordingly, in applications, h_w can be chosen using cross-validation or any other bandwidth selection method for nonparametric density estimation. Methods for selecting h_{yx} and a_n in applications are not yet available.

Assumption 5(a) places tight restrictions on the rate of convergence of λ_j. Among other things, it rules out distributions of (Y, X, W) for which λ_j decreases at an exponential rate as j increases. We suspect that $\|\hat{g} - g\|^2$ converges at a logarithmic rate in this case, but we cannot verify this conjecture because our method for deriving the rate of convergence does not work with exponentially decreasing λ_j's. We note that the restrictions of Assumption 5 are not needed for consistent estimation. It may be possible to relax Assumption 5 by using a non-Tikhonov regularization method. Carrasco, Florens, and Renault (2005), Bissantz, Hohage, Munk, and Ruymgaart (2006), and Kaltenbacher and Neubauer (2006) discussed several such methods for problems in which $\mathcal{T}$ in (1.4) is known, but additional research is needed to determine the extent to which these methods are useful in nonlinear problems with an unknown $\mathcal{T}$.

[3]Assumption 5 is motivated by the regularity conditions of Hall and Horowitz (2005). It implies that g is in the regularity space Φ_μ defined by Darolles, Florens, and Renault (2006) and Carrasco, Florens, and Renault (2005) for $1 \leq \mu < (2\beta - 1)/\alpha$. The numerical analysis literature imposes regularity on g and $\mathcal{T}$ by imposing a "source condition" that requires the existence of a function ν_g such that $g = (T_g^* T_g)^\mu \nu_g$ for some $\mu > 0$. Assumption 5 implies satisfaction of the source condition for $1 \leq \mu < (2\beta - 1)/\alpha$.

Assumptions 2 and 5(a) imply that $T_g^* T_g$ is a compact linear operator, so its eigenvectors form a basis for $L_2[0,1]$. Assumption 2 implies that inequality (3.3) holds for some $L < \infty$, so (3.3) amounts to the definition of L. Inequality (3.4) restricts the norm of the second Fréchet derivative of $\mathcal{T}$. It is shown in the supplement to this paper that a sufficient condition for (3.4) is

$$\sup_{y,x,w}\left|\frac{\partial f_{YXW}(y,x,w)}{\partial y}\right| < \left(\sum_{j=1}^{\infty}\frac{b_j^2}{\lambda_j}\right)^{-1}. \tag{3.5}$$

Restrictions similar to (3.4) are well known in the theory of nonlinear integral equations. It is an open question whether an estimator of g that has our rate of convergence can be achieved without making an assumption similar to (3.4). Assumptions 2 and 5(b) imply that f_W and $\mathcal{T}\varphi$ for any $\varphi \in L_2[0,1]$ can be estimated with a rate of convergence in probability of $O_p[n^{-(\beta-1/2+\alpha/2)/(2\beta+\alpha)}]$ and that f_{YXW} can be estimated with a rate of convergence of $O_p[n^{-(\beta-1/2)/(2\beta+\alpha)}]$. These rates are needed to obtain our rate of convergence of $\hat{g}$.

Let $\mathcal{H} = \mathcal{H}(M, C_0, \alpha, \beta, L)$ be the set of distributions of (Y, X, W) that satisfy Assumptions 1, 2, 5, and 6 with fixed values of M, C_0, α, β, and L. Our rate-of-convergence result is given by the following theorem.

THEOREM 2: *Let Assumptions 1, 2, and 4–7 hold. Then*

$$\lim_{D\to\infty}\limsup_{n\to\infty}\sup_{H\in\mathcal{H}}\mathbf{P}_H[\|\hat{g}-g\|^2 > Dn^{-(2\beta-1)/(2\beta+\alpha)}] = 0.$$

As expected, the rate of convergence of $\|\hat{g}-g\|^2$ decreases as α increases (the ill-posed inverse problem becomes more severe) and increases as β increases (g increasingly resembles a finite-dimensional parametric model).

