Available online at www.sciencedirect.com

ScienceDirect

Journal of Econometrics 139 (2007) 15–34

JOURNAL OF Econometrics

www.elsevier.com/locate/jeconom

# Instrumental values

Andrew Chesher*

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

Available online 20 July 2006

**Abstract**

This paper studies identification of partial differences of nonseparable structural functions. A model is defined which admits structural functions exhibiting a degree of monotonicity with respect to a latent variate. The model identifies partial differences when there are instrumental values of covariates over which the latent variate exhibits a local quantile invariance, and a local order condition holds. The result is useful when covariates exhibit discrete variation, as arises often in practice, and when restricting latent variates and covariates to be statistically independent is unpalatable. The results are illustrated with data from the returns-to-schooling study of Angrist and Krueger [1991. Does compulsory schooling attendance affect schooling and earnings? Quarterly Journal of Economics 106, 979–1014].


*JEL classification:* C14; C20; C30; C51

*Keywords:* Identification; Nonseparable models; Nonparametric identification; Endogeneity; Instrumental variables

## 1. Introduction

### 1.1. Nonseparability

Models that permit latent variates to be nonseparable[1] from a structural function are capable of capturing important aspects of economic behaviour. When data are generated by structures in which there is nonseparability, a possibility which it is difficult to exclude a

*Tel.: +44 2076795857; fax: +44 2079162775.
*E-mail address:* andrew.chesher@ucl.ac.uk.

[1]"Nonseparable" is used as a short-hand for "not additively separable"

0304-4076/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.jeconom.2006.06.003

priori, responses to changes in observable variables have nondegenerate distributions across economic agents. Understanding the distribution of response as well as average response is important in many policy contexts.

Consider for example a model for observable outcomes $Y_1$ and $Y_2$ in which nonseparable structural functions are permitted:

$$Y_1 = h_1(Y_2, Z, \varepsilon_1, \varepsilon_2),$$

$$Y_2 = h_2(Z, \varepsilon_2).$$

Here $Z$ is a list of covariates and $\varepsilon_1$ and $\varepsilon_2$ are latent random variables. Such a model can arise for example in the context of a study of returns to schooling with $Y_1$ denoting the log wage and $Y_2$ denoting a measure of investment in schooling. The model admits structures in which the change in $h_1$ that arises when $Y_2$ alone is altered, the "pure" effect of schooling, is permitted to depend on $\varepsilon_1$ and $\varepsilon_2$ and so to be stochastic.

This paper proposes weak restrictions under which there is identification of characteristics of a nonseparable structural function, such as $h_1$ above, at well-defined values of latent variates. A construction is employed that allows for the possibility that covariates exhibit discrete variation.

Discrete covariates arise frequently in econometric practice, as binary indicators or counts of events. One aim of this paper is to understand the limitations that discrete variation places on the class of structural characteristics that can be identified.

The paper focusses on identification of a particular class of structural characteristics, namely values of *partial differences* of a structural function, that is differences in the values yielded by a structural function when all arguments but one are held constant and the remaining argument takes two distinct values. These are structural characteristics which, unlike partial derivatives, can feasibly be identified without parametric restrictions when covariates exhibit discrete variation. The results also apply when there is continuous variation and limiting arguments produce existing results on the nonparametric identification of partial *derivatives* of structural functions.

### 1.2. Identifying restrictions

A structural characteristic is identified only in the context of a model which places restrictions on admissible structures. Since the interpretation of all econometric analysis is contingent upon identifiability which necessarily rests on some untestable restrictions, it is prudent to base identifiability on weak restrictions. This motivates the conditions proposed here which include neither parametric restrictions nor the requirement that latent variates and covariates be independently distributed.

If, additionally, parametric restrictions are imposed then the value of a local characteristic (for example the slope of a chord of a structural function over some interval) may be equal to the value of a global characteristic (for example the slope of a parametric linear structural function). Then the force of the parametric restriction is to allow identification of the value of the global characteristic from information provided by just local, discrete variation in covariates.

Classical identification conditions impose restrictions on the covariation of latent variates and covariates and limit the sensitivity of structural functions to variation in certain covariates. The focus of this paper on the restrictions on structures required to identify *partial differences* of structural functions allows the role played by these conditions to be seen rather clearly.

Under very weak conditions, which do *not* include these classical identification conditions, differences of structural functions *are* identifiable. The classical identification conditions just described ensure that certain identifiable differences of structural functions are *partial* differences, that is differences obtained by varying just one argument of a structural function.

The classical "rank" condition, when viewed entirely in the context of the study of identification of differences, ensures that an identifiable partial difference is obtained by comparing values of a structural function at two *distinct* points. If the rank condition fails to hold then the partial difference is necessarily zero.

### 1.3. Quantile functions

A key to progress in the study of identification when latent variates may be nonseparable from structural functions is understanding that an analysis that proceeds in terms of *conditional quantile functions* is well suited to the nature of the problem considered.

For $\tau \in (0,1)$ the conditional $\tau$-quantile of $A$ given a vector of covariates $B = b$ is defined as

$$Q_{A|B}(\tau|b) \equiv \inf_q \{q \in \Re : F_{A|B}(q|b) \geqslant \tau\},$$

where $F_{A|B}(a|b) \equiv P[A \leqslant a | B = b]$ is the conditional distribution function of $A$ given $B = b$. For all $b$ at which $h(a,b)$ is a nondecreasing function of $a$ there is the following equivariance property:

$$Q_{h(A,B)|B}(\tau|b) = h(Q_{A|B}(\tau|b), b).$$

Because of this property, restrictions imposed on the covariate driven variation of conditional quantiles of a latent variate given covariates can have implications for the information about a structural function contained in conditional quantiles of outcomes given covariates, as long as the function is restricted to exhibit a degree of monotonic variation with respect to the latent variate. That sort of monotonicity restriction is an essential element in the restrictions that define the identifying model of this paper.

### 1.4. Identifying restrictions (continued)

Consider identification of a partial difference of a nonseparable structural function $h_1$ which delivers the value of an outcome $Y_1$ as follows:

$$Y_1 = h_1(Y_2, Z, \varepsilon_1, \varepsilon_2).$$

The arguments of the function are two latent random variables, $\varepsilon_1$ and $\varepsilon_2$, a list of covariates, $Z$, and a *continuously* distributed outcome $Y_2$ which may not be distributed independently of $\varepsilon_1$. This outcome is determined by a "reduced form" equation

$$Y_2 = h_2(Z, \varepsilon_2) \tag{1}$$

involving the latent variate variate $\varepsilon_2$ which may appear in the structural function $h_1$.

The variate $\varepsilon_2$ is normalised, uniformly distributed on $(0,1)$ independent of $Z$. Under this normalisation:

$$P[\varepsilon_2 \leqslant F_{Y_2|Z}(y_2|z)] = F_{Y_2|Z}(y_2|z) = P[Q_{Y_2|Z}(\varepsilon_2|z) \leqslant y_2] \tag{2}$$

the third expression here arising on inverting the distribution function in the first expression. Since, in view of (1),

$$F_{Y_2|Z}(y_2|z) = P[h_2(z, \varepsilon_2) \leqslant y_2] \tag{3}$$

it is clear on comparing (2) and (3) that, under the normalisation adopted for the distribution of $\varepsilon_2$, the function $h_2$ can be defined to be the conditional quantile function of $Y_2$ given $Z$, with its probability argument replaced by $\varepsilon_2$, that is:

$$h_2(z, \varepsilon_2) \equiv Q_{Y_2|Z}(\varepsilon_2|z).$$

This definition is employed henceforth.

The variates $\varepsilon_1$ and $\varepsilon_2$ are not required to be independently distributed and $\varepsilon_1$ may not be distributed independently of $Z$. $Y_2$ is scalar in the main development of the results. The case in which there may be many endogenous arguments of $h_1$ is addressed in Section 5.8.

Define

$$\bar{e}_1(z) \equiv Q_{\varepsilon_1|\varepsilon_2 Z}(\bar{\tau}_1|\bar{\tau}_2, z),$$

where $\bar{\tau}_1$ and $\bar{\tau}_2$ are probabilities in $(0,1)$. Note that since $\varepsilon_2$ is uniformly distributed on $(0,1)$ $\bar{\tau}_2$ is the $\bar{\tau}_2$-quantile of $\varepsilon_2$.

