

## NOTES AND COMMENTS

### AN IV MODEL OF QUANTILE TREATMENT EFFECTS

BY VICTOR CHERNOZHUKOV AND CHRISTIAN HANSEN<sup>1</sup>

The ability of quantile regression models to characterize the heterogeneous impact of variables on different points of an outcome distribution makes them appealing in many economic applications. However, in observational studies, the variables of interest (e.g., education, prices) are often endogenous, making conventional quantile regression inconsistent and hence inappropriate for recovering the causal effects of these variables on the quantiles of economic outcomes. In order to address this problem, we develop a model of quantile treatment effects (QTE) in the presence of endogeneity and obtain conditions for identification of the QTE without functional form assumptions. The principal feature of the model is the imposition of conditions that restrict the evolution of ranks across treatment states. This feature allows us to overcome the endogeneity problem and recover the true QTE through the use of instrumental variables. The proposed model can also be equivalently viewed as a structural simultaneous equation model with nonadditive errors, where QTE can be interpreted as the structural quantile effects (SQE).

**KEYWORDS:** Endogeneity, quantile regression, simultaneous equations, instrumental regression, identification, nonlinear model, monotone likelihood ratio, bounded completeness, partial identification.

#### 1. INTRODUCTION

THE ABILITY OF QUANTILE REGRESSION models to characterize the heterogeneous impact of variables on different points of an outcome distribution makes them appealing in many economic applications. However, in observational studies, the variables of interest (e.g., education, prices) are often endogenous, making conventional quantile regression inconsistent and hence inappropriate for recovering the causal effects of these variables on the quantiles of economic outcomes. In order to address this problem, we develop a model of quantile treatment effects (QTE) in the presence of endogeneity and obtain conditions for identification of the QTE without functional form assumptions. The principal feature of the model is the imposition of conditions that restrict the evolution of ranks across treatment states. This feature allows us to overcome the endogeneity problem and recover the true QTE through the use of instrumental variables. Hausman and Sidak (2002), Januszewski (2002), and

<sup>1</sup>We benefited from seminars at Cornell, Penn, University of Illinois at Urbana-Champaign (all in 2001), Harvard-MIT in 2002, and the EC2 2001 Conference on Causality and Exogeneity in Econometrics in Louvain-la-Neuve. Conversations with Takeshi Amemiya, Tom MaCurdy and especially Alberto Abadie motivated the line of research taken here. We thank Jerry Hausman and Whitney Newey for careful readings of the paper and for help with presentation. We also thank Josh Angrist, Moshe Buchinsky, Jin Hahn, James Heckman, Guido Imbens, Roger Koenker, Joanna Lahey, Igor Makarov, the coeditor, and anonymous referees for many valuable comments.

Chernozhukov and Hansen (2001a, 2003) present economic applications of the proposed model/method.

Our proposal complements other modern heterogeneous effect models under endogeneity,<sup>2</sup> but is different in that it puts restrictions on the evolution of ranks across treatment states and that it primarily focuses on QTE. Our model also differs from that in Abadie, Angrist, and Imbens (2002), who consider a QTE model for the (unobserved) subpopulation of “compliers,” which applies only to binary treatment variables. The approach in this paper is expressly designed for studying heterogeneous QTE over the entire population and applies to binary, discrete, and continuous treatment variables. As will be discussed, our approach is also different from the control function methods for triangular structural models in Chesher (2003) and Imbens and Newey (2003) that aim at estimating triangular structures using quantile transforms. We instead aim at directly estimating the QTE (SQE) using IV equations. Also, our approach provides a causal model and foundation for estimation methods based on IV quantile-independence conditions and median-independence conditions.<sup>3</sup>

## 2. THE MODEL

In this section we present the model, its main statistical implication, and the principal identification result. In what follows, we focus the discussion on the case where the treatment (endogenous variable) takes on two values,  $D = 0$  and  $D = 1$ , as it simplifies the discussion and best illustrates the conditions required for identification. However, the definition of the model and its main statistical implications do not rely upon this notational simplification. The Appendix contains generalizations to nonbinary treatments.

### 2.1. Framework

Our model is developed within the conventional potential (latent) outcome framework, e.g., Heckman and Robb (1986). Potential real-valued outcomes that vary among individuals or observational units are indexed against potential treatment states  $d \in \{0, 1\}$  and denoted  $Y_d$ . The potential outcomes  $\{Y_d\}$  are latent because, given the selected treatment  $D$ , the observed outcome for each individual or observational unit is only one component

$$(2.1) \quad Y \equiv Y_D$$

<sup>2</sup>E.g., see Heckman and Vytlacil (1999), Imbens and Angrist (1994), and Blundell and Powell (2001).

<sup>3</sup>Chernozhukov and Hansen (2001a, 2001b, 2003), Chernozhukov and Hong (2003), and Chernozhukov, Imbens, and Newey (2003) propose estimation methods based on quantile-independence conditions to estimate parameters of general heterogeneous (nonadditive) models proposed in this paper. Hong and Tamer (2003), Abadie (1995), MaCurdy and Timmins (2001), Honore and Hu (2004), and Sakata (2001) use median-independence and other robust moments to estimate classical additive models.

of the potential outcomes vector  $\{Y_d\}$ . Throughout the paper, capital letters denote random variables, and lower case letters denote the potential values they may take.

The objective of causal (structural) analysis is to learn about features of the distributions of potential outcomes  $Y_d$ . Of primary interest to us are the  $\tau$ th quantiles of potential outcomes under various treatments  $d$ , conditional on observed characteristics  $X = x$ , and denoted as

$$(2.2) \quad q(d, x, \tau).$$

We will refer to the function  $q(d, x, \tau)$  as the quantile treatment response (QTR) function. We are also interested in the quantile treatment effects (QTE), defined as

$$(2.3) \quad q(1, x, \tau) - q(0, x, \tau),$$

that summarize the differences in the impact of treatments on the quantiles of potential outcomes.<sup>4</sup> The related average effects are defined as  $\int_0^1 (q(1, x, \tau) - q(0, x, \tau)) \partial \tau$ .<sup>5</sup>

Typically, the realized treatment  $D$  is selected in relation to potential outcomes inducing selection bias (endogeneity). This makes the conventional quantile regression of observed  $Y$  on observed  $D$  inappropriate for measuring  $q(d, x, \tau)$  and the derived QTE. The model presented next states the conditions under which we can identify and estimate the quantiles of latent outcomes through the nonlinear quantile-type conditional moment restrictions:

$$(2.4) \quad P[Y \leq q(D, X, \tau) | X, Z] = \tau \quad \text{a.s.,}$$

where  $Z$  is an instrument that affects  $D$  but is independent of potential outcomes.

## 2.2. The Instrumental Quantile Treatment Effects (IVQR) Model

Having conditioned on the observed characteristics  $X = x$ , each latent outcome  $Y_d$  can be related to its quantile function  $q(d, x, \tau)$  as<sup>6</sup>

$$(2.5) \quad Y_d = q(d, x, U_d), \quad \text{where} \quad U_d \sim U(0, 1).$$

<sup>4</sup>Early formulations of QTE go back to Lehmann (1974) and Doksum (1974). In cases when  $d$  is continuous, QTE may be defined as the partial effects  $\partial q(d, \tau) / \partial d$ .

<sup>5</sup>In cases when  $d$  is continuous, ATE may be defined as  $\int_0^1 (\partial q(d, \tau) / \partial d) \partial \tau$ .

