INSTRUMENTAL VALUES

Andrew Chesher

THE INSTITUTE FOR FISCAL STUDIES
DEPARTMENT OF ECONOMICS, UCL
cemmap working paper CWP17/02

Instrumental Values

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

August 10th 2002, revised September 17th 2002†

Abstract. This paper studies the identification of partial differences of nonseparable structural functions. The paper considers triangular structures with no more stochastic unobservables than observable outcomes, that exhibit a degree of monotonicity with respect to variation in certain stochastic unobservables. It is shown that, the existence of a set of instrumental values of covariates, over which the stochastic unobservables exhibit local quantile invariance and over which a local order condition holds, defines a model which identifies certain partial differences of structural functions. This result is useful when covariates exhibit discrete variation. The paper also considers the identification of partial derivatives in smooth structures when covariates exhibit continuous variation.

1. Introduction

1.1. Nonseparable structures. Structures in which stochastic unobservables are nonseparable are of interest because they are capable of representing a very wide class of social and economic processes. Further, they allow responses to changes in conditioning variables[1] to exhibit stochastic variation which may be significant in the analysis of the social and economic behaviour of individuals.

Economic theory rarely tells us where the sources of stochastic variation appear in economic models, so conservative analysis of the econometric issues that arise when models are brought to data allows for the possibility that structures are nonseparable.

This paper explores the limits of identification of characteristics of nonseparable structures[2] using a construction which allows for the possibility that covariates and outcomes may exhibit discrete variation. One aim of this paper is to understand the limitations that discrete variation may place on the class of structural characteristics that can be identified by a model.

*I am grateful to Jaap Abbring, Lars Nesheim and Hide Ichimura for helpful comments at a cemmap workshop on July 30th 2002, and to Valérie Lechene and Richard Spady for helpful discussions. This paper formed the basis of an invited address at the 2002 European Econometric Society Meeting held in Venice. I am grateful to the discussant, Whitney Newey, for his insightful comments.

†This revision corrects an error in previous versions in which it was stated that the results of the Theorem of the paper apply when outcomes have discrete distributions of general form conditional on covariate values. In fact the Theorem allows for only rather special types of discrete distributions for outcomes which appear as arguments of structural functions if partial differences of structural functions are to be identifiable under the weak conditions considered in this paper. Section 5.3 contains a discussion of this issue. This revision also corrects the discussion of large structural systems in Section 5.10.

[1] By "conditioning variables", which I shall refer to as "covariates", I mean variables whose values may determine the distribution of a random variable - a distribution I will refer to as a "conditional distribution". The covariates discussed in this paper may not be random variables in the sense that they may not have well defined probability distributions - for example they may be values selected by an experimenter. A consequence of this is that the "distribution" of covariates will convey no information about structures. Of course values taken by covariates may contain information.

[2] Structures, characteristics of structures, models and identification of structural characteristics are defined as in Hurwicz (1950) and Koopmans and Reiersol (1950). Definitions are given in Section 3.

1.2. Discrete variation. Covariates frequently show discrete variation. Covariates may be realisations of discrete random variables such as binary indicators, for example measuring labour force participation, or of integer valued random variables, for example years of schooling, or, because of the granularity of the observation process, show discrete variation in practice, even though continuous variation is possible in principle.

This paper focuses on identification of members of a particular class of structural characteristics, namely values of partial *differences* of structural functions, that is differences obtained when all arguments but one are held constant and the remaining argument takes two distinct values. These structural characteristics are, unlike say derivatives of structural equations, characteristics which could feasibly be identified in the absence of parametric restrictions when covariates exhibit discrete variation. The analysis of this paper permits discrete variation in covariates but the results also apply when there is continuous variation and limiting arguments allow the study of identifiability of partial *derivatives* of structural functions and so a link to the results in Chesher (2001a, 2001b, 2001c, 2002).

Of course conditions must be placed on structures if a structural characteristic is to be identified. The strategy taken in this paper is to seek *weak* identifying conditions. Since the interpretation of all econometric analysis is contingent upon identifiability and identifiability necessarily rests on some untestable restrictions, it is prudent to base identifiability on the weakest possible restrictions.

Weak restrictions may lead to identifiability of only a limited class of structural characteristics and it may be that none of its members is of interest in practice. In this circumstance one may wish to impose further restrictions which lead to identification of interesting characteristics. The impact of additional (for example parametric) restrictions can be examined using the construction developed here.

1.3. Local identification. This paper studies the identifiability of *local* characteristics of structures and there is no attempt to develop conditions under which, say, a complete structural function is identified. This approach is taken because, when there is discrete variation in covariates, and in the absence of parametric restrictions, data may only be informative about local characteristics of structures, for example, the partial difference of a structural function when its arguments are set to particular values.

If 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.

A significant advantage of a focus on identification of local characteristics of structures is that, as shown in Chesher (2001b), restrictions placed on structures to achieve local identification need only be locally valid. For example, to identify a partial difference of a nonseparable function over some interval one need not restrict attention to structures in which stochastic unobservables and covariates are statistically independent[3]. Identification can be achieved if there is dependence but it is limited in extent *at the values of arguments* of the structural function at which knowledge of the value of the structural feature is desired.

Global validity of local identification restrictions may lead to identification of global characteristics of structures, a possibility that can be examined using the construction developed in this paper.

[3]A restriction commonly imposed in the study of identification when structures may be nonseparable.

1.4. Partial differences. Classical identification conditions impose a degree of independence[4] on the variations in stochastic unobservables and covariates and an "order" condition[5] which limits the covariate driven variation in structural functions.

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 ensures that an identfiable partial difference is non-zero.

1.5. Quantiles. A key to progress in the study of identification when structures may be nonseparable is understanding that an analysis that proceeds in terms of *conditional quantile functions* is extremely well suited to the nature of the problem considered.

For $\tau \in (0,1)$ the τ-quantile of a scalar random variable, A, with distribution function F_A is defined as follows[6],

$$Q_A(\tau) = \inf\{q \in \Re | F_A(q) \geq \tau\}$$

and note that such quantiles are equivariant with respect to monotone transformations, that is, if h is a non-decreasing function on $\Re$ then

$$Q_{h(A)}(\tau) = h(Q_A(\tau)).$$

This τ-quantile is well defined whenever A has a proper distribution function, including cases in which A is a discrete random variable and the equivariance property applies in such cases.

The conditional τ-quantile of A given a vector of covariates $B = b$ is analogously defined as

$$Q_{A|B}(\tau, b) = \inf\{q \in \Re | F_{A|B}(q, b) \geq \tau\}$$

where $F_{A|B}$ is the conditional distribution function[7] of A given $B = b$, and the equivariance property

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

applies for all b for which $h(a, b)$ is a nondecreasing function of a.

Because of this equivariance property, restrictions imposed on the covariate driven variation of conditional quantiles of a stochastic unobservable given covariates can be "passed through" a structural function as long as the function is restricted to exhibit a degree of monotonic variation with respect to the unobservable.

That sort of monotonicity restriction is an essential element in the restrictions that define the identifying models of this paper.

1.6. Multiplicity of stochastic unobservables. Another essential element in the restrictions considered in this paper is that the number of unobservables (R) should be *no greater than* the number of observable outcomes (M). This does allow the possibility that a structure involves more than M unobservables, but for the purpose of this paper

[4] For example, conditional mean independence, full independence, or, as considered in this paper, conditional quantile invariance.

[5] In the language of Koopmans, Rubin and Leipnik (1950).

[6] The distribution function is defined as: $F_A(a) = P[A \leq a]$.

[7] That is $F_{A|B}(a, b) = P[A \leq a | B = b]$.

a model contains only such structures in which unobservables coalesce to produce M or fewer unobservables, no more than the number of observable outcomes.

In a nonparametric attack under weak restrictions, imposing this condition is essential, a point easily seen when we consider that otherwise one would be seeking knowledge of characteristics of a structure generated by $R > M$ stochastic unobservables from information contained in a $M < R$ dimensional distribution function for outcomes given covariates.

There are of course many econometric models used in practice that do not embody such a restriction, for example the mixed proportionate hazard models popular in the analysis of durations, measurement error models and models for panel data that incorporate "individual effects". It is notable that in all cases models of this sort gain identifying power from strong restrictions, which usually require additivity (in some specified metric) at key points in admissible structural functions[8].

1.7. Restrictions and instrumental values. Two types of restrictions define the models considered in this paper.

First there are restrictions on admissible structural equations, specifically that they have a triangular structure, that they exhibit monotonic variation with respect to certain unobservables and that there is a specific sort of variation in the values delivered by the structural functions as covariate values vary.

A two equation example of the sort of triangular structure considered here is the following[9].

$$\begin{aligned} Y_1 &= h_1(Y_2, X, \varepsilon_1, \varepsilon_2) \\ Y_2 &= h_2(X, \varepsilon_2) \end{aligned}$$

The monotonicity requirement is that h_1 be non-increasing or non-decreasing in ε_1 and h_2 be strictly increasing or strictly decreasing in ε_2. These conditions can be weakened, for example to "single crossing" conditions as set out in Chesher (2002). The functions h_1 and h_2 are normalised to be respectively non-decreasing and strictly increasing[10].

The second type of restriction limits the variation in the conditional distribution of the unobservables (ε_1 and ε_2) as covariate values (X) vary. Specifically if identification of a structural feature at τ_1- and τ_2- quantiles of respectively ε_1 and ε_2 is required then restrictions are placed on the dependence of those conditional quantiles given $X = x$ as x varies.

The restrictions on X-driven variation in conditional quantiles of unobservables and on X-driven variation in structural functions are both required to hold for variations in X confined to a set of *instrumental values*. This set may be non-denumerable, but in cases in which there is discrete variation it may be denumerable.

