Tighter Bounds in Triangular Systems[*]

Sung Jae Jun,[†] Joris Pinkse[‡] and Haiqing Xu[§]

Center for Auctions, Procurements and Competition Policy

Department of Economics

The Pennsylvania State University

October 2010

We study a nonparametric triangular system with (potentially discrete) endogenous regressors and nonseparable errors. Like in other work in this area, the parameter of interest is the structural function evaluated at particular values. We impose a global exclusion and exogeneity condition, in contrast to Chesher (2005), but develop a rank condition which is weaker than Chesher's. The alternative rank condition can be satisfied for binary endogenous regressors, and it often leads to a tighter identified interval than Chesher (2005)'s minimum length interval. We illustrate the potential of the new rank condition using the Angrist and Krueger (1991) data.

Key Words: Nonparametric Triangular Systems; Control Variables; Weak Monotonicity; Partial Identification; Instrumental Variables; Rank Conditions.

JEL Classification Codes: C14; C30; C31.

[*]This paper is based on research supported by NSF grant SES–0922127. We thank seminar participants at the University of Pittsburgh, the University of British Columbia, the University of Western Ontario, the 2009 North American Econometric Society Summer meeting, the 2009 Canadian Econometrics Study Group meeting, the conference on quantile regression at University College London/CEMMAP, Andrew Chesher, Rosa Matzkin, Jeff Racine, Peter Robinson, Yuanyuan Wan, the associate editor, and the anonymous referees for helpful comments and suggestions. We thank the Human Capital Foundation for their support of CAPCP.

[†](corresponding author) 608 Kern Graduate Building, University Park 16802, sjun@psu.edu

[‡]joris@psu.edu; Joris Pinkse is an extramural fellow at Center, Tilburg University.

[§]hux100@psu.edu

1. INTRODUCTION

The primary objective of our paper is to obtain identification results — that are stronger than are currently available in the literature under alternative conditions — for the nonparametric triangular model

$$\begin{cases} \boldsymbol{y} = g(\boldsymbol{x}, \boldsymbol{u}), \\ \boldsymbol{x} = h(\boldsymbol{z}, \boldsymbol{v}), \end{cases} \tag{1}$$

where $\boldsymbol{y} \in \mathscr{S}_y \subset \mathbb{R}, \boldsymbol{x} \in \mathscr{S}_x \subset \mathbb{R}^d, \boldsymbol{z} \in \mathscr{S}_z \subset \mathbb{R}^{d_z}$ are observables, g, h are unknown functions, and $\boldsymbol{u} \in \mathscr{U} = (0,1], \boldsymbol{v} \in \mathscr{V} \subseteq \mathscr{U}^d$ are errors. We refer to $\boldsymbol{x}$ as endogenous regressors and $\boldsymbol{z}$ as instruments and use bold face symbols to denote random variables and regular face symbols for (nonrandom) values the corresponding random variable can take. Similar to Chesher (2005), the regressors need not be continuous and the objective is identification of the object

$$\psi^* = \psi(x^*, \tau^*, v^*) = g\big(x^*, Q_{\boldsymbol{u}|\boldsymbol{v}}(\tau^*|v^*)\big) \tag{2}$$

for given values of $(\tau^*, x^*, v^*) \in \mathscr{U} \times \mathscr{S}_x \times \mathscr{V}$, where $Q_{\boldsymbol{u}|\boldsymbol{v}}(\tau|v) = \inf\{u : \mathbb{P}[\boldsymbol{u} \leq u|\boldsymbol{v} = v] \geq \tau\}$.

If, for the sake of intuition, one attaches the labels 'earnings' to $\boldsymbol{y}$, 'education' to $\boldsymbol{x}$, 'demographics' to $\boldsymbol{z}$, 'talent' to $\boldsymbol{v}$, and '(market) success' to $\boldsymbol{u}$, then $\psi(x^*, 0.5, 0.5)$ can be interpreted as the (counterfactual) earnings of someone with median success and median talent if she were given education x^*.[1] Identification of marginal effects such as $\psi(x^*, 0.5, 0.5) - \psi(x^{**}, 0.5, 0.5)$ naturally follows from the identification of $\psi(x^*, 0.5, 0.5)$ and $\psi(x^{**}, 0.5, 0.5)$.

A model similar to (1) was studied in Chesher (2003) and Chesher (2005). Chesher (2003) used a strict monotonicity assumption, excluding discrete–valued $\boldsymbol{x}$, to identify the partial derivatives of g with respect to $\boldsymbol{x}$. Ma and Koenker (2006) and Jun (2009) proposed a parametric and a semiparametric estimator of Chesher's 2003 model, respectively. Chesher (2005) is more closely related to our paper in the sense that $\boldsymbol{x}$ is allowed to be discrete and that the object of interest is also ψ^*. The object of estimation in Newey, Powell, and Vella (1999) and Pinkse (2000) is $g\big(x^*, \mathbb{E}(\boldsymbol{u}|\boldsymbol{v} = v^*)\big)$, which is similar to ψ^*, but in those papers the errors are assumed additively separable in both equations in (1).

In Chesher (2003) regressors are assumed to be continuous in which case point identification of the partial derivatives of g can be achieved by using strict monotonicity conditions on the second argument of g and h. However, when $\boldsymbol{x}$ is discrete, as in the example of the years of schooling, strict monotonicity cannot hold. In Chesher (2005) (identified) bounds are obtained for ψ^* under

[1] We use the term 'success' to emphasize the potential dependence between $\boldsymbol{u}, \boldsymbol{v}$.

weak monotonicity, a dependence condition on $\boldsymbol{u}$ and $\boldsymbol{v}$, and 'local exclusion' and 'local exogeneity' conditions on the instrument $\boldsymbol{z}$.[2] We present our results under 'global' rather than local conditions, i.e. we impose a global exclusion restriction ($\boldsymbol{z}$ does not enter g) and assume that $\boldsymbol{z}$ is independent of $\boldsymbol{u}, \boldsymbol{v}$. Global conditions are stronger than local ones, but we note that those conditions are not testable and that global conditions are more common in multi–equations models.[3] Further, our global conditions allow us to replace the rank condition in Chesher (2005) ($\mathscr{R}$) with an alternative, weaker, rank condition ($\mathscr{R}^*$) which allows for the construction of tighter bounds on ψ^* than those obtained in Chesher (2005).[4] Moreover, in the case of binary regressors $\mathscr{R}$ is never satisfied, but $\mathscr{R}^*$ developed in this paper usually holds and in some cases leads to point identification of ψ^*; the example we provide exploits continuous variation in $\boldsymbol{z}$. A more precise and detailed discussion follows in the next section. Section 3 contains an empirical example illustrating the difference between $\mathscr{R}$ and $\mathscr{R}^*$.