The partial differences whose identification is studied have the following form:

$$\triangle = h_1(y_2', z, \bar{e}_1(z), \bar{\tau}_2) - h_1(y_2'', z, \bar{e}_1(z), \bar{\tau}_2).$$

Two types of additional restrictions define the model considered in this paper.

First there are restrictions on admissible structural functions $h_1$, specifically that: admissible $h_1$ exhibit *monotonic* variation with respect to the latent variate $\varepsilon_1$, normalised nondecreasing, and that there is a degree of insensitivity to some variation in values of covariates. The number of latent variates appearing in the structural function is required to be no larger than the number of observable outcomes, two in the case considered here.[2] This restriction is essential if characteristics of a structural function *at particular values* of latent variates are to be identified without imposing strong restrictions. The restriction can be dropped if one is content to identify certain average structural functions.[3]

The second type of restriction limits the variation in the conditional quantile $\bar{e}_1(z)$ as covariate values vary.

The restrictions on $Z$-driven variation in $\bar{e}_1(z)$ and on $Z$ -driven variation in the structural function $h_1$ are both required to hold for variations in $Z$ confined to a set of *instrumental values*. This set may be nondenumerable, but when there are discrete covariates it may be denumerable. The membership of the set of instrumental values may

---

[2]This number of latent variates, $R$, is the irreducible number that remain after combination free of outcomes and covariates. Thus a structural function

$$h(Y_2, Z, \varepsilon_1, \varepsilon_2, \varepsilon_3) = \exp(\theta + \varepsilon_2)Y_2 + \beta Z + \varepsilon_1 + \varepsilon_3$$

involving three latent variables has $R = 2$ because $\varepsilon_1$ and $\varepsilon_3$ can be combined, free of $Y_2$ and $Z$.

[3]Imbens and Newey (2003) study identification of average structural functions in nonseparable models with many latent variates.

depend upon the chosen probabilities, $\bar{\tau}_1$ and $\bar{\tau}_2$, which define the conditional quantiles of $\varepsilon_1$ and $\varepsilon_2$.

The set of instrumental values, $\bar{V} \subseteq \Re^K$, is required to be such that:

1. $\bar{e}_1(z)$ is invariant with respect to variation in $z \in \bar{V}$, and,
2. for any $\{z', z''\} \in \bar{V}$,

$$h_1(y_2', z', \bar{e}_1, \bar{\tau}_2) = h_1(y_2', z'', \bar{e}_1, \bar{\tau}_2),$$

   where $\bar{e}_1$ is the common value of $\bar{e}_1(z)$ for $z \in \bar{V}$. This is like the order condition of Koopmans et al. (1950). It would be satisfied if some elements of $z$ were excluded from $h_1$ and only these elements varied within $\bar{V}$.

In Section 4 a theorem is stated and proved which defines a model and asserts the following:

1. If $z'$, $z''$ and $z^*$ belong to a set of instrumental values, $\bar{V}$, then the model identifies the partial difference:

$$h_1(y_2', z^*, \bar{e}_1, \bar{\tau}_2) - h_1(y_2'', z^*, \bar{e}_1, \bar{\tau}_2), \tag{4}$$

   which is invariant with respect to choice of $z^* \in \bar{V}$. Here $y_2'$ and $y_2''$ are, respectively, $h_2(z', \bar{\tau}_2)$ and $h_2(z'', \bar{\tau}_2)$.
2. All admissible structures in which the partial difference (4) takes a particular value, say $a$, generate distributions for $Y_1$ and $Y_2$ given $Z$ such that the difference in conditional quantiles:

$$Q_{Y_1|Y_2Z}(\bar{\tau}_1|Q_{Y_2|Z}(\bar{\tau}_2|z'), z') - Q_{Y_1|Y_2Z}(\bar{\tau}_1|Q_{Y_2|Z}(\bar{\tau}_2|z''), z'') \tag{5}$$

   takes the same value, $a$, for any $\{z', z''\} \in \bar{V}$. Here $Q_{Y_1|Y_2Z}$ and $Q_{Y_2|Z}$ are conditional quantile functions of, respectively, $Y_1$ given $Y_2$ and $Z$, and of $Y_2$ given $Z$. This identifies the partial difference in Eq. (4).

Eq. (5) immediately suggests an estimator of the value of the partial difference (4) constructed using estimators of the conditional quantile functions that appear there.

### *1.5. Plan of the paper*

Section 2 reviews related literature. Section 3 defines concepts used in the paper and states and proves a lemma which is helpful in determining whether a model identifies a structural characteristic.

Section 4 defines a model and states and proves theorems which assert that the model identifies values of, and partial differences of, structural functions in which one argument is endogenous.

Section 5 examines a number of issues including: the way in which a classical analysis via instrumental variables is subsumed in the analysis of this paper; over- and just-identification; identification of partial *derivatives* of structural functions; identification of partial differences with respect to *covariates* and identification of partial differences when there are many endogenous variables.

Section 6 illustrates the results using some of the data employed in Angrist and Krueger (1991).

## 2. Related literature

The study of identification dates from the start of the discipline of econometrics. Early work on the topic was largely confined to the study of the identifying power of parametric, mainly linear, models under conditions requiring latent variates and covariates to be mean independent. Notable contributions are contained in Working (1925), Working (1927), Tinbergen (1930), Frisch (1934, 1938), Hurwicz (1950), Koopmans and Reiersøl (1950), Koopmans et al. (1950), Fisher (1959, 1961, 1966) and Wegge (1965). Morgan (1990) reviews the early contributions. Hurwicz (1950) is a particularly notable paper because it puts in place, at an early stage in the development of econometrics, the framework used in the modern analysis of nonparametric and partial identification.[4]

Until the early 1970s much econometric analysis dealt with aggregate market or national data. One would not expect such data to be generated by highly nonlinear structures and so the focus of the study of identification on simple parametric models and indeed on linear models was apposite.

The microeconometrics revolution of the 1970s wrought a major change, bringing new interest in the study of the behaviour of individual economic agents who may face wide variations in conditions under which choices are made, leading to consideration of structures in which nonlinearity is an essential and interesting element. Economic theory provides little guidance concerning the precise forms of nonlinear structural equations and responses of heterogeneous agents to changes in values of arguments of structural functions are plausibly heterogeneous. These considerations led to interest in nonparametric identification in nonseparable models which is the subject of this paper.

Roehrig (1988) addressed this topic, studying global identification of smooth structural functions when latent variates and covariates are independently distributed and there is continuous variation in covariates and outcomes.[5] Newey (1999) study global identification under certain conditional mean independence conditions and the restriction that latent variates are additively separable. Imbens and Newey (2003) study global identification in triangular nonseparable models under the restriction that latent variates and covariates are independently distributed. They dispense with the smoothness restrictions imposed in Roehrig (1988). Matzkin (2003), like this paper, considers how the information contained in conditional quantile functions bears on the identifiability of characteristics of nonseparable structures dealing with the case in which arguments of structural functions are exogenous.

Chesher (2003) studies local identification of partial derivatives of nonseparable smooth structural functions in triangular models when there are restrictions on iterated conditional quantiles of latent variates. Independence of latent variates and covariates is not required. When covariates exhibit discrete variation nonparametric identification of partial derivatives is infeasible. This paper focusses on local identification of *partial differences* of nonseparable structural functions in triangular models with discrete variation in

[4]The Cowles Commission Monograph No. 10 in which Hurwicz's remarkable paper appears is available at the Cowles Foundation web site ⟨http://cowles.econ.yale.edu/P/cm/m10/index.htm⟩.

[5]Benkhard and Berry (2006) reports an error in Roehrig (1988) and Matzkin (2005) provides a correction.

covariates. In smooth models with continuous variation in covariates, limiting arguments applied to the results presented here lead to the results given in Chesher (2003). Endogenous variables are required to be continuously distributed. Chesher (2005) gives some interval identification results for discrete endogenous variable models.

## 3. Structures, models and identification

This section makes precise the definition of a structure and of the identification of a structural characteristic and states and proves a lemma which is helpful in determining whether a model identifies a structural characteristic.