<sup>6</sup>This follows by the Skorohod representation of random variables, which states that given a collection of variables  $\{\xi_d\}$ , each variable  $\xi_d$  can be represented as  $\xi_d = q(d, U_d)$  for some  $U_d \sim U(0, 1)$ ; cf. Durrett (1996), where  $q(d, \tau)$  denotes the  $\tau$ -quantile of variable  $\xi_d$ .

We will refer to  $U_d$  as the rank variable, and note that representation (2.5) is essential to what follows.

The rank variable  $U_d$  is responsible for heterogeneity of outcomes among individuals with the same observed characteristics  $x$  and treatment state  $d$ . It also determines their relative ranking in terms of potential outcomes; hence one may think of rank  $U_d$  as representing some unobserved characteristic, e.g., ability or proneness.<sup>7</sup> This interpretation makes quantile analysis an interesting tool for describing and learning the structure of heterogeneous treatment effects and controlling for unobserved heterogeneity.<sup>8</sup>

For example, consider a returns-to-training model, where  $Y_d$ 's are potential earnings under different training levels  $d$ , and  $q(d, x, \tau)$  is the earning function, which describes how an individual having training  $d$  and "ability"  $\tau$  is rewarded by the labor market. The earning function may be different for different levels of  $\tau$ , implying heterogeneous effects of training on earnings of people that have different levels of "ability."

Formally, our model consists of five conditions (some are representations) that hold jointly.

*Main Conditions of the Model:* Given a common probability space  $(\Omega, F, P)$ , the following conditions hold *jointly* with probability one:

A1. POTENTIAL OUTCOMES: *Conditional on  $X = x$ , for each  $d$ ,  $Y_d = q(d, x, U_d)$ , where  $q(d, x, \tau)$  is strictly increasing in  $\tau$  and  $U_d \sim U(0, 1)$ .*

A2. INDEPENDENCE: *Conditional on  $X = x$ ,  $\{U_d\}$  are independent of  $Z$ .*

A3. SELECTION:  *$D \equiv \delta(Z, X, V)$  for some unknown function  $\delta$  and random vector  $V$ .*

A4. RANK INVARIANCE OR RANK SIMILARITY: *Conditional on  $X = x$ ,  $Z = z$ ,*

- (a)  *$\{U_d\}$  are equal to each other; or, more generally,*
- (b)  *$\{U_d\}$  are identically distributed, conditional on  $V$ .*

A5. OBSERVED VARIABLES: *Observed variables consist of  $Y \equiv q(D, X, U_D)$ ,  $D$ ,  $X$ , and  $Z$ .*<sup>9</sup>

The following is the main statistical implication of the model.

<sup>7</sup>Doksum (1974) uses the term proneness as in "prone to learn fast" or "prone to grow taller."

<sup>8</sup>See Doksum (1974), Lehmann (1974), and Koenker and Hallock (2001).

<sup>9</sup>Here  $U_D = D \cdot U_1 + (1 - D)U_0$  when  $D \in \{0, 1\}$  and, more generally,  $U_D \equiv \sum_d \mathbb{1}(D = d) \cdot U_d$ .

**THEOREM 1 (Main Testable Implication):** *Suppose conditions A1–A5 hold. Then, for all  $\tau \in (0, 1)$ , a.s.*

$$(2.6) \quad P[Y \leq q(D, X, \tau) | X, Z] = P[Y < q(D, X, \tau) | X, Z] = \tau,$$

and  $U_D \sim U(0, 1)$  conditional on  $Z$  and  $X$ .

The result is simplest to see under rank invariance A4(a), i.e. when  $U_d = U$  for all  $d$ . Indeed, by A1 under rank invariance, the event

$$(2.7) \quad \{Y \leq q(D, X, \tau)\} \text{ is equivalent to } \{U \leq \tau\},$$

which yields the conclusion given independence condition A3. The proof of Theorem 1 given in Appendix A provides more details and generalizes the result to rank similarity A4(b).

The model and the results of Theorem 1 are important for two reasons. First, Theorem 1 serves as a means of identifying the QTE in a general heterogeneous effects model. Second, by demonstrating that the IVQR model leads to the conditional moment restrictions (2.6), Theorem 1 provides an economic and causal foundation for estimation based on these restrictions; the pertinent estimation approaches are referenced in the Introduction. It should be noted that conditioning on the instrument as in (2.6) may appear to be a natural strategy for estimating the QTE.<sup>10</sup> However, this strategy will typically fail outside of the developed IVQR model, where the quantiles of potential outcomes  $q(d, x, \tau)$  will generally not satisfy equation (2.6). Thus, the IVQR model provides conditions under which one can recover the quantiles of potential outcomes from statistical equations (2.6) in applications.

### 2.3. Discussion of the Model

Condition A1 restates the basic Skorohod representation (2.5) of potential outcomes with strict monotonicity imposed on the QTR function. The imposition of strict monotonicity rules out discrete response cases that we hope to explore in future work.

Condition A2 states that potential outcomes are independent of  $Z$ , given  $X$ , which is a conventional independence restriction. Condition A3 is a convenient representation of a treatment selection mechanism, stated for the purposes of discussion. In A3 the unobserved random vector  $V$  is responsible for the difference in treatment choices  $D$  across observationally identical individuals. The independence condition in A2 and A3 is significantly weaker than the commonly made assumption—that both the disturbances  $\{U_d\}$  in the outcome

<sup>10</sup>In general, variation of treatment effects across individuals creates many problems for “conditioning on  $Z$ ” approaches, as explained in Heckman and Robb (1986). This problem is resolved here by imposing rank similarity conditions and using quantile transforms.

equations and the disturbances  $V$  in the selection equation are *jointly* independent of the instrument  $Z$ ; e.g., Heckman and Robb (1986) and Imbens and Angrist (1994). The latter assumption may be violated when the instrument is measured with error (cf. Hausman (1977)) or the instrument is not assigned exogenously relative to the selection equation (cf. Example 2 in Imbens and Angrist (1994)).

Condition A4 is probably the most important assumption. Its simplest, though strongest, form is rank invariance, A4(a), when ranks  $U_d$  do not vary with potential treatment states  $d$ .<sup>11</sup>

$$(2.8) \quad U_d = U \quad \text{for each } d.$$

For example, under rank invariance, people who are strong (highly ranked) earners without a training program remain strong earners having done the training. Indeed, the earning of a person with rank  $U = \tau$  in the training state “0” is  $Y_0 = q(0, x, \tau)$  and in the state “1” is  $Y_1 = q(1, x, \tau)$ .<sup>12</sup> Thus, rank invariance implies that a common unobserved factor  $U$ —say, innate ability—determines the ranking of a given person across treatment states. Conditioning on appropriate covariates  $X$  may be important to achieve rank invariance.

However, rank invariance implies that the potential outcomes  $\{Y_d\}$  are not truly multivariate, being jointly degenerate, which may be implausible on logical grounds as noted in Heckman, Smith, and Clements (1997). Also, the rank variables  $U_d$  may be determined by many unobserved factors. Thus, it is desirable to allow the rank  $U_d$  to change across  $d$ , reflecting some unobserved, unsystematic variation. Rank similarity A4(b) achieves this property, thus accommodating general multivariate outcomes, while managing to preserve the useful moment restriction (2.6).