Results similar to those developed in this paper can in principle be established under local conditions, also. However, obtaining much tighter bounds under conditions that are meaningfully different from the global ones results in conditions that are exceedingly difficult to interpret; see Jun, Pinkse, and Xu (2009), which is available on our website. Chesher (2005) establishes that his bounds are tight in a point identification example; our paper does not provide insights as to whether or not Chesher's bounds under his conditions are tight more generally.

Alternatively, one can conduct the analysis conditional on a subset $\mathscr{S}_z^*$ of instrument values. Our results go through without modification provided that all conditions and results are interpreted conditional on $\boldsymbol{z} \in \mathscr{S}_z^*$. Since conditioning on $\mathscr{S}_z^*$ amounts to throwing away information, doing so generally yields wider bounds than if global exclusion/exogeneity can be assumed to hold without such conditioning. But it is weaker than global exclusion/exogeneity and it does allow for instruments to enter into the g–function directly, albeit subject to the strong condition that the g–function value is the same for all $z \in \mathscr{S}_z^*$.

The methodology developed in this paper can be applied in other settings. For instance, Jun, Pinkse, and Xu (2010) provide an extension of the identification method proposed in this paper, which is applied to the model of Vytlacil and Yildiz (2007); the Vytlacil–Yıldız results for binary endogenous regressors are extended to cover discrete endogenous regressors that can take more than two values and their support restrictions are relaxed in the case of binary regressors.

---

[2] Starting from $\boldsymbol{y} = \tilde{g}(\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{u})$, Chesher (2005) assumed that there exist $z_1, z_2 \in \mathscr{S}_z$ such that for some r^*, $Q_{\boldsymbol{u}|\boldsymbol{v},\boldsymbol{z}}(\tau^*|v^*, z_1) = Q_{\boldsymbol{u}|\boldsymbol{v},\boldsymbol{z}}(\tau^*|v^*, z_2) = r^*$ (local exogeneity) and $\tilde{g}(x^*, z_1, r^*) = \tilde{g}(x^*, z_2, r^*)$ (local exclusion).

[3] This is true for traditional linear simultaneous equations models, as well as for the bulk of the modern literature, e.g. **?**.

[4] Cases exist in which the bounds are the same.

The proof of our theorem relies on an inversion of the conditional distribution function $\Pi(y|x^*, v^*) = \mathbb{P}[\boldsymbol{y} \leq y|\boldsymbol{x} = x^*, \boldsymbol{v} = v^*]$. Therefore, our methodology can be used to derive bounds on other functionals of Π such as the mean $\delta(x^*, v^*) = \mathbb{E}[\boldsymbol{y}|\boldsymbol{x} = x^*, \boldsymbol{v} = v^*]$, which is in fact the quantity of interest in Manski and Tamer (2002).

There are certain similarities between Manski and Tamer (2002) and Chesher (2005), and indeed our paper. There are however several differences besides the difference in object of interest (mean versus quantile) noted earlier. First, the primary objective in Manski and Tamer (2002) is estimation, whereas in Chesher (2005) and here it is identification. Second, in Manski and Tamer (2002) upper and lower bounds $(\boldsymbol{v}_0^{MT}, \boldsymbol{v}_1^{MT})$ on $\boldsymbol{v}$ are assumed to be available while Chesher (2005) provides conditions (involving instrumental variables) under which such bounds are available and can be used. Finally, if the Manski and Tamer (2002) bounds were used in the quantile context, the bounds that obtain after inversion of Π would be the ones obtained by Chesher (2005), not the ones provided in this paper; see appendix C.2 for details.

Although we only provide identification results in this paper, the identification approach here can be implemented in practice. We are currently developing an estimator for ψ^* in a separate paper. This estimator assumes the existence of continuous instruments, which we do not assume for our identification result in the present paper. Developing an estimator which takes full advantage of the weakest set of identification results contained in this paper could be challenging.

Our paper is organized as follows. Section 2 contains the main results established in this paper. In section 3 we illustrate our proposal using the Angrist and Krueger (1991) data set.

2. Main Results

2.1. **Assumptions.** Consider again the model in (1). The objective remains to find identifiable bounds on ψ^* defined in (2) for given values of τ^*, x^*, v^*. We make the following assumptions.

Assumption A. *$\boldsymbol{u}, \boldsymbol{v}_1, \ldots, \boldsymbol{v}_d$ have (marginal) $U(0,1]$–distributions.*

Assumption B. *g is nondecreasing in u for all values of x and $h(z, v) = \big[h_1(z, v_1), \ldots, h_d(z, v_d)\big]^\intercal$, where h_j is nondecreasing and left–continuous in v_j for all values of z for $j = 1, \ldots, d$.*

Assumption C. *$\boldsymbol{u}, \boldsymbol{v}$ are independent of $\boldsymbol{z}$.*

Assumption D. *$\boldsymbol{u}$ is positive regression dependent on $\boldsymbol{v}$, i.e. $Q_{\boldsymbol{u}|\boldsymbol{v}}(\tau|v)$ is nondecreasing in v for all values of τ.*

Assumption E. *$\mathscr{X}(x^*, v^*) = \{z \in \mathscr{S}_z :\ h(z, v^*) = x^*\}$ is nonempty.*

Given that g, h are unknown, the distributional conditions in assumption A plus the weak monotonicity and left–continuity conditions in assumption B by themselves amount to normalizations; see appendix C.1. The assumption that only one error enters each h_j–equation is general since no dependence conditions are imposed between the $\boldsymbol{v}_j$'s. Assumptions A and B do become restrictive, however, when paired with the positive regression–dependence condition in assumption D. Assumption C is restrictive, as was discussed in the introduction. In the context of (1) and imposing C, the only addition in assumptions A, B and D over what is assumed in Chesher (2005) is that the direction of monotonicity of $Q_{\boldsymbol{u}|\boldsymbol{v}}(\tau|v)$ is specified. This is innocuous, because the same analysis can be repeated under the assumption of the other direction of monotonicity after which one can compare the resulting bounds with the bounds based on assumption D.

Assumption E requires that the type of individual for which bounds are desired exists. If there are no demographic characteristics z that yield an education level x^* for someone of talent v^*, then our procedure does not yield meaningful bounds for the earnings of someone with education level x^* and talent v^* for any level of success τ. Assumption E could be restrictive if one conditions on a subset $\mathscr{S}_z^*$ of demographic profiles, as discussed in the introduction.

2.2. **Basics.** We start by stating a lemma, which shows that if $\boldsymbol{v}$ were observable then ψ^* would be directly estimable from the data; $\boldsymbol{v}$ then plays the role of a control variable. The assumptions made above are presumed to hold for all lemmas.

