

Chesher, Andrew

Working Paper

Local identification in nonseparable models

cemmap working paper, No. CWP05/02

Provided in Cooperation with:

The Institute for Fiscal Studies (IFS), London

Suggested Citation: Chesher, Andrew (2002) : Local identification in nonseparable models, cemmap working paper, No. CWP05/02, Centre for Microdata Methods and Practice (cemmap), London, <https://doi.org/10.1920/wp.cem.2002.0502>

This Version is available at:

<https://hdl.handle.net/10419/79308>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

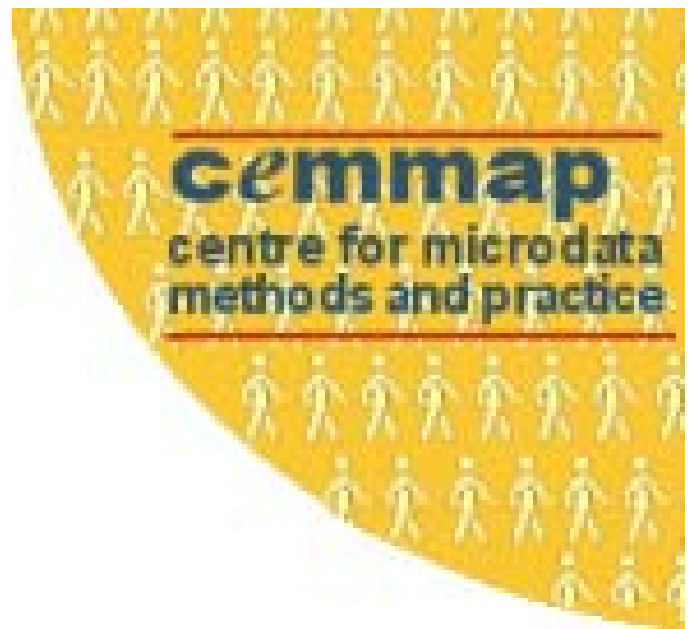
Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



LOCAL IDENTIFICATION IN NONSEPARABLE MODELS

Andrew Chesher

THE INSTITUTE FOR FISCAL STUDIES
DEPARTMENT OF ECONOMICS, UCL
cemmap working paper CWP05/02

Local Identification in Nonseparable Models

ANDREW CHESHER*

CENTRE FOR MICRODATA METHODS AND PRACTICE
INSTITUTE FOR FISCAL STUDIES AND UNIVERSITY COLLEGE LONDON

March 11th 2002, revised March 20th 2002

ABSTRACT. Conditions are derived under which there is local nonparametric identification of values of structural functions and of their derivatives in potentially nonlinear nonseparable models. The attack on this problem is via conditional quantile functions and exploits local quantile independence conditions. The identification conditions include local analogues of the order and rank conditions familiar in the analysis of linear simultaneous equations models. The derivatives whose identification is sought are derivatives of structural equations at a point defined by values of covariates and quantiles of the distributions of the stochastic drivers of the system. These objects convey information about the distribution of the exogenous impact of changes in variables potentially endogenous in the data generating process. The identification conditions point directly to analogue estimators of derivatives of structural functions which are functionals of quantile regression function estimators.

1. IDENTIFICATION IN NONSEPARABLE MODELS

1.1. Introduction. Econometric models in which the unobservable terms that induce stochastic variation are potentially nonseparable are of fundamental interest. First because they constitute a very general form of econometric model, a construction sympathetic to the qualitative nature of the information on economic processes that economic theory provides. Second, because in nonseparable models, sensitivities to changes in conditioning and other variables exhibit stochastic variation. In a microeconomic context this admits the possibility that a policy intervention may have effects which vary across individuals that, measured by covariates, are identical.