Rank similarity A4(b) relaxes exact rank invariance by allowing unsystematic deviations, “slippages” in one’s rank away from some common level  $U$ . Rank similarity requires that, conditional on  $U$ , which may enter disturbance  $V$  in the selection equation, the slippages<sup>13</sup>

$$(2.9) \quad U_d - U \text{ are identically distributed across } d \in \{0, 1\}.$$

In this formulation, we implicitly make the assumption that one selects the treatment without knowing the exact potential outcomes; i.e., one may know  $U$  and even the distribution of slippages, but does not know the exact slippages  $U_d - U$ . This assumption is consistent with many empirical situations where the exact latent outcomes are not known beforehand.

<sup>11</sup>Notice that under rank invariance, condition A3 is a pure representation, not a restriction, since nothing restricts the unobserved information component  $V$ .

<sup>12</sup>Rank invariance is used in many interesting models without endogeneity. See, e.g., Doksum (1974), Heckman, Smith, and Clements (1997), and Koenker and Geling (2001).

<sup>13</sup>Formally, conditioning is required to be on all components of  $V$  in the selection equation A3.

In summary, rank similarity is the main restriction of the IVQR model that allows us to address endogeneity. This restriction is absent in conventional endogenous heterogeneous treatment effect models. However, rank similarity enables a more general selection mechanism, A3, that requires neither the monotonicity assumptions of the LATE approach (Imbens and Angrist (1994)) nor the stronger independence assumptions of the conventional models listed earlier. The main force of rank similarity and the other stated assumptions is the implied moment restriction (2.6) of Theorem 1, which is useful for estimation and identification of the quantile treatment effects.

#### 2.4. Identification

As in Newey and Powell (2003), we focus on obtaining point-identifying assumptions. Newey and Powell (2003) show that a necessary and sufficient condition for nonparametric identification of a function  $\mu$  under the conventional linear IV condition

$$(2.10) \quad E(Y - \mu(D)|Z) = 0 \quad \text{a.s.}$$

is that the Jacobian of the vector of moment equations should be of full rank when  $D$  is binary or discrete. They also demonstrate that for continuous  $D$  the full rank condition generalizes to an instrument completeness condition (see the Appendix).

Our identification conditions are also formulated in terms of full rank and completeness conditions. However, the conditions differ from those in Newey and Powell (2003) in that they reflect the specific nature of our problem and, due to the nonlinear nature of the IV condition (2.6), are not minimal. In the main text, we focus the discussion on the binary case, which is the most relevant case for program evaluation, where  $D$ 's  $\in \{0, 1\}$  are participation states and  $Z$ 's  $\in \{0, 1\}$  are offers of participation. Generalizations to discrete and continuous  $D$  and  $Z$  follow analogously, and formal results are stated in the Appendix.

The following analysis is all conditional on  $X = x$  and for a given quantile  $\tau \in (0, 1)$ , but we suppress this dependence for ease of notation. Under the conditions of Theorem 1 we know that there is at least one function  $q(d) \equiv q(d, x, \tau)$  that solves  $P[Y \leq q(D)|Z] = \tau$  a.s. The function  $q(\cdot)$  can be equivalently represented by a vector of its values  $q = (q(0), q(1))'$ . Therefore, for vectors of the form  $y = (y_0, y_1)$  and the vector of moment equations

$$(2.11) \quad \Pi(y) \equiv (P[Y \leq y_D|Z = 0] - \tau, P[Y \leq y_D|Z = 1] - \tau)',$$

where  $y_D \equiv (1 - D) \cdot y_0 + D \cdot y_1$ , the identification question is whether  $y = q$  uniquely solves  $\Pi(y) = 0$ . Next we specify a parameter space  $\mathcal{L}$  consisting of vectors that may also potentially solve  $\Pi(y) = 0$ . Identification will hold when a rank condition is imposed on all elements of  $\mathcal{L}$ .

Fix some small constants  $\delta > 0$  and  $\underline{f} > 0$ , and define  $\mathcal{L}$  as a closed rectangle containing all vectors  $(y_0, y_1)$  that satisfy

$$(2.12) \quad \begin{aligned} & \text{(i) for each } z, P[Y < y_D | Z = z] \in [\tau - \delta, \tau + \delta] \text{ and} \\ & \text{(ii) for each } d, y_d \in s_d \equiv \{\lambda : f_Y(\lambda | d, z) \geq \underline{f} \text{ for all } z \text{ with} \\ & \quad P[D = d | Z = z] > 0\}. \end{aligned}$$

Condition (i) defines the parameter space  $\mathcal{L}$  as a set of potential solutions to the moment equations  $\Pi(y) = 0$ , while condition (ii) requires these solutions to be in the support of the response variable. Define the Jacobian of the moment equations  $\Pi(y)$  with respect to  $y = (y_0, y_1)'$  as

$$(2.13) \quad \begin{aligned} \Pi'(y) &\equiv \begin{bmatrix} f_Y(y_0 | D = 0, Z = 0)P[D = 0 | Z = 0] \\ f_Y(y_0 | D = 0, Z = 1)P[D = 0 | Z = 1] \\ f_Y(y_1 | D = 1, Z = 0)P[D = 1 | Z = 0] \\ f_Y(y_1 | D = 1, Z = 1)P[D = 1 | Z = 1] \end{bmatrix} \\ &\equiv \begin{bmatrix} f_{Y,D}(y_0, 0 | Z = 0) & f_{Y,D}(y_1, 1 | Z = 0) \\ f_{Y,D}(y_0, 0 | Z = 1) & f_{Y,D}(y_1, 1 | Z = 1) \end{bmatrix}. \end{aligned}$$

**THEOREM 2 (Global Identification by Full Rank Conditions):** *Suppose A1–A5 hold, the support of  $D$  is  $\{0, 1\}$ , and the support of  $Z$  is  $\{0, 1\}$ . Assume that for the sets  $\mathcal{L}$  and  $s_d$  specified above, (i)  $\Pi'(y)$  is continuous for all  $y \in \mathcal{L}$ , and (ii)  $q(d) \in s_d$  for each  $d$ . Then the  $\tau$ -quantiles of potential outcomes,  $q = (q(0), q(1))$ , are identified if*

$$(2.14) \quad \text{rank}[\Pi'(y)] \text{ is full for any } y \in \mathcal{L}.$$

Condition (i) imposes continuity on the conditional density and (ii) is an inclusion assumption that requires that the realized outcome  $Y$  take on values around  $q(d)$ . This condition implies that  $q \in \mathcal{L}$ . Therefore, to check identification one needs to show that  $y = q$  is the only solution to  $\Pi(y) = 0$  among all  $y \in \mathcal{L}$ . The main condition, (2.14), insures that  $y = q$  is the unique solution by imposing that the Jacobian  $\Pi'(y)$  is of full rank for all  $y \in \mathcal{L}$ . The full rank condition requires the impact of instrument  $Z$  on the joint distribution of  $(Y, D)$  to be sufficiently rich; in particular, the instrument  $Z$  should not be independent of the endogenous variable  $D$ . The proposed global identification condition is stronger but is almost as reasonable as the local identification condition:  $\text{rank}[\Pi'(y)]$  is full for  $y = q$ .

In applications, the identification conditions may also be weakened by restricting the parameter space  $\mathcal{L}$  using further economic restrictions. If economic restrictions impose that  $(q(0), q(1))' \in \mathcal{L}^R$  for some convex closed



set  $\mathcal{L}^R$  in  $\mathbb{R}^2$ , then  $\mathcal{L}$  can be replaced by  $\mathcal{L}^R \cap \mathcal{L}$ , provided conditions of Theorem 3 stated in the Appendix hold.