Lemma 1. *For all* $\tau \in \mathscr{U}$, $\psi(x^*, \tau, v^*) = Q_{\boldsymbol{y}|\boldsymbol{x},\boldsymbol{v}}(\tau|x^*, v^*)$.

Proof. See appendix A. □

Lemma 1 implies that ψ^* can alternatively be interpreted as the τ^*–quantile of the earnings distribution of individuals with education x^* and talent v^*. Note that

$$h(z,v) = \begin{bmatrix} h_1(z,v_1) \\ \vdots \\ h_d(z,v_d) \end{bmatrix} = \begin{bmatrix} h_1\big(z, Q_{\boldsymbol{v}_1}(v_1)\big) \\ \vdots \\ h_d\big(z, Q_{\boldsymbol{v}_d}(v_d)\big) \end{bmatrix} = \begin{bmatrix} h_1\big(z, Q_{\boldsymbol{v}_1|\boldsymbol{z}}(v_1|z)\big) \\ \vdots \\ h_d\big(z, Q_{\boldsymbol{v}_d|\boldsymbol{z}}(v_d|z)\big) \end{bmatrix} = \begin{bmatrix} Q_{\boldsymbol{x}_1|\boldsymbol{z}}(v_1|z) \\ \vdots \\ Q_{\boldsymbol{x}_d|\boldsymbol{z}}(v_d|z) \end{bmatrix}, \tag{3}$$

where the second to fourth equalities follow from assumptions A to C, respectively. Hence, if the conditional distribution of $\boldsymbol{x}_j$ given $\boldsymbol{z} = z$ is continuous for all z then lemma 1 implies that h_j is invertible in its second argument and that $\boldsymbol{v}_j = F_{\boldsymbol{x}_j|\boldsymbol{z}}(\boldsymbol{x}_j|\boldsymbol{z})$, where $F_{\boldsymbol{x}_j|\boldsymbol{z}}$ is the conditional distribution function of $\boldsymbol{x}_j$ given $\boldsymbol{z}$. Therefore, the $\boldsymbol{v}_j$'s that correspond to continuous $\boldsymbol{x}_j$'s can be recovered from the data. For this reason we only discuss the case in which the elements of $\boldsymbol{x}$ are all discrete from hereon.

Let

$$V_j(x_j, z) = \left(\mathbb{P}[\boldsymbol{x_j} < x_j | \boldsymbol{z} = z],\ \mathbb{P}[\boldsymbol{x_j} \leq x_j | \boldsymbol{z} = z]\right] \tag{4}$$

for $j = 1, 2, \cdots, d$. Then, for any $v \in V(x, z) \equiv V_1(x_1, z) \times \cdots \times V_d(x_d, z)$ we have

$$h(z, v) \overset{3}{=} \left[Q_{\boldsymbol{x_1}|\boldsymbol{z}}(v_1|z),\ \cdots, Q_{\boldsymbol{x_d}|\boldsymbol{z}}(v_d|z)\right]^\intercal = x, \tag{5}$$

where the last equality follows from the definition of $V_j(x_j, z)$ in (4). Therefore, $V(x, z)$ is a set of talent levels for which individuals with demographics z achieve education level x,

$$V(x, z) = \{v \in \mathscr{U}^d :\ h(z, v) = x\}. \tag{6}$$

Please note that since $V(x, z)$ depends only on (a conditional distribution function of) observables, it is identified for all $(x, z) \in \mathscr{S}_x \times \mathscr{S}_z$.

2.3. **Basic Rank Condition.** Let

$$\begin{cases} \mathscr{G}^+(x, v) = \left\{V \in \mathscr{U}^d : \exists z \in \mathscr{S}_z : V = V(x, z) \geq v\right\}, \\ \mathscr{G}^-(x, v) = \left\{V \in \mathscr{U}^d : \exists z \in \mathscr{S}_z : V = V(x, z) \leq v\right\}, \end{cases} \tag{7}$$

where $V \geq v$ ($V \leq v$) means that no vectors in V have elements that are strictly less (greater) than the corresponding element of v. Intuitively, for $V(x, z)$ to belong to $\mathscr{G}^+(x, v)$, demographics z must be so unfavorable as to ensure that anyone with demographics z but talent less than v would not be able to achieve education x.

We now turn to the first of the two rank conditions mentioned in the introduction, namely $\mathscr{R}$.

Condition 1 ($\mathscr{R}$)**.** *Neither $\mathscr{G}^+(x^*, v^*)$ nor $\mathscr{G}^-(x^*, v^*)$ is empty.* □

$\mathscr{R}$ is due to Chesher (2005) as, under local conditions, is lemma 3 below. $\mathscr{R}$ requires the instrument to be strong enough to ensure that

$$\mathbb{P}\left[\boldsymbol{x_j} \leq x_j^* | \boldsymbol{z} = z\right] \leq v^* \leq \mathbb{P}\left[\boldsymbol{x_j} < x_j^* | \boldsymbol{z} = \tilde{z}\right], \quad j = 1, \ldots, d, \tag{8}$$

for some $z, \tilde{z} \in \mathscr{S}_z$. If $\mathscr{R}$ is satisfied, then the result of lemma 3 below follows almost immediately, using lemma 2 along the way. Let $Q_{\boldsymbol{u}|\boldsymbol{v}}(\tau|V)$ denote the τ quantile of the conditional distribution of $\boldsymbol{u}$ given that $\boldsymbol{v} \in V$.

Lemma 2. *For all $\tau \in \mathscr{U}$ and all $(x, z) \in \mathscr{S}_x \times \mathscr{S}_z$, if $V(x, z) \neq \emptyset$, then* $Q_{\boldsymbol{y}|\boldsymbol{x},\boldsymbol{z}}(\tau|x, z) = g\{x, Q_{\boldsymbol{u}|\boldsymbol{v}}(\tau|V(x, z))\}$.