The next theorem shows that the convergence rate in Theorem 2 is optimal in a minimax sense under our assumptions. Let $\{\tilde{g}_n\}$ denote any sequence of estimators of g satisfying $\|\tilde{g}_n\|^2 \le M$ for some $M < \infty$.

THEOREM 3: *Let Assumptions 1, 2, and 4–7 hold. Then for every finite $D > 0$,*

$$\liminf_{n\to\infty}\sup_{H\in\mathcal{H}}\mathbf{P}_H[\|\tilde{g}_n-g\|^2 > Dn^{-(2\beta-1)/(2\beta+\alpha)}] > 0.$$

4. MULTIVARIATE MODEL

This section extends the results of Sections 2 and 3 to a multivariate model in which X is a vector that may have some exogenous components. We rewrite model (1.1)–(1.2) as

$$Y = g(X, Z) + U \tag{4.1}$$

$$\mathbf{P}(U \le 0 | W = w, Z = z) = q, \tag{4.2}$$

where Y is the dependent variable, $X \in \mathbb{R}^c$ $(c \geq 1)$ is a vector of possibly endogenous explanatory variables, $Z \in \mathbb{R}^m$ $(m \geq 0)$ is a vector of exogenous explanatory variables, $W \in \mathbb{R}^\ell$ $(\ell \geq c)$ is a vector of instruments for X, U is an unobserved random variable, and q is a known constant satisfying $0 < q < 1$. If $m = 0$, then Z is not in the model. The problem is to estimate g nonparametrically from data consisting of the independent random sample $\{Y_i, X_i, W_i, Z_i : i = 1, \ldots, n\}$.

The estimator is obtained by applying the method of Section 2 after conditioning on Z. To do this, let $K_{\ell,h}(\nu) = \prod_{j=1}^{\ell} K_h(\nu^j/h)$, where h is a bandwidth and ν^j is the jth component of the ℓ-vector ν. Define $K_{m,h}(\nu)$ for an m-vector ν similarly. Let f_{WZ} and f_{YXWZ}, respectively, denote the probability density functions of (W, Z) and (Y, X, W, Z). Define

$$F_{YXWZ}(y, x, w, z) = \int_{-\infty}^{y} f_{YXWZ}(\nu, x, w, z)\, d\nu.$$

For bandwidths h_{wz} and h_{yx}, define kernel estimators of f_{WZ}, f_{YXWZ}, and F_{YXWZ} as

$$\hat{f}_{WZ}(w, z) = \frac{1}{nh_{wz}^{\ell+m}} \sum_{i=1}^{n} K_{\ell,h_{wz}}(w - W_i) K_{m,h_{wz}}(z - Z_i),$$

$$\hat{f}_{YXWZ}(w, z) = \frac{1}{nh_{yx}^{c+1} h_{wz}^{\ell+m}} \sum_{i=1}^{n} K_{h_{yx}}(y - Y_i) K_{c,h_{yx}}(x - X_i) \times K_{\ell,h_{wz}}(w - W_i) K_{m,h_{wz}}(z - Z_i),$$

and

$$\hat{F}_{YXWZ}(y, x, w, z) = \int_{-\infty}^{y} \hat{f}_{YXWZ}(\nu, x, w, z)\, d\nu.$$

For each $z \in [0, 1]^m$, define the operators $\mathcal{T}_z$ and $\hat{\mathcal{T}}_z$ on $L_2[0, 1]^c$ by

$$(\mathcal{T}_z \varphi_z)(w) = \int_{[0,1]^c} F_{YXWZ}[\varphi_z(x), x, w, z]\, dx$$

and

$$(\hat{\mathcal{T}}_z \varphi_z)(w) = \int_{[0,1]^c} \hat{F}_{YXWZ}[\varphi_z(x), x, w, z]\, dx,$$

where φ_z is any function on $L_2[0, 1]^c$.