A simple example of the type of model considered in this paper is the following stylised model for data which are realisations of a jointly determined labour market outcome, W , for this example the *log wage*, and a measure of *schooling* (S), given

*I am grateful to three referees and an editor, to Christian Dustmann, Art Goldberger, Hidehiko Ichimura, Tony Lancaster, Charles Manski, Whitney Newey, Peter Phillips, Richard Smith, Richard Spady, Tieman Woutersen and participants at the Malinvaud Seminar at CREST and at seminars at Nuffield College Oxford, the European University Institute and University College London for comments on earlier versions of this and related papers.

values of observable covariates, Z . Stochastic variation is caused by two unobservable variables, F and A , which are here thought of as capturing respectively fortune in the labour market and ability.

$$\left. \begin{aligned} W &= w(S, Z, F, A) \\ S &= s(Z, A) \end{aligned} \right\} \quad (1)$$

This paper addresses the problem of identifying first partial derivatives of equations in potentially nonseparable systems like this one, at chosen quantiles of the distributions of unobservable stochastic drivers, F and A , and at chosen values of the covariates, Z .

Of particular interest in this example is the “returns to schooling”, the first partial derivative of the function w with respect to S . The specification (1) permits the returns to schooling to vary with Z (a consequence of allowing nonlinearity in S and Z) and with F and A (a consequence of allowing F and A to appear in nonseparable form). David Card’s 1998 Fisher-Schultz lecture (Card (2001)) develops a model in which returns to schooling vary with unobserved ability.

This paper develops weak conditions under which this derivative is locally identifiable at a chosen value of Z and at chosen quantiles of the distributions of F and A . It is shown that the local identifiability of such a derivative can be achieved under weak local restrictions on the model. For example independence of $\{F, A\}$ and Z is not required. The identification conditions are developed in the context of a nonparametric specification but in a way which allows the impact of semiparametric and parametric restrictions to be assessed. The identification conditions are constructive in the sense that they point directly to analogue estimators of the values of derivatives of structural equations at particular quantiles.

Knowledge of conditions sufficient to ensure *nonparametric* identification of a model feature is useful even if, in estimation, a less flexible, semiparametric or parametric specification is employed. With nonparametric identifiability assured one can regard less flexible specifications employed during estimation as approximations to the data generating process, content that details of the approximation (for example linearity, a single index condition) are not the driving force achieving identification. This is a point stressed by Roehrig (1988). Nonparametric identification is of particular interest in nonseparable models because in such models, in which there are potentially heterogeneous responses to changes in conditioning and other variables, it may be difficult to specify a sufficiently rich parametric form of the data generating process.

This paper develops local conditions under which there is local identifiability of model features. In the example above identification can be achieved if the degree of dependence between $\{F, A\}$ and Z is limited at the values of Z , F and A of interest though there may be strong dependence at other values. A “local instrument” is required to have no first order impact on the wage equation at the values of Z , F and A of interest though it may have strong impact at other values.

In a policy context this is an issue of practical importance. For example, when seeking to understand the impact on a labour market outcome of a change in the amount of schooling an individual receives, one may be mainly interested in the impact on people who, pre-intervention, would have had low levels of schooling, and perhaps on people at the lower end of the distribution of “ability”. Similarly, when seeking to understand the impact of policies that raise, prices, say of foods, one may be particularly interested in the impacts of the policy on the poor and on households with relatively high demands for food due to unobserved variations in tastes. In both these cases a focused, local analysis of the impact of a policy change is appropriate. This paper provides local conditions under which local features of models are identifiable.

Happily the emphasis on local identifiability brings with it very substantial analytic benefits. In sufficiently “smooth” problems of the sort studied here, nonlinear equation systems have locally linear representations. This allows the armoury of linear algebra to be deployed and leads to identification conditions which are essentially local versions of the rank and order conditions developed by Frisch (1938) and Koopmans, Rubin and Leipnik (1950), familiar territory in introductory econometric courses.

The approach of this paper to the problem of developing conditions under which model features are identifiable is unconventional. Before outlining the approach it is helpful to define the concept of identification of a feature of a model. An example of a model feature is the value of a derivative of a structural equation evaluated at chosen quantiles of the distributions of unobservable stochastic drivers and at chosen values of covariates.

Following Hurwicz (1950), let a model, $\mathbf{m} \in \mathfrak{M}$, consist of (a) equations determining outcomes as functions of covariates and unobserved stochastic drivers, and (b) a probability distribution for the stochastic drivers conditional on covariate values such that (a) and (b) together imply a probability distribution of outcomes conditional on values of covariates. Let $\mathfrak{M}^A \subseteq \mathfrak{M}$ be a set of admissible models, that is models whose components (a) and (b) satisfy *a priori* restrictions. The *a priori* restrictions that limit the class of admissible models to \mathfrak{M}^A are sufficient to identify a model feature $\mathcal{F}(\mathbf{m})$ when that feature takes the value \mathcal{F}^* , if every admissible model, $\mathbf{m} \in \mathfrak{M}^A$ with $\mathcal{F}(\mathbf{m}) = \mathcal{F}^*$ generates a distribution function of outcomes given covariates such that a well defined functional applied to that conditional distribution function yields one and only one value, namely \mathcal{F}^* .