To illustrate the above condition, note that  $\text{rank } \Pi'(y)$  is full is equivalent to  $\det \Pi'(y) \neq 0$ , which implies

$$(2.15) \quad \frac{f_{Y,D}(y_1, 1|Z=1)}{f_{Y,D}(y_0, 0|Z=1)} > \frac{f_{Y,D}(y_1, 1|Z=0)}{f_{Y,D}(y_0, 0|Z=0)} \quad \text{for all } y = (y_0, y_1) \in \mathcal{L}$$

(or the same condition but with  $>$  replaced with  $<$ ). Inequality (2.15) may be interpreted as a *monotone likelihood ratio condition*. That is, the instrument  $Z$  should have a monotonic impact on the likelihood ratio specified in (2.15), which in general is a stronger condition than the usual condition that  $D$  is correlated with  $Z$ . Nevertheless, the condition (2.15) will be trivially satisfied in many useful contexts. For instance, if we impose monotonicity of the treatment impact on the outcome quantiles, so that  $q(1) \geq q(0)$ , i.e.,  $q \in \mathcal{L}^R = \{y \in \mathbb{R}^2 : y_1 \geq y_0\}$ , then condition (2.15) has to hold only for  $y \in \mathcal{L}^R \cap \mathcal{L}$ . That is, it simply suffices that the instrument  $Z$  (e.g., the offer of treatment) increases the relative joint likelihood of receiving the treatment and having a weakly higher outcome, which is a plausible condition in many program evaluation applications. In other cases, we also have that  $P[D=1|Z=0] = 0$  (e.g., those not offered a treatment do not receive that treatment), so that the right-hand side of (2.15) equals 0, which makes the identification condition (2.15) satisfied trivially, without monotonicity restrictions.<sup>14</sup>

### 2.5. Partial Identification and the Identification Region

When the rank conditions do not hold, point identification may not hold generally, just as in the additive IV models studied in Newey and Powell (2003). In that case, given a quantile of interest  $\tau$ , the identification region of models for  $q(d, x, \tau)$  can be stated as the set  $\mathcal{M}$  of functions  $m(d, x)$  that satisfy the following relations:

$$(2.16) \quad P[Y \leq m(D, X)|X, Z] = P[Y < m(D, X)|X, Z] = \tau \quad \text{a.s.}$$

This representation of the identification region  $\mathcal{M}$  is implicit, since a closed-form representation of  $\mathcal{M}$  does not exist generally.<sup>15</sup> Nevertheless, statistical inference about  $\mathcal{M}$  can be based on (2.16) and can be carried out using the set-inference approach described in Manski and Tamer (2002) and Chernozhukov, Hong, and Tamer (2003).

<sup>14</sup>However, this implicitly requires the technical condition that the support sets  $s_0$  and  $s_1$  specified in (2.12) are connected.

<sup>15</sup>This implicit nature of the identification region is not special to the present problem, and would hold generally for problems where parameters solve nonlinear moment equations.

## 3. COMPARISON WITH CONTROL FUNCTION APPROACH

It is useful to compare our approach with the recent literature on Roehrig (1988) type triangular simultaneous equations, e.g., Chesher (2003) and Imbens and Newey (2003). In order to do so, recall that the observational equations of our model (A5) combined with the main implication of the model (Theorem 1), suppressing dependence on covariates  $X$ , yield the following relations:

$$(3.1) \quad \begin{cases} Y = q(D, U_D), \\ D = \delta(Z, V), \end{cases} \quad U_D \text{ is a scalar independent of } Z,$$

where  $q(d, \tau)$  is the quantile treatment response function. The disturbances in the selection equation  $V$  are possibly multivariate and some components of  $V$  may depend on  $Z$ .

The model studied in Chesher (2003) and Imbens and Newey (2003) takes the following form:

$$(3.2) \quad \begin{cases} Y = t(D, \nu, \eta), \\ D = \varphi(Z, \nu), \end{cases} \quad \eta, \nu \text{ are scalars jointly independent of } Z,$$

where  $D$  and  $Z$  are both continuous and  $t(d, \nu, \eta)$  is strictly increasing in  $\eta$  and  $\varphi(z, \nu)$  is strictly increasing in  $\nu$ .<sup>16</sup>

Since Chesher (2003) and Imbens and Newey (2003) require  $D$  and  $Z$  to be continuous, their approach does not apply to typical returns-to-education examples or program evaluation studies. In contrast, our approach allows for both discrete and continuous  $D$  and  $Z$ . Moreover, our outcome equation is stated in terms of the quantile treatment response function  $q(d, \tau)$ , which is a conceptually different object than the triangular structural function  $t(d, \nu, \eta)$ . Hence, our prime object of interest is also different from that in those papers. However, since (3.2) necessarily implies (3.1), by defining the quantile response function  $q(d, \tau)$  as the quantile function of  $t(d, \nu, \eta)$  for fixed  $d$ , our approach is always applicable to (3.2).<sup>17</sup> At the same time, our approach does not impose the strong independence and continuity assumptions employed in (3.2).

The joint independence condition and scalar nature of  $\nu$  in (3.2) are essential for identification in the control function strategies used by Chesher (2003) and Imbens and Newey (2003). However, this condition expressly requires the instrument  $Z$  to be independent of the disturbances in the whole system, thus

<sup>16</sup>Chesher (2003) actually employs a local independence condition that suffices for local identification. Imbens and Newey (2003) also analyze identification of average derivatives and other functionals of  $t(d, \nu, \eta)$  without the condition that  $\eta$  is a scalar.

<sup>17</sup>This works irrespective of the dimension of  $\eta$ . Even if  $\eta$  is multidimensional, our approach models the identifiable object  $q(d, \tau)$  defined as the  $\tau$ -quantile function of  $t(d, \nu, \eta)$ .

ruling out potentially useful instruments that will remain valid in our approach. For example, Imbens and Angrist (1994) provide important examples in which the instrument  $Z$  is assigned depending on  $D$ , though independently of potential outcomes. In another example, Hausman (1977) shows that when  $Z$  is measured with error the triangular structure (3.2) becomes inapplicable.<sup>18</sup> Our approach accommodates these situations, since it allows  $V$  in  $D = \varphi(Z, V)$  to be of arbitrary dimension and some components of  $V$  to be correlated with  $Z$ .

#### 4. CONCLUSION

In many observational studies, the variables of interest are endogenous, making quantile regression inappropriate for recovering the causal effects of these variables on the quantiles of economic outcomes. In this paper, we have developed a model of quantile treatment effects in the presence of endogeneity and obtained conditions for identification of the QTE without functional form assumptions. The key feature of the model is the imposition of conditions that restrict the evolution of the distribution of the ranks across treatment states. This feature allows us to overcome the endogeneity problem and recover the true QTE through the use of instrumental variables. The proposed model can also be equivalently viewed as a structural simultaneous equation model with nonadditive errors.

There are several interesting directions for further work. The previous discussion of identification of QTE relates to the broader issue of identification of the joint distribution of outcomes. Our approach identifies the marginal quantiles of potential outcomes, as is typically required in welfare analysis, but does not identify the joint distribution of outcomes, unless the extreme case of rank similarity (rank invariance) is imposed. However, one may potentially adopt the approach of Heckman, Smith, and Clements (1997) to put bounds on the joint distributions even under rank similarity. In their approach, developed for the exogenous setting, one first assumes a bound on the degree of slippage in the ranks, and then derives bounds on the joint distributions from the marginal distributions. Implementation of this approach in the present endogenous setting is an interesting direction for further work.