The function g satisfies

$$(\mathcal{T}_z g)(w, z) = q f_{WZ}(w, z). \tag{4.3}$$

The function $g(\cdot, z)$ is identified if (4.3) has a unique solution (up to a set of x values of Lebesgue measure 0) for the specified z. Define $\mathcal{G} = \{\varphi \in L_2[0,1]^c : \|\varphi\|^2 \le M\}$ for some constant $M < \infty$. For each $z \in [0,1]^m$, the estimator of $g(\cdot, z)$ is any solution to the problem

$$\hat{g}(\cdot, z) = \arg\min_{\varphi_z \in \mathcal{G}} \left\{ \int_{[0,1]^\ell} [(\hat{\mathcal{T}}_z \varphi_z)(w) - q\hat{f}_{WZ}(w,z)]^2 \, dw + a_n \int_{[0,1]^\ell} \varphi_z(w)^2 \, dw \right\}. \tag{4.4}$$

4.1. *Consistency*

This section gives conditions under which $\mathbf{E}\|\hat{g} - g\|^2 \to 0$ as $n \to \infty$ in model (4.1)–(4.2). Assume that f_{YXZW} has r_{yx} bounded, continuous derivatives with respect to its first two arguments, r_{wz} bounded, continuous derivatives with respect to its second two arguments, and all lower-order mixed partial derivatives. Define $\delta_n = h_{yx}^{2r_{yx}} + h_{wz}^{2r_{wz}} + 1/(nh_{wz}^{\ell+m})$. Make the following assumptions, which are extensions of the assumptions of Section 3.1.

ASSUMPTION 1′: (a) *The function g is identified*. (b) $\int_{[0,1]^c} g(x,z)^2 \, dx \le M$ *for each* $z \in [0,1]^m$ *and some constant* $M < \infty$.

ASSUMPTION 2′: $r_{yx} > 0$ *and* $r_{wz} > 0$. *The corresponding derivatives of* f_{YXWZ} *and* f_{YXWZ} *are bounded in absolute value by* M.

ASSUMPTION 3′: *As* $n \to \infty$, $a_n \to 0$, $\delta_n \to 0$, *and* $\delta_n / a_n \to 0$.

THEOREM 4: *Let Assumptions 1′–3′ and 4 hold. Then for each* $z \in [0,1]^m$,

$$\lim_{n \to \infty} \mathbf{E} \int_{[0,1]^\ell} [\hat{g}(x,z) - g(x,z)]^2 \, dx = 0.$$

4.2. *Rate of Convergence*

As when X and W are scalars, the rate of convergence of $\hat{g}$ in the multivariate model depends on the rate of convergence of the singular values of the Fréchet derivative of $\mathcal{T}_z$. Accordingly, let T_{gz} denote the Fréchet derivative of $\mathcal{T}_z$ at g and let T_{gz}^* denote the adjoint of T_{gz}. Then T_{gz} and T_{gz}^*, respectively, are the operators defined by

$$(T_{gz}\varphi_z)(w) = \int_{[0,1]^c} f_{YXWZ}[g(x,z), x, w, z]\varphi_z(x) \, dx$$

and

$$(T_{gz}^{*}\varphi_z)(w) = \int_{[0,1]^c} f_{YXWZ}[g(w,z), w, x, z]\varphi_z(x)\,dx.$$

Assume that for each $z \in [0,1]^\ell$, $T_{gz}^{*}T_{gz}$ is nonsingular. Let $\{(\lambda_{zj}, \phi_{zj}) : j = 1, 2, \ldots\}$ denote the eigenvalues and eigenvectors of $T_{gz}^{*}T_{gz}$ ordered so that $\lambda_{z1} \geq \lambda_{z2} \geq \cdots > 0$. The eigenvectors $\{\phi_{zj}\}$ form a complete, orthonormal basis for $L_2[0,1]^\ell$. Thus, for each $z \in L_2[0,1]^\ell$ we can write

$$g(x,z) = \sum_{j=1}^{\infty} b_{zj}\phi_{zj}(x),$$

where

$$b_{zj} = \int_{[0,1]^c} g(x,z)\phi_z(x)\,dx.$$

Now make the following additional assumptions, which extend those of Section 3.2.

