NONPARAMETRIC IDENTIFICATION WITH DISCRETE ENDOGENOUS VARIABLES

Andrew Chesher

Nonparametric identification with discrete endogenous variables

Andrew Chesher[*]

Centre for Microdata Methods and Practice
Institute for Fiscal Studies and University College London

July 29th 2003, revised September 8th 2003[†]

ABSTRACT. This paper studies the nonparametric identification of partial differences of a nonseparable structural function that determines the value of a discrete or continuous random variable when the function depends on endogenous variates which have discrete probability distributions. Weak conditions are developed under which a partial difference of a structural function with respect to a discrete endogenous variable is nonparametrically identified up to an interval. The interval is short when there are many points of support of the endogenous variable. The interval has finite length if there are at least three points of support but not when there are only two. The interval can be estimated using quantile regression estimation methods.

1. INTRODUCTION

1.1. Nonseparable models and identification of partial differences. This paper considers the identification of a partial difference of a nonseparable structural function that yields the value of a discrete or continuous random variable, a function which includes among its arguments endogenous variates which have discrete probability distributions.

Weak conditions are developed under which a partial difference of a structural function with respect to a discrete endogenous variable is nonparametrically identified to within an interval. The interval can be estimated using quantile regression estimation procedures.

As the granularity of the support of a discrete endogenous variable diminshes, the interval shrinks to a point, yielding a point identification result when the endogenous variable is continuously distributed.

To motivate the study of this problem consider the following stylised Becker-Chiswick-Mincer model[1] of the determination of the log wage (W) and completed years of schooling (S).

$$
\begin{aligned}
W &= h_1(S, X, \rho_1) \\
S &= h_2(X, \rho_2)
\end{aligned}
$$

Here X denotes a list of covariates measuring characteristics of the individual and of the environment in which decisions are made and outcomes determined, and ρ_1 and ρ_2 are unobserved, continuously, and possibly dependently distributed random variables. The variate ρ_2 will be interpreted as a measure of ability.

---

[*] I am grateful to Roger Koenker and to seminar participants at cemmap, Northwestern University, University of Chicago and at the Harvard-MIT econometrics workshop for comments on the related paper Chesher (2002) and for discussion of the problem considered in this paper.

[†] This revision corrects errors in the presentation of the proof of Lemma 2 in which, in the July 29th version, some inequalities were incorrectly oriented.

[1] See Becker and Chiswick (1966), Chiswick and Mincer (1972), Chiswick (1974), Mincer (1974), Card (1995), Card (2001).

In this context there is interest in partial differences like

$$\Delta \equiv h_1(s', x, r_1) - h_1(s'', x, r_1)$$

which measure the "pure" effect of schooling on the wage.

When ρ_1 and ρ_2 are dependently distributed these "pure" effects cannot be identified without adding further restrictions to the model and (W, S, X) data can only reveal the combined effect on the wage of schooling (*via* the S-argument of h_1) and ability (*via* the ρ_1-argument of h_1 through its dependence on ρ_2).

Chesher (2003) develops weak conditions under which partial derivatives of structural functions like h_1 are nonparametrically point identified, but those results are not applicable to the problem studied here because, when S is a discrete random variable, there cannot be nonparametric identification of the S-partial *derivative* of the function h_1. That paper also surveys the related literature.

Weak conditions under which there is nonparametric point identification of a partial difference, like Δ, when endogenous variables (S above) are *continuously* distributed but covariates vary only *discretely* are developed in Chesher (2002). Discrete variation in X prohibits nonparametric identification of the S-partial derivative of the function h_1, hence the focus in that paper on identification of partial differences.

The method employed in Chesher (2002) is not directly applicable when S has a discrete distribution. The discrete endogenous variable case was left in that paper as a puzzle which this paper solves by developing weak conditions under which an *interval* containing a structural partial difference like Δ can be nonparametrically identified.

1.2. Interval identification of a partial difference. The interval identification result is now introduced in the context of a simple restricted version of the model at the start of this Section in which the covariates, X, are excluded from the wage structural function, h_1, as follows.

$$
\begin{aligned}
W &= h_1(S, \rho_1) \\
S &= h_2(X, \rho_2)
\end{aligned}
$$

Let $\{s_m\}_{m=1}^M$ be the points of support of the discrete distribution of S and consider two distinct points of support, s_i and s_j and the neighbouring points of support below them, s_{i-1} and s_{j-1}.

The variate ρ_2 is normalised to be uniformly distributed on $(0,1)$ independent of X. The function $h_2(X, \rho_2)$ is therefore the conditional quantile function of S given X, a weakly increasing caglad step function with steps whose lengths are equal to the probability masses on the points of support of the distribution of S given X.

Let r_1 be a value of continuously distributed ρ_1, to be specified. Identification of the structural partial difference

$$\Delta \equiv h_1(s_j, r_1) - h_1(s_i, r_1) \tag{1}$$

is considered.

Sufficient conditions for identification of Δ are as follows.[2]

1. **Monotonicity**. The function h_1 is strictly monotonically varying with ρ_1. The conditional τ_1-quantile of ρ_1 given ρ_2 and X is weakly monotonically varying with ρ_2.[3]

---

[2] Weaker conditions are set down in Section 2.

[3] The second monotonicity condition is only required to hold over intervals of values of ρ_2.

2. **Rank condition**. There exist four distinct values of X: x^i, x^{i-1}, x^j and x^{j-1} such that for a chosen value $\tau_2 \in (0, 1)$

$$\tau_2 = P[S = s^i | x^i] = P[S = s^{i-1} | x^{i-1}] = P[S = s^j | x^j] = P[S = s^{j-1} | x^{j-1}]. \quad (2)$$

These covariate values are referred to as *instrumental values*.[4]

3. **Covariation**. The value of the τ_1-quantile of ρ_1 given $\rho_2 = \tau_2$ (see equation (2)) and X does not vary across the four instrumental values. Denote the common value by r_1 - this is the value of ρ_1 that appears in the definition of the partial difference Δ, given in equation (1).[5]

The interval identification result makes use of conditional quantile functions. The conditional τ-quantile of a random variable A given a (possibly vector) random variable $B = b$ is denoted by $Q_{A|B}(\tau|b)$,[6] defined as

$$Q_{A|B}(\tau|b) \equiv \inf\{q : F_{A|B}(q|b) \geq \tau\}$$

where $F_{A|B}$ is the conditional distribution function of A given B.

Under the conditions just stated there is the following inequality

$$\min \left(\begin{array}{c} Q_{W|SX}(\tau_1|s_j, x^j) - Q_{W|SX}(\tau_1|s_i, x^{i-1}), \\ Q_{W|SX}(\tau_1|s_j, x^{j-1}) - Q_{W|SX}(\tau_1|s_i, x^i) \end{array} \right)$$

$$\leq \Delta \leq$$

$$\max \left(\begin{array}{c} Q_{W|SX}(\tau_1|s_j, x^j) - Q_{W|SX}(\tau_1|s_i, x^{i-1}), \\ Q_{W|SX}(\tau_1|s_j, x^{j-1}) - Q_{W|SX}(\tau_1|s_i, x^i) \end{array} \right) \quad (3)$$

where the values of S that appear in (3) are conditional τ_2-quantiles of the distribution of S given X, as follows.[7]

$$\left. \begin{array}{cc} s_i = Q_{S|X}(\tau_2|x^i) & s_{i-1} = Q_{S|X}(\tau_2|x^{i-1}) \\ s_j = Q_{S|X}(\tau_2|x^j) & s_{j-1} = Q_{S|X}(\tau_2|x^{j-1}) \end{array} \right\} \quad (4)$$

This inequality identifies Δ up to an interval which can be estimated *via* the analogue principle (Manski (1988)), replacing the conditional quantiles in (3) by estimated conditional quantiles evaluated at estimates of the instrumental values of X. The inequality (3) holds when W is discretely or continuously distributed.

This Section concludes with a discussion of some issues raised by the result (3). Section 2 gives a formal statement of the identifying restrictions and a Theorem stating the interval identification result.