In Section 4 a theorem is stated and proved which defines a model such that, for two equation admissible structures as set out above, if x' and x'' belong to a set of instrumental values, $V^* \subseteq \Re^K$, then the partial difference:

$$h_1(y_2', x^*, e_1^*, e_2^*) - h_1(y_2'', x^*, e_1^*, e_2^*) \tag{1}$$

[8] For example measurement error models typically have measurement error additive in some specified metric, mixed proportionate hazard models typically have the unobserved heterogeneity term additive in the log hazard function, panel data models typically have individual effects additive with the other unobservables.

[9] Note that ε_2 need not be present in h_1 and that ε_1 and ε_2 can be jointly dependent and dependent upon X.

[10] The requirement that h_2 be *strictly* monotonic with respect to variation in ε_2 is significant, restricting the stochastic variation in Y_2 (which note appears in the function h_1) to take place across a support which has a one-to-one correspondence with the support of ε_2. The issue is addressed further in Section 5.3.

is identifiable. Here e_2^* is the τ_2^*-quantile of the conditional distribution of ε_2 given X, e_1^* is the τ_1^*-quantile of the conditional distribution of ε_1 given ε_2 and X, y_2' and y_2'' are respectively $h_2(x', \varepsilon_2^*)$ and $h_2(x'', \varepsilon_2^*)$ and x^* is any value of X in the set of instrumental values.

The set of instrumental values is required to be such that:

1. the τ_2^*-quantile of the conditional distribution of ε_2 given $X = x$, e_2^*, and the τ_1^*-quantile of the conditional distribution of ε_1 given $\varepsilon_2 = e_2^*$ and $X = x$, e_1^*, are invariant with respect to variation in x within the set of instrumental values, and,

2. for any x' and x'' in the set of instrumental values,

$$h_1(y_2', x', e_1^*, e_2^*) = h_1(y_2', x'', e_1^*, e_2^*)$$

   which is in the nature of an order condition.

The membership of the set of instrumental values may depend upon the probabilities, τ_1^* and τ_2^*, which define the conditional quantiles of ε_1 and ε_2. The "$*$" in "V^*" is intended to indicate this dependence. A more expansive, but cumbersome, notation would write a set of instrumental values associated with τ_1^* and τ_2^* as $V^{(\tau_1^*, \tau_2^*)}$. The membership of the set of instrumental values may also depend upon the partial difference whose identification is sought.

It is shown in Section 4 that in this two equation problem:

1. the partial difference (1) is uniformly identified[11] by a model embodying restrictions as set out above, restrictions made precise in the statement of a Theorem set out in Section 4, and,

2. all structures in which the partial difference (1) takes a particular value, say a, generate conditional distributions for Y given X such that the difference in conditional quantiles:

$$Q_{Y_1|Y_2X}(\tau_1^*, Q_{Y_2|X}(\tau_2^*, x'), x') - Q_{Y_1|Y_2X}(\tau_1^*, Q_{Y_2|X}(\tau_2^*, x''), x'') \tag{2}$$

   takes the same value, a. Here $Q_{Y_1|Y_2X}$ and $Q_{Y_2|X}$ are conditional quantile functions of respectively Y_1 given Y_2 and X, and of Y_2 given X.

The analog principle[12] suggests an estimator of the value of the partial difference, namely the expression (2) applied to estimates of these conditional quantile functions.

1.8. Plan of this paper. Section 2 briefly reviews the 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 states and proves a Theorem which defines a model that identifies values of, and partial differences of, structural functions in two equation systems. The model embodies restrictions of the sort described above.

The ten sub-sections of Section 5 consider the following issues.

1. The requirement that a rank condition hold (that is $y_2' \neq y_2''$ in the example above) if the identification result is to be useful, the impact that discrete variation in covariates has on the information that data can provide about differences of structural functions, and the utilty of parametric restrictions when covariates show discrete variation.

[11] In the sense of Koopmans and Reiersol (1950), see Section 3.
[12] Manski (1988).

2. The impact of weak instruments on the ability to identify interesting partial differences.
3. Discrete variation in outcomes.
4. Analog estimation of partial differences.
5. The way in which a classical analysis *via* instrumental variables is subsumed in the analysis of this paper.
6. The concept of overidentification in this analysis *via* instrumental values and the use of overidentifying restrictions in analog estimation.
7. Smooth structures and the identification of partial *derivatives* of structural functions.
8. Identification of partial differences with respect to *covariates*.
9. Identification of partial differences with respect to *stochastic unobservables*.
10. Identification of partial differences in M equation structures.

Section 6 concludes.

2. Related literature

The study of identification has a long history with early contributions by Working (1925, 1927), and Frisch (1934, 1938) and with notable developments by, among others, Haavelmo (1944), Hurwicz (1950), Koopmans and Reiersol (1950), Koopmans, Rubin and Leipnik (1950), Wald (1950), Fisher (1959, 1961, 1966), Wegge (1965) and Rothenberg (1971). One product of this research was the order and rank conditions in linear models, local versions of which feature in the results of this paper.

Most of this work was cast in the context of *parametric* models although Koopmans and Reiersol (1950) understood that identification could be achieved with less restricted models and the definitions provided by Hurwicz (1950), adopted by Koopmans and Reiersol (1950), and used in this paper, were designed to apply in the consideration of identification in the absence of parametric restrictions.

Until the early 1970's 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 1970's 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 an interesting element. Economic theory provides little guidance concerning the precise forms of nonlinear structural equations and so interest in *nonparametric* identification, that is identification of structural characteristics in the absence of parametric restrictions, was rekindled.

Charles Roehrig (1988), extending the work of Brown (1983), re-stimulated interest in *nonparametric* identification. Roehrig (1988) is concerned with global identification of structural functions for models in which stochastic unobservables are restricted to be statistically independent of covariates. Most of the discussion in Roehrig (1988) is for the case in which the stochastic unobservables are separable, appearing additively in the structural equations of the model. Newey and Powell (1988), Newey, Powell and Vella (1999), Pinkse (2000), Darolles, Florens and Renault (2000) study identification using such models with stochastic unobservables which satisfy mean independence conditions of various types.

There has been recent interest in determining when global identification can be achieved in structures with *nonseparable* disturbances. Brown and Matzkin (1996) consider the identification of nonparametric primitive nonseparable structural functions (for example production or utility functions) under the restriction that stochastic unobservables and covariates are independently distributed. Altonji and Matzkin (2001) study panel data structures restricted by conditional exchangeability conditions. Imbens and Newey (2001) study triangular, nonseparable structures similar to those addressed in this paper[13]. They determine conditions under which there is global identification of structural functions when stochastic unobservables are restricted to be statistically independent of covariates, and they develop an ingenious estimator of a structural function and provide conditions under which it is consistent.

Statistical independence of stochastic unobservable and covariates is a very strong condition which we might not expect to hold in practice, particularly when working with microdata, which may exhibit heteroskedastic variation.

One of the aims of this paper is to develop identification conditions which do not require full statistical independence of unobservables and covariates while still allowing identification of pertinent structural characteristics. This is achieved by placing restrictions on the covariate-driven variation of conditional *quantiles* of the unobservables. Such restrictions can be tailored to suit the case under study. For example where heteroskedastic variation is considered likely one might only be prepared to place restrictions on covariate-driven variation in conditional *medians*, allowing other conditional quantiles to depend upon the values of covariates.

One of the few papers to consider identification from a conditional quantile perspective is Matzkin (1999) which considers a model in which a structural function takes the form $Y = m(X, \varepsilon)$ with ε distributed independently of X and $m(X, \varepsilon)$ strictly monotonic in ε. Conditions under which the function $m(\cdot, \cdot)$ and the distribution function of ε are identifiable are obtained. The value of $m(\cdot, \cdot)$ at a point (x, e) is shown, under suitable conditions, to be identifiable as the value of the conditional τ-quantile of Y given $X = x$ where τ is such that e is the τ-quantile of the marginal distribution of ε.

There is a large recent literature concerning the identifying power of treatment effect models[14]. The structures admitted in these models have two potential outcomes only one of which is observed depending on whether a treatment is assigned or not. These structures contain more sources of stochastic variation than observable stochastic outcomes and so the analysis of this paper is not applicable.

The identification conditions of this paper include local quantile independence restrictions. A number of papers have used quantile independence restrictions as the basis for developing estimators including Amemiya (1982), Powell (1983), Newey and Powell (1990), Chaudhuri, Doksum and Samarov (1997), Kahn (2001) and Chernozhukov and Hansen (2001).

This paper extends the research reported in Chesher (2001a, 2001b, 2001c, 2002) to problems in which there is discrete variation in covariates or outcomes. The results of this paper can be specialised to yield those given in the earlier papers, as indicated in Section 5.7. As is often the case, viewing a problem, as here, from a more general standpoint creates great simplification, so the results of this paper shed light on the results contained

[13]In the Imbens-Newey model each structural equation contains exactly one stochastic unobservable. In the model of this paper more than one stochastic unobservable may appear in a structural equation as long as the unobservables appear in triangular form.

[14]See for example many contributions by James Heckman including Heckman (1990) and Heckman, Smith and Clements (1997), and Heckman and Vytlacil (2001) and the papers referenced therein, and Imbens and Angrist (1994), Das (2000), Chernozhukov and Hansen (2001), Abadie, Angrist and Imbens (2002), and Vytlacil (2002).

in these earlier papers, and more generally on a number of results in the extensive literature on identification of structures and their characteristics.