ASSUMPTION 5′: (a) *There are constants* $\alpha > 1$, $\beta > 1$, *and* $C_0 < \infty$ *such that* $\beta - 1/2 \leq \alpha < 2\beta - 1$, $j^{-\alpha} \leq C_0\lambda_{zj}$, *and* $|b_{zj}| \leq C_0 j^{-\beta}$ *uniformly in* $z \in [0,1]^m$ *for all* $j \geq 1$. (b) *Moreover*, $r_{wz} \geq \ell(2\beta + \alpha - 1)/2$ *and* $r_{yx} > (c+1)(2\beta + \alpha - 1)/(2\alpha)$.

ASSUMPTION 6′: *There is a finite constant* $L > 0$ *such that*

$$\|\mathcal{T}_z(g_1) - \mathcal{T}_z(g_2) - T_{g_2 z}(g_1 - g_2)\| \leq 0.5L\|g_1 - g_2\|^2$$

for any $g_1, g_2 \in L_2[0,1]$ *uniformly in* $z \in [0,1]^m$ *and*

$$\sum_{j=1}^{\infty} \frac{b_{zj}^2}{\lambda_{zj}} < \frac{1}{L}$$

uniformly in $z \in [0,1]^m$.

ASSUMPTION 7′: *The tuning parameters* h_{wz} *and* a_n *satisfy* $h_{wz} = C_{wz} \times n^{-1/(2r_{wz}+\ell+m)}$ *and* $a_n = C_a n^{-\tau\alpha/(2\beta+\alpha)}$, *where* $\tau = 2r_{wz}/(2r_{wz} + m)$ *and* C_{wz} *and* C_a *are positive, finite constants. Moreover*, $h_{yz} = C_{yz} n^{-\gamma}$, *where*

$$\frac{\tau}{2r_{yx}} \frac{2\beta + \alpha - 1}{2\beta + \alpha} < \gamma < \frac{1}{c+1} \frac{\tau\alpha}{2\beta + \alpha}.$$

Let $\mathcal{H}_M = \mathcal{H}_M(M, C_0, \alpha, \beta, L, r_{yx}, r_{wz}, c, \ell, m)$ be the set of distributions of (Y, X, Z, W) that satisfy Assumptions 1′, 2′, 5′, and 6′ with fixed values of $M, C_0, \alpha, \beta, L, r_{yx}, r_{wz}, c, \ell$, and m. The multivariate extension of Theorems 2 and 3 is the following theorem.

THEOREM 5: *Let Assumptions 1′, 2′, 4, and 5′–7′ hold. Then for each* $z \in [0,1]^m$,

$$\lim_{D\to\infty}\limsup_{n\to\infty}\sup_{H\in\mathcal{H}_M}\mathbf{P}_H\left\{\int_{[0,1]^c}[\hat{g}(x,z)-g(x,z)]^2\,dx > Dn^{-\tau(2\beta-1)/(2\beta+\alpha)}\right\} = 0. \tag{4.5}$$

Moreover, for each $z \in [0,1]^m$ *and any* $D > 0$,

$$\liminf_{n\to\infty}\sup_{H\in\mathcal{H}_M}\mathbf{P}_H\left\{\int_{[0,1]^c}[\tilde{g}(x,z)-g(x,z)]^2\,dx > Dn^{-\tau(2\beta-1)/(2\beta+\alpha)}\right\} > 0. \tag{4.6}$$

The rate of convergence in Theorem 5 is the same as that in Theorem 3 if $\ell = 1$ and $m = 0$. The theorem shows that increasing m decreases the rate of convergence of $\hat{g}$ for any fixed r_{wz}. This is the familiar curse of dimensionality of nonparametric estimation.