1.3. Estimation and overidentification. To estimate the bounding interval calculate an estimate $\hat{F}_{S|X}$ and then find values, $\{\hat{x}^i, \hat{x}^{i-1}, \hat{x}^j, \hat{x}^{j-1}\}$ of X such that

$$\tau_2 = \hat{F}_{S|X}(s_i | \hat{x}^i) = \hat{F}_{S|X}(s_{i-1} | \hat{x}^{i-1}) = \hat{F}_{S|X}(s_j | \hat{x}^j) = \hat{F}_{S|X}(s_{j-1} | \hat{x}^{j-1}).$$

---

[4] See Chesher (2002). Note that the existence of such instrumental values for any chosen τ_2 will depend on the nature and extent of the support of X.

[5] This would of course be satisfied if ρ_1 were distributed independently of X given ρ_2, but is far less restrictive than an independence condition being specific to τ_1, τ_2 and the values of X considered.

[6] For example $Q_{A|B_1B_2}(\tau|b_1, b_2)$ is the τ-quantile of A given $B_1 = b_1$ and $B_2 = b_2$.

[7] Since S has a discrete distribution these values of S are also conditional τ-quantiles for some values of $\tau \neq \tau_2$.

Finally calculate estimates, $\hat{Q}_{W|SX}(\tau_1|s_i,x)$ for $x \in \{\hat{x}^i, \hat{x}^{i-1}\}$ and $\hat{Q}_{W|SX}(\tau_1|s_j,x)$ for $x \in \{\hat{x}^j, \hat{x}^{j-1}\}$ and substitute in (3). At each stage the estimates could be calculated using a parametric, semi- or nonparametric estimator.[8]

When X contains many covariates there may be many choices of instrumental values which yield the same identifying interval. Then the interval is overidentified and the procedure described above will yield many estimates of the identifying interval. One could perhaps combine these using a minimum distance procedure to produce an asymptotically efficient estimator.

In the just identified case, if $F_{S|X}$ does not vary monotonically with x there may be distinct sets of instrumental values leading to different identifying intervals. One would naturally be lead to use the set of instrumental values yielding the shortest interval although accuracy of estimation will be a factor to bear in mind when choosing among sets of instrumental values.

1.4. Rich support. Suppose that the granularity of the support of S is very fine, so that S is "nearly" continuously distributed. Then if $F_{S|X}(s|x)$ is smoothly varying with x it will be possible to choose x^{i-1} close to x^i and x^{j-1} close to x^j and then, if $Q_{\rho_1|\rho_2 X}$ is smoothly varying with ρ_2, the separation between the bounds on Δ will be small.

If in fact S is continuously distributed then, if there exist instrumental values x^i and x^j, one can choose $x^{i-1} = x^i$ and $x^{j-1} = x^j$ and the inequality (3) collapses, yielding the point identification result given in Chesher (2002)

$$\Delta = Q_{W|SX}(\tau_1|s_j,x^j) - Q_{W|SX}(\tau_1|s_i,x^i)$$

where

$$\begin{aligned} s_i &= Q_{S|X}(\tau_2|x^i) \\ s_j &= Q_{S|X}(\tau_2|x^j). \end{aligned}$$

When S is continuously distributed, τ_2 is the *unique* probability value which satisfies these equations.

1.5. Discrete covariates. If X has limited variation, for example when X is discrete, it may not be possible to find any instrumental values that satisfy (2) for a particular, or perhaps for any choice of s_i and s_j at any value of τ_2. It is sometimes possible to develop wider identifying intervals in this case as explained in Section 3.3.

1.6. Binary endogenous variables. If $j = i \pm 1$ the interval identifying inequality (3) involves three points of support of the discrete endogenous variable, otherwise it involves four points of support. Suppose the endogenous variable has just two points of support, without loss of generality $\{0, 1\}$. Then (3) does not apply in full. However, following the argument in the Appendix, it can be shown that if the conditional τ_1-quantile of ρ_1 given ρ_2 and X is non-decreasing with ρ_2 there is the inequality:

$$\Delta \geq Q_{W|SX}(\tau_1|1,x^0) - Q_{W|SX}(\tau_1|0,x^0)$$

with the inequality reversed if the conditional τ_1 quantile of ρ_1 given ρ_2 and X is non-increasing with ρ_2.

So, in the binary endogenous variable case, any knowledge of Δ under the weak conditions set out here requires prior knowledge of the direction of the dependence of ρ_1 on

---

[8]Regarding the quantile function estimation: for parametric estimation, see Koenker and Bassett (1978) and Koenker and d'Orey (1987); for semiparametric estimation see Chaudhuri, Doksum and Samarov (1997), Kahn (2001) and Lee (2003); for nonparametric estimation, see Chaudhuri (1991).

ρ_2. Even with that knowledge, Δ can only be nonparametrically identified in this non-separable model up to an interval with infinite length. More informative identification in the presence of binary endogenous variables requires stronger restrictions than those considered here.[9]

1.7. Plan of the remainder of the paper. The next Section provides a full set of identifying restrictions, rather weaker than those used in introducing the result and a Theorem stating an interval identification result for a structural partial difference with respect to an endogenous variate for the case in which the structural function depends on covariates and just one endogenous variate. The proof is contained in the Appendix.

Section 3 considers three extensions: interval identification when the structural function depends on *more than one* endogenous variate; interval identification of a structural partial difference with respect to a *covariate*; and interval identification when covariates have sparse support insufficient to allow the existence of instrumental values of X satisfying all the equalities in condition (2). The final Section concludes.

2. INTERVAL IDENTIFICATION OF STRUCTURAL PARTIAL DIFFERENCES

Consider a model for two outcomes, Y_1 and Y_2 with covariates $X = \{X_k\}_{k=1}^K$ and two continuously distributed latent random variables ρ_1 and ρ_2 with structural equations as follows.

$$
\begin{aligned}
Y_1 &= h_1(Y_2, X, \rho_1) & (5) \\
Y_2 &= h_2(X, \rho_2) & (6)
\end{aligned}
$$

The conditional distribution function of ρ_1 given ρ_2 and X is denoted by $F_{\rho_1|\rho_2 X}$ and $Q_{\rho_1|\rho_2 X}$ denotes the associated conditional quantile function. The distribution of ρ_2 is normalised as uniform on $(0,1)$ independent of X.

2.1. The discrete distribution of the endogenous variable. Let the points of support of Y_2 be $\{y_2^m\}_{m=1}^M$, independent of X, with M not necessarily finite. The associated positive probability masses are $\{p_m(x)\}_{m=1}^M$ with $\sum_{m=1}^M p_m(x) = 1$. Define the cumulative probability sums $\{p^m(x)\}_{m=1}^M$ as follows.

$$
p^m(x) \equiv \sum_{n=1}^m p_n(x), \qquad m \in \{1, \dots, M\}
$$

The distribution and quantile functions of Y_2 given $X = x$ are as follows.

$$
F_{Y_2|X}(y_2|x) = \begin{cases} 0 & -\infty < y_2 < y_2^1 \\ p^m(x) & y_2^m \le y_2 < y_2^{m+1}, \\ 1 & y_2^M \le y_2 < \infty \end{cases} \qquad m \in \{2, \dots, M-1\}
$$

$$
Q_{Y_2|X}(\tau|x) = \begin{cases} y_2^1 & 0 < \tau \le p^1(x) \\ y_2^m & p^{m-1}(x) < \tau \le p^m(x), \\ y_2^M & p^m(x) < \tau \le 1 \end{cases} \qquad m \in \{2, \dots, M-1\}
$$

Since ρ_2 is normalised as uniformly distributed on $(0,1)$ independent of X, the structural function h_2 is the quantile function of Y_2 given X, that is:

$$
h_2(x, \rho_2) = Q_{Y_2|X}(\rho_2|x).
$$

---

[9]For example parametric restrictions.

2.2. Interval identifying restrictions and a Theorem. Consider probabilities $\{\tau_1, \tau_2\} \in (0,1)^2$ and two points of support of Y_2, $y_2^i \neq y_2^j$. There are the following conditions.

I. **Triangularity**. At any value of X, and of continuously distributed ρ_1 and ρ_2, the values of the outcomes Y_1 and Y_2 are uniquely determined by equations (5) and (6).