Following Hurwicz (1950), a *structure* is defined as:[6]

1. a system of equations delivering a unique value of a vector outcome, $Y \equiv \{Y_m\}_{m=1}^{M}$, given a value of a vector covariate, $Z \equiv \{Z_k\}_{k=1}^{K}$ and a value of a vector of latent variates, $\varepsilon \equiv \{\varepsilon_r\}_{r=1}^{R}$, and,
2. a conditional distribution function, $F_{\varepsilon|Z}$ for absolutely continuously distributed latent variates given the covariates.

Each such structure implies a conditional distribution for $Y$ given $Z$: $F_{Y|Z}$. Identification issues arise because distinct structures may generate the same distribution function $F_{Y|X}$. Such structures are termed observationally equivalent. No amount of data can distinguish amongst observationally equivalent structures.

A model comprises restrictions on admissible structures. A *structural characteristic*[7] is a functional $\theta(S)$ of a structure, $S$. A model identifies a structural characteristic $\theta(S)$ in a structure $S_0$ if that characteristic has the same value in all structures admitted by the model and observationally equivalent to $S_0$ (Koopmans and Reiersøl, 1950). A structural characteristic $\theta(S)$ is *uniformly identified* by a model if it is identifiable for every structure $S$ admitted by the model.

It is helpful to have a simple means of determining whether a model uniformly identifies a structural characteristic. This is provided by the following lemma.

**Lemma.** *Consider a model, let $S^a$ be the set of structures admitted by the model such that $\theta(S) = a$ and let $A$ be the set of all values of $\theta(S)$ generated by admissible structures. Let $F_{Y|Z}^{S}$ denote the conditional distribution function generated by a structure $S$. Suppose there exists a single valued functional of the conditional distribution function of $Y$ given $Z$, $\mathscr{G}(F_{Y|Z})$, such that for each $a \in A$, $\mathscr{G}(F_{Y|Z}^{S}) = a$ for all $S \in S^a$. Then $\theta(S)$ is uniformly identified by the model.*

**Proof.** Consider any value $a_0 \in A$ and any structure $S_0$ with $\theta(S_0) = a_0$ and let $S_0^e$ be the set of structures observationally equivalent to $S_0$. Consider any $S' \in S_0^e$ and let $\theta(S') = a'$. If a functional $\mathscr{G}$ with the stated property exists then $\mathscr{G}(F_{Y|Z}^{S'}) = a'$ and $\mathscr{G}(F_{Y|Z}^{S_0}) = a_0$. Since $S'$ and $S_0$ are observationally equivalent $F_{Y|Z}^{S'} = F_{Y|Z}^{S_0}$ and therefore $a' = a_0$. Therefore, if a functional $\mathscr{G}$ with the stated property exists then, for any $a_0 \in A$, all structures

[6]This departs slightly from Hurwicz (1950) in that the role of covariates, $Z$, is made explicit and latent variates are required to have absolutely continuous distributions.

[7]The term "structural characteristic" is due to Koopmans and Reiersøl (1950). Hurwicz (1950) used the term "criterion".

observationally equivalent to any structure $S_0$ with $\theta(S_0) = a_0$ have the same value, $a_0$, of the structural characteristic, and so $\theta(S)$ is uniformly identified by the model. $\square$

If, for a model and structural characteristic, such a functional can be found then uniform identification of the structural feature by the model is assured and there is a clear route to estimation via the analog principle.[8]

## 4. Identification of a partial difference

This section defines a model that identifies a partial difference of a nonseparable structural function with respect to an endogenous variable.

In the case considered in this section only one of the arguments of the structural function is endogenous. The extension to the case in which there are many endogenous variables is discussed in Section 5.8. Identification of partial differences with respect to a covariate is discussed in Section 5.7.

Four restrictions are now introduced and then two theorems are stated and proved.

First there are some definitions. Let $Y_1$ and $Y_2$ be scalar random variables; let $Z \equiv \{Z_k\}_{k=1}^{K}$ be a list of covariates; let $\varepsilon_1$ and $\varepsilon_2$ be continuously distributed unobservable scalar random variables with $\varepsilon_2$ distributed uniformly on $(0, 1)$ independently of $Z$. Let $Q_{\varepsilon_1|\varepsilon_2 Z}(\tau|e_2, z)$ denote the conditional $\tau$-quantile of $\varepsilon_1$ given $\varepsilon_2 = e_2$ and $Z = z$.

Consider $\bar{\tau} \equiv \{\bar{\tau}_1, \bar{\tau}_2\} \in (0, 1) \times (0, 1)$, and a set of *instrumental values* of the covariates $Z$, $\bar{V} \subseteq \Re^K$, whose membership may depend upon the value of $\bar{\tau}$.

Define the following iterated conditional quantile:

$$\bar{e}_1(z) \equiv Q_{\varepsilon_1|\varepsilon_2 Z}(\bar{\tau}_1|\bar{\tau}_2, z). \tag{6}$$

The restrictions are now introduced.

**Restriction 1.** *The value of the outcome $Y_1$ is uniquely determined by the structural equation*

$$Y_1 = h_1(Y_2, Z, \varepsilon_1, \varepsilon_2) \tag{7}$$

*and $Y_2$ is continuously distributed given $Z = z$ for all $z \in \bar{V}$.*

**Remark.** The continuously distributed outcome $Y_2$ is determined by the equation:

$$Y_2 = h_2(Z, \varepsilon_2), \tag{8}$$

where $\varepsilon_2$ is normalised uniformly distributed on $(0, 1)$ independently of $Z$ and $h_2(Z, \varepsilon_2) \equiv Q_{Y_2|Z}(\varepsilon_2|Z)$. Necessarily $Y_2$ conditional on $Z$ has the distribution function $F_{Y_2|Z}$ which is the inverse function of $Q_{Y_2|Z}$. Eq. (8) constitutes a "reduced form" equation for $Y_2$ suitable for consideration of identification when latent variables may be nonseparable. Define the following values of $Y_1$ and $Y_2$:

$$\bar{y}_2(z) \equiv h_2(z, \bar{\tau}_2), \tag{9}$$

$$\bar{y}_1(z) \equiv h_1(\bar{y}_2(z), z, \bar{e}_1(z), \bar{\tau}_2). \tag{10}$$

**Restriction 2** (*Monotonicity*). *For $z \in \bar{V}$, $h_1(\bar{y}_2(z), z, e_1, \bar{\tau}_2)$ is a monotonic function of $e_1$, normalised nondecreasing.*

[8]See Manski (1988). For example one could use $\mathcal{G}(\hat{F}^{S}_{Y|Z})$ or $\mathcal{G}(\widehat{F^{S}_{Y|Z}})$ as estimates of $\theta(S)$.

**Restriction 3** (*Quantile invariance*). *For all* $\{z', z''\} \in \bar{V}$

$$\bar{e}_1(z') = \bar{e}_1(z'') \equiv \bar{e}_1.$$

**Restriction 4** (*Order condition*). *For all* $\{z', z''\} \in \bar{V}$

$$h_1(\bar{y}_2(z'), z', \bar{e}_1, \bar{\tau}_2) = h_1(\bar{y}_2(z'), z'', \bar{e}_1, \bar{\tau}_2).$$

Theorem 1 asserts the identifiability of the value delivered by the structural function $h_1$ at well-defined values of its arguments.

**Theorem 1.** *Under Restrictions* 1 *and* 2, *for* $z \in \bar{V}$ *and any* $a$:

$$\bar{y}_1(z) = a \Rightarrow Q_{Y_1|Y_2Z}(\bar{\tau}_1 | Q_{Y_2|Z}(\bar{\tau}_2|z), z) = a$$

*and the model defined by Restrictions* 1 *and* 2 *uniformly identifies* $\bar{y}_1(z)$ *for* $z \in \bar{V}$.