5. MONTE CARLO EXPERIMENTS

This section reports the results of Monte Carlo simulations that illustrate the finite-sample performance of the estimator that is obtained by solving (2.5). Samples of (Y, X, W) were generated from the model

$$Y = g(X) + U,$$

where

$$g(x) = x - x^2$$

and (U, X, W) is sampled from the distribution whose density is

$$f_{UXW}(u,x,w) = C_\alpha \sum_{j=1}^{\infty} a_j G_j(u)\sin(j\pi x)\sin(j\pi w) + 0.25\sum_{j=1}^{\infty} d_j \sin[j\pi(u+1)]\sin(2j\pi x)\sin(j\pi w)$$

for $(U, X, W) \in [-1,1]\times[0,1]^2$. In this density, $a_j = j^{-2.45}$, $d_j = j^{-10}$,

$$G_j(u) = \frac{2j+1}{4j}(1-u^{2j})I(|u|\le 1),$$

and

$$C = \left\{ \sum_{j=1}^{\infty} \frac{4a_{2j-1}}{[(2j-1)\pi]^2} \right\}^{-1}.$$

For computational purposes, the infinite series were truncated at $j = 100$. The density f_{UXW} is illustrated in Figure 1. The density is cumbersome algebraically, but has the advantage of making the Fourier coefficients b_j and eigenvalues λ_j easy to calculate.

The kernel function used for density estimation is $K(x) = (15/16)(1 - x^2)^2 I(|x| \leq 1)$. The estimates of g were computed by using the Levenberg–Marquardt method (Engl, Hanke, and Neubauer (1996, p. 285)). The starting function was $g(x) = 0$. The estimation results are not sensitive to the choice of starting function. The numerical approximations can be made arbitrarily accurate, so numerical approximation error does not affect the results.

Each experiment consisted of estimating g at the 99 points $x = 0.01, 0.02, \ldots, 0.99$. The experiments were carried out in GAUSS using GAUSS pseudo-random number generators. There were 500 Monte Carlo replications in each experiment.

Figure 2 summarizes the results for $n = 200$ for several different values of the tuning parameters h_w, h_{yx}, and a_n. Further results are available electronically in Horowitz and Lee (2007). Figure 2 shows $g(x)$ (dashed line), the Monte Carlo approximation to $E[\hat{g}(x)]$ (solid line), and the estimates, $\hat{g}$, whose integrated mean-square errors (IMSEs) are the 25th, 50th, and 75th percentiles of the IMSEs of the 500 Monte Carlo replications. The estimates, $\hat{g}$, vary, but their shapes are similar to that of g.

6. CONCLUSIONS

This paper has presented a nonparametric instrumental variables estimator of a quantile regression model, derived the estimator's rate of mean-square convergence in probability, and given conditions under which this rate is the fastest possible in a minimax sense. Several topics remain for future research. The problem of deriving the asymptotic distribution of $\hat{g}$ appears to be quite difficult. As was explained in Section 1, asymptotic normality cannot be obtained by using a Taylor series approximation to linearize the first-order condition for (2.5). This problem does not arise with the mean-regression estimator of Hall and Horowitz (2005) and Horowitz (2007), because the first-order condition in the mean regression is a linear equation for $\hat{g}$. Finding methods to choose the tuning parameters h_{yx} and a_n in applications is another topic for future research.