*Dept. Economics, Massachusetts Institute of Technology, E52-262F, 50 Memorial Drive, Cambridge, MA 02142, U.S.A.; vchern@mit.edu; www.mit.edu/~vchern*  
and

*Graduate School of Business, University of Chicago, 1101 East 58th Street, Chicago, IL 60637, U.S.A.; chansen1@gsb.uchicago.edu.*

*Manuscript received January, 2002; final revision received March, 2004.*

<sup>18</sup>In this case, the selection equation contains the original structural disturbance and an additional measurement error correlated with  $Z$ , violating model (3.2).

## APPENDIX A: PROOF OF THEOREM 1

Conditioning on  $X = x$  is suppressed. We first show the result under rank invariance A4(a) and then generalize to rank similarity A4(b). Under rank invariance,  $U_d = U$  for each  $d$ , so that a.s.

$$(A.1) \quad P[Y \leq q(D, \tau)|Z] \stackrel{(1)}{=} P[q(D, U) \leq q(D, \tau)|Z] \stackrel{(2)}{=} P[U \leq \tau|Z] \stackrel{(3)}{=} \tau,$$

where equality (1) follows from representation A1, A5, and rank invariance. Define the inverse of  $q(d, \tau)$  in its second argument as  $q_2^{-1}(D, y) \equiv \inf\{p: q(D, p) \geq y\}$ . Because  $q(d, \tau)$  is strictly increasing in  $\tau$ , it is one-to-one and its inverse is also strictly increasing by A1. Hence, applying the inverse to both sides of the inequality  $q(D, U) \leq q(D, \tau)$ , we have the equivalence of two events

$$(A.2) \quad \{q(D, U) \leq q(D, \tau)\} = \{U \leq \tau\},$$

which shows (2). Lastly, (3) follows by the independence condition A2 and rank invariance A4(a). A similar argument shows that  $P[Y < q(D, \tau)|Z] = \tau$  a.s.

Now let us relax the rank invariance assumption A4(a), and assume rank similarity A4(b). For  $P$ -a.e. value  $z$  of  $Z$ ,

$$\begin{aligned} (A.3) \quad P[Y \leq q(D, \tau)|Z = z] &\stackrel{(1)}{=} P[q(D, U_D) \leq q(D, \tau)|Z = z] \\ &\stackrel{(2)}{=} P[U_D \leq \tau|Z = z] \\ &\stackrel{(3)}{=} \int P[U_D \leq \tau|Z = z, V = v] dP[V = v|Z = z] \\ &\stackrel{(4)}{=} \int P[U_{\delta(z, v)} \leq \tau|Z = z, V = v] dP[V = v|Z = z] \\ &\stackrel{(5)}{=} \int P[U_0 \leq \tau|Z = z, V = v] dP[V = v|Z = z] \\ &\stackrel{(6)}{=} P[U_0 \leq \tau|Z = z] \\ &\stackrel{(7)}{=} \tau. \end{aligned}$$

Equality (1) is by A1 and A5. Equality (2) is immediate from the equivalence relation  $\{q(D, U_D) \leq q(D, \tau)\} = \{U_D \leq \tau\}$ , which follows similarly to (A.2). Equality (3) is by definition. Equality (4) is by the representation A3. Equality (5) is by the similarity assumption A4(b) and representation A3: Conditional on  $(V = v, X = x, Z = z)$ ,  $D = \delta(z, v)$  is a constant, so that by A4(b)

$$(A.4) \quad U_{\delta(z, v)} \text{ equals in distribution } U_0,$$

where “0” denotes any fixed value of  $D$ . Equality (6) is by definition, and equality (7) is by the independence assumption A2. Similarly, we conclude  $P[Y < q(D, \tau)|Z] = \tau$  a.s.

Finally, the conclusion that  $U_D$  is  $U(0, 1)$  conditional on  $Z$  follows from (A.3).

## APPENDIX B: PROOF OF THEOREM 2

The result can be deduced from the high-level technical results used in Theorem 3, but this case is simple enough to give an elementary proof that highlights the essence of what is required for point identification. Consider the two nonlinear curves (iso-probability curves) defined as

$$(B.1) \quad \mathcal{Y} = \{(y_0, y_1) : P[Y \leq y_D|Z = 0] = \tau\} \quad \text{and} \quad \tilde{\mathcal{Y}} = \{(y_0, y_1) : P[Y \leq y_D|Z = 1] = \tau\}.$$

By Theorem 1 and by condition (ii),  $q = (q(0), q(1)) \in \mathcal{L}$ . Hence, we need to check whether  $y = q$  is the only solution to  $\Pi(y) = 0$  over  $\mathcal{L}$ . The common set of solutions of the two moment

equations defined in (B.1) over  $\mathcal{L}$  is given by  $(\mathcal{Y} \cap \tilde{\mathcal{Y}} \cap \mathcal{L})$ . By Theorem 1 and by condition (ii),  $q \in (\mathcal{Y} \cap \tilde{\mathcal{Y}} \cap \mathcal{L})$ . Point identification requires that  $q$  be the only point in  $(\mathcal{Y} \cap \tilde{\mathcal{Y}} \cap \mathcal{L})$ .

Under the stated conditions, curves  $\mathcal{Y}$  and  $\tilde{\mathcal{Y}}$  satisfy the differential equations:

$$(B.2) \quad f_{Y,D}(y_0, 0|Z=0) dy_0 + f_{Y,D}(y_1, 1|Z=0) dy_1 = 0 \quad \text{and} \\ f_{Y,D}(y_0, 0|Z=1) dy_0 + f_{Y,D}(y_1, 1|Z=1) dy_1 = 0.$$

Thus the slopes of the curves  $\mathcal{Y}$  and  $\tilde{\mathcal{Y}}$  at  $y = (y_0, y_1)$  in  $\mathcal{L}$  are given by

$$(B.3) \quad \left( \frac{dy_0}{dy_1} \right) \Big|_{(y_0, y_1)} = -\frac{f_{Y,D}(y_1, 1|Z=0)}{f_{Y,D}(y_0, 0|Z=0)} \quad \text{and} \quad \left( \widetilde{\frac{dy_0}{dy_1}} \right) \Big|_{(y_0, y_1)} = -\frac{f_{Y,D}(y_1, 1|Z=1)}{f_{Y,D}(y_0, 0|Z=1)}.$$

The slopes take values only in  $[-\infty, 0]$  since entries of  $\Pi'(y)$  are nonnegative and rank  $\Pi'(y)$  is full. Moreover, the slopes are not equal to each other when evaluated at the same point  $(y_0, y_1)$  in  $\mathcal{L}$ , if rank  $\Pi'(y)$  is full. It is intuitively clear that if one slope is bigger than the other at some relevant points, the curves  $\mathcal{Y}$  and  $\tilde{\mathcal{Y}}$  intersect over  $\mathcal{L}$  only once, hence the solution set  $(\mathcal{Y} \cap \tilde{\mathcal{Y}} \cap \mathcal{L})$  should be a singleton. The following argument shows that it suffices that the slopes are different only when evaluated at any *same* point  $(y_0, y_1)$  in  $\mathcal{L}$ .