Proof. See appendix A. □

Lemma 3. *Under $\mathscr{R}$ (condition 1),*

$$\sup_{\{z \in \mathscr{S}_z : V(x^*, z) \leq v^*\}} Q_{\boldsymbol{y}|\boldsymbol{x},\boldsymbol{z}}(\tau^*|x^*, z) \leq \psi^* \leq \inf_{\{z \in \mathscr{S}_z : V(x^*, z) \geq v^*\}} Q_{\boldsymbol{y}|\boldsymbol{x},\boldsymbol{z}}(\tau^*|x^*, z), \tag{9}$$

or equivalently

$$\sup_{V \in \mathscr{G}^-(x^*, v^*)} g\big(x^*, Q_{\boldsymbol{u}|\boldsymbol{v}}(\tau^*|V)\big) \leq \psi^* \leq \inf_{V \in \mathscr{G}^+(x^*, v^*)} g\big(x^*, Q_{\boldsymbol{u}|\boldsymbol{v}}(\tau^*|V)\big). \tag{10}$$

Proof. See appendix A. □

Lemma 3 is a sensible result, which has the following intuition. Find a demographic profile z such that individuals must have talent no less (greater) than v^* to achieve education x^*. Individuals with demographics z and education x^* then have a success distribution no less (more) favorable than those of individuals with talent equal to v^* by assumptions C and D and the same level of education. Hence, by assumption B ψ^* must be no less (greater) than the τ^* quantile of the earnings distribution of individuals with demographics z and education x^*. Out of all such profiles z, select the one resulting in the tightest upper (lower) bound.

Two problems with $\mathscr{R}$ (condition 1) are that (i) it may not hold and (ii) the classes $\mathscr{G}^+(x^*, v^*)$, $\mathscr{G}^-(x^*, v^*)$, even when nonempty, may not be large. In fact, $\mathscr{R}$ cannot be satisfied if x^* is a scalar and equals the highest or lowest value possible. For instance, if $\boldsymbol{x}$ is binary (college–educated or not) then $V(0, z) = \big(0, \mathbb{P}[\boldsymbol{x} = 0|\boldsymbol{z} = z]\big]$ and $V(1, z) = \big(\mathbb{P}[\boldsymbol{x} = 0|\boldsymbol{z} = z], 1\big]$ for all z; $V(1, z)$ has upper limit equal to one since there is no nontrivial upper bound to the talent of individuals with a college education. For vector–valued x^*, the problem is still more severe.

Further, note that each value of z generates at most one element in either $\mathscr{G}^+(x^*, v^*)$ or $\mathscr{G}^-(x^*, v^*)$. This fact, together with the global exogeneity of $\boldsymbol{z}$, suggests that $\mathscr{G}^+(x^*, v^*)$ and $\mathscr{G}^-(x^*, v^*)$ may be too small; a new rank condition is needed. A more detailed discussion follows in the next subsection.

2.4. **New Rank Condition.** We now develop our new, weaker, rank condition $\mathscr{R}^*$. It is based on the idea that the collection $\{V(x^*, z) : z \in \mathscr{S}_z\}$ can in fact generate larger classes of sets that are useful for bounding ψ^* than $\mathscr{G}^-(x^*, v^*)$ and $\mathscr{G}^+(x^*, v^*)$. To be more specific, consider the example of binary $\boldsymbol{x}$ (college–educated or not) again. In this example $V(1, z_1) - V(1, z_2)$ is the set of talent–levels for which individuals would attend college with demographics z_1 but not with z_2. If

$V(1, z_1) - V(1, z_2) \leq v^*$ then the success distribution of college–educated individuals whose talent is in the range $V(1, z_1) - V(1, z_2)$ is no more favorable than those of college–educated individuals with talent v^*. This can be the case even when neither $V(1, z_1) \leq v^*$ nor $V(1, z_2) \leq v^*$.

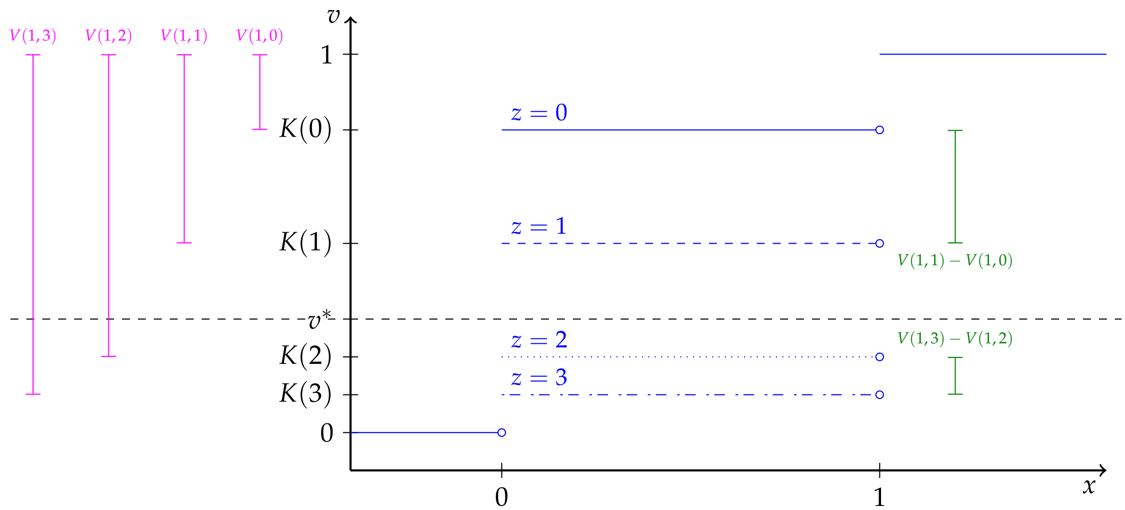

FIGURE 1. Example of how sets are combined.

The situation is illustrated in figure 1 in which $\boldsymbol{x}$ is binary, $\mathscr{S}_z = \{0, 1, 2, 3\}$ and $K(z) = \mathbb{P}[\boldsymbol{x} = 0 | \boldsymbol{z} = z]$. In the graphed example, demographic profiles 0 and 1 ensure that all those with a college education must have talent no less than v^*, but there is no demographic value that makes individuals with talent no better than v^* attend college. So $\mathscr{R}$ is not satisfied; $\mathscr{G}^+(1, v^*) = \{V(1, 1), V(1, 0)\}$ and $\mathscr{G}^-(1, v^*) = \emptyset$. Therefore, the method used in lemma 3 provides an upper bound for ψ^*, but it does not provide a meaningful lower bound. As will be shown below, there is information available for the construction of an upper bound that is not contained in $\mathscr{G}^+(1, v^*)$. Indeed, we can also construct an upper bound by looking at the group of individuals who would attend college with $z = 1$ but not with $z = 0$. The bound provided by $V(1, 1) - V(1, 0)$ may well be tighter than the bound provided by either $V(1, 1)$ or $V(1, 0)$. Likewise, $V(1, 3) - V(1, 2)$, the group of talent levels that would result in a college education with $z = 3$ but not with $z = 2$, can be used to construct a lower bound. A more complicated example, involving vector–valued $\boldsymbol{x}$, can be found at the end of this subsection.