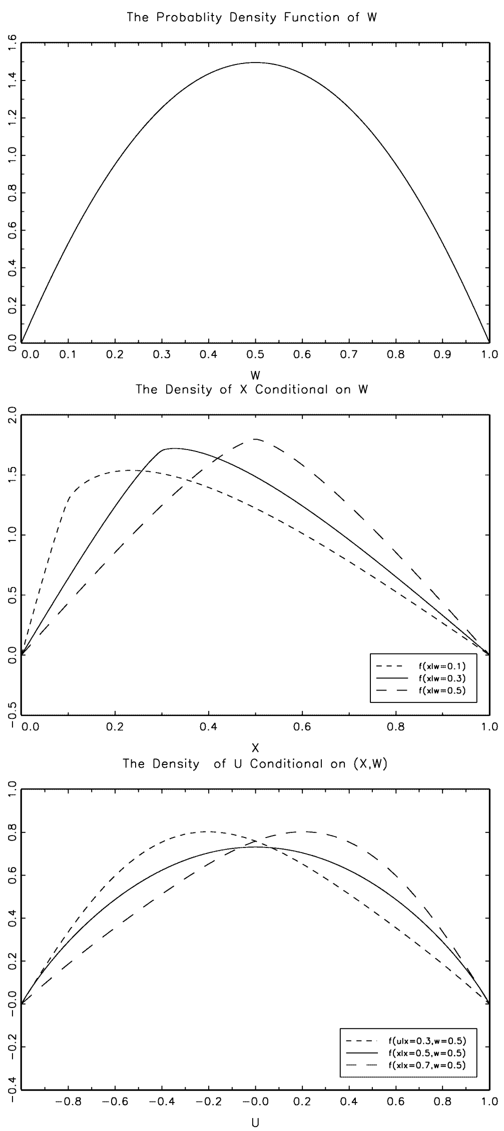

FIGURE 1.—Illustration of the density of (U, X, W) in the Monte Carlo experiments. The top panel of the figure shows the marginal density of W; the middle panel shows the density of X conditional on $W = 0.1, 0.3, 0.5$; and the bottom panel shows the density of U conditional on $(X, W) = (0.3, 0.5), (0.5, 0.5), (0.7, 0.5)$.