Since we suppose that the Jacobian is continuous, the determinant of  $\Pi'(y)$  is continuous in  $y = (y_0, y_1)$  over  $\mathcal{L}$ . Hence the assumption rank  $\Pi'(y) = 2$  for all  $y \in \mathcal{L}$  is equivalent to  $\det \Pi'(y) > (\text{or } <) 0$  for all  $y \in \mathcal{L}$ . This is equivalent to the condition

$$(B.4) \quad \frac{f_{Y,D}(y_1, 1|Z=0)}{f_{Y,D}(y_0, 0|Z=0)} > \frac{f_{Y,D}(y_1, 1|Z=1)}{f_{Y,D}(y_0, 0|Z=1)} \quad \text{for all } y = (y_0, y_1) \in \mathcal{L}, \quad \text{or} \\ \frac{f_{Y,D}(y_1, 1|Z=0)}{f_{Y,D}(y_0, 0|Z=0)} < \frac{f_{Y,D}(y_1, 1|Z=1)}{f_{Y,D}(y_0, 0|Z=1)} \quad \text{for all } y = (y_0, y_1) \in \mathcal{L}.$$

In the main text, this was interpreted as the monotone likelihood ratio property.

By  $\mathcal{L}$  being compact and Milnor (1965, p. 8), the set  $(\mathcal{Y} \cap \tilde{\mathcal{Y}} \cap \mathcal{L})$  is finite. Denote its points as a collection of vectors  $(y^{(j)}, j = 1, \dots, k)$  for  $k < \infty$ , where each  $y^{(j)}$  is of the form  $y^{(j)} = (y_0^{(j)}, y_1^{(j)}) \in \mathcal{L}$ . If  $k > 1$ , there must be at least two solutions  $y^{(j)}$  and  $y^{(j')}$  in  $\mathcal{L}$  such that the slopes of the iso-probability curves satisfy the relations

$$(B.5) \quad \left( \frac{dy_0}{dy_1} \right) \Big|_{y^{(j)}} > \left( \widetilde{\frac{dy_0}{dy_1}} \right) \Big|_{y^{(j)}} \quad \text{and} \quad \left( \frac{dy_0}{dy_1} \right) \Big|_{y^{(j')}} < \left( \widetilde{\frac{dy_0}{dy_1}} \right) \Big|_{y^{(j')}}.$$

That is, if there exist multiple crossings in  $\mathcal{L}$  of curve  $\mathcal{Y}$  by curve  $\tilde{\mathcal{Y}}$ , for any crossing point  $y^{(j)}$  of  $\mathcal{Y}$  by  $\tilde{\mathcal{Y}}$  from above, there must be another crossing point  $y^{(j')}$  where the direction of crossing is the opposite—from below. (If  $k > 1$ , crossings cannot be all from above or all from below since the slopes of the curves are restricted to  $[-\infty, 0]$ .) Crossing from opposite directions implies (B.5), posing a contradiction to (B.4) and (B.3).

## APPENDIX C: GENERALIZATIONS

In this section we generalize the identification work to nonbinary treatments. Note that the definition of the model in A1–A5 and Theorem 1 do not depend on treatments being binary. The following analysis is all conditional on  $X = x$  and for a given quantile  $\tau \in (0, 1)$ , but we suppress this dependence for ease of notation. First consider the case when  $D$  has the support  $\{1, \dots, l\}$  and  $Z$  has the support  $\{1, \dots, r\}$  ( $l \leq r < \infty$ ). Note that function  $q(\cdot)$  can be represented by a vector  $q = (q(1), \dots, q(l)) \in \mathbb{R}^l$ . Under conditions of Theorem 1 there is at least one function  $q(d)$  that solves  $P[Y \leq q(D)|Z] = \tau$  a.s. Therefore, for vectors of the form  $y = (y_1, \dots, y_l)$  and the vector of moment equations

$$(C.1) \quad \Pi(y) = (P[Y \leq y_D | Z = z] - \tau, z = 1, \dots, r)',$$

where  $y_D \equiv \sum_d \mathbb{1}[D = d] \cdot y_d$ , the identification question is whether  $y = q$  uniquely solves  $\Pi(y) = 0$ .

Fix some small constants  $\delta > 0$  and  $\underline{f} > 0$ , and define the set  $\mathcal{L}$  as the convex hull of all vectors  $y = (y_1, \dots, y_l)$  that satisfy (i) for each  $z$ ,  $P[Y \leq y_D | Z = z] \in [\tau - \delta, \tau + \delta]$ , where  $y_D \equiv \sum_{d=1}^l y_d \cdot \mathbb{1}(D = d)$ , and (ii) for each  $d$ ,  $y_d \in s_d \equiv \{\lambda : f_Y(\lambda | d, z) > \underline{f} > 0 \text{ for all } z \text{ with } P[D = d | Z = z] > 0\}$ , where  $d$  denotes elements in the support of  $D$  and  $z$  denotes elements in the support of  $Z$ . The parameter space  $\mathcal{L}$  contains potential solutions to the equations  $\Pi(y) = 0$ . Imposition of further economic restrictions  $\mathcal{L}^R$  on  $\mathcal{L}$  can be useful in applications (as discussed in the main text). In that case, provided  $\mathcal{L}^R \cap \mathcal{L}$  is convex,  $\mathcal{L}$  can be replaced by  $\mathcal{L}^R \cap \mathcal{L}$  without affecting the argument given below.

Identification will hold when a rank condition is imposed on each element of  $\mathcal{L}$ . Define

$$\Pi'(y) = \frac{d}{dy'} \Pi(y)$$

as the Jacobian matrix of  $\Pi(\cdot)$  with a typical  $(z, d)$ th element given by  $f_Y(y_d | D = d, Z = z) \times P[D = d | Z = z]$ , where  $d = 1, \dots, l$ ,  $z = 1, \dots, r$ . Let  $\Pi_m(y)$  denote the  $m$ th  $(l \times 1)$  sub-block of  $\Pi(y)$  for some ordering  $1, \dots, M$  of all  $(l \times 1)$  sub-blocks, and let  $\Pi'_m(y)$  denote the corresponding Jacobian. The following theorem generalizes Theorem 2.

**THEOREM 3 (Identification for Discrete  $D$ ):** *Suppose A1–A5 hold, and the support of  $D$  is  $\{1, \dots, l\}$  and of  $Z$  is  $\{1, \dots, r\}$ . Assume that for the sets  $\mathcal{L}$  and  $s_d$  specified above, (i)  $q(d) \in s_d$  for each  $d$ , and (ii)  $\Pi'(y)$  is continuous for all  $y \in \mathbb{R}^l$ . Then  $q = (q(1), \dots, q(l))'$  is identified, if  $\mathcal{L}$  can be covered by convex compact sets  $\{\mathcal{L}_j\}$  with smooth boundary such that for each  $j$ ,  $q \in \mathcal{L}_j$  and there is  $m(j)$  such that*

$$(C.2) \quad \text{rank}[\Pi'_{m(j)}(y)] \text{ is full for all } y \in \mathcal{L}_j,$$

*and either  $\Pi'_{m(j)}(y)$  is positive quasi-definite along the boundary of  $\mathcal{L}_j$ , in the sense of Mas-Colell (1979a), or, more generally,  $\Pi_{m(j)}(\mathcal{L}_j)$  is simply connected.*

**COMMENT C.1:** The proof uses variants of Hadamard's global inverse function theorem given by Ambrosetti and Prodi (1995) and Mas-Colell (1979a); also see Mas-Colell (1979b) for useful discussion. Here, application of Theorem 2 from Mas-Colell (1979a) requires the technical condition on principal minors, while application of Theorem 1.8 from Ambrosetti and Prodi (1995) requires the simple connectivity of the image set  $\Pi_{m(j)}(\mathcal{L}_j)$ .