3. STRUCTURES, MODELS AND IDENTIFICATION

This Section makes precise the definitions of various concepts used in this paper 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:

1. a system of equations delivering a value of a vector outcome, $Y = \{Y_m\}_{m=1}^M$ given a value of a vector covariate, $X = \{X_k\}_{k=1}^K$ and a value of a vector of unobservable random variables, $\varepsilon = \{\varepsilon_r\}_{r=1}^R$, and,
2. a conditional distribution function, $F_{\varepsilon|X}$ for the unobservables given the covariates,
3. such that, the conditional distribution function of outcomes given covariates, $F_{Y|X}$ is well defined.

Note that the definition of a particular structure requires a complete (i.e. numerical) specification of a system of equations and a conditional distribution $F_{\varepsilon|X}$.

A *structural characteristic*[15] is a functional $\theta(S)$ of a structure, S, for example the value of a partial derivative of a structural function at a given point or of a partial difference calculated at a given pair of points. Data are generated by some structure, we know not which, and we wish to discover the value of a characteristic of the data generating structure. Many structures with different values of a structural characteristic may generate identical conditional distribution functions, $F_{Y|X}$. Structures which generate the same conditional distribution function for Y given X are said to be *observationally equivalent.*

Data generated by a structure are informative about $F_{Y|X}$, but cannot alone distinguish one observationally equivalent structure from another. If the value of a structural characteristic varies within observationally equivalent structures then that value cannot be identfied. So, in order to identify the value of a structural characteristic the class of admissible structures must be restricted so that there is no variation in the value of the characteristic within observationally equivalent structures.

The term "*model*" is used to describe a set of restrictions defining admissible structures. A model is a proper subset of the class of all structures, for example all structures in which the equations are restricted to be linear and $F_{\varepsilon|X}$ is multivariate normal independent of X.

A model *identifies* a characteristic, $\theta(S)$ in a structure S_0 if that characteristic is the same in all structures which are admitted by the model and observationally equivalent to S_0 (Koopmans and Reiersol (1950)). A 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 admissible structures such that $\theta(S) = a$ and let A be the set of all values of $\theta(S)$ generated by admissible structures. Let $F^S_{Y|X}$ denote the conditional distribution function generated by a structure S. Suppose there exists a functional of the conditional distribution function of Y given X, $\mathcal{G}(F_{Y|X})$, such that for each $a \in A$, $\mathcal{G}(F^S_{Y|X}) = a$ for all $S \in S^a$. Then $\theta(S)$ is uniformly identified by the model.

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

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

In practice it may not be possible to find a functional of $F_{Y|X}$ with the required property even though the structural characteristic is uniformly identified. If, for some model, 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 using $\hat{\theta}(S) = \mathcal{G}(\hat{F}_{Y|X}^S)$.

4. A two equation model

This Section considers two equation structures and states and proves a Theorem which provides restrictions on structures, that is *defines a model*, under which certain partial differences of a structural function are uniformly identified. Issues arising from the result of the theorem and extension to structures with more than two equations are discussed in Section 5.

The conditions of the Theorem require structural equations to be triangular, complete, and to exhibit a degree of monotonicity with respect to variation in certain stochastic unobservables. From these conditions alone, two of the four results of the Theorem follow, namely that values delivered by the two structural functions at a point of interest are identifiable. From this we can immediately conclude that certain differences of structural functions are identifiable under these three conditions.

The remaining condition of the Theorem ensures that certain identifiable differences of structural functions are *partial* differences, a conclusion expressed in the final two results of the Theorem. This condition posits the existence of a set of instrumental values of the covariates such that variation in covariate values within this set:

1. results in no change in τ_1^*- and τ_2^*- conditional quantiles of ε_1 and ε_2, and,
2. results in no variation in the values delivered by the structural function h_1 through its X argument.

The Theorem is now stated and proved.

Theorem

Let Y_1 and Y_2 be scalar random variables, let $X = \{X_k\}_{k=1}^K$ be a list of covariates and let ε_1 and ε_2 be unobservable scalar random variables.

Let $Q_{\varepsilon_2|X}(\tau, x)$ denote the conditional τ-quantile of ε_2 given $X = x$, and let $Q_{\varepsilon_1|\varepsilon_2 X}(\tau, e_2, x)$ denote the conditional τ-quantile of ε_1 given $\varepsilon_2 = e_2$ and $X = x$.

Consider $\{\tau_i^*\}_{i=1}^2 \in (0,1) \times (0,1)$, and a set of *instrumental values* $V^* \subseteq \Re^K$ of the conditioning variables whose membership may depend upon the value of τ^*. Define

$$\begin{aligned} e_2^*(x) &\equiv Q_{\varepsilon_2|X}(\tau_2^*, x) \\ e_1^*(x) &\equiv Q_{\varepsilon_1|\varepsilon_2 X}(\tau_1^*, e_2^*(x), x). \end{aligned}$$

There are the following assumptions.

A1 **Triangularity**. Y_1 and Y_2 are determined by the following structural equations.

$$Y_1 = h_1(Y_2, X, \varepsilon_1, \varepsilon_2) \quad (3)$$
$$Y_2 = h_2(X, \varepsilon_2) \quad (4)$$

A2 **Completeness**. For each $x \in V^*$, the equations

$$Y_1 = h_1(Y_2, x, e_1^*(x), e_2^*(x))$$
$$Y_2 = h_2(x, e_2^*(x))$$

have a unique solution for Y_1 and Y_2, denoted by $y_1^*(x)$ and $y_2^*(x)$.

A3 **Monotonicity**.

(a) For all $x \in V^*$, the function $h_2(x, \varepsilon_2)$ is strictly monotonic (either decreasing for all $x \in V^*$ or increasing for all $x \in V^*$) with respect to variation in ε_2. Normalise h_2 to be increasing with respect to variation in ε_2.

(b) For all $x \in V^*$ and, at $Y_2 = y_2^*(x)$, $\varepsilon_2 = e_2^*(x)$ the function $h_1(y_2^*(x), x, \varepsilon_1, e_2^*(x))$ is weakly monotonic (either non-decreasing for all $x \in V^*$ or non-increasing for all $x \in V^*$) with respect to variation in ε_1. Normalise h_1 to be non-decreasing with respect to variation in ε_1.

A4 **Quantile invariance**. For all $\{x', x''\} \in V^*$

$$e_1^*(x') = e_1^*(x'')$$
$$e_2^*(x') = e_2^*(x'').$$

Denote the common values by e_1^* and e_2^*.

A5 **Order condition**. For all $\{x', x''\} \in V^*$

$$h_1(y_2^*(x'), x', e_1^*, e_2^*) = h_1(y_2^*(x'), x'', e_1^*, e_2^*)$$

Consider $\{x', x''\} \in V^*$ and define

$$\triangle_Q^*(x', x'') \equiv Q_{Y_1|Y_2X}(\tau_1^*, Q_{Y_2|X}(\tau_2^*, x'), x') - Q_{Y_1|Y_2X}(\tau_1^*, Q_{Y_2|X}(\tau_2^*, x''), x''). \quad (5)$$

Consider a third value of x, x^*, possibly distinct from x and x' with $\{x', x'', x^*\} \in V^*$ and define

$$\triangle_{h_1}^*(x', x'', x^*) \equiv h_1\left(y_2^*(x'), x^*, e_1^*, e_2^*\right) - h_1\left(y_2^*(x''), x^*, e_1^*, e_2^*\right).$$

Four results follow.

(a). Under conditions (A1) - (A3), for any $x \in V^*$ and any a:

$$y_2^*(x) = a \implies Q_{Y_2|X}(\tau_2^*, x) = a$$

and the model defined by conditions (A1) - (A3) uniformly identifies $y_2^*(x)$ for $x \in V^*$.

(b). Under conditions (A1) - (A3), for any $x \in V^*$ and any a:

$$y_1^*(x) = a \implies Q_{Y_1|Y_2X}(\tau_1^*, Q_{Y_2|X}(\tau_2^*, x), x) = a$$

and the model defined by conditions (A1) - (A3) uniformly identifies $y_1^*(x)$ for $x \in V^*$.

(c). Under conditions (A1) - (A5) for any $\{x', x^*, x^+\} \in V^*$

$$h_1(y_2^*(x'), x^*, \varepsilon_1^*, \varepsilon_2^*) = h_1(y_2^*(x'), x^+, \varepsilon_1^*, \varepsilon_2^*),$$

for any $\{x', x^*, x^+\} \in V^*$ and any a:

$$h_1(y_2^*(x'), x^*, \varepsilon_1^*, \varepsilon_2^*) = a \implies Q_{Y_1|Y_2X}(\tau_1^*, Q_{Y_2|X}(\tau_2^*, x'), x') = a$$

and the model defined by conditions (A1) - (A5) uniformly identifies $h_1(y_2^*(x'), x^*, \varepsilon_1^*, \varepsilon_2^*)$ for $\{x', x^*\} \in V^*$.

(d). Under conditions (A1) - (A5) for any $\{x', x'', x^*\} \in V^*$ and any a

$$\triangle_{h_1}^*(x', x'', x^*) = a \Longrightarrow \triangle_Q^*(x', x'') = a$$

and the model defined by conditions (A1) - (A5) uniformly identifies the partial difference $\triangle_{h_1}^*(x', x'', x^*)$, for any $\{x', x'', x^*\} \in V^*$.