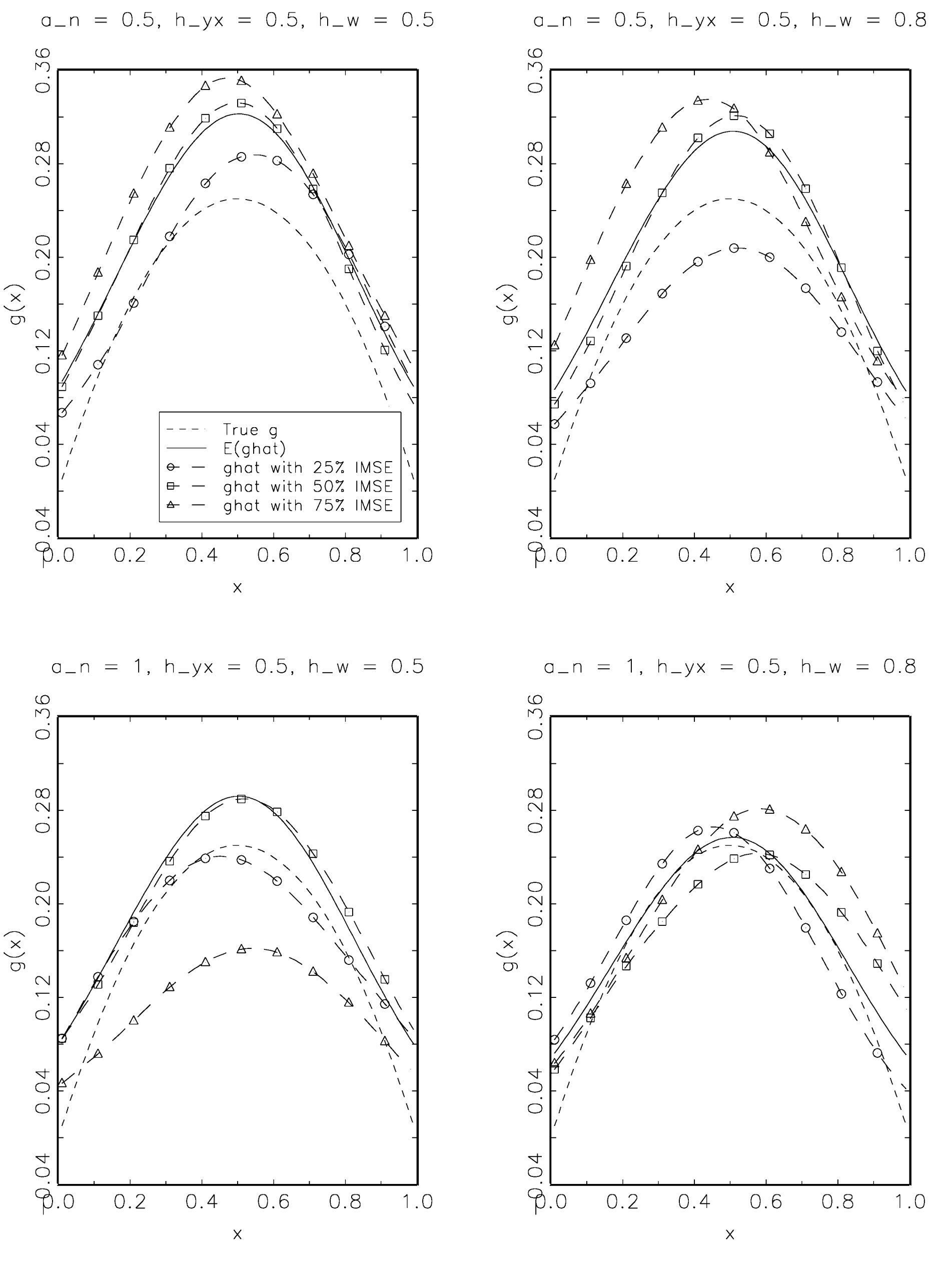

FIGURE 2.—Monte Carlo results for $n = 200$.

Dept. of Economics, Northwestern University, 2001 Sheridan Road, Evanston, IL 60208, U.S.A.; joel-horowitz@northwestern.edu

and

Dept. of Economics, University College London, Gower Street, London WC1E 6BT, U.K.; l.simon@ucl.ac.uk.

Manuscript received May, 2006; final revision received February, 2007.

REFERENCES

AMEMIYA, T. (1982): "Two Stage Least Absolute Deviations Estimators," *Econometrica*, 50, 689–711. [1192]

BISSANTZ, N., T. HOHAGE, AND A. MUNK (2004): "Consistency and Rates of Convergence of Nonlinear Tikhonov Regularization with Random Noise," *Inverse Problems*, 20, 1773–1789. [1195]

BISSANTZ, N., T. HOHAGE, A. MUNK, AND F. RUYMGAART (2006): "Convergence Rates of General Regularization Methods for Statistical Inverse Problems and Applications," Working Paper, Institut für Numerische und Angewandte Mathematik Georg-August-Universität, Göttingen, Germany. [1198]

BLUNDELL, R., AND J. L. POWELL (2003): "Endogeneity in Nonparametric and Semiparametric Regression Models," in *Advances in Economics and Econometrics: Theory and Applications*, Vol. 2, ed. by M. Dewatripont, L.-P. Hansen, and S. J. Turnovsky. Cambridge, U.K.: Cambridge University Press, 312–357. [1192]

——— (2007): "Censored Regression Quantiles with Endogenous Regressors," *Journal of Econometrics*, forthcoming. [1192]

CARRASCO, M., J.-P. FLORENS, AND E. RENAULT (2005): "Linear Inverse Problems in Structural Econometrics: Estimation Based on Spectral Decomposition and Regularization," in *Handbook of Econometrics*, Vol. 6, ed. by E. E. Leamer and J. J. Heckman. Amsterdam: North-Holland. [1192,1198]

CHEN, L., AND S. PORTNOY (1996): "Two-Stage Regression Quantiles and Two-Stage Trimmed Least Squares Estimators for Structural Equation Models," *Communications in Statistics, Theory and Methods*, 25, 1005–1032. [1192]

CHERNOZHUKOV, V., AND C. HANSEN (2004): "The Effects of $401(k)$ Participation on the Wealth Distribution: An Instrumental Quantile Regression Analysis," *Review of Economics and Statistics*, 86, 735–751. [1192]

——— (2005): "An IV Model of Quantile Treatment Effects," *Econometrica*, 73, 245–261. [1192,1194]