II. **Rank condition**. There exists a set of instrumental values of X:

$$\mathcal{X} = \{x^i, x^j, x^{i-1}, x^{j-1}\}$$

such that

$$p^i(x^i) = p^j(x^j) = p^{i-1}(x^{i-1}) = p^{j-1}(x^{j-1}) = \tau_2.$$

III. **Monotonicity**. The function h_1 is weakly monotonic with respect to variation in ρ_1, normalised non-decreasing and caglad. For $x \in \mathcal{X}$, and

$$r_2 \in \mathcal{R}_2(x) \equiv [p^{i-1}(x), p^i(x)] \cup [p^{j-1}(x), p^j(x)]$$

the conditional quantile function $Q_{\rho_1|\rho_2 X}(\tau_1|r_2, x)$ is either a non-increasing or non-decreasing function of r_2.[10]

IV. **Covariation**. For $\{x', x''\} \in \mathcal{X}$ and $r_2 \in \mathcal{R}_2(x') \cup \mathcal{R}_2(x'')$,

$$Q_{\rho_1|\rho_2 X}(\tau_1|r_2, x') = Q_{\rho_1|\rho_2 X}(\tau_1|r_2, x'')$$

and ρ_2 is normalised to be uniformly distributed on $(0,1)$ independent of X.

V. **Order condition**. For $x \in \mathcal{X}$, define $r_1 \equiv Q_{\rho_1|\rho_2 X}(\tau_1|\tau_2, x)$.[11] For $\{x', x''\} \in \mathcal{X}$ and $l \in \{i, i-1, j, j-1\}$

$$h_1(y_2^l, x', r_1) = h_1(y_2^l, x'', r_1).$$

For $x \in \mathcal{X}$ define the following structural partial difference.

$$\Delta_{\mathcal{X}}(y_2^i, y_2^j, \tau_1, \tau_2) \equiv h_1(y_2^j, x, r_1) - h_1(y_2^i, x, r_1)$$

The Theorem stating the set identification of $\Delta_{\mathcal{X}}(y_2^i, y_2^j, \tau_1, \tau_2)$ is as follows.

Theorem 1. Conditions I - V imply the following inequality.

$$\min \left(\begin{array}{l} Q_{Y_1|Y_2 X}(\tau_1|y_2^j, x^j) - Q_{Y_1|Y_2 X}(\tau_1|y_2^i, x^{i-1}), \\ Q_{Y_1|Y_2 X}(\tau_1|y_2^j, x^{j-1}) - Q_{Y_1|Y_2 X}(\tau_1|y_2^i, x^i) \end{array} \right)$$

$$\leq \Delta_{\mathcal{X}}(y_2^i, y_2^j, \tau_1, \tau_2) \leq$$

$$\max \left(\begin{array}{l} Q_{Y_1|Y_2 X}(\tau_1|y_2^j, x^j) - Q_{Y_1|Y_2 X}(\tau_1|y_2^i, x^{i-1}), \\ Q_{Y_1|Y_2 X}(\tau_1|y_2^j, x^{j-1}) - Q_{Y_1|Y_2 X}(\tau_1|y_2^i, x^i) \end{array} \right)$$

The Theorem is proved in the Appendix. The method of proof is as follows.

---

[10] This is somewhat stronger then is required but has the virtue of simplicity - see the Appendix for discussion.

[11] Note r_1 is invariant with respect to $x \in \mathcal{X}$ by virtue of Condition IV.

2.3.   The method of proof.   First an expression is obtained for the conditional distribution function of Y_1 given $Y_2 = y_2^m$ (one of the points of support) and $X = x$. This leads directly to an equation whose solution is the conditional quantile function of Y_1 given $Y_2 = y_2^m$ and $X = x$.

When $F_{\rho_1|\rho_2 X}$ is *non-decreasing* in ρ_2 for $\rho_2 \in [p^{m-1}(x), p^m(x)]$ the solution can be bounded, as follows:

$$h_1(y_2^m, x, Q_{\rho_1|\rho_2 X}(\tau_1|p^m(x), x)) \leq Q_{Y_1|Y_2 X}(\tau_1|y_2^m, x) \leq h_1(y_2^m, x, Q_{\rho_1|\rho_2 X}(\tau_1|p^{m-1}(x), x))$$
$$(7)$$

with the inequalities reversed when $F_{\rho_1|\rho_2 X}$ is *non-increasing* in ρ_2 for $\rho_2 \in [p^{m-1}(x), p^m(x)]$.

In the *non-decreasing* $F_{\rho_1|\rho_2 X}$ case, for $x \in \{x^i, x^{i-1}\}$ with m set equal to i there is, under the conditions of Theorem 1, from (7)

$$h_1(y_2^i, x, r_1) \quad \leq \quad Q_{Y_1|Y_2 X}(\tau_1|y_2^i, x^i) \tag{8}$$
$$Q_{Y_1|Y_2 X}(\tau_1|y_2^i, x^{i-1}) \quad \leq \quad h_1(y_2^i, x, r_1) \tag{9}$$

and for $x \in \{x^j, x^{j-1}\}$ with m set equal to j

$$h_1(y_2^j, x, r_1) \quad \leq \quad Q_{Y_1|Y_2 X}(\tau_1|y_2^j, x^j) \tag{10}$$
$$Q_{Y_1|Y_2 X}(\tau_1|y_2^j, x^{j-1}) \quad \leq \quad h_1(y_2^j, x, r_1). \tag{11}$$

Adding (8) to (11) and rearranging and adding (9) to (10) and rearranging gives

$$Q_{Y_1|Y_2 X}(\tau_1|y_2^j, x^{j-1}) - Q_{Y_1|Y_2 X}(\tau_1|y_2^i, x^i)$$

$$\leq h_1(y_2^j, x, r_1) - h_1(y_2^i, x, r_1) \equiv \Delta_{\mathcal{X}}(y_2^i, y_2^j, \tau_1, \tau_2) \leq$$

$$Q_{Y_1|Y_2 X}(\tau_1|y_2^j, x^j) - Q_{Y_1|Y_2 X}(\tau_1|y_2^i, x^{i-1})$$

In the *non-increasing* $F_{\rho_1|\rho_2 X}$ case the inequalities are reversed and the result of Theorem 1 follows on combining the two cases.

3.   Extensions

3.1.   Many endogenous variables.   The identification result of Theorem 1 is for the case in which there is a single endogenous variable in the structural function of interest. Extension to the case with more than one endogenous variable is straightforward.

Consider the case in which there are two endogenous variates, Y_2 and Y_3 and in place of condition (I), condition (I').

I'. **Triangularity**. At any value of X, and of continuously distributed ρ_1, ρ_2 and ρ_3, the values of the outcomes Y_1, Y_2 and Y_3 are uniquely determined by the following equations.

$$\begin{aligned} Y_1 &= h_1(Y_2, Y_3, X, \rho_1) \\ Y_2 &= h_2(X, \rho_2) \\ Y_3 &= h_3(X, \rho_3) \end{aligned}$$

Let y_3^g be a point of support of the distribution of Y_3 and consider interval identification of the following partial difference with respect to Y_2

$$\Delta_{\mathcal{X}}(y_2^i, y_2^j, y_3^g, \tau_1, \tau_2) \equiv h_1(y_2^j, y_3^g, x, r_1) - h_1(y_2^i, y_3^g, x, r_1)$$

for $x \in \mathcal{X}$ a set of instrumental values satisfying condition (II'). Now cumulative probabilities are subscripted to indicate the endogenous variate to which they refer.

II′. **Rank condition**. There exists a set of instrumental values of X:

$$\mathcal{X} = \{x^i, x^j, x^{i-1}, x^{j-1}\}$$

such that

$$
\begin{array}{rcl}
p_2^i(x^i) & = & p_2^j(x^j) = p_2^{i-1}(x^{i-1}) = p_2^{j-1}(x^{j-1}) = \tau_2 \\
p_3^g(x^i) & = & p_3^g(x^j) = p_3^g(x^{i-1}) = p_3^g(x^{j-1}) = \tau_3.
\end{array}
$$

The monotonicity, covariation and order conditions are amended as follows.