**Proof of Theorem 1.** Substitute for $Y_2$ in Eq. (7) giving

$$Y_1 = h_1(h_2(Z, \varepsilon_2), Z, \varepsilon_1, \varepsilon_2)$$

and evaluate the right-hand side at $\varepsilon_2 = \bar{\tau}_2$ and $Z = z$ and consider variations in $\varepsilon_1$. Restriction 2 and the equivariance property of quantiles imply that, for any $z \in \bar{V}$, since $\bar{e}_1(z)$ is the $\bar{\tau}_1$-quantile of $\varepsilon_1$ given $\varepsilon_2 = \bar{\tau}_2$ and $Z = z$,

$$h_1(h_2(z, \bar{\tau}_2), z, \bar{e}_1(z), \bar{\tau}_2) = Q_{Y_1|\varepsilon_2 Z}(\bar{\tau}_1 | \bar{\tau}_2, z). \tag{11}$$

Consider the right-hand side of Eq. (11). Since $\bar{y}_2(z) \equiv h_2(z, \bar{\tau}_2)$ and $Y_2$ is continuously distributed given $Z = z \in \bar{V}$, the events $(\varepsilon_2 = \bar{\tau}_2 \cap Z = z)$ and $(Y_2 = \bar{y}_2(z) \cap Z = z)$ are identical, so conditioning on $\varepsilon_2 = \bar{\tau}_2$ and $Z = z$ is the same as conditioning on $Y_2 = \bar{y}_2(z)$ and $Z = z$.[9] Therefore

$$Q_{Y_1|\varepsilon_2 Z}(\bar{\tau}_1 | \bar{\tau}_2, z) = Q_{Y_1|Y_2 Z}(\bar{\tau}_1 | \bar{y}_2(z), z). \tag{12}$$

Considering the left-hand side of Eq. (11) there is by virtue of (9) and (10)

$$h_1(h_2(z, \bar{\tau}_2), z, \bar{e}_1(z), \bar{\tau}_2) = h_1(\bar{y}_2(z), z, \bar{e}_1(z), \bar{\tau}_2) \equiv \bar{y}_1(z) \tag{13}$$

and therefore

$$\bar{y}_1(z) = Q_{Y_1|Y_2Z}(\bar{\tau}_1 | \bar{y}_2(z), z)$$

in which $\bar{y}_2(z)$ can be replaced by $Q_{Y_2|Z}(\bar{\tau}_2|z)$ giving:

$$\bar{y}_1(z) = Q_{Y_1|Y_2Z}(\bar{\tau}_1 | Q_{Y_2|Z}(\bar{\tau}_2|z), z).$$

It follows that, for any $z \in \bar{V}$ and any $a$,

$$\bar{y}_1(z) = a \Rightarrow Q_{Y_1|Y_2Z}(\bar{\tau}_1 | Q_{Y_2|Z}(\bar{\tau}_2|z), z) = a.$$

Applying the lemma of Section 3 gives the result of Theorem 1. This ends the Proof of Theorem 1. □

Theorem 2 concerns the identification of a partial difference of the structural function $h_1$ with respect to a finite variation in $Y_2$.

[9]Note that if $Y_2$ has a discrete distribution this conclusion would *not* be true because there could be many values of $\bar{\tau}_2$ yielding the same value of $\bar{y}_2(z)$. The case in which $Y_2$ is discrete is considered in Chesher (2005).

Let $z'$, $z''$ and $z^*$ be values of the covariates $Z$. Define the quantile function difference, $\bar{\Delta}_Q(z', z'')$, and structural function difference, $\bar{\Delta}_{h_1}(z', z'', z^*)$ as follows:

$$\bar{\Delta}_Q(z', z'') \equiv Q_{Y_1|Y_2Z}(\bar{\tau}_1|Q_{Y_2|Z}(\bar{\tau}_2|z'), z') - Q_{Y_1|Y_2Z}(\bar{\tau}_1|Q_{Y_2|Z}(\bar{\tau}_2|z''), z''), \tag{14}$$

$$\bar{\Delta}_{h_1}(z', z'', z^*) \equiv h_1(\bar{y}_2(z'), z^*, \bar{e}_1, \bar{\tau}_2) - h_1(\bar{y}_2(z''), z^*, \bar{e}_1, \bar{\tau}_2). \tag{15}$$

Note that $\bar{\Delta}_{h_1}(z', z'', z^*)$ is a *partial* difference of the structural function $h_1$ with respect to the finite variation $\bar{y}_2(z'') \to \bar{y}_2(z')$.

**Theorem 2.** *Under Restrictions* 1–4, *for* $\{z', z'', z^*\} \in \bar{V}$, *the structural function difference*, $\bar{\Delta}_{h_1}(z', z'', z^*)$, *is invariant with respect to* $z^*$. *For any a*

$$\bar{\Delta}_{h_1}(z', z'', z^*) = a \Rightarrow \bar{\Delta}_Q(z', z'') = a$$

*and the model defined by Restrictions* 1–4 *uniformly identifies the structural function partial difference* $\bar{\Delta}_{h_1}(z', z'', z^*)$.

**Proof of Theorem 2.** Restriction 3 and Eq. (13) imply that for $z \in \bar{V}$

$$\bar{y}_1(z) = h_1(\bar{y}_2(z), z, \bar{e}_1, \bar{\tau}_2). \tag{16}$$

Restriction 4 implies that, for any $z \in \bar{V}$ the second appearance of $z$ in Eq. (16) can be replaced by any $z^* \in \bar{V}$ and so for all $\{z, z^*\} \in \bar{V}$ there is

$$\bar{y}_1(z) = h_1(\bar{y}_2(z), z^*, \bar{e}_1, \bar{\tau}_2).$$

Therefore the structural function difference $\bar{\Delta}_{h_1}(z', z'', z^*)$ is invariant with respect to $z^*$, that is:

$$\bar{\Delta}_{h_1}(z', z'', z^*) = \bar{y}_1(z') - \bar{y}_1(z'').$$

Theorem 1 implies that each component of the structural function difference $\bar{\Delta}_{h_1}(z', z'', z^*)$ is identified as follows:

$$\bar{y}_1(z') = a' \Rightarrow Q_{Y_1|Y_2Z}(\bar{\tau}_1|Q_{Y_2|Z}(\bar{\tau}_2|z'), z') = a',$$

$$\bar{y}_1(z'') = a'' \Rightarrow Q_{Y_1|Y_2Z}(\bar{\tau}_1|Q_{Y_2|Z}(\bar{\tau}_2|z''), z'') = a''$$

and therefore,

$$\bar{\Delta}_{h_1}(z', z'', z^*) = b \Rightarrow \bar{\Delta}_Q(z', z'') = b,$$

which, using Lemma 1, completes the proof of Theorem 2. □

## 5. Remarks and extensions

### *5.1. Rank condition, weak instruments and parametric restrictions*

The identification result of Theorem 2 is useful only if the "rank condition", $\bar{y}_2(z') \neq \bar{y}_2(z'')$ holds, for otherwise the partial difference considered in the theorem is necessarily zero.

The partial differences that can be identified depend on the content of the set of instrumental values and the strength of the relationship between $Y_2$ and $Z$. If there are just two admissible instrumental values, for example if $Z$ is a single binary covariate, then for any choice of $\bar{\tau}_2$ only two values of $Y_2$ can be generated and only one partial difference can be identified. Even if there are many instrumental values in $\bar{V}$, if $Z$ has only a weak effect on $Y_2$ then the partial differences of $h_1$ that can be identified will pertain to small finite variations in $Y_2$ and will be restricted to a small range of values of $Y_2$. These issues are revisited in Section 6.

If $h_1$ is a smooth function of $Y_2$ then the slope of a $Y_2$-chord of the structural function, that is:

$$\frac{\bar{\Delta}_{h_1}(z', z'')}{\bar{y}_2(z') - \bar{y}_2(z'')},$$

can be identified as long as the rank condition is satisfied.[10]

Additional parametric restrictions are helpful but will rarely have secure foundation in economic theory. If $h_1$ is restricted to be linear in $Y_2$, with a coefficient that may depend upon $Z$ and $\varepsilon_2$, then just two instrumental values are sufficient to globally identify the value of this coefficient at a value of $Z$ and a value of $\varepsilon_2$. By extension, if $h_1$ is restricted to be a polynomial function of $Y_2$ of degree $G$ then $G+1$ distinct instrumental values may be sufficient to identify the coefficients of the polynomial.