Proof

(a). First consider the identification of the value of $y_2^*(x)$, defined (see conditions (A1) and (A2)) as follows.

$$y_2^*(x) \equiv h_2(x, e_2^*(x)) \tag{6}$$

The monotonicity condition (A3) and the equivariance property of quantiles imply that, for any $x \in V^*$, since $e_2^*(x)$ is the τ_2^*-quantile of ε_2 given $X = x$,

$$h_2(x, e_2^*(x)) = Q_{Y_2|X}(\tau_2^*, x).$$

Therefore, for any a,

$$y_2^*(x) = a \implies Q_{Y_2|X}(\tau_2^*, x) = a. \tag{7}$$

Applying the Lemma of Section 3 gives the result that the model defined by (A1) - (A3) uniformly identifies the value of $y_2^*(x)$ for all $x \in V^*$ since $Q_{Y_2|X}(\tau_2^*, x)$ is a well defined functional of the conditional distribution of Y given X satisfying the condition of the Lemma. This completes the proof of part (a) of the Theorem[16].

(b). Now consider identification of the value of $y_1^*(x)$ defined (see conditions (A1) and (A2)) as follows.

$$y_1^*(x) \equiv h_1(y_2^*(x), x, e_1^*(x), e_2^*(x)) \tag{8}$$

Substitute for Y_2 in equation (3) giving

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

and evaluate the right hand side at $\varepsilon_2 = e_2^*(x)$ which gives the expression:

$$g(x, \varepsilon_1) = h_1(h_2(x, e_2^*(x)), x, \varepsilon_1, e_2^*(x)).$$

Considering variations in ε_1, the monotonicity condition (A3) and the equivariance property of quantiles imply that, for any $x \in V^*$, since $e_1^*(x)$ is the τ_1^*-quantile of ε_1 given $\varepsilon_2 = e_2^*(x)$ and $X = x$,

$$h_1(h_2(x, e_2^*(x)), x, e_1^*(x), e_2^*(x)) = Q_{Y_1|\varepsilon_2 X}(\tau_1^*, e_2^*(x), x) \tag{9}$$

where the left hand side here is $g(x, e_1^*(x))$.

[16] Note that this conclusion of the Theorem follows when h_2 is *weakly* monotonic with respect to variation in ε_2.

Consider the right hand side of equation (9). The definition of $y_2^*(x)$ given in equation (6) implies that the events $(\varepsilon_2 = e_2^*(x) \cap X = x)$ and $(Y_2 = y_2^*(x) \cap X = x)$ are identical, so conditioning on $\varepsilon_2 = e_2^*(x)$ and $X = x$ is the same as conditioning on $Y_2 = y_2^*(x)$ and $X = x$. Therefore[17]

$$Q_{Y_1|\varepsilon_2 X}(\tau_1^*, e_2^*(x), x) = Q_{Y_1|Y_2 X}(\tau_1^*, y_2^*(x), x). \tag{10}$$

Consider the left hand side of equation (9). From the definition of $y_2^*(x)$ given in equation (6)

$$h_1(h_2(x, e_2^*(x)), x, e_1^*(x), e_2^*(x)) = h_1(y_2^*(x), x, e_1^*(x), e_2^*(x)) \tag{11}$$

and using the definition of $y_1^*(x)$ given in equation (8), on combining (9), (10) and (11), there is the following equation.

$$y_1^*(x) = Q_{Y_1|Y_2 X}(\tau_1^*, y_2^*(x), x) \tag{12}$$

Equation (7) implies that $y_2^*(x)$ in (12) can be replaced by $Q_{Y_2|X}(\tau_2^*, x)$ giving:

$$Q_{Y_1|Y_2 X}(\tau_1^*, y_2^*(x), x) = Q_{Y_1|Y_2 X}(\tau_1^*, Q_{Y_2|X}(\tau_2^*, x), x)$$

and so, for any $x \in V^*$ and any a,

$$y_1^*(x) = a \implies Q_{Y_1|Y_2 X}(\tau_1^*, Q_{Y_2|X}(\tau_2^*, x), x) = a. \tag{13}$$

Applying the Lemma of Section 3 gives the result that the model defined by (A1) - (A3) uniformly identifies the value of $y_1^*(x)$ for all $x \in V^*$ since $Q_{Y_1|Y_2 X}(\tau_1^*, Q_{Y_2|X}(\tau_2^*, x), x)$ is a well defined functional of the conditional distribution of Y given X satisfying the condition of the Lemma. This completes the proof of part (b) of the Theorem.

(c). The quantile invariance condition (A4) implies that for all $x \in V^*$ the terms $e_1^*(x)$ and $e_2^*(x)$ in equation (8) can be replaced by respectively e_1^* and e_2^* giving the following.

$$y_1^*(x) = h_1(y_2^*(x), x, e_1^*, e_2^*) \tag{14}$$

The order condition (A5) implies that, for any $x \in V^*$ the second appearance of x in equation (14) can be replaced by x^* for any $x^* \in V^*$ which gives the following.

$$y_1^*(x) = h_1(y_2^*(x), x^*, e_1^*, e_2^*) \tag{15}$$

Therefore, for all $\{x', x^*, x^+\} \in V^*$, setting $x = x'$ in equation (15), and considering alternative values x^* and x^+ for the second argument of h_1 gives the following.

$$h_1(y_2^*(x'), x^*, e_1^*, e_2^*) = h_1(y_2^*(x'), x^+, e_1^*, e_2^*)$$

It follows from (13) and (15) that, for $\{x', x^*\} \in V^*$ and any a,

$$h_1(y_2^*(x'), x^*, e_1^*, e_2^*) = a \implies Q_{Y_1|Y_2 X}(\tau_1^*, Q_{Y_2|X}(\tau_2^*, x'), x') = a \tag{16}$$

Applying the Lemma of Section 3 gives the result that the model defined by (A1) - (A5) uniformly identifies the value of $h_1(y_2^*(x'), x^*, \varepsilon_1^*, \varepsilon_2^*)$, which is invariant with respect to $x^* \in V^*$, since $Q_{Y_1|Y_2 X}(\tau_1^*, Q_{Y_2|X}(\tau_2^*, x'), x')$ is a well defined functional of the conditional distribution of Y given X satisfying the condition of the Lemma. This completes the proof of part (c) of the Theorem.

[17]Note that if h_2 were only *weakly* monotonic with respect to variation in ε_2 this conclusion would *not* follow because there could be many values of $e_2^*(x)$ implying the same value of $y_2^*(x)$.

(d). First recall that the partial difference $\triangle^*_{h_1}(x', x'', x^*)$, which is invariant to choice of $x^* \in V^*$, is defined as follows.

$$\triangle^*_{h_1}(x', x'', x^*) \equiv h_1\left(y_2^*(x'), x^*, e_1^*, e_2^*\right) - h_1\left(y_2^*(x''), x^*, e_1^*, e_2^*\right)$$

It follows directly from (16) that for any $\{x', x'', x^*\} \in V^*$ and any a,

$$\triangle^*_{h_1}(x', x'', x^*) = a \implies \triangle^*_Q(x', x'') = a.$$

Applying the Lemma of Section 3 gives the result that the model defined by (A1) - (A5) uniformly identifies the value of $\triangle^*_{h_1}(x', x'', x^*)$, since $\triangle^*_Q(x', x'')$ is a well defined functional of the conditional distribution of Y given X satisfying the condition of the Lemma. This completes the proof of the final part of the Theorem.

5. Remarks and extensions

5.1. Rank condition, discreteness and parametric restrictions. Part (d) of the Theorem is only of interest if the "rank condition", $y_2^*(x') \neq y_2^*(x'')$ holds.

Even when this condition holds, the extent to which $y_2^*(x)$ can be varied may be severely limited by the nature of the set of instrumental values. To take an extreme example, if there are just two admissible instrumental values (perhaps because X is a single binary instrument), then for any choice of τ_2^*, and thus of e_2^*, only two values of Y_2 can be generated and only one partial difference can be identified. Whether or not that is an interesting partial difference is a matter for case-by-case consideration.

If X does not show continuous variation but h_1 is a smooth function of Y_2 then the slope of a Y_2-chord of the structural function, that is:

$$\triangle^*_{h_1}(x', x'')/\left(y_2^*(x') - y_2^*(x'')\right),$$

can be identified as long as the rank condition is satisfied.

If h_1 is restricted to be linear in Y_2, with a coefficient that may depend upon X and ε, then just two instrumental values are sufficient to globally identify the value of this coefficient at a value of X and ε. By extension, if h_1 is restricted to be a degree M polynomial function of Y_2, with coefficients possibly depending on X and ε, then, as long as the conditions of the Theorem are satisfied, $M + 1$ distinct instrumental values are sufficient to identify the $M + 1$ coefficients of the polynomial.

5.2. Weak instruments. Even when the set of instrumental values has extensive coverage, variation across the set of instrumental values may still induce only limited variation in $y_2^*(x)$. This will be the case when h_2 transmits the effect of variation in X only weakly. In this situation extensive understanding of the impact of Y_2 on h_1 cannot be obtained without further, for example, parametric, restrictions.

This "weak instrument" problem is additional to the weak instrument problem commonly discussed in the context of estimation which arises from the possibly poor quality of asymptotic approximations to the distributions of estimators based on estimation using numerous weak instruments.

5.3. Discrete outcomes. The results of the Theorem apply when the outcomes Y_1 and Y_2 have discrete, continuous or mixed distributions. However, results (b) - (d) of the Theorem are not likely to be useful when Y_2 is not continuously distributed in a neighbourhood of the quantile $e_2(x)$ of interest. Although the Theorem does apply when the outcome Y_2 has a discrete distribution, the impact of the covariates, X, on the distribution of Y_2 allowed by the Theorem is extremely limited when Y_2 has a discrete distribution.

To see this, first note that since h_2 is required (at the values of X of interest) to be strictly monotonic with respect to variation in ε_2, Y_2 can only have a discrete distribution if ε_2 has a discrete distribution.