III′. **Monotonicity**. The function h_1 is weakly monotonic with respect to variation in ρ_1, normalised non-decreasing and caglad. For $x \in \mathcal{X}$, and

$$r_2 \in \mathcal{R}_2(x) \equiv [p_2^{i-1}(x), p_2^i(x)] \cup [p_2^{j-1}(x), p_2^j(x)]$$

the conditional quantile function $Q_{\rho_1|\rho_2\rho_3 X}(\tau_1|r_2, \tau_3, x)$ is either a non-increasing or non-decreasing function of r_2.

IV′. **Covariation**. For $\{x', x''\} \in \mathcal{X}$ and $r_2 \in \mathcal{R}_2(x') \cup \mathcal{R}_2(x'')$,

$$Q_{\rho_1|\rho_2\rho_3 X}(\tau_1|r_2, \tau_3, x') = Q_{\rho_1|\rho_2\rho_3 X}(\tau_1|r_2, \tau_3, x'')$$

and ρ_2 and ρ_3 are normalised to be independently uniformly distributed on $(0,1)^2$ independent of X.

V′. **Order condition**. For $x \in \mathcal{X}$ define $r_1 \equiv Q_{\rho_1|\rho_2\rho_3 X}(\tau_1|\tau_2, \tau_3, x)$. For $\{x', x''\} \in \mathcal{X}$ and $l \in \{i, i-1, j, j-1\}$

$$h_1(y_2^l, y_3^g, x', r_1) = h_1(y_2^l, y_3^g, x'', r_1).$$

Under conditions I′ - V′ there is the following interval identifying inequality.

$$\min \left(\begin{array}{l} Q_{Y_1|Y_2Y_3X}(\tau_1|y_2^j, y_3^g, x^j) - Q_{Y_1|Y_2Y_3X}(\tau_1|y_2^i, y_3^g, x^{i-1}), \\ Q_{Y_1|Y_2Y_3X}(\tau_1|y_2^j, y_3^g, x^{j-1}) - Q_{Y_1|Y_2Y_3X}(\tau_1|y_2^i, y_3^g, x^i) \end{array} \right)$$

$$\leq \Delta_{\mathcal{X}}(y_2^i, y_2^j, y_3^g, \tau_1, \tau_2) \leq$$

$$\max \left(\begin{array}{l} Q_{Y_1|Y_2Y_3X}(\tau_1|y_2^j, y_3^g, x^j) - Q_{Y_1|Y_2Y_3X}(\tau_1|y_2^i, y_3^g, x^{i-1}), \\ Q_{Y_1|Y_2Y_3X}(\tau_1|y_2^j, y_3^g, x^{j-1}) - Q_{Y_1|Y_2Y_3X}(\tau_1|y_2^i, y_3^g, x^i) \end{array} \right)$$

The proof is similar to the proof of the Theorem 1 given in the Appendix.

3.2. Partial differences with respect to covariates. Consider a simple case in which the structural function of interest, h_1, depends on a single discrete endogenous variable and a single covariate $X_\blacklozenge$ with $X = (X_\blacklozenge, X_\lozenge)$.

Interval identification of the structural partial difference,

$$\Lambda_{\mathcal{X}}(y_2^i, x_\blacklozenge^a, x_\blacklozenge^b, \tau_1, \tau_2) = h_1(y_2^i, x_\blacklozenge^a, r_1) - h_1(y_2^i, x_\blacklozenge^b, r_1)$$

requires the existence of four instrumental values, $\mathcal{X} = \{x_a^i, x_a^{i-1}, x_b^i, x_b^{i-1}\}$ as in condition (II) but with the requirement that within the set of instrumental values, $x_\blacklozenge = x_\blacklozenge^a$ for $x \in \{x_a^i, x_a^{i-1}\}$, $x_\blacklozenge = x_\blacklozenge^b$ for $x \in \{x_b^i, x_b^{i-1}\}$ and the probability of obtaining the value, y_2^i, of the endogenous variable given $X = x$ is equal to τ_2 for all x within the set of instrumental values. Condition (V) is modified, now requiring that the structural function

is insensitive, only *via* its $X_\diamond$ arguments, to variation in x within the set of instrumental values. Conditions (I), (III) and (IV) are maintained.

The interval identifying inequality is:

$$\min\left(Q_{Y_1|Y_2X}(\tau_1|y_2^i, x_a^i) - Q_{Y_1|Y_2X}(\tau_1|y_2^i, x_b^{i-1}), Q_{Y_1|Y_2X}(\tau_1|y_2^i, x_a^{i-1}) - Q_{Y_1|Y_2X}(\tau_1|y_2^i, x_b^i)\right)$$

$$\leq \Lambda_{\mathcal{X}}(y_2^i, x_\blacklozenge^a, x_\blacklozenge^b, \tau_1, \tau_2) \leq$$

$$\max\left(Q_{Y_1|Y_2X}(\tau_1|y_2^i, x_a^i) - Q_{Y_1|Y_2X}(\tau_1|y_2^i, x_b^{i-1}), Q_{Y_1|Y_2X}(\tau_1|y_2^i, x_a^{i-1}) - Q_{Y_1|Y_2X}(\tau_1|y_2^i, x_b^i)\right)$$

a result that can be proved by minor modification of the argument of the proof of Theorem 1.

3.3. Sparse support. The rank condition (II) requires that there exist a set of instrumental values of X

$$\mathcal{X} = \{x^i, x^j, x^{i-1}, x^{j-1}\}$$

such that

$$p^i(x^i) = p^j(x^j) = p^{i-1}(x^{i-1}) = p^{j-1}(x^{j-1}) = \tau_2.$$

When X has sparse support, for example when X does not exhibit continuous variation it may not be possible to find such a set $\mathcal{X}$.

In this case, modification of the rank condition (II) can lead to interval identification. Replace condition (II) by the following condition (II$''$).

II$''$. **Rank condition**. There exists a set of distinct instrumental values of X:

$$\mathcal{X} = \{x^i, x^j, x^{i-1}, x^{j-1}\}$$

such that

$$p^i(x^i) \leq \tau_2 \qquad p^{i-1}(x^{i-1}) \geq \tau_2 \qquad p^j(x^j) \leq \tau_2 \qquad p^{j-1}(x^{j-1}) \geq \tau_2 \qquad (12)$$

The other conditions are unchanged and there is the following Theorem.

Theorem 2. Conditions (I), (II$''$), (III), (IV) and (V) imply the inequality stated in Theorem 1.

A proof is given in the Appendix. When one or more strong inequalities hold in (12) the bounds in the inequality of Theorem 1 may be wider than those obtained when equalities hold throughout. Further, the conditions of Theorem 2 are more restrictive in this case since they require, that the covariation and monotonicity conditions hold over a larger set of values of ρ_2 than when strict equalities hold throughout in (12).

4. CONCLUDING REMARKS

The interval identification result obtained here suggests that *point* nonparametric identification of Δ is not possible in a nonseparable model.

If Y_2 is discrete and h_2 is strictly increasing then ρ_2 must be discrete, and then, if h_2 depends on X, changes in X change the locations of the points of support. This is not a case of real econometric interest.

The more interesting case considered here is that in which discrete Y_2 is generated by *continuous* ρ_2 which requires that h_2 is a step function. But in this case, each value of Y_2 is associated with a interval of values of ρ_2 and the data generating value in this interval cannot be identified.

Therefore it is not possible to "control" precisely for variation in ρ_2 by fixing Y_2 at some quantile of its conditional distribution given X and considering alternative instrumental values of X, as was proposed in Chesher (2002) for the case in which Y_2 is continuously distributed.

However the monotonicity condition on $Q_{\rho_1|\rho_2 X}$ invoked here allows the impact of ρ_2 on ρ_1, and thus on the value delivered by h_1, to be bounded as ρ_2 varies within the interval associated with the points of support of Y_2 that are of interest, and this leads to the interval identification result of this paper.

The result for binary endogenous variables presents an interesting puzzle. Under the conditions proposed there can be interval identification of a structural partial difference with respect to a binary endogenous variable, but the interval does not have finite length. What minimal additional restrictions would lead to identification up to an interval of finite length?