### 5.2. Discrete outcomes

The identification result of Theorem 2 holds when the outcome $Y_1$ is discrete or continuous. The case in which $Y_2$ is discrete is not covered by the theorem. When $Y_2$ is discrete the function $h_2$ is constant over ranges of $\varepsilon_2$ and so the relationship between $\bar{\tau}_2$ and $y_2(z)$ is not one-to-one and Eq. (12), essential in the proof of Theorem 1, does not hold. Chesher (2005) shows that under an additional monotonicity restriction there can be interval identification of a structural partial difference in this case.

### 5.3. Estimation

Quantile regression estimation methods reviewed in Koenker (2005) can be employed to estimate the values of partial differences identified by the theorem.

Estimation could be parametric, semi- or nonparametric depending on the extent of additional restrictions imposed. For parametric estimation, see Koenker and Bassett (1978), Koenker and d'Orey (1987); for semiparametric estimation see for example Chaudhuri et al. (1997), Kahn (2001) and Lee (2003); for nonparametric estimation, see for example Chaudhuri (1991). The sampling properties of the chosen estimator will depend upon restrictions on structures additional to those considered in this paper.

The software suite `R` (Ihaka and Gentleman, 1996) contains the `quantreg` library written by Koenker which implements parametric linear and nonlinear quantile regression procedures. The facility offered by `quantreg` to calculate weighted quantile regression estimators allows easy calculation of locally linear quantile regression estimators and thus a route to calculation of nonparametric quantile regression estimates.

[10] Here and later the argument $z^*$ of $\bar{\Delta}_{h_1}$ is suppressed because under Restrictions 1–4 $\bar{\Delta}_{h_1}$ is invariant with respect to $z^*$.

### 5.4. Instrumental variables

In the classical analysis of the identifying power of parametric models it is common to consider "exclusion" restrictions, the variables excluded from the structural function of interest being termed "instrumental variables". This case falls within the scope of Theorem 2.

Partition $Z$ into two subsets, $Z_{\text{inc}}$ and $Z_{\text{exc}}$ and suppose there are restrictions excluding $Z_{\text{exc}}$ (the instrumental variables) from the structural function $h_1$. Define

$$z' = \{z^{\dagger}_{\text{inc}}, z'_{\text{exc}}\}, \quad z'' = \{z^{\dagger}_{\text{inc}}, z''_{\text{exc}}\}$$

where, note, the value of $Z_{\text{inc}}$ is $z^{\dagger}_{\text{inc}}$ in both $z'$ and $z''$ which ensures that the order Restriction 4 is satisfied.

If the quantile invariance Restriction 3 is satisfied, that is:

$$Q_{\varepsilon_1|\varepsilon_2 Z}(\bar{\tau}_1|\bar{\tau}_2, z') = Q_{\varepsilon_1|\varepsilon_2 Z}(\bar{\tau}_1|\bar{\tau}_2, z'') \equiv \bar{e}_1$$

then the partial difference

$$h_1(\bar{y}_2(z'), z^*, \bar{e}_1, \bar{\tau}_2) - h_1(\bar{y}_2(z''), z^*, \bar{e}_1, \bar{\tau}_2)$$

is identified for any $z^* \in \bar{V}$. Note that in this nonparametric analysis of identification of a partial difference the sensitivity of $Q_{\varepsilon_1|\varepsilon_2 Z}$ to variations in $z_{\text{inc}}$ is not subject to any restrictions because $z_{\text{inc}}$ is held constant at $z^{\dagger}$. However, dependence of $Q_{\varepsilon_1|\varepsilon_2 Z}$ on $z_{\text{inc}}$ does bear on the interpretation of $\bar{e}_1$ and thus of the structural partial difference.

### 5.5. Over- and just-identification

There is over-identification of the value of a structural partial difference, $\bar{\triangle}_{h_1}(z', z'', z^*)$, when there are two or more distinct pairs of instrumental values, say: $\{z', z''\}$ and $\{z^+, z^{++}\}$, for which Restrictions 1–4 are satisfied, such that $\bar{y}_2(z') = \bar{y}_2(z^+)$ and $\bar{y}_2(z'') = \bar{y}_2(z^{++})$.[11] Otherwise the partial difference is just-identified in the sense that weakening any of Restrictions 1–4 results in loss of identification.

When there is over-identification there is scope for improving the efficiency of estimation by combining alternative estimates, based on different pairs of instrumental values, for example using a minimum distance estimator. There is also scope for testing a subset of the identifying restrictions.

### 5.6. Smooth structures and identification of derivatives

If the $Y_2$ partial derivative of the structural function exists, and $Z$ can vary continuously in the set of instrumental values inducing continuous variation in $Y_2$ then, by considering the limiting behaviour of $\bar{\triangle}_{h_1}(z', z'')/(\bar{y}_2(z') - \bar{y}_2(z''))$, the identification of a value of the $Y_2$-derivative of $h_1$ can be achieved, yielding the result given in Chesher (2003).

To see this, consider the case in which all elements of $z'$ and $z''$ are identical except for one, denoted by $z_{\triangle}$. Consider the slope of a $Y_2$-chord of $h_1$ obtained by moving from $z'$ to $z''$, inducing a movement from $\bar{y}_2(z')$ to $\bar{y}_2(z'')$, and suppose the rank condition, $\bar{y}_2(z') \neq$

[11] At least one, but not necessarily both of: $z' \neq z^+$ and $z'' \neq z^{++}$ must hold.

$\bar{y}_2(z'')$ is satisfied. The slope of the chord is

$$A(z', z'') \equiv \frac{\bar{\Delta}_{h_1}(z', z'')}{\bar{y}_2(z') - \bar{y}_2(z'')}$$

and consider its limit (assumed to exist) as $z' \to z''$. Let the limiting value be denoted by $\bar{z}$ and suppose that all values $z$ encountered on passing to the limit lie in a set of instrumental values.

Define

$$B(\bar{z}) \equiv \lim_{z' \to z'' = \bar{z}} A(z', z'').$$

If the limit exists then

$$B(\bar{z}) = \nabla_{y_2} h_1(y_2, z, \varepsilon_1, \varepsilon_2)|_{y_2 = \bar{y}_2(\bar{z}), z = \bar{z}, \varepsilon_1 = \bar{e}_1, \varepsilon_2 = \bar{\tau}_2}, \tag{17}$$

which is the $Y_2$-partial derivative of the structural function $h_1$ evaluated at the point indicated.

Write $B(\bar{z})$ as:

$$B(\bar{z}) = \lim_{z' \to z'' = \bar{z}} \frac{\bar{\Delta}_{h_1}(z', z'')/(z'_\triangle - z''_\triangle)}{(\bar{y}_2(z') - \bar{y}_2(z''))/(z'_\triangle - z''_\triangle)}.$$

Then, if the required limits exist,

$$B(\bar{z}) = \frac{\lim_{z' \to z'' = \bar{z}} (\bar{\Delta}_{h_1}(z', z'')/(z'_\triangle - z''_\triangle))}{\lim_{z' \to z'' = \bar{z}} ((\bar{y}_2(z') - \bar{y}_2(z''))/(z'_\triangle - z''_\triangle))}$$

and therefore:

$$B(\bar{z}) = \frac{\nabla_{z_\triangle} h_1\left(\bar{y}_2(z), z, \bar{e}_1, \bar{\tau}_2\right)|_{z = \bar{z}}}{\nabla_{z_\triangle} \bar{y}_2(z)|_{z = \bar{z}}} \tag{18}$$

$$= \nabla_{y_2} h_1\left(y_2, z, \bar{e}_1, \bar{\tau}_2\right)|_{y_2 = y_2(\bar{z}), z = \bar{z}} + \frac{\nabla_{z_\triangle} h_1\left(y_2, z, \bar{e}_1, \bar{\tau}_2\right)|_{y_2 = y_2(\bar{z}), z = \bar{z}}}{\nabla_{z_\triangle} \bar{y}_2(z)|_{z = \bar{z}}} \tag{19}$$

the second line, following on applying the chain rule, noting that $z_\triangle$ affects $h_1$ in (18) directly and via $\bar{y}_2(z)$.