Lemma 4 is our starting point. Let $\phi^*(V) = \phi(\tau^*, V) = g\big(x^*, Q_{\boldsymbol{u}|\boldsymbol{v}}(\tau^* | V)\big)$. Our main result, theorem 1, below is based on the fact that

$$\phi^*(V) \geq \psi^* \tag{11}$$

whenever $V \geq v^*$; this is a direct implication of assumption D. Lemma 4 shows how sets can be combined. Let $\mu(V) = \mathbb{P}[\boldsymbol{v} \in V]$ and let $\mathscr{K} = \{V \subset \mathscr{U}^d : V \neq \emptyset,\ \mu(V) \text{ and } \phi(\tau, V) \text{ are identified for all } \tau \in \mathscr{U}\}$.

Lemma 4. *For any $V_1, V_2 \in \mathscr{K}$,*

(i) *If $V_1 \subset V_2, \mu(V_2 - V_1) > 0$, then $V_2 - V_1 \in \mathscr{K}$.*

(ii) *If $V_1 \cap V_2 = \emptyset, \mu(V_1 \cup V_2) > 0$, then $V_1 \cup V_2 \in \mathscr{K}$.*

Proof. The proof is in appendix A. □

Lemma 4 can be applied to $V(x,z)$–sets because $\phi(\tau, V(x,z))$ is identified for all τ by lemma 2 and because $\mu(V(x,z)) = \mathbb{P}[\boldsymbol{x} = x | \boldsymbol{z} = z]$ is identified.[5] If one applies either operation described in lemma 4 to $V_1 = V(x, z_1)$ and $V_2 = V(x, z_2)$ then the resulting set V_3 belongs to $\mathscr{K}$, and the procedure can be iterated. Doing so ultimately leads to a *Dynkin system* or *λ system* (Billingsley, 1995, p.41) of measurable sets.

Let $\mathscr{V}(x) = \{V : V \neq \emptyset,\ \exists z \in \mathscr{S}_z : V(x,z) = V\}$. In definition 1 below one can take $\mathscr{D}_0 = \mathscr{A} = \mathscr{V}(x^*)$, $\mathscr{D}_1$ to be the collection of sets that contains all sets in $\mathscr{A}$ plus all sets that arise when one applies lemma 4 to all combinations of elements in $\mathscr{A}$, $\mathscr{D}_2$ to be the collection of all sets in $\mathscr{D}_1$ plus all sets that arise when one applies lemma 4 to all combinations of elements in $\mathscr{D}_1$, and so forth. Ultimately, one ends up with $\mathscr{D} = \mathscr{D}_\infty$.

Definition 1. *Let $\mathscr{A}$ be a collection of measurable subsets of $\mathscr{U}^d$. Then $\mathscr{D} = \mathscr{D}(\mathscr{A})$ is the collection $\mathscr{D}_\infty$ in the following iterative scheme. Let $\mathscr{D}_0 = \mathscr{A}$. Then for all $t \geq 0$, $\mathscr{D}_{t+1}$ consists of all sets A^* such that at least one of the following three conditions is satisfied.*

(i) $A^* \in \mathscr{D}_t$,

(ii) $\exists A_1, A_2 \in \mathscr{D}_t : A_1 \subset A_2,\ \mu(A_2 - A_1) > 0,\ A^* = A_2 - A_1$,

(iii) $\exists A_1, A_2 \in \mathscr{D}_t : A_1 \cap A_2 = \emptyset,\ \mu(A_1 \cup A_2) > 0,\ A^* = A_1 \cup A_2$. □

We will use $\mathscr{D}(x)$ in lieu of $\mathscr{D}(\mathscr{V}(x))$ to emphasize its dependence on x. We are now in a position to state our rank condition. Let

$$\begin{cases} \mathscr{J}^-(x,v) = \{V \in \mathscr{D}(x) : V \leq v\}, \\ \mathscr{J}^+(x,v) = \{V \in \mathscr{D}(x) : V \geq v\}. \end{cases} \tag{12}$$

Condition 2 ($\mathscr{R}^*$)**.** *Neither $\mathscr{J}^-(x^*, v^*)$ nor $\mathscr{J}^+(x^*, v^*)$ is empty.* □

[5] Since $V(x,z)$ depends only on the conditional distribution of $\boldsymbol{x}$ given $\boldsymbol{z}$, $V(x,z)$ itself is also identified.

If one compares $\mathscr{R}^*$ to $\mathscr{R}$ (condition 2 to condition 1), $\mathscr{R}^*$ is weaker since $\mathscr{J}^-, \mathscr{J}^+$ contain all elements of $\mathscr{G}^-, \mathscr{G}^+$, respectively. Only in rare circumstances are $\mathscr{R}$ and $\mathscr{R}^*$ the same.

Theorem 1. *Under assumptions A–E, if $\mathscr{R}^*$ is satisfied then*

$$\sup_{V \in \mathscr{J}^-(x^*, v^*)} g\big(x^*, Q_{u|v}(\tau^*|V)\big) \le \psi^* \le \inf_{V \in \mathscr{J}^+(x^*, v^*)} g\big(x^*, Q_{u|v}(\tau^*|V)\big), \tag{13}$$

where the bounds are identified.

Proof. The proof is in appendix B. □

It is instructive to compare the bounds in (13) to those resulting from $\mathscr{R}$ in (10). Because $\mathscr{J}^-, \mathscr{J}^+$ are larger classes than $\mathscr{G}^-, \mathscr{G}^+$, the bounds in (13) are generally tighter than those in (10), and hence also than those in (9).

An example of the difference between the bounds in (10) and (13) arises when we consider figure 1; note for instance that $\mathscr{G}^-(1, v^*) = \emptyset$, whereas $\mathscr{J}^-(1, v^*) = \{V(1,3) - V(1,2)\}$. The difference between the bounds arising from $\mathscr{R}$ and $\mathscr{R}^*$ becomes extreme when the instrument has continuous variation. Consider again figure 1, but with $\mathscr{S}_z = \mathbb{R}$ and for the special case that $h(z, v) = I\big(v > H(z)\big)$ for all v, z, where H is a continuous distribution function. $\mathscr{R}$ is as before not satisfied, so it produces no bounds. Now $\mathscr{R}^*$. Note that $V(0,z) = \big(0, H(z)\big]$ and $V(1,z) = \big(H(z), 1\big]$. Take $z^* = H^{-1}(v^*)$. If $Q_{u|v}(\tau^*|v)$ is continuous at $v = v^*$, then

$$\lim_{t\to\infty} Q_{u|v}\big(\tau^*|V(1, z^* - 1/t) - V(1, z^*)\big) = \lim_{t\to\infty} Q_{u|v}\big(\tau^*|V(1, z^*) - V(1, z^* + 1/t)\big) = Q_{u|v}(\tau^*|v^*),$$

and using $\mathscr{R}^*$ ψ^* is then point–identified for $x^* = 1$ (and similarly for $x^* = 0$) since $V(1, z^* - 1/t) - V(1, z^*) \in \mathscr{J}^-(1, v^*)$ and $V(1, z^*) - V(1, z^* + 1/t) \in \mathscr{J}^+(1, v^*)$ for all $t \ge 1$.