References

Becker, Gary S., and Barry R. Chiswick (1966): "Education and the distribution of earnings," *American Economic Review*, 56, 358-369.

Card, David (2001): "Estimating the returns to schooling: Progress on some persistent econometric problems," *Econometrica*, 69, 1127-1160.

Card, David (1995): "Earnings, ability and schooling revisited," in *Research in Labour Economics, Volume 14*, ed., S. Polachek. Greenwich, Conn: JAI Press.

Chaudhuri, Probal (1991): "Nonparametric estimation of regression quantiles and their local Bahadur representation," *Annals of Statistics*, 19, 760-777.

Chaudhuri, Probal, Kjell Doksum and Alexander Samarov (1997): "On average derivative quantile regression," *Annals of Statistics*, 25, 715-744.

Chesher, Andrew D. (2002): "Instrumental values," Centre for Microdata Methods and Practice Working Paper 17/02.

Chesher, Andrew D. (2003): "Identification in nonseparable models," forthcoming in *Econometrica*.

Chiswick, Barry R. (1974): *Income inequality: regional analyses within a human capital framework*. New York: Columbia University Press.

Chiswick, Barry R., and Jacob Mincer (1972): "Time series changes in personal income inequality," *Journal of Political Economy*, 80, S34-S66.

Kahn, Shakeeb (2001): "Two-stage rank estimation of quantile index models," *Journal of Econometrics*, 100, 319-355.

Koenker, Roger W., and Gilbert W. Bassett Jr. (1978): "Regression quantiles," *Econometrica*, 46, 33-50.

Koenker, Roger W. and Vasco d'Orey (1987): "Computing regression quantiles," *Journal of the Royal Statistical Society, Series C*, 36, 383-393.

Lee, Sokbae (2003): "Efficient semiparametric estimation of a partially linear quantile regression model," *Econometric Theory*, 19, 1-31.

Manski, Charles F. (1988b): *Analog estimation methods in econometrics*, New York: Chapman and Hall.

Mincer, Jacob (1974): *Schooling experience and earnings*. New York: Columbia University Press.

APPENDIX: PROOFS OF THEOREMS 1 AND 2

First, two lemmata employed in the proofs of the theorems are stated and proved. Lemma 1 provides an expression for the distribution function of Y_1 conditional on $Y_2 = y_2^m$ and $X = x$. Lemma 2 places bounds on the associated conditional quantile function. The proofs of Theorems 1 and 2 follow.

Lemma 1

Under condition (I) and the weak monotonicity (normalised non-decreasing) of h_1 with respect to ρ_1 required by condition (III), the conditional distribution function of Y_1 given $Y_2 = y_2^m$ and $X = x$ is as follows.

$$F_{Y_1|Y_2X}(y_1|y_2^m,x) = \frac{1}{p_m(x)} \int_{p^{m-1}(x)}^{p^m(x)} F_{\rho_1|\rho_2X}(\sup_p\{p : h_1(y_2^m,x,p) \le y_1\}|\rho_2,x)d\rho_2$$

If h_1 is strictly monotonic (normalised increasing) with respect to ρ_1 then

$$F_{Y_1|Y_2X}(y_1|y_2^m,x) = \frac{1}{p_m(x)} \int_{p^{m-1}(x)}^{p^m(x)} F_{\rho_1|\rho_2X}(h_1^{-1}(y_2^m,x,y_1)|\rho_2,x)d\rho_2$$

where $h_1^{-1}(y_2^m,x,y_1)$ satisfies

$$y_1 = h_1(y_2^m,x,h_1^{-1}(y_2^m,x,y_1)). \tag{A1}$$

Proof of Lemma 1

The conditional distribution function of Y_1 given $Y_2 = y_2^m$ and $X = x$ is defined as follows.

$$F_{Y_1|Y_2X}(y_1|y_2^m,x) \equiv \frac{P[Y_1 \le y_1 \cap Y_2 = y_2^m|X = x]}{P[Y_2 = y_2^m|X = x]}$$

Given $X = x$, in terms of events, since h_1 is a non-decreasing function of ρ_1,

$$\{Y_1 \le y_1 \cap Y_2 = y_2^m\} = \{\rho_1 \le \sup_p\{p : h_1(y_2^m,x,p) \le y_1\} \cap \{p^{m-1}(x) < \rho_2 \le p^m(x)\}$$

and so

$$P[Y_1 \le y_1 \cap Y_2 = y_2^m|X = x] = \int_{p^{m-1}(x)}^{p^m(x)} F_{\rho_1|\rho_2X}(\sup_p\{p : h_1(y_2^m,x,p) \le y_1\}|\rho_2,x)d\rho_2$$
$$\tag{A2}$$

and the first result of the Lemma follows on dividing by $P[Y_2 = y_2^m|X = x] = p_m(x)$.

If h_1 is an increasing function of ρ_1 then

$$\sup_p\{p : h_1(y_2^m,x,p) \le y_1\} = h_1^{-1}(y_2^m,x,y_1) \tag{A3}$$

and substituting in (A2) leads to the second result of the Lemma. □

The conditional τ_1-quantile of Y_1 given $Y_2 = y_2^m$ and $X = x$ is

$$Q_{Y_1|Y_2X}(\tau_1|y_2^m,x) \equiv \inf_q\{q : \frac{1}{p_m(x)} \int_{p^{m-1}(x)}^{p^m(x)} F_{\rho_1|\rho_2X}(\sup_p\{p : h_1(y_2^m,x,p) \le q\}|\rho_2,x)d\rho_2 \ge \tau_1\}$$

and when h_1 is increasing with respect to ρ_1:

$$Q_{Y_1|Y_2X}(\tau_1|y_2^m, x) \equiv \inf_q \left\{ q : \frac{1}{p_m(x)} \int_{p^{m-1}(x)}^{p^m(x)} F_{\rho_1|\rho_2X}(h_1^{-1}(y_2^m, x, q)|\rho_2, x) d\rho_2 \geq \tau_1 \right\}.$$

Lemma 2 places bounds on this conditional quantile.

Lemma 2

Under conditions (I) and (III), if $F_{\rho_1|\rho_2X}$ is non-decreasing in ρ_2 for $\rho_2 \in [p^{m-1}(x), p^m(x)]$ then

$$h_1(y_2^m, x, Q_{\rho_1|\rho_2X}(\tau_1|p^m(x), x)) \leq Q_{Y_1|Y_2X}(\tau_1|y_2^m, x) \leq h_1(y_2^m, x, Q_{\rho_1|\rho_2X}(\tau_1|p^{m-1}(x), x))$$

with the inequalities reversed if $F_{\rho_1|\rho_2X}$ is non-increasing in ρ_2 for $\rho_2 \in [p^{m-1}(x), p^m(x)]$.

Proof of Lemma 2

If $F_{\rho_1|\rho_2X}$ is non-decreasing in ρ_2 for $\rho_2 \in [p^{m-1}(x), p^m(x)]$ then, recalling the result of Lemma 1, replacing ρ_2 in the integrand first by $p^{m-1}(x)$ and then by $p^m(x)$ yields:

$$\frac{1}{p_m(x)} \int_{p^{m-1}(x)}^{p^m(x)} F_{\rho_1|\rho_2X}(\sup_p\{p : h_1(y_2^m, x, p) \leq y_1\}|p^{m-1}(x), x) d\rho_2$$

$$\leq F_{Y_1|Y_2X}(y_1|y_2^m, x) \leq$$

$$\frac{1}{p_m(x)} \int_{p^{m-1}(x)}^{p^m(x)} F_{\rho_1|\rho_2X}(\sup_p\{p : h_1(y_2^m, x, p) \leq y_1\}|p^m(x), x) d\rho_2$$

and since, for example

$$\frac{1}{p_m(x)} \int_{p^{m-1}(x)}^{p^m(x)} F_{\rho_1|\rho_2X}(\sup_p\{p : h_1(y_2^m, x, p) \leq y_1\}|p^m(x), x) d\rho_2$$

$$= F_{\rho_1|\rho_2X}(\sup_p\{p : h_1(y_2^m, x, p) \leq y_1\}|p^m(x), x)$$

there is, for all y_1, the following.

$$F_{\rho_1|\rho_2X}(\sup_p\{p : h_1(y_2^m, x, p) \leq y_1\}|p^{m-1}(x), x)$$

$$\leq F_{Y_1|Y_2X}(y_1|y_2^m, x) \leq$$

$$F_{\rho_1|\rho_2X}(\sup_p\{p : h_1(y_2^m, x, p) \leq y_1\}|p^m(x), x) \tag{A4}$$

If h_1 is an increasing function of ρ_1 then in (A4) $\sup_p\{p : h_1(y_2^m, x, p) \leq y_1\}$ can be replaced by $h_1^{-1}(y_2^m, x, y_1)$.