Suppose that ε_2 is discrete. Since h_2 is required to be strictly monotonic, there is a one-to-one correspondence between the points of support of the distribution of ε_2 and the points of support of the distribution of Y_2. Consider a particular value which ε_2 can take, say $e_2^{(i)}(x)$. The probability that Y_2 take the value, $y_2^{(i)}(x) = h_2(x, e_2^{(i)}(x))$ is equal to the probability that $\varepsilon_2 = e_2^{(i)}(x)$ because h_2 is required to be strictly monotonic. Variation in x via the first argument of h_2 can change the *locations* of the points of support of Y_2 but cannot alter the *probabilities* with which these points of support occur. Variation in these probabilities could arise because $e_2^{(i)}(x)$ varies with x, but such variation is ruled out for values of x in a set of instrumental values.

This issue is now explored a little further. Any discrete distribution can be arbitrarily closely approximated by a continuous distribution. As an example, consider the case in which Y_2 has *exactly* a Poisson distribution with mean $\lambda(x)$. This arises when $h_2(x, \varepsilon_2)$ is defined as the non-decreasing function:

$$h_2(x, \varepsilon_2) = i, \quad \varepsilon_2 \in (F(i-1, \lambda(x)), F(i, \lambda(x))] \tag{17}$$

where ε_2 is uniformly distributed on $(0,1)$ and

$$\begin{aligned} F(-1, \lambda(x)) &= 0 \\ F(i, \lambda(x)) &= e^{-\lambda(x)} \sum_{j=0}^{i} \frac{\lambda(x)^j}{j!}, \quad i \geq 0. \end{aligned}$$

Consider the function $h_2^{(\beta)}(x, \varepsilon_2)$ defined for positive β as follows.

$$h_2^{(\beta)}(x, \varepsilon_2) = i + 1 - \left(\frac{F(i, \lambda(x)) - \varepsilon_2}{F(i, \lambda(x)) - F(i-1, \lambda(x))} \right)^{\beta}, \quad \varepsilon_2 \in (F(i-1, \lambda(x)), F(i, \lambda(x))]$$

With ε_2 uniformly distributed on $(0,1)$ this generates a variate $Y_2 = h_2^{(\beta)}(x, \varepsilon_2)$ which is the sum of a Poisson variate with mean $\lambda(x)$ and an independently distributed variate, V, which has a Beta distribution on $(0,1)$ with distribution function $F_V(v) = 1 - (1-v)^{1/\beta}$.

As β approaches zero the probability mass of this Beta variate comes to be concentrated closer and closer to zero and $h_2^{(\beta)}(x, \varepsilon_2)$ approaches $h_2(x, \varepsilon_2)$ defined in equation (17). The Theorem of Section 4 applies when structures have Y_2 generated in this fashion with any *positive* value of β but the Theorem does not apply when $\beta = 0$ at which point there is a fundamental discontinuity.

For $\beta > 0$ there is a one-to-one correspondence between a quantile[18] $e_2(x)$ of the distribution of ε_2 given $X = x$ and $y_2(x) = h_2^{(\beta)}(x, e_2(x))$ and so, in the proof of the Theorem we can use the identity

$$Q_{Y_1|\varepsilon_2 X}(\tau_1^*, e_2(x), x) = Q_{Y_1|Y_2 X}(\tau_1^*, y_2(x), x)$$

where $y_2(x) = h_2^{(\beta)}(x, e_2(x))$.

When $\beta = 0$, $Q_{Y_1|Y_2 X}(\tau_1^*, y_2(x), x)$ is a conditional quantile of the distribution of Y_1 given Y_2 and X which arises from the conditional distribution, $F_{Y_1|\varepsilon_2 X}$, as the "mixture":

$$F_{Y_1|Y_2 X}(y_1|y_2(x), x) = \int_{e_2 \in A(y_2(x), x)} F_{Y_1|\varepsilon_2 X}(y_1|e_2, x) dF_{\varepsilon_2|X}(e_2|x)$$

[18]Note that in this example ε_2 is uniformly distributed on $(0,1)$, independent of X and so $e_2(x)$ does not in fact depend upon x.

where

$$A(y_2(x), x) = \{e_2 : h_2(x, e_2) = y_2(x)\}.$$

This observation suggests that the identification of characteristics of structures in which Y_2 has non-trivial discrete variation is similar to the problem of identification of characteristics of structures with more sources of variation than outcomes and, as in those cases, requires restrictions of types different to, and in a sense stronger than, those considered in this paper.

5.4. Estimation. Quantile regression estimation methods (see Koenker and Bassett (1978)) can be employed to estimate the values of the structural characteristics identified by the Theorem.

Estimation could be parametric, semi- or non-parametric depending on the extent of additional structural restrictions one cares to impose. For parametric estimation, see Koenker and Bassett (1978), Koenker and d'Orey (1987); for semiparametric estimation see for example Chaudhuri, Doksum and Samarov (1997), Kahn (2001) and Lee (2002); 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.

Quantile regression estimation is well understood so estimation issues are not considered further in this paper, except in Section 5.6.

5.5. Instrumental variables. Regarding the order condition, (A5), the special case, familiar in the classical analysis of identification in parametric models with "exclusion" restrictions arises when X is partitioned into two subsets, X_{inc} and X_{exc} and X_{exc} does not feature in the h_1 equation.

We then often talk of X_{exc} as *instrumental variables* and it will commonly be the case that within the set of instrumental values, $x_{inc} = x^*_{inc}$, some common value for all $x \in V^*$. Typical pairs of instrumental values in this case would have the form:

$$x' = \{x^*_{inc}, x'_{exc}\}, \quad x'' = \{x^*_{inc}, x''_{exc}\}.$$

In the absence of parametric restrictions we could allow X_{exc} to feature in the h_1 equation but maintain the "order" restriction that h_1 is insensitive to variations in X_{exc} at values of X in the set of instrumental values.

5.6. Overidentification. There is overidentification of the value of $\triangle^*_{h_1}(x', x'', x^*)$ when there exists more than one pair $\{x', x''\} \in V^*$ yielding a common value of $\triangle^*_{h_1}(x', x'', x^*)$. Then the efficiency of estimation will be enhanced if alternative estimates, based on different just identifying pairs, $\{x', x''\}$, are combined, for example using a minimum distance estimator. There is also scope for testing some of the overidentifying restrictions.

In the classical analysis of parametric identification with exclusion restrictions, overidentification arises in this two equation model when X_{exc} contains more than one covariate. This can be set in the context of the "instrumental values" of this paper by writing

$$X_{exc} = \{X_{exc,1}, X_{exc,2}\}$$

and noting that

$$\{x', x''\} = \{(x^*_{inc}, x'_{exc,1}, x^+_{exc,2}), (x^*_{inc}, x''_{exc,1}, x^+_{exc,2})\}$$

and

$$\{x', x''\} = \{(x^*_{inc}, x^+_{exc,1}, x'_{exc,2}), (x^*_{inc}, x^+_{exc,1}, x''_{exc,2})\}$$

are then overidentifying pairs of instrumental values provided that both pairs produce *identical* values $\{y_2(x'), y_2(x'')\}$ and that the two values of x' and x'' both fall in V^*.

5.7. Smooth structures. If the Y_2 derivative of the structural function exists, and X can vary continuously in the set of instrumental values inducing continuous variation in Y_2 then, by considering[19] the limiting behaviour of $\triangle^*_{h_1}(x',x'')/\left(y_2^*(x')-y_2^*(x'')\right)$, the identification of the Y_2 derivative of h_1 can be achieved, yielding the result given in Chesher (2001b).

To see this, consider the case in which all elements of x' and x'' are identical except for one, denoted by $x_\triangledown$. Consider the slope of a Y_2-chord of h_1 obtained by moving from x' to x'', inducing a movement from $y_2^*(x')$ to $y_2^*(x'')$, and suppose the rank condition, $y_2^*(x') \neq y_2^*(x'')$ is satisfied. The slope of the chord is

$$A(x',x'') = \frac{\triangle^*_{h_1}(x',x'')}{y_2^*(x')-y_2^*(x'')}$$

and consider its limit (assumed to exist) as $x' \to x''$. Let the limiting value of x' and x'' be denoted by x^*.

Let

$$A(x^*) = \lim_{x' \to x'' = x^*} A(x',x'').$$

If the limit exists then

$$A(x^*) = \nabla_{Y_2} h_1(y_2,x,\varepsilon_1,\varepsilon_2)|_{y_2=y_2^*(x),x=x^*,\varepsilon_1=e_1^*(x),\varepsilon_2=e_2^*(x)} \qquad (18)$$

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

Write $A(x^*)$ as:

$$A(x^*) = \lim_{x' \to x'' = x^*} \frac{\triangle^*_{h_1}(x',x'')/\left(x'-x''\right)}{\left(y_2^*(x')-y_2^*(x'')\right)/\left(x'-x''\right)}.$$

Then, if the required limits exist,

$$A(x^*) = \frac{\lim_{x' \to x'' = x^*}\left(\triangle^*_{h_1}(x',x'')/\left(x'-x''\right)\right)}{\lim_{x' \to x'' = x^*}\left(\left(y_2^*(x')-y_2^*(x'')\right)/\left(x'-x''\right)\right)}$$

and therefore:

$$A(x^*) = \frac{\nabla_{x_\triangledown} h_1\left(y_2^*(x^*),x^*,e_1^*,e_2^*\right)}{\nabla_{x_\triangledown} y_2^*(x^*)} \qquad (19)$$

$$= \nabla_{Y_2} h_1\left(y_2^*(x^*),x^*,e_1^*,e_2^*\right) + \frac{\nabla_{X_\triangledown} h_1\left(y_2^*(x^*),x^*,e_1^*,e_2^*\right)}{\nabla_{X_\triangledown} y_2^*(x^*)}. \qquad (20)$$

In (19) $\nabla_{x_\triangledown} h_1\left(y_2^*(x^*),x^*,e_1^*,e_2^*\right)$ is the partial derivative of $h_1\left(y_2^*(x),x,e_1^*,e_2^*\right)$ with respect to $x_\triangledown$, and $\nabla_{x_\triangledown} y_2^*(x^*)$ is the partial derivative of $y_2^*(x) \equiv h_2(x,e_2^*)$ with respect to $x_\triangledown$, both evaluated at $x = x^*$.