**PROOF:** By Theorem 1 and by condition (i),  $q \in \mathcal{L}$ . Hence, we need to check whether  $y = q$  is the only solution to  $\Pi(y) = 0$  over  $\mathcal{L}$ . Consider a covering set  $\mathcal{L}_j$ . By A1–A5, Theorem 1 implies that  $\Pi_{m(j)}(q) = 0$ , where  $\Pi_{m(j)}(y)$  denotes the  $m(j)$ th  $(l \times 1)$  sub-block of  $\Pi(y)$ . By assumption  $q \in \mathcal{L}_j$ . The stated rank conditions and compactness and convexity of  $\mathcal{L}_j$  imply that the mapping  $y \rightarrow \Pi_{m(j)}(y)$  is a homeomorphism (one-to-one) between  $\mathcal{L}_j$  and  $\Pi_{m(j)}(\mathcal{L}_j)$ , by the global inverse function theorem; cf. Theorem 2 in Mas-Colell (1979a) or Theorem 1.8 in Ambrosetti and Prodi (1995). Thus,  $y = q$  is the unique solution of  $\Pi_{m(j)}(y) = 0$  over  $\mathcal{L}_j$ . Since this argument applies to every  $j$  and  $\{\mathcal{L}_j\}$  cover  $\mathcal{L}$ , it follows that  $y = q$  is the unique solution of  $\Pi(y) = 0$  over  $\mathcal{L}$ . *Q.E.D.*

Finally we consider continuous  $D$ . Recall that in the IV equation  $E(Y - \mu(D) | Z) = 0$  a.s. the condition for identification of a bounded function  $\mu(\cdot)$  is the bounded completeness condition (cf. Newey and Powell (2003)).<sup>19</sup> For any bounded function  $\Delta(d) = m(d) - \mu(d)$ ,

<sup>19</sup>Actually, L1 is weaker than the unbounded completeness used by Newey and Powell (2003); see Mattner (1993).

L1:  $E[\Delta(D)|Z] = 0$  a.s.  $\Rightarrow \Delta(D) = 0$  a.s.

Lehmann (1959) provided a fundamental sufficient condition for L1:

L2:  $f_D(d|z)$  is a full rank exponential or any other boundedly-complete family.

For instance, the full rank exponential family condition means that  $f_D(d|z) \propto \exp(\eta(z)'T(d) + h(d) + \lambda(z))$ , where  $\eta(z)$  must vary over an open rectangle in  $\mathbb{R}^{\dim(T(d))}$ , which imposes that the number of continuous, jointly nondegenerate instruments should be larger than or equal to  $\dim(T(d)) \geq \dim(d)$ ; cf. Newey and Powell (2003). Condition L2 is reasonable because the exponential families include or approximate a broad variety of nonparametric distributions; cf. Stone (1991).

In our case, we can transform both  $Y$  and  $D$  to have bounded support, without loss of generality. Fix some small constants  $\delta, \underline{f} > 0$ , and define the relevant parameter space  $\mathcal{L}$  as the convex hull of functions  $m(\cdot)$  that satisfy: (i) for each  $z$ ,  $P[Y \leq m(D)|z] \in [\tau - \delta, \tau + \delta]$  and (ii) for each  $d$ ,  $m(d) \in s_d \equiv \{y: f_Y(y|d, z) \geq \underline{f} > 0 \text{ for all } z \text{ with } f_D(d|z) > 0\}$ , where  $d$  denotes elements in the support of  $D$  and  $z$  denotes elements in the support of  $Z$ . For any bounded  $\Delta(d) = m(d) - q(d)$  with  $m(\cdot) \in \mathcal{L}$  and  $\epsilon \equiv Y - q(D)$ , consider two conditions:

L1\*:  $E[\Delta(D) \cdot \omega(D, Z)|Z] = 0$  a.s.  $\Rightarrow \Delta(D) = 0$  a.s., for  $\omega(D, Z) = \int_0^1 f_\epsilon(\delta \Delta(D)|D, Z) d\delta > 0$ .

L2\*:  $\varphi(d|z) \equiv c(z) \cdot \omega(d, z) \cdot f_D(d|z)$  is a full rank exponential or other boundedly-complete family.<sup>20</sup>

COMMENT C.2: Condition L1\* is a *bounded completeness condition*, an analog of L1; and L2\* is an analog of L2. These conditions suffice for global identification. The *local bounded completeness conditions* (L1' and L2') are obtained simply by replacing  $\omega(D, Z)$  with  $\omega'(D, Z) = f_\epsilon(0|D, Z)$  in L1\* and L2\*. These conditions suffice only for local identification, but are easier to check.<sup>21</sup>

THEOREM 4 (Identification for Continuous  $D$ ): Suppose A1–A5 hold, and both  $Y$  and  $D$  have bounded support. Assume that (i) the density  $f_\epsilon(e|D, Z)$  is continuous and bounded in  $e$  over  $\mathbb{R}$ , a.s., and (ii)  $q(d) \in s_d$  for each  $d$ . Then  $q(\cdot)$  is identified,<sup>22</sup> if L1\* (or L2\*) holds.

PROOF: By Theorem 1,  $q$  solves  $P[Y \leq q(D)|Z] = \tau$  a.s., and by condition (ii)  $q \in \mathcal{L}$ . Hence, we need to check whether  $q$  is the only solution to  $P[Y \leq q(D)|Z] = \tau$  a.s. in  $\mathcal{L}$ . Suppose there is  $m \in \mathcal{L}$  such that  $P[Y \leq m(D)|Z] = \tau$  a.s. Define  $\Delta(d) \equiv m(d) - q(d)$ , and write

$$\begin{aligned} \text{(C.3)} \quad P[Y \leq m(D)|Z] - P[Y \leq q(D)|Z] &\stackrel{(1)}{=} E \left[ E \left[ \int_0^1 f_\epsilon(\delta \Delta(D)|D, Z) \Delta(D) d\delta \middle| D, Z \right] \middle| Z \right] \\ &\stackrel{(2)}{=} E \left[ \int_0^1 f_\epsilon(\delta \Delta(D)|D, Z) \Delta(D) d\delta \middle| Z \right] \\ &\stackrel{(3)}{=} E[\Delta(D) \cdot \omega(D, Z)|Z]. \end{aligned}$$

<sup>20</sup>The constant  $c(z) > 0$  is chosen so that  $\varphi(d|z)$  integrates to one over the support of  $D$  given  $Z = z$ .

<sup>21</sup>See Chernozhukov, Imbens, and Newey (2003).

<sup>22</sup>I.e., for any other  $m(\cdot) \in \mathcal{L}$  such that  $P[Y \leq m(D)|Z] = \tau$  a.s.,  $m(D) = q(D)$  a.s.

Noting that  $\Delta(D)$  and  $f_\epsilon(\delta\Delta(D)|D, Z)$  are bounded, observe that (1) follows by the fundamental theorem of calculus, (2) by the law of iterated expectations, and (3) by linearity of integration. For uniqueness, we need that  $(C.3) = 0$  a.s.  $\Rightarrow \Delta(D) = 0$  a.s. This shows sufficiency of  $L1^*$ .