If $F_{\rho_1|\rho_2X}$ is non-increasing in ρ_2 for $\rho_2 \in [p^{m-1}(x), p^m(x)]$ the inequality (A4) is reversed.[12]

---

[12]Monotonicity of $F_{\rho_1|\rho_2X}$ with respect to ρ_2 is a stronger condition than required but it has the advantage that it is easily interpreted. The inequality (A4) holds if for all $\rho_2 \in [p^{m-1}(x), p^m(x)]$,

$$F_{\rho_1|\rho_2X}(\sup_p\{p : h_1(y_2^m, x, p) \leq y_1\}|\rho_2, x) \in [\min(A_m^{m-1}, A_m^m), \max(A_m^{m-1}, A_m^m)]$$

$$A_m^{m-1} \equiv F_{\rho_1|\rho_2X}(\sup_p\{p : h_1(y_2^m, x, p) \leq y_1\}|p^{m-1}(x), x)$$

$$A_m^m \equiv F_{\rho_1|\rho_2X}(\sup_p\{p : h_1(y_2^m, x, p) \leq y_1\}|p^m(x), x)]$$

for which the assumed monotonicity condition is sufficient.

Since the functions on the left and the right of the inequality (A4) bound $F_{Y_1|Y_2X}(y_1|y_2^m, x)$ respectively *below* and *above*, the solutions, q_{m-1} and q_m to

$$q_{m-1} = \inf_q\{q : F_{\rho_1|\rho_2X}(\sup_p\{p : h_1(y_2^m, x, p) \leq q\}|p^{m-1}(x), x) \geq \tau_1\}$$

$$q_m = \inf_q\{q : F_{\rho_1|\rho_2X}(\sup_p\{p : h_1(y_2^m, x, p) \leq q\}|p^m(x), x) \geq \tau_1\}$$

bound

$$\inf_q\{q : F_{Y_1|Y_2X}(q|y_2^m, x) \geq \tau_1\} \equiv Q_{Y_1|Y_2X}(\tau_1|y_2^m, x)$$

respectively *above* and *below*, that is:

$$q_m \leq Q_{Y_1|Y_2X}(\tau_1|y_2^m, x) \leq q_{m-1}.$$

To aid understanding, consider Figure 1 where the distribution function $F_{Y_1|Y_2X}(y_1|y_2^m, x)$ is the solid cadlag step function and the upper (u) and lower (l) bounding functions in (A4) are drawn respectively dashed and dotted. A value of τ_1 is shown as a solid horizontal line. In the example drawn, $q_m < q < q_{m-1}$. If τ_1 was slightly smaller then q and q_{m-1} would coincide, and a further small decrease would result in coincidence of q_m and q.

The relationships between the values q_{m-1} and q_m and the structural function h_1 are now determined.

Consider q_m. For any choice of q,

$$F_{\rho_1|\rho_2X}(\sup_p\{p : h_1(y_2^m, x, p) \leq q\}|p^m(x), x) \geq \tau_1$$

if and only if

$$Q_{\rho_1|\rho_2X}(\tau_1|p^m(x), x) \leq \sup_p\{p : h_1(y_2^m, x, p) \leq q\}$$

because $F_{\rho_1|\rho_2X}$ is strictly increasing in its first argument; recall ρ_1 is required to be continuously distributed. Therefore

$$q_m = \inf_q\{q : Q_{\rho_1|\rho_2X}(\tau_1|p^m(x), x) \leq \sup_p\{p : h_1(y_2^m, x, p) \leq q\}\}.$$

Consider

$$q^* = h_1(y_2^m, x, Q_{\rho_1|\rho_2X}(\tau_1|p^m(x), x))$$

which is a candidate value for q_m because,

$$Q_{\rho_1|\rho_2X}(\tau_1|p^m(x), x) \leq \sup_p\{p : h_1(y_2^m, x, p) \leq q^*\}. \tag{A5}$$

This holds as an equality when h_1 is a strictly increasing function of ρ_1 - this follows directly from equation (A3) and the definition, (A1), of the inverse function $h_1^{-1}(y_2^m, x, y_1)$ - and in that case,

$$q_m \equiv \inf_q\{q : Q_{\rho_1|\rho_2X}(\tau_1|p^m(x), x) \leq h_1^{-1}(y_2^m, x, q)\} = q^*.$$

The case in which h_1 is a non-decreasing caglad step function of ρ_1 is drawn in Figure 2. Ends of steps at which h_1 is continuous from the left are drawn as filled circles. The graph shows a value of $Q_{\rho_1|\rho_2X}(\tau_1|p^m(x), x)$ and the candidate value q^*. The value of

$$p^* \equiv \sup_p\{p : h_1(y_2^m, x, p) \leq q^*\}$$

is indicated in Figure 2 and clearly $p^* \geq Q_{\rho_1|\rho_2 X}(\tau_1|p^m(x),x)$, demonstrating that the inequality (A5) does hold.

It is evident from Figure 2 that for $q < q^*$ there is no $p \geq Q_{\rho_1|\rho_2 X}(\tau_1|p^m(x),x)$ such that

$$Q_{\rho_1|\rho_2 X}(\tau_1|p^m(x),x) \leq \sup_p \{p : h_1(y_2^m,x,p) \leq q\}$$

and so $q_m = q^*$.

A similar argument shows that q_{m-1} is as follows.

$$q_{m-1} = h_1(y_2^m,x,Q_{\rho_1|\rho_2 X}(\tau_1|p^{m-1}(x),x))$$

Therefore, in the case in which $F_{\rho_1|\rho_2 X}$ is non-decreasing in ρ_2 for $\rho_2 \in [p^{m-1}(x),p^m(x)]$, there is the following inequality.

$$h_1(y_2^m,x,Q_{\rho_1|\rho_2 X}(\tau_1|p^m(x),x)) \leq Q_{Y_1|Y_2 X}(\tau_1|y_2^m,x) \leq h_1(y_2^m,x,Q_{\rho_1|\rho_2 X}(\tau_1|p^{m-1}(x),x))$$

A similar argument for the case in which $F_{\rho_1|\rho_2 X}$ is non-increasing in ρ_2 for $\rho_2 \in [p^{m-1}(x),p^m(x)]$ produces the reverse of this inequality

$$h_1(y_2^m,x,Q_{\rho_1|\rho_2 X}(\tau_1|p^{m-1}(x),x)) \leq Q_{Y_1|Y_2 X}(\tau_1|y_2^m,x) \leq h_1(y_2^m,x,Q_{\rho_1|\rho_2 X}(\tau_1|p^m(x),x))$$

and combining the two preceding inequalities gives the result of the Lemma. □

Proof of Theorem 1

Consider the case in which, for $x \in \mathcal{X}$, $Q_{\rho_1|\rho_2 X}$ is non-increasing in ρ_2 over the intervals $\mathcal{R}_2(x)$ defined in condition (III). Since $F_{\rho_1|\rho_2 X}$ is the inverse function of $Q_{\rho_1|\rho_2 X}$ this implies that $F_{\rho_1|\rho_2 X}$ is non-decreasing over the intervals $\mathcal{R}_2(x)$ defined in condition (III). In the inequality of Lemma 2 set $m = i$, and set $x = x^i$ giving

$$h_1(y_2^i,x^i,Q_{\rho_1|\rho_2 X}(\tau_1|p^i(x^i),x^i)) \leq Q_{Y_1|Y_2 X}(\tau_1|y_2^i,x^i)$$

and set $m = j$ and $x = x^{j-1}$ giving

$$Q_{Y_1|Y_2 X}(\tau_1|y_2^j,x^{j-1}) \leq h_1(y_2^j,x^{j-1},Q_{\rho_1|\rho_2 X}(\tau_1|p^{j-1}(x^{j-1}),x^{j-1})).$$

Invoking the rank and order conditions (II and V) these inequalities simplify to

$$h_1(y_2^i,x,Q_{\rho_1|\rho_2 X}(\tau_1|\tau_2,x^i)) \leq Q_{Y_1|Y_2 X}(\tau_1|y_2^i,x^i)$$

$$Q_{Y_1|Y_2 X}(\tau_1|y_2^j,x^{j-1}) \leq h_1(y_2^j,x,Q_{\rho_1|\rho_2 X}(\tau_1|\tau_2,x^{j-1}))$$

where x is any member of $\mathcal{X}$, the set of instrumental values and h_1 is insensitive through its x argument to choice of $x \in \mathcal{X}$.