The second line, (20), follows on applying the chain rule, noting that $x_\triangledown$ affects h_1 in (19) directly and *via* $y_2^*(x)$.

In (20) $\nabla_{Y_2} h_1\left(y_2^*(x^*),x^*,e_1^*,e_2^*\right)$ is the partial derivative of $h_1\left(Y_2,X,\varepsilon_1,\varepsilon_2\right)$ with respect to Y_2, $\nabla_{X_\triangledown} h_1\left(y_2^*(x^*),x^*,e_1^*,e_2^*\right)$ is the partial derivative of $h_1\left(Y_2,X,\varepsilon_1,\varepsilon_2\right)$ with respect to $X_\triangledown$, both evaluated at $y_2^*(x^*)$, x^*, e_1^*, e_2^*, and $\nabla_{X_\triangledown} y_2^*(x^*)$ is the partial derivative of[20] $h_2(X,\varepsilon_2)$ with respect to $X_\triangledown$ evaluated at x^*, e_2^*.

[19] Henceforth the argument x^* of $\triangle^*_{h_1}$ is suppressed since in the model considered here (defined by conditions (A1) - (A5) of the Theorem) $\triangle^*_{h_1}$ is invariant with respect to x^*.

[20] Recall $y_2^*(x) \equiv h_2(x,e_2^*)$.

Under certain conditions[21] the values of the derivatives in (20) and the value of the Y_2-partial derivative of the structural function evaluated at the point indicated in equation (18) are uniformly identified and for any a,

$$A = a \implies \nabla_{Y_2} Q_{Y_1|Y_2X}(\tau_1^*, Q_{Y_2|X}(\tau_2^*, x^*), x^*) + \frac{\nabla_{X_\triangledown} Q_{Y_1|Y_2X}(\tau_1^*, Q_{Y_2|X}(\tau_2^*, x^*), x^*)}{\nabla_{X_\triangledown} Q_{Y_2|X}(\tau_2^*, x^*)} = a$$

where e.g. $\nabla_{Y_2} Q_{Y_1|Y_2X}(\tau_1^*, Q_{Y_2|X}(\tau_2^*, x^*), x^*)$ is the y_2-derivative of $Q_{Y_1|Y_2X}(\tau_1^*, y_2, x)$ evaluated at $y_2 = Q_{Y_2|X}(\tau_2^*, x^*)$, $x = x^*$.

5.8. Partial differences with respect to covariates. Partial differences of h_1 with respect to an element of X can be identified in a similar fashion. Consider an element $X_\Diamond$, denote remaining elements of X by $X_\blacklozenge$, and suppose there exists a set of instrumental values V^* with elements x written as

$$x = (x_\Diamond, x_\blacklozenge)$$

such that:

1. for all $\{x', x''\} \in V^*$,

$$\begin{aligned} y_2^*(x') &= y_2^*(x'') \\ e_1^*(x') &= e_1^*(x'') \\ e_2^*(x') &= e_2^*(x'') \end{aligned}$$

   with common values denoted by y_2^*, $e_1^*$and e_2^*, and,

2. for all $\{x', x''\} \in V^*$,

$$h_1(y_2^*, x'_\Diamond, x'_\blacklozenge, e_1^*, e_2^*) = h_1(y_2^*, x'_\Diamond, x''_\blacklozenge, e_1^*, e_2^*).$$

   where the dependence of h_1 on $x_\Diamond$ and $x_\blacklozenge$ is made explicit in the notation.

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

$$\triangle^*_{h_1, X_\Diamond}(x'_\Diamond, x''_\Diamond, x^*_\blacklozenge) \equiv h_1(y_2^*, x'_\Diamond, x^*_\blacklozenge, e_1^*, e_2^*) - h_1(y_2^*, x''_\Diamond, x^*_\blacklozenge, e_1^*, e_2^*).$$

Then it can be shown that the model defined by (A1) - (A3) of the Theorem of Section 4 and conditions (1) and (2) above uniformly identifies the partial difference $\triangle^*_{h_1, X_\Diamond}(x'_\Diamond, x''_\Diamond, x^*_\blacklozenge)$, which, note is invariant with respect to $x^*_\blacklozenge$, and that for any a,

$$\triangle^*_{h_1, X_\Diamond}(x'_\Diamond, x''_\Diamond, x^*_\blacklozenge) = a \Longrightarrow \triangle^*_Q(x', x'') = a$$

where $\triangle^*_Q(x', x'')$ is the difference of conditional quantile functions already defined in (5).

5.9. Partial differences with respect to unobservables. Now consider identification of partial differences of h_1 with respect to variation in the stochastic unobservables. First consider differences with respect to variation in ε_1.

Choose a value of X, x, and a probability τ_2^*, define

$$\begin{aligned} e_2^*(x) &\equiv Q_{\varepsilon_2|X}(\tau_2^*, x) \\ y_2^*(x) &\equiv h_2(x, e_2^*(x)) \end{aligned}$$

[21] See Chesher (2002).

choose two probability levels, τ^*_{11} and τ^*_{12}, define

$$\begin{aligned} e^*_{11}(x) &\equiv Q_{\varepsilon_1|\varepsilon_2 X}(\tau^*_{11}, e^*_2(x), x) \\ e^*_{12}(x) &\equiv Q_{\varepsilon_1|\varepsilon_2 X}(\tau^*_{12}, e^*_2(x), x) \end{aligned}$$

and consider

$$\triangle^*_{h_1,\varepsilon_1}(x) \equiv h_1(y^*_2(x), x, e^*_{11}(x), e^*_2(x)) - h_1(y^*_2(x), x, e^*_{12}(x), e^*_2(x))$$

which is clearly a partial difference with respect to variation in ε_1.

Results (a) and (b) of the Theorem imply that under conditions (A1) - (A3) (that is, triangularity, completeness and monotonicity) $\triangle^*_{h_1,\varepsilon_1}(x)$ is uniformly identified and its value is delivered by the follwing difference of conditional quantile functions.

$$Q_{Y_1|Y_2X}(\tau^*_{11}, Q_{Y_2|X}(\tau^*_2, x), x) - Q_{Y_1|Y_2X}(\tau^*_{12}, Q_{Y_2|X}(\tau^*_2, x), x)$$

Now consider differences with respect to variation in ε_2.

Choose two values of X, x' and x'', and two pairs of probabilities $\{\tau^*_{11}, \tau^*_{12}\}$ and $\{\tau^*_{21}, \tau^*_{22}\}$. For any x and $i \in \{1, 2\}$ define

$$\begin{aligned} e^*_{2i}(x) &\equiv Q_{\varepsilon_2|X}(\tau^*_{2i}, x) \\ y^*_{2i}(x) &\equiv h_2(x, e^*_{2i}(x)) \end{aligned}$$

and

$$e^*_{1i}(x) \equiv Q_{\varepsilon_1|\varepsilon_2 X}(\tau^*_{1i}, e^*_{2i}(x), x)$$

and consider

$$\triangle^*_{h_1,\varepsilon_2}(x', x'') \equiv h_1(y^*_{21}(x'), x', e^*_{11}(x'), e^*_{21}(x')) - h_1(y^*_{22}(x''), x'', e^*_{12}(x''), e^*_{22}(x''))$$

Assume that x' and x'' are members of a set of instrumental values, V^*, which has the following properties.

1. For all $\{x', x''\} \in V^*$

$$\begin{aligned} y^*_{21}(x') &= y^*_{22}(x'') && \text{(a)} \\ e^*_{11}(x') &= e^*_{12}(x'') && \text{(b)} \end{aligned}$$

   with common values denoted by y^*_2 and e^*_1.

2. For all $\{x', x''\} \in V^*$, one or both of the following conditions hold

$$\begin{aligned} h_1(y^*_2, x', e^*_1, e^*_{21}(x')) &= h_1(y^*_2, x'', e^*_1, e^*_{21}(x')) && \text{(a)} \\ h_1(y^*_2, x', e^*_1, e^*_{22}(x')) &= h_1(y^*_2, x'', e^*_1, e^*_{22}(x')) && \text{(b)} \end{aligned}$$

If conditions (1) and (2a) hold then $\triangle^*_{h_1,\varepsilon_2}(x', x'')$ can be written as

$$\begin{aligned} \triangle^*_{h_1,\varepsilon_2}(x', x'') &= h_1(y^*_2, x', e^*_1, e^*_{21}(x')) - h_1(y^*_2, x'', e^*_1, e^*_{22}(x'')) \\ &= h_1(y^*_2, x'', e^*_1, e^*_{21}(x')) - h_1(y^*_2, x'', e^*_1, e^*_{22}(x'')), \end{aligned} \tag{21}$$

for any $\{x', x''\} \in V^*$. If conditions (1) and (2b) hold then $\triangle^*_{h_1,\varepsilon_2}(x', x'')$ can be written as

$$\begin{aligned} \triangle^*_{h_1,\varepsilon_2}(x', x'') &= h_1(y^*_2, x', e^*_1, e^*_{21}(x')) - h_1(y^*_2, x'', e^*_1, e^*_{22}(x'')) \\ &= h_1(y^*_2, x', e^*_1, e^*_{21}(x')) - h_1(y^*_2, x', e^*_1, e^*_{22}(x'')), \end{aligned} \tag{22}$$

for any $\{x', x''\} \in V^*$. If condition (1) and *both* of (2a) and (2b) hold then $\triangle^*_{h_1,\varepsilon_2}(x', x'')$ can be written as

$$\triangle^*_{h_1,\varepsilon_2}(x', x'') = h_1(y_2^*, x^*, e_1^*, e_{21}^*(x')) - h_1(y_2^*, x^*, e_1^*, e_{22}^*(x'')) \quad (23)$$

for any $\{x', x'', x^*\} \in V^*$.