Under conditions given in Chesher (2003) the values of the derivatives in (19) are uniformly identified by derivatives of conditional quantile functions of outcomes given covariates, and so the structural partial derivative (17) is identified because for any $b$,

$$B(\bar{z}) = b \Rightarrow \nabla_{y_2} Q_{Y_1|Y_2Z}(\bar{\tau}_1 | Q_{Y_2|Z}(\bar{\tau}_2|\bar{z}), \bar{z}) + \frac{\nabla_{z_\triangle} Q_{Y_1|Y_2Z}(\bar{\tau}_1 | Q_{Y_2|Z}(\bar{\tau}_2|\bar{z}), \bar{z})}{\nabla_{z_\triangle} Q_{Y_2|Z}(\bar{\tau}_2|\bar{z})} = b,$$

where these quantile function derivatives are defined as follows:

$$\nabla_{y_2} Q_{Y_1|Y_2Z}(\bar{\tau}_1 | Q_{Y_2|Z}(\bar{\tau}_2|\bar{z}), \bar{z}) \equiv \nabla_{y_2} Q_{Y_1|Y_2Z}(\bar{\tau}_1 | y_2, z)|_{y_2 = Q_{Y_2|Z}(\bar{\tau}_2|\bar{z}), z = \bar{z}},$$

$$\nabla_{z_\triangle} Q_{Y_1|Y_2Z}(\bar{\tau}_1 | Q_{Y_2|Z}(\bar{\tau}_2|\bar{z}), \bar{z}) \equiv \nabla_{z_\triangle} Q_{Y_1|Y_2Z}(\bar{\tau}_1 | y_2, z)|_{y_2 = Q_{Y_2|Z}(\bar{\tau}_2|\bar{z}), z = \bar{z}},$$

$$\nabla_{z_\triangle} Q_{Y_2|Z}(\bar{\tau}_2|\bar{z}) \equiv \nabla_{z_\triangle} Q_{Y_2|Z}(\bar{\tau}_2|z)|_{z = \bar{z}}.$$

### 5.7. Partial differences with respect to covariates

Partial differences of the structural function $h_1$ with respect to a covariate, that is an element of $Z$, are identified by a model similar to that used earlier with slightly modified Restrictions.

Consider a single element $Z_\diamond$, denote remaining elements of $Z$ by $Z_\blacklozenge$, and suppose there exists a set of instrumental values $\bar{V}$ with elements $z$ written as

$$z = (z_\diamond, z_\blacklozenge)$$

such that the following two conditions hold:

1. For all $\{z', z''\} \in \bar{V}$,
$$\bar{y}_2(z') = \bar{y}_2(z'') \equiv \bar{y}_2,$$
$$\bar{e}_1(z') = \bar{e}_1(z'') \equiv \bar{e}_1.$$

2. For all $\{z', z''\} \in \bar{V}$,
$$h_1(\bar{y}_2, z'_\diamond, z'_\blacklozenge, \bar{e}_1, \bar{e}_2) = h_1(\bar{y}_2, z'_\diamond, z''_\blacklozenge, \bar{e}_1, \bar{e}_2).$$
Here the dependence of $h_1$ on $z_\diamond$ and $z_\blacklozenge$ is made explicit in the notation.

For $\{z', z'', z^*\} \in \bar{V}$ define:

$$\bar{\Delta}_{h_1, Z_\diamond}(z'_\diamond, z''_\diamond, z^*_\blacklozenge) \equiv h_1(\bar{y}_2, z'_\diamond, z^*_\blacklozenge, \bar{e}_1, \bar{\tau}_2) - h_1(\bar{y}_2, z''_\diamond, z^*_\blacklozenge, \bar{e}_1, \bar{\tau}_2).$$

Then it can be shown that the model defined by Restrictions 1–4 and conditions (1) and (2) above uniformly identifies the partial difference $\bar{\Delta}_{h_1, Z_\diamond}(z'_\diamond, z''_\diamond, z^*_\blacklozenge)$, which is invariant with respect to $z^*_\blacklozenge$, and that for any $a$,

$$\bar{\Delta}_{h_1, Z_\diamond}(z'_\diamond, z''_\diamond, z^*_\blacklozenge) = a \Rightarrow \bar{\Delta}_Q(z', z'') = a,$$

where $\bar{\Delta}_Q(z', z'')$ is the difference of conditional quantile functions defined in Eq. (14) in Section 4.

### 5.8. Many endogenous variables

This section sketches the development of an identification result when a structural function has many endogenous arguments.

Consider a structural equation with $M-1$ continuously distributed endogenous variables:

$$Y_1 = h_1(Y_2, \ldots, Y_M, Z, \varepsilon_1, \ldots, \varepsilon_M).$$

Define $M-1$ "reduced form" equations

$$Y_i = h_i(Z, \varepsilon_i, \ldots, \varepsilon_M), \quad i = 2, \ldots, M$$

in which $\varepsilon_2, \ldots, \varepsilon_M$ are independently uniformly distributed and distributed independently of $Z$.

The reduced form functions $h_2, \ldots, h_M$ are iterated conditional quantile functions in which the uniformly distributed $\varepsilon_2, \ldots, \varepsilon_M$ appear in place of the probability arguments of the quantile functions. The joint distribution of $Y_2, \ldots, Y_M$ given $Z$ under this definition is $F_{Y_2 \ldots Y_M|Z}$ which is the joint distribution function associated with the iterated conditional quantile functions that define the structural functions $h_2, \ldots, h_M$.

For example with $M = 3$ there is

$$h_3(Z, \varepsilon_3) \equiv Q_{Y_3|Z}(\varepsilon_3|Z),$$

$$h_2(Z, \varepsilon_2, \varepsilon_3) \equiv Q_{Y_2|Y_3Z}(\varepsilon_2|Q_{Y_3|Z}(\varepsilon_3|Z), Z)$$

and the joint density function of $Y_2$ and $Y_3$ given $Z$ is

$$f_{Y_2Y_3|Z} = f_{Y_2|Y_3Z} \times f_{Y_3|Z},$$

where $f_{Y_2|Y_3Z}$ and $f_{Y_3|Z}$ are the density functions associated with the distribution functions $F_{Y_2|Y_3Z}$ and $F_{Y_3|Z}$ whose inverse functions are the conditional quantile functions $Q_{Y_2|Y_3Z}$ and $Q_{Y_3|Z}$.

Choose $\bar{\tau} = \{\bar{\tau}_i\}_{i=1}^{M}$, consider a set of instrumental values of covariates, $\bar{V} \subseteq \Re^K$, define

$$\bar{e}_1(z) \equiv Q_{\varepsilon_1|\varepsilon_2 \ldots \varepsilon_M Z}(\bar{\tau}_1|\bar{\tau}_2, \ldots, \bar{\tau}_M, z) \tag{20}$$

and for $i \in \{1, \ldots, M\}$, $\bar{y}_i(z)$, $i > 1$, and $\bar{y}_1(z)$

$$\bar{y}_i(z) \equiv h_i(z, \bar{\tau}_i, \ldots, \bar{\tau}_M), \quad i = 2, \ldots, M,$$

$$\bar{y}_1(z) \equiv h_1(\bar{y}_2(z), \ldots, \bar{y}_M(z), z, \bar{e}_1(z), \bar{\tau}_2, \ldots, \bar{\tau}_M).$$

Restriction 2 (Monotonicity) is maintained with $h_1$ required to be monotonically varying in its $\varepsilon_1$ argument, normalised nondecreasing. Restriction 3 (quantile invariance) is maintained with $\bar{e}_1 = \bar{e}_1(z)$ defined as in Eq. (20).

Suppose a $Y_j$-partial difference of $h_1$ is of interest, defined for some $\{z', z'', z^*\}$ as follows:

$$\begin{aligned}\bar{\triangle}_{h_1,Y_j}(z', z'', z^*) \equiv{} & h_1(\bar{y}_2(z^*), \ldots, \bar{y}_{j-1}(z^*), \bar{y}_j(z'), \bar{y}_{j+1}(z^*), \ldots, \bar{y}_M(z^*), z^*, \bar{e}_1, \bar{\tau}_2, \ldots, \bar{\tau}_M) \\ & - h_1(\bar{y}_2(z^*), \ldots, \bar{y}_{j-1}(z^*), \bar{y}_j(z''), \bar{y}_{j+1}(z^*), \ldots, \bar{y}_M(z^*), z^*, \bar{e}_1, \bar{\tau}_2, \ldots, \bar{\tau}_M).\end{aligned}$$

Note that *only* the $Y_j$ argument varies here.