The covariation condition (IV) implies that the conditional quantiles of ρ_1 that appear in the two inequalities take the same value, namely r_1. Thus for $x \in \mathcal{X}$ there is the following.

$$h_1(y_2^i,x,r_1) \leq Q_{Y_1|Y_2 X}(\tau_1|y_2^i,x^i)$$

$$Q_{Y_1|Y_2 X}(\tau_1|y_2^j,x^{j-1}) \leq h_1(y_2^j,x,r_1)$$

Adding the two inequalities and rearranging there is, for $x \in \mathcal{X}$:

$$h_1(y_2^j,x,r_1) - h_1(y_2^i,x,r_1) \geq Q_{Y_1|Y_2 X}(\tau_1|y_2^j,x^{j-1}) - Q_{Y_1|Y_2 X}(\tau_1|y_2^i,x^i). \qquad \text{(A6)}$$

Setting $m = j$ and $x = x^j$ and then $m = i$ and $x = x^{i-1}$ in the inequality of Lemma 2, and arguing along similar lines yields the following inequality.

$$Q_{Y_1|Y_2 X}(\tau_1|y_2^j,x^j) - Q_{Y_1|Y_2 X}(\tau_1|y_2^i,x^{i-1}) \geq h_1(y_2^j,x,r_1) - h_1(y_2^i,x,r_1) \qquad \text{(A7)}$$

Combining (A6) and (A7) gives the following for $x \in \mathcal{X}$.

$$Q_{Y_1|Y_2X}(\tau_1|y_2^j, x^{j-1}) - Q_{Y_1|Y_2X}(\tau_1|y_2^i, x^i)$$

$$\leq h_1(y_2^j, x, r_1) - h_1(y_2^i, x, r_1) \leq$$

$$Q_{Y_1|Y_2X}(\tau_1|y_2^j, x^j) - Q_{Y_1|Y_2X}(\tau_1|y_2^i, x^{i-1})$$

When, for $x \in \mathcal{X}$, $Q_{\rho_1|\rho_2X}$ is non-decreasing in ρ_2 over the intervals $\mathcal{R}_2(x)$ defined in condition (III) (which implies $F_{\rho_1|\rho_2X}$ is non-increasing in ρ_2 over these intervals) these inequalities are reversed and combining the two sets of inequalities gives the result stated in the theorem. $\square$

Proof of Theorem 2

As in the proof of Theorem 1, first consider the case in which, for $x \in \mathcal{X}$, $Q_{\rho_1|\rho_2X}$ is non-increasing in ρ_2 over the intervals $\mathcal{R}_2(x)$ defined in condition (III). Since $F_{\rho_1|\rho_2X}$ is the inverse function of $Q_{\rho_1|\rho_2X}$ this implies that $F_{\rho_1|\rho_2X}$ is non-decreasing over the intervals $\mathcal{R}_2(x)$ defined in condition (III).

Recall that in Theorem 2 the instrumental values satisfy the following weak inequalities.

$$\begin{array}{rclcrcl} p^i(x^i) & \leq & \tau_2 & & p^{i-1}(x^{i-1}) & \geq & \tau_2 \\ p^j(x^j) & \leq & \tau_2 & & p^{j-1}(x^{j-1}) & \geq & \tau_2 \end{array}$$

Lemma 2 implies

$$\begin{array}{rcl} h_1(y_2^i, x^i, Q_{\rho_1|\rho_2X}(\tau_1|p^i(x^i), x^i)) & \leq & Q_{Y_1|Y_2X}(\tau_1|y_2^i, x^i) \\ Q_{Y_1|Y_2X}(\tau_1|y_2^j, x^{j-1}) & \leq & h_1(y_2^j, x^{j-1}, Q_{\rho_1|\rho_2X}(\tau_1|p^{j-1}(x^{j-1}), x^{j-1})) \end{array}$$

and since $Q_{\rho_1|\rho_2X}$ is a non-*increasing* function of ρ_2 (recall h_1 is normalised non-decreasing in its final argument) there is the following pair of inequalities.

$$\begin{array}{rcl} h_1(y_2^i, x^i, Q_{\rho_1|\rho_2X}(\tau_1|\tau_2, x^i)) & \leq & h_1(y_2^i, x^i, Q_{\rho_1|\rho_2X}(\tau_1|p^i(x^i), x^i)) \\ h_1(y_2^j, x^{j-1}, Q_{\rho_1|\rho_2X}(\tau_1|p^{j-1}(x^{j-1}), x^{j-1})) & \leq & h_1(y_2^j, x^{j-1}, Q_{\rho_1|\rho_2X}(\tau_1|\tau_2, x^{j-1})) \end{array}$$

Combining the preceding two pairs of inequalities yields

$$\begin{array}{rcl} h_1(y_2^i, x^i, Q_{\rho_1|\rho_2X}(\tau_1|\tau_2, x^i)) & \leq & Q_{Y_1|Y_2X}(\tau_1|y_2^i, x^i) \\ Q_{Y_1|Y_2X}(\tau_1|y_2^j, x^{j-1}) & \leq & h_1(y_2^j, x^{j-1}, Q_{\rho_1|\rho_2X}(\tau_1|\tau_2, x^{j-1})) \end{array}$$

and adding these inequalities and rearranging there is the following.

$$Q_{Y_1|Y_2X}(\tau_1|y_2^j, x^{j-1}) - Q_{Y_1|Y_2X}(\tau_1|y_2^i, x^i)$$

$$\leq h_1(y_2^j, x^{j-1}, Q_{\rho_1|\rho_2X}(\tau_1|\tau_2, x^{j-1})) - h_1(y_2^i, x^i, Q_{\rho_1|\rho_2X}(\tau_1|\tau_2, x^i))$$

Setting $m = j$ and $x = x^j$ and then $m = i$ and $x = x^{i-1}$ in the inequality of Lemma 2 and arguing as above yields the following inequality.

$$h_1(y_2^j, x^j, Q_{\rho_1|\rho_2X}(\tau_1|\tau_2, x^j)) - h_1(y_2^i, x^{i-1}, Q_{\rho_1|\rho_2X}(\tau_1|\tau_2, x^{i-1}))$$

$$\leq Q_{Y_1|Y_2X}(\tau_1|y_2^j, x^j) - Q_{Y_1|Y_2X}(\tau_1|y_2^i, x^{i-1}).$$

Combining the last pair of inequalities and invoking the order and covariation restrictions and using the definition of r_1, there is, for $x \in \mathcal{X}$, the following inequality.