Note that each of (21), (22) and (23) is a *partial* difference with respect to variation in ε_2. Arguing as in the proof of the Theorem of Section 4, each of these partial differences is uniformly identified with a value delivered by the following difference of conditional quantile functions

$$Q_{Y_1|Y_2X}(\tau^*_{11}, Q_{Y_2|X}(\tau^*_{21}, x'), x') - Q_{Y_1|Y_2X}(\tau^*_{12}, Q_{Y_2|X}(\tau^*_{22}, x''), x'')$$

where note that, by virtue of Condition (1a),

$$Q_{Y_2|X}(\tau^*_{21}, x') = Q_{Y_2|X}(\tau^*_{22}, x'').$$

At the point of estimation this latter condition may be easy to achieve in the sense that, with probability levels τ^*_{21} and τ^*_{22} chosen, and estimated conditional quantiles for Y_2 given X to hand, one may be able to find x' and x'' such that

$$\hat{Q}_{Y_2|X}(\tau^*_{21}, x') = \hat{Q}_{Y_2|X}(\tau^*_{22}, x'').$$

Achieving Condition (1b) is more problematic.

To satisfy Condition (1b) we must find probability levels, $\{\tau^*_{11}, \tau^*_{12}\}$, such that for the chosen probabilty levels $\{\tau^*_{21}, \tau^*_{22}\}$ and instrumental values $\{x', x''\}$ the following condition holds.

$$Q_{\varepsilon_1|\varepsilon_2X}(\tau^*_{11}, Q_{\varepsilon_2|X}(\tau^*_{21}, x'), x') = Q_{\varepsilon_1|\varepsilon_2X}(\tau^*_{12}, Q_{\varepsilon_2|X}(\tau^*_{22}, x''), x'')$$

It is not obvious how this could be done without additional structural restrictions. A restriction requiring ε_1 and ε_2 to be independently distributed given X with quantile invariance holding for variations of x in $\{x', x''\}$ would suffice. Note that in that case we would require τ^*_{11} to equal τ^*_{12}. It is now demonstrated that if $\tau^*_{11} = \tau^*_{12}$ then (21), (22) and (23) can be interpreted as a partial differences of a *normalised* version of the structural function h_1.

First note that a structure in which there is strongly monotonic variation of h_1 with respect to ε_1 and of h_2 with respect to ε_2 can always be written in terms of independently distributed stochastic unobservables which will be denoted by η_1 and η_2. Let $F_{\varepsilon_1|\varepsilon_2X}$ and $F_{\varepsilon_2|X}$ denote the conditional distribution functions of respectively ε_1 given ε_2 and X and of ε_2 given X, and define the random variables

$$\begin{aligned} \eta_2 &= F_{\varepsilon_2|X}(\varepsilon_2|x) \\ \eta_1 &= F_{\varepsilon_1|\varepsilon_2X}(\varepsilon_1|\varepsilon_2, x) \end{aligned}$$

so that

$$\varepsilon_2 = Q_{\varepsilon_2|X}(\eta_2|x) \quad (24)$$

$$\varepsilon_1 = Q_{\varepsilon_1|\varepsilon_2X}(\eta_1|Q_{\varepsilon_2|X}(\eta_2|x), x). \quad (25)$$

Then $\{\eta_1, \eta_2\}$ are independently distributed random variables each uniformly distributed on $(0, 1)$, distributed independently of X.[22]

[22]This normalisation plays a central role in Imbens and Newey (2001). The transformations (24) and (25) on which it is based are familar in the context of the generation of psuedo-random numbers, ε_1 and ε_2 with distributions $F_{\varepsilon_1|\varepsilon_2X}$ and $F_{\varepsilon_2|X}$ employing independently uniformly distributed psuedo-random numbers, η_1 and η_2.

The structural equations h_1 and h_2 can be rewritten in terms of independently uniformly distributed η_1 and η_2 as

$$\begin{aligned} Y_1 &= h_1(Y_2, X, Q_{\varepsilon_1|\varepsilon_2 X}(\eta_1|Q_{\varepsilon_2|X}(\eta_2|X), X), Q_{\varepsilon_2|X}(\eta_2|X)) \\ Y_2 &= h_2(X, Q_{\varepsilon_2|X}(\eta_2|X)) \end{aligned}$$

alternatively as

$$\begin{aligned} Y_1 &= h_1^N(Y_2, X, \eta_1, \eta_2) \\ Y_2 &= h_2^N(X, \eta_2) \end{aligned}$$

where h_1^N and h_2^N are normalised structural functions. Note that if the quantiles of ε_1 and ε_2 vary with X then the dependence of h_1^N on X through its second argument will differ from the dependence of h_1 on X through its second argument.

In terms of η_1 and η_2 we have, for $i \in \{1, 2\}$,

$$\begin{aligned} \varepsilon_2 &= e_{2i}^*(x) \Longrightarrow \eta_2 = \tau_{2i}^* \\ \varepsilon_1 &= e_{1i}^*(x) \Longrightarrow \eta_1 = \tau_{1i}^* \end{aligned}$$

and so, if $\tau_{11}^* = \tau_{12}^*$ with common value τ_1^*, and Condition (1a) and, for example, Condition (2a) hold, then, for any $\{x', x''\} \in V^*$,

$$\triangle_{h_1,\varepsilon_2}^*(x', x'') = h_1^N(y_2^*, x'', \tau_1^*, \tau_{21}^*) - h_1^N(y_2^*, x'', \tau_1^*, \tau_{22}^*)$$

which is a partial difference of the *normalised* function h_1^I with respect to variation in η_2. Similarly if Condition (2b) holds then for any $\{x', x''\} \in V^*$,

$$\triangle_{h_1,\varepsilon_2}^*(x', x'') = h_1^N(y_2^*, x', \tau_1^*, \tau_{21}^*) - h_1^N(y_2^*, x', \tau_1^*, \tau_{22}^*)$$

and if both Conditions (2a) and (2b) hold then

$$\triangle_{h_1,\varepsilon_2}^*(x', x'') = h_1^N(y_2^*, x^*, \tau_1^*, \tau_{21}^*) - h_1^N(y_2^*, x^*, \tau_1^*, \tau_{22}^*)$$

for any $\{x', x'', x^*\} \in V^*$.

It follows that under the triangularity, completeness and (strong) monotonicity conditions and if Condition (1a) and one or both of Conditions (2a) and (2b) hold then a partial difference of the normalised structural function h_1^N with respect to variation in η_2 is uniformly identified and its value is delivered by the following difference in conditional quantile functions

$$Q_{Y_1|Y_2X}(\tau_1^*, Q_{Y_2|X}(\tau_{21}^*, x'), x') - Q_{Y_1|Y_2X}(\tau_1^*, Q_{Y_2|X}(\tau_{22}^*, x''), x'')$$

where, again note that $Q_{Y_2|X}(\tau_{21}^*, x') = Q_{Y_2|X}(\tau_{22}^*, x'')$ by virtue of condition (1a). If the "rank condition" $\tau_{12}^* \neq \tau_{22}^*$ is not satisfied then the partial difference is trivially zero.

5.10. Larger structural systems. The Theorem of Section 4 is easily extended to larger systems. The basic steps are outlined now.

Consider a single equation from an M equation structure

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

and $M - 1$ "reduced form" equations[23]

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

[23] These can be thought of as arising from a structural triangular system of equations in which each h_i involves Y_j, $j > i$, and these Y_j's have been recursively substituted out.

As before choose probability levels $\tau^* = \{\tau_i^*\}_{i=1}^M$, consider a set of instrumental values of covariates, $V^* \subseteq \Re^K$, for $i = 1, \ldots, M$, recursively define $e_i^*(x)$ as the conditional τ_i^*-quantile of ε_i given $X = x$ and $\varepsilon_j = e_j^*(x)$, $j > i$, and define $y_i^*(x)$, $i > 1$, and $y_1^*(x)$ as follows.

$$y_i^*(x) \equiv h_i(x, e_i^*(x), \ldots, e_M^*(x)), \quad i = 2, \ldots, M.$$

$$y_1^*(x) \equiv h_1(y_2^*(x), \ldots, y_M^*(x), x, e_1^*(x), \ldots, e_M^*(x))$$

The triangularity condition (A1) of the Theorem is satisfied and assume that the completeness condition (A2) is satisfied.

Assume that an extended version of the monotonicity condition (A3) holds, namely that each function h_i, $i > 1$, is strictly monotonic (normalised to be increasing) in ε_i when other arguments are evaluated at x, $e_j^*(x)$, $j = i, \ldots, M$, with $x \in V^*$. The function h_1 is required to be non-decreasing or non-increasing with respect to variation in ε_1 and is normalised to be non-decreasing.

Assume that the **quantile invariance** condition holds for each $e_i^*(x)$ which take values e_i^*, $i = 1, \ldots, M$, invariant with respect to $x \in V^*$.