Restriction 4 (order condition) is modified requiring for all $(z', z'') \in \bar{V}$

$$\begin{aligned}& h_1(\bar{y}_2(z'), \ldots, \bar{y}_{j-1}(z'), \bar{y}_j(z'), \bar{y}_{j+1}(z'), \ldots, \bar{y}_M(z'), z', \bar{e}_1, \bar{\tau}_2, \ldots, \bar{\tau}_M) \\ & \quad = h_1(\bar{y}_2(z''), \ldots, \bar{y}_{j-1}(z''), \bar{y}_j(z'), \bar{y}_{j+1}(z''), \ldots, \bar{y}_M(z''), z'', \bar{e}_1, \bar{\tau}_2, \ldots, \bar{\tau}_M).\end{aligned}$$

Note that the $z$ argument of $\bar{y}_j(z)$ does *not* vary here.

Recursively define

$$\begin{aligned}\bar{Q}_M(z) &\equiv Q_{Y_M|Z}(\bar{\tau}_M|z), \\ \bar{Q}_{M-1}(z) &\equiv Q_{Y_{M-1}|Y_MZ}(\bar{\tau}_{M-1}, |\bar{Q}_M(z), z), \\ \bar{Q}_{M-2}(z) &\equiv Q_{Y_{M-2}|Y_{M-1}Y_MZ}(\bar{\tau}_{M-2}|\bar{Q}_{M-1}(z), \bar{Q}_M(z), z), \\ \vdots &= \vdots\end{aligned}$$

so that $\bar{Q}_1(z)$ is the iterated conditional $\bar{\tau}_1$-quantile of $Y_1$ given $Y_2, \ldots, Y_M$ in which each $Y_i$, $i > 1$, is evaluated at its iterated conditional $\bar{\tau}_i$-quantile given $Y_{i+1}, \ldots, Y_M$ and $Z$.

Define the difference in the iterated conditional quantile function of $Y_1$

$$\bar{\Delta}_{Q_1}(z', z'') \equiv \bar{Q}_1(z') - \bar{Q}_1(z'').$$

Then an argument similar to that employed in Section 4 leads to the result that the model defined by the modified restrictions uniformly identifies $\bar{\Delta}_{h_1, Y_j}(z', z'', z^*)$ for any $\{z', z'', z^*\} \in \bar{V}$, and $\bar{\Delta}_{Q_1}(z', z'')$ delivers the value of this partial difference. A rank condition: $\bar{y}_j(z') \neq \bar{y}_j(z'')$ ensures the identified partial difference compares values of the structural function at distinct values of its arguments.

## 6. Illustration

To conclude some of the data used in the analysis of Angrist and Krueger (1991) (AK) is used to illustrate issues arising from the results given earlier. AK employed discrete quarter-of-birth instruments ($Z$) to estimate the impact of schooling ($Y_2$) on log wages ($Y_1$) using a parametric linear model with structural equations:

$$Y_1 = \theta_0 + \theta_1 Y_2 + \varepsilon_1,$$

$$Y_2 = \sum_{i=1}^{4} \gamma_i Z_i + \varepsilon_2,$$

where $Z_i = 1$ signifies birth in quarter $i$. Their model is completed by the following covariation restrictions:

$$E[\varepsilon_1 | Z_1, \ldots, Z_4] = E[\varepsilon_2 | Z_1, \ldots, Z_4] = 0.$$

The steps in a nonparametric analysis using the identification correspondences given earlier are now set out. A nonseparable model is employed

$$Y_1 = h_1(Y_2, \varepsilon_1, \varepsilon_2),$$

$$Y_2 = h_2(Z_1, \ldots, Z_4, \varepsilon_2)$$

with Restrictions 1–4 of Section 4 imposed.

The data employed are 329,509 records of log wages and years of schooling for a sample of the 1930–1939 birth cohort of US residents drawn and observed in 1980.[12] The schooling data are integer-valued. Purely for the purpose of illustrating the results given earlier the data are rendered "continuous" by adding noise, uniformly distributed on $(-0.5, 0.5)$.[13]

The quarter of birth instruments have a weak association with achieved amounts of schooling. Following the critique in Bound et al. (1995), there was much discussion of and research on the "many weak instrument" problem they identified.[14] This weak association is revealed in Table 1 which shows estimates of $\bar{\tau}_2$-quantiles of schooling for each quarter

[12]There is a full description of the data in Angrist and Krueger (1991). The data are also used in Angrist et al. (1999) and are available at the Journal of Applied Econometrics Data Archive ⟨http://qed.econ.queensu.ca/jae/1999-v14.1/angrist-imbens-krueger/⟩.

[13]Chesher (2005) studies the identifying power of nonparametric models using the original discrete years of schooling data.

[14]In some of the AK analysis quarter-of-birth was interacted with US state-of-birth indicators resulting in over 200 instrumental variables which in combination had a rather small association with achieved amounts of schooling.

Table 1
Estimated quantiles of schooling by quarter of birth ($z$)

| $\bar{\tau}_2 =$ | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 |
|---|---|---|---|---|---|
| $z=1$ | 8.42 | 10.58 | 11.66 | 11.93 | 12.20 |
| $z=2$ | 8.48 | 10.64 | 11.67 | 11.95 | 12.22 |
| $z=3$ | 8.66 | 10.95 | 11.71 | 11.95 | 12.22 |
| $z=4$ | 8.75 | 11.06 | 11.71 | 11.98 | 12.25 |

Table 2
Estimated returns to schooling by quantiles of schooling using alternative quarter of birth ($z$) comparisons

| $\bar{\tau}_2 =$ | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 |
|---|---|---|---|---|---|
| $z=1$ | 0.065 | 0.012 | −0.508 | −0.210 | −0.176 |
| vs | (0.121) | (0.090) | (0.393) | (0.218) | (0.228) |
| $z=2$ | | | | | |
| $z=1$ | 0.070 | 0.045 | 0.048 | 0.064 | 0.060 |
| vs | (0.030) | (0.014) | (0.095) | (0.106) | (0.111) |
| $z=3$ | | | | | |
| $z=1$ | 0.065 | 0.060 | 0.083 | 0.068 | 0.043 |
| vs | (0.022) | (0.011) | (0.079) | (0.083) | (0.089) |
| $z=4$ | | | | | |

Estimated standard errors in parentheses.

of birth and $\bar{\tau}_2 \in \{0.1, 0.2, 0.3, 0.4, 0.5\}$. For $\bar{\tau}_2 > 0.2$ the differences in schooling quantiles across quarter of birth are less than 0.05 years (18 days). For smaller values of $\bar{\tau}_2$ the differences are of the order of 0.3–0.5 years for quarter 1–4 comparisons. The entries in Table 1 are estimates of the objects $Q_{Y_2|Z}(\bar{\tau}_2|z)$ which identify the values of schooling $\bar{y}_2(z)$ that feature in Theorems 1 and 2. The data suggest that the "rank condition" holds with any strength only at low values of $\bar{\tau}_2$ and only for comparison of people born in quarters 1 or 2 with people born in quarters 3 or 4.

Table 2 shows nonparametric estimates of the returns to schooling with $\bar{\tau}_1 = 0.5$ and $\bar{\tau}_2 \in \{0.1, 0.2, 0.3, 0.4, 0.5\}$ (varying across columns) got by comparing, in turn, people born in quarters 2–4 (varying across rows) with people born in quarter 1. Each entry is of the form:

$$\frac{\hat{Q}_{Y_1|Y_2Z}(\bar{\tau}_1|\hat{Q}_{Y_2|Z}(\bar{\tau}_2|z=1), z=1) - \hat{Q}_{Y_1|Y_2Z}(\bar{\tau}_1|\hat{Q}_{Y_2|Z}(\bar{\tau}_2|z=j), z=j)}{\hat{Q}_{Y_2|Z}(\bar{\tau}_2|z=1) - \hat{Q}_{Y_2|Z}(\bar{\tau}_2|z=j)},$$

where $z=1$ signifies birth in quarter 1 and $z=j$ indicates birth in quarter $j \in \{2,3,4\}$.[15] This estimator follows directly from the identifying correspondence of Theorem 2.