$$Q_{Y_1|Y_2X}(\tau_1|y_2^j, x^{j-1}) - Q_{Y_1|Y_2X}(\tau_1|y_2^i, x^i)$$

$$\leq h_1(y_2^j, x, r_1) - h_1(y_2^i, x, r_1) \leq$$

$$Q_{Y_1|Y_2X}(\tau_1|y_2^j, x^j) - Q_{Y_1|Y_2X}(\tau_1|y_2^i, x^{i-1}).$$

In the case in which, for $x \in \mathcal{X}$, $F_{\rho_1|\rho_2 X}$ is a *non-increasing* function of ρ_2 for $\rho_2 \in R_2(x)$, which implies that $Q_{\rho_1|\rho_2 X}$ is a *non-decreasing* function of ρ_2 for $\rho_2 \in R_2(x)$, applying the same argument leads to the reverse of the preceding inequality. Combining the results yields the inequality as stated in Theorem 1 and thus the required result. $\square$
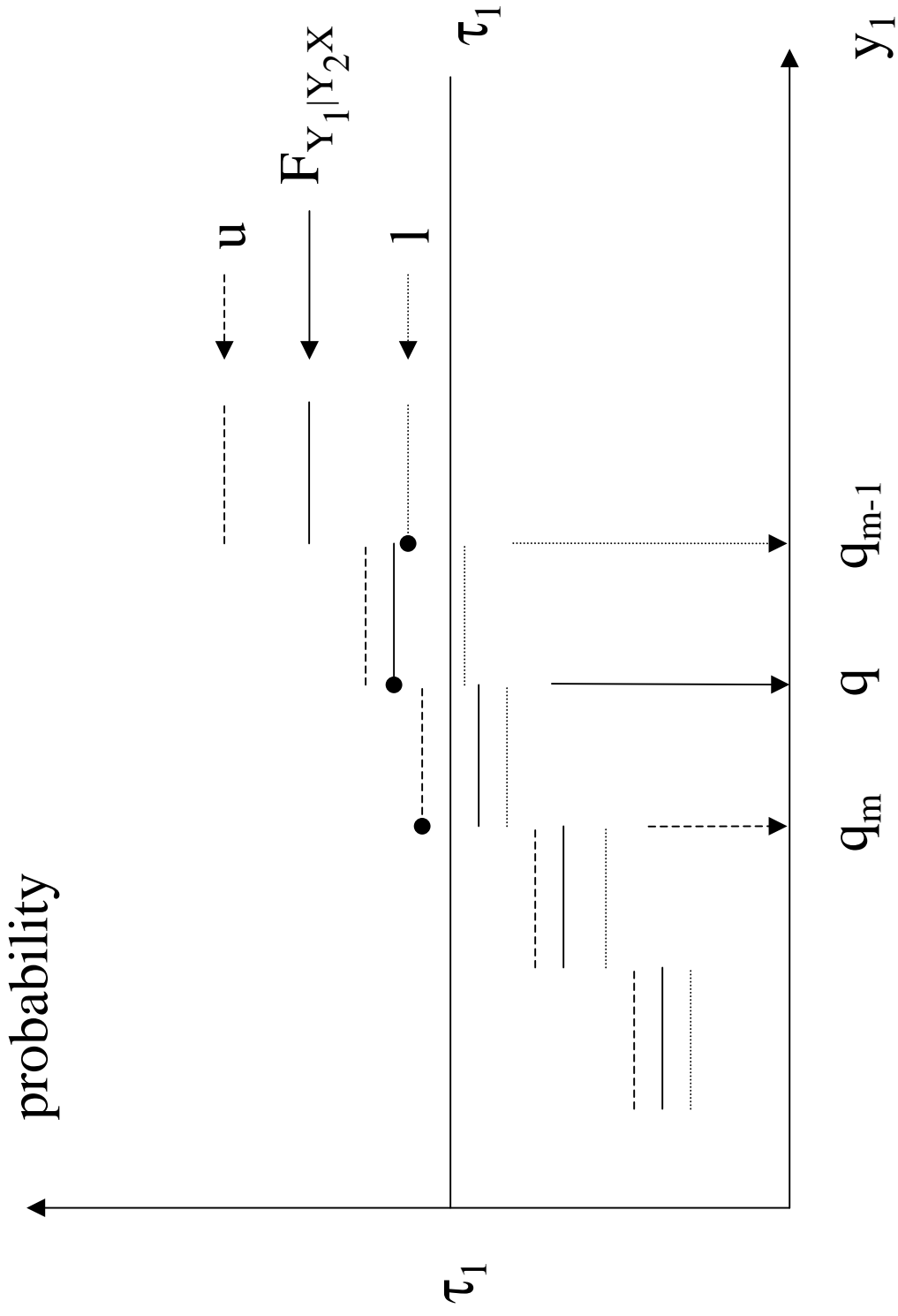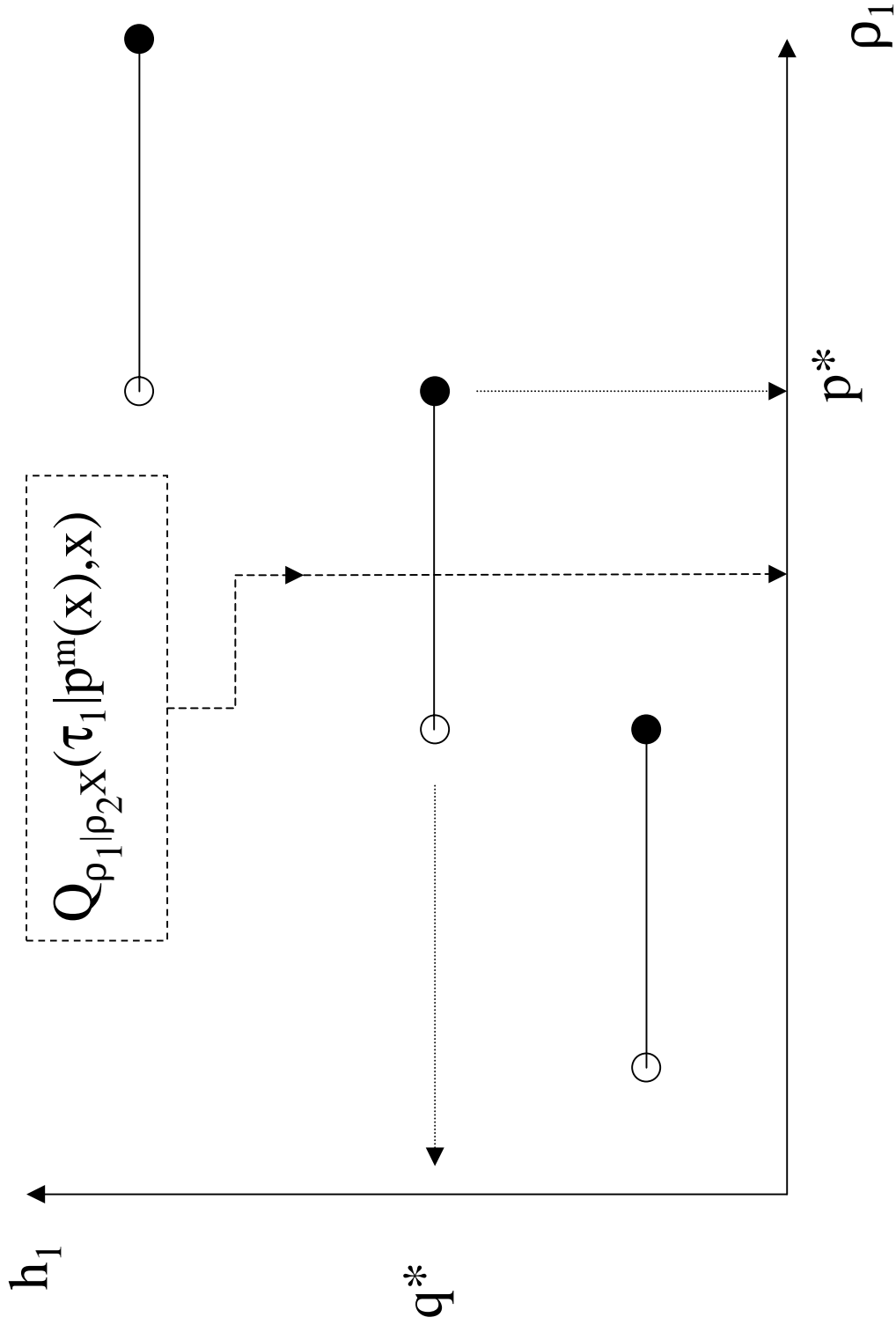
Figure 1: Bounding distribution and quantile functions

Figure 2: Proof of Lemma 2 – discontinuous h_1