Suppose the Y_j-partial difference of h_1 is of interest, defined for some $\{x', x'', x^*\}$ as follows.

$$\begin{aligned} \triangle_{h_1,Y_j}^*(x', x'') \equiv\; & h_1(y_2^*(x^*), \ldots, y_{j-1}^*(x^*), y_j^*(x'), y_{j+1}^*(x^*), , \ldots, y_M^*(x^*), x^*, e_1^*, \ldots, e_M^*) \\ & - h_1(y_2^*(x^*), \ldots, y_{j-1}^*(x^*), y_j^*(x''), y_{j+1}^*(x^*), \ldots, y_M^*(x^*), x^*, e_1^*, \ldots, e_M^*) \end{aligned}$$

Impose the **order condition**: for all $(x', x'') \in V^*$

$$\begin{aligned} & h_1(y_2^*(x'), \ldots, y_{j-1}^*(x'), y_j^*(x'), y_{j+1}^*(x'), \ldots, y_M^*(x'), x', e_1^*, \ldots, e_M^*) \\ = \; & h_1(y_2^*(x''), \ldots, y_{j-1}^*(x''), y_j^*(x'), y_{j+1}^*(x''), \ldots, y_M^*(x''), x'', e_1^*, \ldots, e_M^*). \end{aligned}$$

Recursively define

$$\begin{aligned} Q_M^*(x) & \equiv Q_{Y_M|X}(\tau_M^*, x) \\ Q_{M-1}^*(x) & \equiv Q_{Y_{M-1}|Y_M X}(\tau_{M-1}^*, Q_M^*(x), x) \\ Q_{M-2}^*(x) & \equiv Q_{Y_{M-2}|Y_{M-1} Y_M X}(\tau_{M-2}^*, Q_{M-1}^*(x), Q_M^*(x), x) \\ \vdots \quad & = \quad \vdots \end{aligned}$$

so that $Q_1^*(x)$ is the iterated conditional τ_1^*-quantile of Y_1 given $Y_2, \ldots, Y_M$ in which each Y_i, $i > 1$, is evaluated at its iterated conditional quantile given $Y_{i+1}, \ldots, Y_M$.

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

$$\triangle_{Q_1}^*(x', x'') \equiv Q_1^*(x') - Q_1^*(x'')$$

Then an argument as in the proof of the Theorem of Section 4 lead to the result that the model uniformly identifies $\triangle_{h_1,Y_j}^*(x', x'')$ for any $\{x', x'', x^*\} \in V^*$, and $\triangle_{Q_1}^*(x', x'')$ delivers the value of this partial difference.

Chesher (2001b, 2002) shows how in smooth structures some or all first partial *derivatives* of structural functions in M equation systems can be identified as functionals of conditional quantile functions and how, by considering a local linearisation of the structural functions about a point at which identification of partial derivatives is required, the manipulations involved reduce to linear algebra similar to that introduced in Koopmans, Rubin and Leipnik (1950).

6. Conclusion

This paper has explored the limits to identification of characteristics, specifically partial differences of structural functions, in possibly nonseparable structures. This exploration has been limited to:

1. complete triangular structures,
2. structures exhibiting a degree of monotonicity in the effect of stochastic unobservables on the values produced by structural functions,
3. structures with at least as many observable outcomes as there are stochastic unobservables.

Under these conditions differences of structural functions are identifiable. If there exists a sufficiently rich set of instrumental values of covariates within which local quantile invariance and order conditions hold then the resulting model identifies *partial* differences of structural functions and the partial differences are non-trivial if a local rank condition holds.

If any of the three fundamental conditions enumerated above are weakened, it does not seem possible to achieve identification of partial differences without bringing additional restrictions to bear, even if the local quantile invariance and order conditions are maintained.

However there may be similarly weak sets of restrictions which can result in identification of partial differences of structural functions that do not involve all of (1) - (3) above. It seems unlikely that these sets of restrictions will be nested within those give here[24].

This paper has studied a class of structural characteristics - partial differences of structural functions - that can be identified under weak conditions. Whether or not in any particular problem the members of this class are of interest is a matter for case by case consideration. In some cases it may be necessary to impose further restrictions if this class of identifiable structural characteristics is to contain members of interest.

[24] In the sense that the class of admissable structures defined by the conditions of this paper will not be a proper subset of the class of admissable structures defined by alternative conditions that do not include all of (1)-(3).

References

Abadie, A., Angrist, J., and G. Imbens (2002): "Instrumental variables estimates of the effect of subsidized training on the quantiles of trainee earnings," *Econometrica*, 70, 91-117.

Altonji, J.G., and R.L. Matzkin, (2001): "Panel data estimators for nonseparable models with endogenous regressors," NBER Technical Working Paper 267.

Amemiya, T., (1982): "Two stage least absolute deviations estimators," *Econometrica*, 50, 689-711.

Brown, B.W., (1983): "The identification problem in systems nonlinear in the variables," *Econometrica*, 51,175-196.

Brown, B.W., and R.L. Matzkin (1996): "Estimation of nonparametric functions in simultaneous equations models, with an application to consumer demand," mimeo, Department of Economics, Northwestern University.

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

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

Chernozhukov, V., and C. Hansen (2001): "An IV model of quantile treatment effects," MIT Department of Economics Working Paper Series No. 02-06, December 2001.

Chesher, A.D., (2001a): "Parameter approximations in econometrics," presented at the 2001 NFS Symposium on Identification and Inference for Econometric Models, University of California Berkeley, August 2-7, 2001.

Chesher, A.D., (2001b): "Exogenous impact and conditional quantile functions," Centre for Microdata Methods and Practice Working Paper 01/01.

Chesher, A.D., (2001c): "Quantile driven identification of structural derivatives," Centre for Microdata Methods and Practice Working Paper 08/01.

Chesher, A.D., (2002): "Local identification in nonseparable models," Centre for Microdata Methods and Practice Working Paper 05/02.

Darolles, S., J-P Florens and E. Renault, (2000): "Nonparametric instrumental regression," CREST Documents de Travail 2000-17.

Das, M., (2000): "Nonparametric instrumental variable estimation with discrete endogenous regressors," presented at the 2000 World Congress of the Econometric Society, Seattle.

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*, New York: McGraw Hill.

Frisch, R., (1934): *Statistical confluence analysis by means of complete regression systems*, Publication No. 5, Oslo: Universitets Økonomiske Institutt.

Frisch, R., (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.

Haavelmo, T., (1944): "The probability approach in econometrics," *Econometrica*, 12, Supplement, July 1944, 118 pp.

Heckman, J.J., (1990): "Varieties of selection bias," *American Economic Review, Papers and Proceedings*, 80, 313-318.

Heckman, J.J., and E.Vytlacil (2001): "Structural equations, treatment effects and econometric polivy evaluation," Fisher-Schultz Lecture at the 8th World Congress of the Econometric Society, Seattle, Washinton, August 13th 2000, revised April 2001.

Heckman, J.J., J. Smith and N. Clements (1997): "Making the most out of programme evaluations and social experiments: accounting for heterogeneity in programme impacts," *The Review of Economic Studies*, 64, 487-535.

Hurwicz, L. (1950): "Generalization of the concept of identification," in *Statistical inference in dynamic economic models.* Cowles Commission Monograph 10, New York, John Wiley.

Imbens, G.W., and J. Angrist (1994): "Identification and estimation of average treatment effects," *Econometrica*, 62, 467-476.

Imbens, G.W., and W.K. Newey (2001): "Identification and estimation of triangular simultaneous equations models without additivity," unpublished manuscript.

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

Koenker, R., and G. Bassett Jr. (1978): "Regression quantiles," *Econometrica*, 46, 33-50.

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

Koopmans, T.C., H. Rubin and R.B. Leipnik (1950): "Measuring the equation systems of dynamic economics," in *Statistical inference in dynamic economic models.* Cowles Commission Monograph 10, New York, John Wiley.

Koopmans, T.C., and O. Reiersol (1950): "The identification of structural characterstics,"*Annals of Mathematical Statistics*, 21, 165-181.

Lee, S., (2002): "Efficient semiparametric estimation of a partially linear quantile regression model," forthcoming *Econometric Theory.*

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

Matzkin, R.L., (1999): "Nonparametric estimation of nonadditive random functions," Invited lecture on New Development in the Estimation of Preferences and Production Functions, 1999 Latin American Meeting of the Econometric Society.

Newey, W.K., and J.L. Powell (1988): "Nonparametric Instrumental Variables Estimation," mimeo, Department of Economics, MIT.

Newey, W.K., and J.L. Powell (1990): "Efficient estimation of linear and Type-1 censored regression models under conditional quantile restrictions," *Econometric Theory*, 6, 295-317.

Newey, W.K., J.L. Powell, and F. Vella (1999): "Nonparametric Estimation of Triangular Simultaneous Equations Models," *Econometrica* 67, 565-603.

Pinkse, J., (2000): "Nonparametric two-step regression functions when regressors and errors are dependent," *Canadian Journal of Statistics*, 28, 289-300.

Powell, J.L., (1983): "The asymptotic normality of two-stage least absolute deviations estimators," *Econometrica*, 51, 1569-1576.

Roehrig, C.S., (1988): "Conditions for identification in nonparametric and parametric models", *Econometrica* 56, 433-447.

Rothenberg, T.J., (1971): "Identification in parametric models," *Econometrica*, 39, 577-591.

Vytlacil, E., (2002): "Independence, monotonicity and latent index models: an equivalence result," *Econometrica*, 70, 331-342.

Wald, A., (1950): "Note on identification of economic relations," in *Statistical inference in dynamic economic models.* Cowles Commission Monograph 10, New York, John Wiley.

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., (1925): "The statistical determination of demand curves," *Quarterly Journal of Economics*, 39, 503-543.
Working, E.J., (1927): "What do statistical 'demand curves' show?" *Quarterly Journal of Economics*, 41,212-235.