The nonparametric estimates of the conditional log wage quantiles for $z = i \in \{1,2,3,4\}$ are obtained by locally linear quantile regression of log wage on schooling using data for

[15] The similarity of this estimator to the Wald (1940) estimator is striking. The Wald estimator, which could be justified in a linear parametric model, would employ *average* log wages at *average* values of schooling by quarter of birth.

people born in quarter $i$ and Gaussian weights centered at the values of schooling given by the appropriate values of $\hat{Q}_{Y_2|Z}(\bar{\tau}_2|z=i)$.[16]

The estimates of returns to schooling are inaccurate except at low quantiles of schooling ($\bar{\tau}_2 \leqslant 0.2$) and for comparisons of quarter of birth 1 with quarters of birth 3 and 4. The low estimated values of returns to schooling in these cases are not inconsistent with the values reported by AK.[17]

The following points emerge from this illustrative nonparametric analysis:

1. Binary instruments place limitations on the characteristics of a structural function which can be nonparametrically identified. Only partial differences for particular pairs of values of an endogenous variable can be identified.
2. When binary instruments are weak the pairs of values of an endogenous variable at which partial differences can be identified span a small range of values. Parametric restrictions allow: (i) extrapolation of estimates of structural function sensitivity from estimates obtained using variation over small and specific ranges of values of an endogenous variable and (ii) combination of these values which, under linear parametric restrictions, all bear on the same structural feature, $\theta_1$.
3. The nonseparable model employed here allows the possibility that an endogenous variable only has an effect on outcomes at particular quantiles of its distribution. This seems quite appropriate in the problem studied here. It is clear from the discussion in AK that the quarter of birth instruments are expected to have a significant effect on achieved schooling only for those who choose to leave school at the earliest opportunity. These are people who at any value of $Z$ lie in the lower tail of the schooling distribution. The analysis of this paper shows how such people can be "isolated" in order to gain understanding of the effect of schooling on wages. It would be incorrect to conduct comparisons based on particular values of achieved schooling—that would involve selection on the basis of the value of an endogenous variable. The analysis of this paper shows that selection can be based on *quantiles* of achieved schooling.

**Acknowlegdements**

I am grateful to Jaap Abbring, James Heckman, Hide Ichimura, Roger Koenker, Valérie Lechene, Chuck Manski, Lars Nesheim, Richard Spady and Elie Tamer for helpful comments and discussion and to two referees for helpful comments. This paper builds on Chesher (2002) which formed the basis of a plenary lecture at the European Meeting of the Econometric Society, Venice 2002. I thank the discussant, Whitney Newey, for his insightful comments. I am grateful for generous support from the Leverhulme Trust through their funding of the Centre for Microdata Methods and Practice and the research project "Evidence, Inference and Enquiry".

[16]Estimates were computed using the quantreg package of R (Ihaka and Gentleman, 1996).

[17]It would be unwise to place great weight on the results of this illustrative nonparametric analysis. To allow a fully nonparametric attack the wage equation has been restricted to a simple form with, for example, no control for within-cohort variation in year of birth. A referee notes that estimated returns could be higher were there to be control for such covariates.

## References


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

Angrist, J.D., Imbens, G.W., Krueger, A.B., 1999. Jackknife instrumental variables estimation. Journal of Applied Econometrics 14, 57–67.

Benkhard, C.L., Berry, S., 2006. On the nonparametric identification of nonlinear simultaneous equations models: comment on B. Brown (1983) and Roehrig (1988). Econometrica, forthcoming.

Bound, J., Jaeger, D.A., Baker, R.M., 1995. Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. Journal of the American Statistical Association 90, 443–450.

Chaudhuri, P., 1991. Nonparametric estimation of regression quantiles and their local Bahadur representation. Annals of Statistics 19, 760–777.

Chaudhuri, P., Doksum, K., Samarov, A., 1997. On average derivative quantile regression. Annals of Statistics 25, 715–744.

Chesher, A.D., 2002. Instrumental values. Centre for Microdata Methods and Practice CEMMAP, Working Paper CWP17/02. Invited address at the 2002 European Meeting of the Econometric Society, Venice, under the title: Local identification in nonseparable models.

Chesher, A.D., 2003. Identification in nonseparable models. Econometrica 71, 1404–1441.

Chesher, A.D., 2005. Nonparametric identification under discrete variation. Econometrica 73, 1525–1550.

Fisher, F.M., 1959. Generalization of the rank and order conditions for identifiability. Econometrica 27, 431–447.

Fisher, F.M., 1961. Identifiability criteria in nonlinear systems. Econometrica 29, 574–590.

Fisher, F.M., 1966. The Identification Problem in Econometrics. McGraw-Hill, New York.

Frisch, R.A.K., 1934. Statistical confluence analysis by means of complete regression systems. Publication No. 5, Universitets Økonomiske Institutt, Oslo.

Frisch, R.A.K., 1938. Statistical versus theoretical relations in economic macrodynamics. Memorandum prepared for a conference in Cambridge, England, July 18–20, 1938, to discuss drafts of Tinbergen's League of Nations publications; mimeographed.

Hurwicz, L., 1950. Generalization of the concept of identification. In: Koopmans, T.C. (Ed.), Statistical Inference in Dynamic Economic Models. Cowles Commission Monograph 10. Wiley, New York.

Ihaka, R., Gentleman, R., 1996. R: A language for data analysis and graphics. Journal of Computational and Graphical Statistics 5, 299–314.

Imbens, G.W., Newey, W.K., 2003. Identification and estimation of triangular simultaneous equations models without additivity. Manuscript, Presented at the 14th $EC^2$ Meeting, December 12–13, 2003, London.

Kahn, S., 2001. Two-stage rank estimation of quantile index models. Journal of Econometrics 100, 319–355.

Koenker, R.W., 2005. Quantile Regression, Econometric Society Monograph Series. Cambridge University Press, New York.

Koenker, R.W., Bassett Jr., G.W., 1978. Regression quantiles. Econometrica 46, 33–50.

Koenker, R.W., d'Orey, V., 1987. Computing regression quantiles. Journal of the Royal Statistical Society Series C 36, 383–393.

Koopmans, T.C., Reiersøl, O., 1950. The identification of structural characteristics. Annals of Mathematical Statistics 21, 165–181.

Koopmans, T.C., Rubin, H., Leipnik, R.B., 1950. Measuring the equation systems of dynamic economics. In: Koopmans, T.C. (Ed.), Statistical Inference in Dynamic Economic Models. Cowles Commission Monograph 10. Wiley, New York.

Lee, S., 2003. Efficient semiparametric estimation of a partially linear quantile regression model. Econometric Theory 19, 1–31.

Manski, C.F., 1988. Analog Estimation Methods in Econometrics. Chapman & Hall, New York.

Matzkin, R.L., 2003. Nonparametric estimation of nonadditive random functions. Econometrica 71, 1339–1376.

Matzkin, R.L., 2005. Identification in nonparametric simultaneous equations. Manuscript, Northwestern University.

Morgan, M.S., 1990. The History of Econometric Ideas. Cambridge University Press, Cambridge, UK.

Newey, W.K., Powell, J.L., Vella, F., 1999. Nonparametric estimation of triangular simultaneous equations models. Econometrica 67, 565–603.

Roehrig, C.S., 1988. Conditions for identification in nonparametric and parametric models. Econometrica 56, 433–447.

Tinbergen, J., 1930. Bestimmung und Deutung von Angebotskurven: Ein Beispiel. Zeitschrift für Nationalökonomie 1, 669–679.

Wald, A., 1940. The fitting of straight lines if both variables are subject to error. Annals of Mathematical Statistics 11, 284–300.

Wegge, L.L., 1965. Identifiability criteria for a system of equations as a whole. The Australian Journal of Statistics 7, 67–77.

Working, E.J., 1927. What do statistical 'demand curves' show? Quarterly Journal of Economics 41, 212–235.

Working, H., 1925. The statistical determination of demand curves. Quarterly Journal of Economics 39, 503–543.