The bounds in theorem 1 are sharp under the stated conditions.[6] To see this, note that the inequalities $\phi^*(V_0) \le \psi^* \le \phi^*(V_1)$ for all $V_0, V_1 \subset \mathscr{U}$ such that $V_0 \le v^* \le V_1$ cannot be improved on and that $\phi^*(V)$ is not identified unless it is expressed as a mapping from one of $\mathscr{P}, \mathscr{P}^2, \cdots, \mathscr{P}^\infty$ to $\mathbb{R}$, where $\mathscr{P} = \{p : \mathbb{R} \to [0,1] : p(y) = \mathbb{P}[\boldsymbol{y} \le y | \boldsymbol{x} = x^*, \boldsymbol{z} = z] \text{ for some } z \in \mathscr{S}_z\}$. Because of the limitations of local identification conditions, Chesher's (2005) analysis is restricted to using $\mathscr{P}$ only but assumption C enables us to use all of $\mathscr{P}, \mathscr{P}^2, \ldots, \mathscr{P}^\infty$.

As mentioned earlier, we conclude with an example for vector–valued $\boldsymbol{x}$. In our examples $\boldsymbol{x}$ contains two binary variables: a college education dummy and a field of specialization dummy (marketable or not). See figures 2 and 3. For $x^* = (1, 1)$, the $V(x^*, z)$–sets are rectangles in

[6] I.e. the bounds cannot be improved without further restrictions.

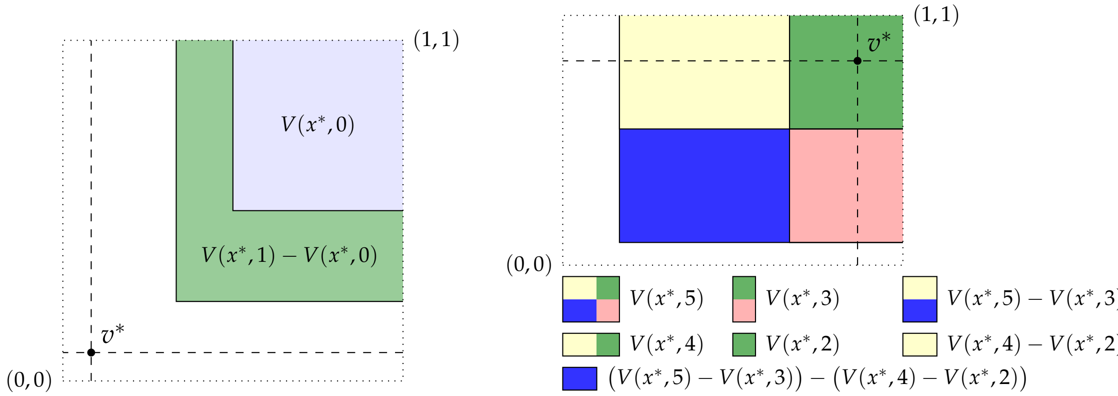

FIGURE 2. How to obtain upper and lower bounds when $x^* = (1,1)$.

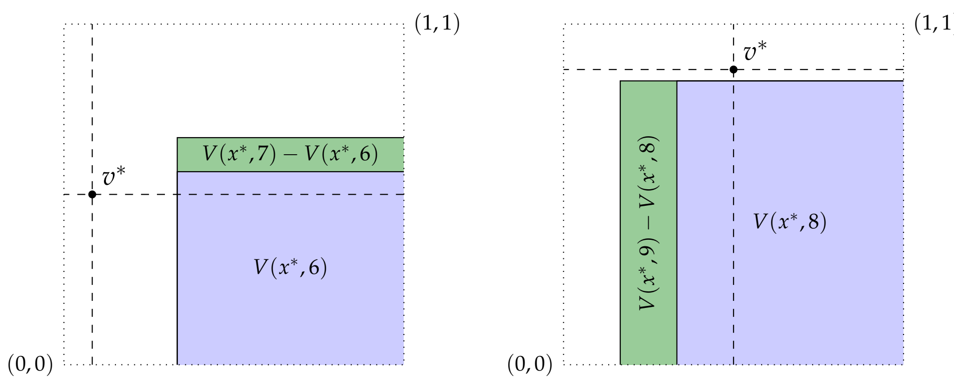

FIGURE 3. How to obtain upper and lower bounds when $x^* = (1,0)$.

$\mathscr{U}^2$ including $(1,1)$ and $V(x^*,1) - V(x^*,0)$ is simply the difference between two such rectangles. $V(x^*,0), V(x^*,1), V(x^*,1) - V(x^*,0)$ can all be used to construct upper bounds. But to obtain a lower bound at all, one must use at least four $V(x^*,z)$–rectangles, as indicated in the second graph in figure 2. The rectangle below v^* is the collection of $v = (v_1, v_2)$–points for which v_1 is a level of (college–)talent sufficient to obtain a college degree with demographics $z = 4$ or $z = 5$ but not with $z = 2$ or $z = 3$ and for which v_2 is a level of (marketable field–)talent sufficient to have a marketable field of specialization with demographics $z = 3$ or $z = 5$ but not with $z = 2$ or $z = 4$. As figure 3 illustrates, finding upper and lower bounds for $x^* = (1,0)$ is easier than finding a lower bound for $x^* = (1,1)$.

3. REVISITING ANGRIST AND KRUEGER (1991)

We now illustrate the difference between $\mathscr{R}$ and $\mathscr{R}^*$ by using the Angrist and Krueger (1991) data. Angrist and Krueger (1991) estimated a wage equation with years of schooling as a(n endogenous) regressor. They used quarter of birth dummies as instruments. Chesher (2005) concluded that the Angrist and Krueger (1991) instruments do not satisfy $\mathscr{R}$ for any value of years of schooling and for any level of talent. In this section we determine whether $\mathscr{R}^*$ is satisfied.

Before we proceed, we comment on the plausibility of global exclusion, exogeneity and rank conditions. Exclusion restrictions and invariance restrictions of the distribution of latent variables given instruments are at best partially testable[7] so they are usually justified by economic reasoning. For instance, the wage equation in Angrist and Krueger (1991) does not include the birth–quarter variables, the exogeneity of which the authors justified by arguing the independence of ability and birth–quarters. Rank conditions, on the other hand, are restrictions on the joint distribution of observables. So they are generally testable once exogeneity of instruments is assumed. Potential failure of rank conditions has received more attention than failure of exclusion restrictions, for instance in the weak instrument literature.