Since  $E[\Delta(D) \cdot \omega(D, z)|z] \propto E_{\varphi(\cdot|z)}[\Delta(D)]$  and by condition (ii)  $\{\varphi(d|z) = 0\} \Leftrightarrow \{f_D(d|z) = 0\}$ , it follows by Lehmann (1959) that  $E_{\varphi(\cdot|z)}[\Delta(D)] = 0$   $P$ -a.e.  $\Rightarrow \Delta(D) = 0$   $P$ -a.s. Here  $E_{\varphi(\cdot|z)}$  denotes the expectation with  $\varphi(d|z)$  used as a density for  $d$ . Hence,  $L2^*$  is sufficient for  $L1^*$ .  $Q.E.D.$

## REFERENCES

- ABADIE, A. (1995): "Changes in Spanish Labor Income Structure During the 1980s: A Quantile Regression Approach," CEMFI Working Paper No. 9521.
- ABADIE, A., J. ANGRIST, AND G. IMBENS (2002): "Instrumental Variables Estimates of the Effect of Subsidized Training on the Quantiles of Trainee Earnings," *Econometrica*, 70, 91–117.
- AMBROSETTI, A., AND G. PRODI (1995): *A Primer of Nonlinear Analysis*, Vol. 34 of *Cambridge Studies in Advanced Mathematics*. Cambridge: Cambridge University Press.
- BLUNDELL, R., AND J. L. POWELL (2001): "Endogeneity in Semi-Parametric and Non-Parametric Regression Models," Working Paper, University College London and University of California at Berkeley.
- CHERNOZHUKOV, V., AND C. HANSEN (2001a): "Instrumental Quantile Regression Inference for Structural and Treatment Effect Models," *Journal of Econometrics*, forthcoming.
- (2001b): "An IV Model of Quantile Treatment Effects," Massachusetts Institute of Technology, Department of Economics, Working Paper No. 02-06.
- (2003): "The Impact of 401(K) on Savings: An Instrumental Quantile Regression Analysis," *Review of Economics and Statistics*, forthcoming.
- CHERNOZHUKOV, V., AND H. HONG (2003): "An MCMC Approach to Classical Estimation," *Journal of Econometrics*, 115, 293–346.
- CHERNOZHUKOV, V., H. HONG, AND E. TAMER (2003): "Parameter Set Inference in a Class of Econometric Models," Working Paper, Department of Economics, Massachusetts Institute of Technology.
- CHERNOZHUKOV, V., G. IMBENS, AND W. NEWEY (2003): "Instrumental Variable Estimation of Nonseparable Models," Working Paper, Department of Economics, Massachusetts Institute of Technology.
- CHESHER, A. (2003): "Identification in Nonseparable Models," *Econometrica*, 71, 1405–1441.
- DOKSUM, K. (1974): "Empirical Probability Plots and Statistical Inference for Nonlinear Models in the Two-Sample Case," *The Annals of Statistics*, 2, 267–277.
- DURRETT, R. (1996): *Probability: Theory and Examples* (Second Ed.). Belmont, CA: Duxbury Press.
- HAUSMAN, J. A. (1977): "Errors in Variables in Simultaneous Equation Models," *Journal of Econometrics*, 5, 389–401.
- HAUSMAN, J. A., AND J. G. SIDAK (2002): "Do the Poor and the Less-Educated Pay Higher Prices for Long-Distance Calls?" Working Paper, Department of Economics, Massachusetts Institute of Technology.
- HECKMAN, J., AND R. ROBB (1986): "Alternative Methods for Solving the Problem of Selection Bias in Evaluating the Impact of Treatments on Outcomes," in *Drawing Inference from Self-Selected Samples*, ed. by H. Wainer. New York: Springer-Verlag, 63–107.
- HECKMAN, J., J. SMITH, AND N. CLEMENTS (1997): "Making the Most out of Programme Evaluations and Social Experiments: Accounting for Heterogeneity in Programme Impacts," *Review of Economic Studies*, 64, 487–535.
- HECKMAN, J., AND E. J. VYTLACIL (1999): "Local Instrumental Variables and Latent Variable Models for Identifying and Bounding Treatment Effects," *Proceeding of the National Academy of Sciences USA*, 96, 4730–4734 (electronic).



- HONG, H., AND E. TAMER (2003): "Inference in Censored Models with Endogenous Regressors," *Econometrica*, 71, 905–932.
- HONORE, B., AND L. HU (2004): "On the Performance of Some Robust Instrumental Variables Estimators," *Journal of Business and Economic Statistics*, forthcoming.
- IMBENS, G. W., AND J. D. ANGRIST (1994): "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62, 467–475.
- IMBENS, G. W., AND W. K. NEWEY (2003): "Estimation and Identification of Triangular Equations Models Without Additivity," Working Paper, Department of Economics, Massachusetts Institute of Technology.
- JANUSZEWSKI, S. I. (2002): "The Effect of Air Traffic Delays on Airline Prices," Working Paper, Department of Economics, University of California, San-Diego.
- KOENKER, R., AND O. GELING (2001): "Reappraising Medfly Longevity: A Quantile Regression Survival Analysis," *Journal of American Statistical Association*, 96, 458–468.
- KOENKER, R., AND K. HALLOCK (2001): "Quantile Regression: An Introduction," *Journal of Economic Perspectives*, 15, 143–156.
- LEHMANN, E. L. (1959): *Testing Statistical Hypotheses*. New York: John Wiley & Sons.
- (1974): *Nonparametrics: Statistical Methods Based on Ranks*. San Francisco, CA: Holden-Day.
- MACURDY, T., AND C. TIMMINS (2001): "Bounding the Influence of Attrition on the Intertemporal Wage Variation in the NLSY," Working Paper, Department of Economics, Yale University.
- MANSKI, C. F., AND E. TAMER (2002): "Inference on Regressions with Interval Data on a Regressor or Outcome," *Econometrica*, 70, 519–546.
- MAS-COLELL, A. (1979a): "Homeomorphisms of Compact, Convex Sets and the Jacobian Matrix," *SIAM Journal of Mathematical Analysis*, 10, 1105–1109.
- (1979b): "Two Propositions on the Global Univalence of Systems of Cost Function," in *General Equilibrium, Growth, and Trade*. New York: Academic Press, 323–331.
- MATTNER, L. (1993): "Some Incomplete but Boundedly Complete Location Families," *The Annals of Statistics*, 21, 2158–2162.
- MILNOR, J. (1965): *Topology from the Differentiable Viewpoint. Based on Notes by David W. Weaver*. Charlottesville, VA: University Press of Virginia.
- NEWEY, W. K., AND J. L. POWELL (2003): "Instrumental Variable Estimation of Nonparametric Models," *Econometrica*, 71, 1565–1578.
- ROEHRLIG, C. S. (1988): "Conditions for Identification in Nonparametric and Parametric Models," *Econometrica*, 56, 433–447.
- SAKATA, S. (2001): "Instrumental Variable Estimation Based on the Least Absolute Deviation Estimator," Working Paper, Department of Economics, University of Michigan.
- STONE, C. J. (1991): "Asymptotics for Doubly Flexible Logspline Response Models," *The Annals of Statistics*, 19, 1832–1854.