——— (2006): "Instrumental Quantile Regression Inference for Structural and Treatment Effect Models," *Journal of Econometrics*, 132, 491–525. [1192]

CHERNOZHUKOV, V., G. W. IMBENS, AND W. K. NEWEY (2007): "Instrumental Variable Identification and Estimation of Nonseparable Models via Quantile Conditions," *Journal of Econometrics*, 139, 4–14. [1192]

CHESHER, A. (2003): "Identification in Nonseparable Models," *Econometrica*, 71, 1405–1441. [1192]

DAROLLES, S., J.-P. FLORENS, AND E. RENAULT (2006): "Nonparametric Instrumental Regression," Working Paper, GREMAQ, University of Social Science, Toulouse, France. [1192,1193, 1198]

ENGL, H. W., M. HANKE, AND A. NEUBAUER (1996): *Regularization of Inverse Problems*. Dordrecht: Kluwer. [1193,1204]

FLORENS, J.-P. (2003): "Inverse Problems and Structural Econometrics: The Example of Instrumental Variables," in *Advances in Economics and Econometrics: Theory and Applications,* *Dewatripont*, Vol. 2, ed. by L.-P. Hansen and S. J. Turnovsky. Cambridge, U.K.: Cambridge University Press, 284–311. [1192]

GROETSCH, C. (1984): *The Theory of Tikhonov Regularization for Fredholm Equations of the First Kind*. London: Pitman. [1193]

HALL, P., AND J. L. HOROWITZ (2005): "Nonparametric Methods for Inference in the Presence of Instrumental Variables," *The Annals of Statistics*, 33, 2904–2929. [1192,1193,1197,1198,1204]

HOROWITZ, J. L. (2007): "Asymptotic Normality of a Nonparametric Instrumental Variables Estimator," *International Economic Review*, forthcoming. [1192,1204]

HOROWITZ, J. L., AND S. LEE (2007): "Supplement to 'Nonparametric Instrumental Variables Estimation of a Quantile Regression Model,'" *Econometrica Supplementary Material*, 75, http://www.econometricsociety.org/ecta/supmat/6471proofs.pdf. [1194,1204]

JANUSZEWSKI, S. I. (2002): "The Effect of Air Traffic Delays on Airline Prices," Working Paper, Department of Economics, University of California at San Diego, La Jolla, CA. [1192]

KALTENBACHER, B., AND A. NEUBAUER (2006): "Convergence of Projected Iterative Regularization Methods for Nonlinear Problems with Smooth Solutions," *Inverse Problems*, 22, 1105–1119. [1198]

KRESS, R. (1999): *Linear Integral Equations* (Second Ed.) New York: Springer. [1193]

LEE, S. (2007): "Endogeneity in Quantile Regression Models: A Control Function Approach," *Journal of Econometrics*, forthcoming. [1192]

MA, L., AND R. KOENKER (2006): "Quantile Regression Methods for Recursive Structural Equation Models," *Journal of Econometrics*, 134, 471–506. [1192]

NEWEY, W. K., AND J. L. POWELL (2003): "Instrumental Variable Estimation of Nonparametric Models," *Econometrica*, 71, 1565–1578. [1192]

NEWEY, W. K., J. L. POWELL, AND F. VELLA (1999): "Nonparametric Estimation of Triangular Simultaneous Equations Models," *Econometrica*, 67, 565–603. [1192]

POWELL, J. L. (1983): "The Asymptotic Normality of Two-Stage Least Absolute Deviations Estimators," *Econometrica*, 50, 1569–1575. [1192]

SAKATA, S. (2007): "Instrumental Variable Estimation Based on Conditional Median Restriction," *Journal of Econometrics*, forthcoming. [1192]

TIKHONOV, A. N. (1963a): "Solution of Incorrectly Formulated Problems and the Regularization Method," *Soviet Mathematics Doklady*, 4, 1035–1038. [1193]

——— (1963b): "Regularization of Incorrectly Posed Problems," *Soviet Mathematics Doklady*, 4, 1624–1627. [1193]