To simplify our discussion we define $\boldsymbol{x}$ to be limited to the values $\{0, 1, 2\}$; no more than 6 years, 7–12 years, and more than 12 years of education, respectively. The instrument $\boldsymbol{z}$ equals the quarter of birth (1–4). Table 1 summarizes the effect of the instruments, where we pretend that the estimated probabilities equal the true probabilities.

	z			
$x \downarrow$	1	2	3	4
0	0.0317	0.0315	0.0280	0.0270
1	0.6119	0.6038	0.5977	0.5946
2	1.0000	1.0000	1.0000	1.0000
n_z	81,671	80,138	86,856	80,844

TABLE 1. The effect of the birth quarter instrument on the education variable $\boldsymbol{x}$.

Because $\mathbb{P}[\boldsymbol{x} < x | \boldsymbol{z} = z] < \mathbb{P}[\boldsymbol{x} \leq x | \boldsymbol{z} = \tilde{z}]$ for all $x, z, \tilde{z}$, $\mathscr{R}$ (or equivalently (8)) is not satisfied, irrespective of the value of x^* and v^*. The birth quarter instruments are hence too weak for $\mathscr{R}$ to be satisfied.

$\mathscr{R}^*$ is different, however. For instance, consider $x^* = 2$ and $v^* = 0.6$. Since $V(2,3) \subset V(2,4)$ and $V(2,4) - V(2,3) = \big(0.5946, 0.5977\big] \leq v^* \leq V(2,1) = \big(0.6119, 1.0000\big]$, $\mathscr{R}^*$ is satisfied for $x^* = 2$ and $v^* = 0.6$.

[7] E.g. by means of a test of overidentifying restrictions

The above example is flawed in two respects. First, we used estimated rather than true probabilities. Note however that the sample size is large and that this is just an example to illustrate the possibility of using $\mathscr{R}^*$ when $\mathscr{R}$ is not satisfied. Second, in the above example, $\boldsymbol{z}$ can only take a small number of different values. With more variation in the instrument, the difference in identifying potential of $\mathscr{R}$ and $\mathscr{R}^*$ increases exponentially.

References Cited

Angrist, J. D., and A. B. Krueger (1991): "Does compulsory school attendance affect schooling and earnings?," *Quarterly Journal of Economics*, 106, 979–1014.

Billingsley, P. (1995): *Probability and measure*. Wiley, third edn.

Chesher, A. (2003): "Identification in nonseparable models," *Econometrica*, 71(5), 1405–1441.

——— (2005): "Nonparametric identification under discrete variation," *Econometrica*, 73, 1525–1550.

Jun, S. J. (2009): "Local structural quantile effects in a model with a nonseparable control variable," *Journal of Econometrics*, 151, 82–97.

Jun, S. J., J. Pinkse, and H. Xu (2009): "Tighter bounds in triangular models: supplementary material," Discussion paper, The Pennsylvania State University.

——— (2010): "Discrete endogenous variables in weakly separable models," Discussion paper, Penn State.

Ma, L., and R. Koenker (2006): "Quantile regression methods for recursive structural equation models," *Journal of Econometrics*, 134, 471–506.

Manski, C., and E. Tamer (2002): "Inference on regressions with interval data on a regressor or outcome," *Econometrica*, 70(2), 519–546.

Newey, W. K., J. L. Powell, and F. Vella (1999): "Nonparametric estimation of triangular simultaneous equations models," *Econometrica*, 67(3), 565–603.

Pinkse, J. (2000): "Nonparametric two–step regression estimation when regressors and error are dependent," *Canadian Journal of Statistics*, 28(2), 289–300.

Vytlacil, E., and N. Yildiz (2007): "Dummy endogenous variables in weakly separable models," *Econometrica*, 75(1), 757–779.

Appendix A. Proofs of lemmas

Proof of lemma 1. By assumption B, $Q_{\boldsymbol{y}|\boldsymbol{x},\boldsymbol{v}}(\tau|x^*,v^*) = g\big(x, Q_{\boldsymbol{u}|\boldsymbol{x},\boldsymbol{v}}(\tau|x^*,v^*)\big)$. Recall that by assumption E, $\mathscr{Z}(x^*,v^*)$ is nonempty. Thus, $\boldsymbol{x} = x^*, \boldsymbol{v} = v^* \Leftrightarrow \boldsymbol{z} \in \mathscr{Z}(x^*,v^*), \boldsymbol{v} = v^*$, such that $Q_{\boldsymbol{u}|\boldsymbol{x},\boldsymbol{v}}(\tau|x^*,v^*) = Q_{\boldsymbol{u}|\boldsymbol{v},\boldsymbol{z}}\big(\tau|v^*,\mathscr{Z}(x^*,v^*)\big)$, which equals $Q_{\boldsymbol{u}|\boldsymbol{v}}(\tau|v^*)$ by assumption C. □

Proof of lemma 2. By assumption B, $Q_{\boldsymbol{y}|\boldsymbol{x},\boldsymbol{z}}(\tau|x,z) = g\big(x, Q_{\boldsymbol{u}|\boldsymbol{x},\boldsymbol{z}}(\tau|x,z)\big)$. Since $\boldsymbol{x} = x, \boldsymbol{z} = z \Leftrightarrow \boldsymbol{v} \in V(x,z), \boldsymbol{z} = z$ and by assumption C,

$$Q_{\boldsymbol{u}|\boldsymbol{x},\boldsymbol{z}}(\tau|x,z) = Q_{\boldsymbol{u}|\boldsymbol{v},\boldsymbol{z}}\big(\tau|V(x,z),z\big) = Q_{\boldsymbol{u}|\boldsymbol{v}}\big(\tau|V(x,z)\big). \qquad \square$$

Proof of lemma 3. We establish the upper bound; the argument for the lower bound is virtually identical. Let z be such that $V(x^*,z) \geq v^*$. Then by lemma 2,

$$Q_{\boldsymbol{y}|\boldsymbol{x},\boldsymbol{z}}(\tau^*|x^*,z) = g\big\{x^*, Q_{\boldsymbol{u}|\boldsymbol{v}}\big(\tau^*|V(x^*,z)\big)\big\} \geq g\big\{x^*, Q_{\boldsymbol{u}|\boldsymbol{v}}(\tau^*|v^*)\big\} = \psi^*,$$

where the weak inequality follows from assumption B and the fact that for any $u \in \mathscr{U}$

$$\begin{aligned}\mathbb{P}\big[\boldsymbol{u} \leq u|\boldsymbol{v} \in V(x^*,z)\big]\mathbb{P}\big[\boldsymbol{v} \in V(x^*,z)\big] &= \int_{V(x^*,z)} \mathbb{P}\big[\boldsymbol{u} \leq u|\boldsymbol{v} = v\big]dv \\ &\leq \int_{V(x^*,z)} \mathbb{P}\big[\boldsymbol{u} \leq u|\boldsymbol{v} = v^*\big]dv = \mathbb{P}\big[\boldsymbol{u} \leq u|\boldsymbol{v} = v^*\big]\mathbb{P}\big[\boldsymbol{v} \in V(x^*,z)\big],\end{aligned}$$

which implies that $Q_{\boldsymbol{u}|\boldsymbol{v}}\big(\tau^*|V(x^*,z)\big) \geq Q_{\boldsymbol{u}|\boldsymbol{v}}(\tau^*|v^*)$. $\square$

Proof of lemma 4. We show (i) where (ii) follows similarly. Note that for any y by the conditions on V_1, V_2,

$$\mathbb{P}\big[g(x^*,\boldsymbol{u}) \leq y|\boldsymbol{v} \in V_2 - V_1\big] = \frac{\mathbb{P}\big[g(x^*,\boldsymbol{u}) \leq y|\boldsymbol{v} \in V_2\big]\mu(V_2) - \mathbb{P}\big[g(x^*,\boldsymbol{u}) \leq y|\boldsymbol{v} \in V_1\big]\mu(V_1)}{\mu(V_2 - V_1)}. \tag{14}$$

Now $\mathbb{P}\big[g(x^*,\boldsymbol{u}) \leq y|\boldsymbol{v} \in V_j\big]$ is identified for $j = 1,2$ and all y because $\phi(\tau, V_j)$ is identified for $j = 1,2$ and all $\tau \in \mathscr{U}$. Further, since V_1, V_2 are disjoint and $\mu(V_1), \mu(V_2)$ are identified by assumption, so is $\mu(V_2 - V_1) = \mu(V_2) - \mu(V_1)$. So the left hand side in (14) is identified for all y. Invert the conditional distribution function to obtain the conditional quantile. $\square$

Appendix B. Proof of theorem

Proof of Theorem 1. Recall that $\mathscr{V}(x^*) \subset \mathscr{K}$ by the discussion following lemma 4. Therefore, $\mathscr{D}(x^*) \subset \mathscr{K}$ by lemma 4 and by construction of $\mathscr{D}(x^*)$. Combining this with (11) (and its converse when $V \leq v^*$) concludes the proof. $\square$

Appendix C. Miscellaneous

C.1. **Normalization.** Consider an arbitrary function $g^*(x,u^*)$ with arbitrarily distributed $\boldsymbol{u}^*$, where g^* is weakly increasing in its second argument. Letting ω be the quantile function of $\boldsymbol{u}^*$, there exists a uniform random variable $\boldsymbol{u}$ such that $u^* = \omega(\boldsymbol{u})$. Therefore, $g(x,u) = g^*\big(x,\omega(u)\big)$, is still weakly increasing in u. Since h is weakly increasing in its second argument and $\boldsymbol{v}$ is uniform, we have $h(z,v) = Q_{\boldsymbol{x}|\boldsymbol{z}}(v|z)$. Left–continuity holds, because $\mathbb{P}[\boldsymbol{x} \leq x|\boldsymbol{z} = z]$ is a CADLAG function of x and $Q_{\boldsymbol{x}|\boldsymbol{z}}(v|z) = \inf\{x : \mathbb{P}[\boldsymbol{x} \leq x|\boldsymbol{z} = z] \geq v\}$.

C.2. **Manski and Tamer.** Below is a somewhat more detailed discussion of the relationship between Manski and Tamer (2002) and the present paper.

Manski and Tamer (2002) assume that $\boldsymbol{v}$ is scalar–valued and that upper and lower bounds $(\boldsymbol{v}_0^{MT}, \boldsymbol{v}_1^{MT})$ on $\boldsymbol{v}$ are observed. They further assume monotonicity of $\delta(x^*, v)$ in v and that $\mathbb{E}[\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{v}, \boldsymbol{v}_0^{MT}, \boldsymbol{v}_1^{MT}] = \mathbb{E}[\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{v}]$ a.s..

Replacing $\boldsymbol{y}, \boldsymbol{v}_0^{MT}, \boldsymbol{v}_1^{MT}$ with $\mathbb{1}(\boldsymbol{y} \leq y), \mathbb{P}[\boldsymbol{x} < x^*|\boldsymbol{z}], \mathbb{P}[\boldsymbol{x} \leq x^*|\boldsymbol{z}]$, respectively,[8] facilitates a comparison of Manski and Tamer (2002) with Chesher (2005) and the current paper. The availability of instruments in the triangular model provides us with more structure to be exploited. To simplify the discussion, suppose that both $\mathbb{P}[\boldsymbol{x} < x^*|\boldsymbol{z} = z]$ and $\mathbb{P}[\boldsymbol{x} \leq x^*|\boldsymbol{z} = z]$ are invertible in z, such that $(\boldsymbol{v}_0^{MT}, \boldsymbol{v}_1^{MT}) = \big(\mathbb{P}[\boldsymbol{x} < x^*|\boldsymbol{z} = z], \mathbb{P}[\boldsymbol{x} \leq x^*|\boldsymbol{z} = z]\big)$ if and only if $\boldsymbol{z} = z$. Then, equation (A3) in the proof of proposition 1 in Manski and Tamer (2002) would become

$$\begin{aligned} \mathbb{P}\big[\boldsymbol{y} \leq y|\boldsymbol{x} = x^*, \boldsymbol{v} = \mathbb{P}[\boldsymbol{x} < x^*|\boldsymbol{z} = z]\big] &\leq \mathbb{P}[\boldsymbol{y} \leq y|\boldsymbol{x} = x^*, \boldsymbol{z} = z] \\ &\leq \mathbb{P}\big[\boldsymbol{y} \leq y|\boldsymbol{x} = x^*, \boldsymbol{v} = \mathbb{P}[\boldsymbol{x} \leq x^*|\boldsymbol{z} = z]\big]. \quad (15) \end{aligned}$$

Note here that the monotonicity assumption of Manski and Tamer (2002)[9] in combination with the inversion of the distribution functions in (15) leads to the quantile version of proposition 1 of Manski and Tamer (2002), which coincides with the bounds in (9) that our main theorem, theorem 1, improves upon.

[8] $\mathbb{1}$ is the indicator function.

[9] I.e. $\mathbb{E}[\mathbb{1}\{\boldsymbol{y} \leq y\}|\boldsymbol{x} = x^*, \boldsymbol{v} = v]$ is weakly decreasing in v.