Identification conditions are naturally of two types: restrictions on model equations, for example order and rank type conditions, and restrictions on the conditional distribution of the stochastic drivers given covariates. In the analysis of separable models it is common to find the latter restrictions couched in terms of conditional moment restrictions, for example requiring stochastic drivers to be mean independent of covariates. In nonseparable models such restrictions are unhelpful without severe restrictions on the model equations (see for example Card (2001)) and the solution

commonly adopted¹ when model equations are relatively unrestricted is to restrict stochastic drivers to be distributed *independently* of covariates. This is a very strong restriction which might be thought too demanding in many problems that arise in practice.

Aiming, as in this paper, at *local* identification of model features allows progress to be made under far weaker conditions than full independence. The key to achieving this is to refocus attention from direct to *inverse* conditional distribution functions of the stochastic drivers (F and A above) given covariates (Z), that is to *conditional quantile functions*.

This paper considers local restrictions on conditional τ -quantile functions² of ε given Z , local in the sense that they may apply at some values of τ and Z but not at others. Weak conditions on the model equations are developed, sufficient to ensure that conditional quantile functions of outcomes (Y) given covariates (Z), carry enough information about the model equations to enable a functional of conditional quantile functions of Y given Z to identify the features of the model that are of interest. The conditions on the model equations include smoothness restrictions, restrictions to classes of functions which exhibit a degree of monotonic variation with respect to certain of the stochastic drivers, and local order and rank conditions. The latter restrict certain covariates to have no local influence in some equations and to have local influence in other equations.

An attack via conditional quantile functions is particularly well suited to models with potentially nonseparable disturbances because of the equivariance property of quantiles, namely that the τ -quantile of a strictly increasing function of a random variable is the result of applying that function to the τ -quantile of the random variable. Under suitable conditions this allows a direct link to be forged between conditional quantiles of disturbances and conditional quantiles of outcomes.

A result of the analysis of this paper is a set of rather weak conditions sufficient to ensure that well defined functionals of conditional quantile functions associated with the distribution of outcomes given covariates identify derivatives of model equations evaluated at chosen values of covariates and quantile probabilities. The result points directly to analogue estimators of the model derivatives obtained by applying the identifying functionals to estimated conditional quantile functions of outcomes given covariates. The properties of estimators of conditional quantile functions are well

¹See for example Roehrig (1988) and Imbens and Newey (2001).

²Let A be a scalar, continuously distributed random variable with strictly increasing conditional distribution function given a vector of covariates $B = \{B_i\}_{i=1}^K$

$$P[A < a|B = b] = F_{A|B}(a|b).$$

The conditional τ -quantile function for A given B is the inverse function associated with $F_{A|B}(\cdot|b)$ relative to its first argument, the solution to

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

understood and the analogue estimator is an elementary manipulation of derivatives of the estimated conditional quantile functions. As a result there is no new estimation technology required to apply the results of this paper and widely available freeware³ can be used to calculate estimates.

The nature of the results of this paper and their implications are easiest understood in the context of the simple returns to schooling example introduced earlier. A brief review of related literature follows the example and the Section concludes with a plan of the remainder of the paper.

1.2. Example: Identification of returns to schooling. Consider again the model for a jointly determined labour market outcome, W , and a measure of *schooling* (S), uniquely determined given a value of observable covariates, Z , and unobservable continuously distributed disturbances, F and A as follows.

$$\begin{aligned} W &= w(S, Z, F, A) \\ S &= s(Z, A) \end{aligned}$$

Let z^* , be a value of Z and let q_F^* and q_A^* be the τ_A^* - and τ_F^* -quantiles of the distributions of A and F . These are respectively quantiles of the conditional distribution of A given $Z = z^*$ and of F given $A = q_A^*$ and $Z = z^*$. A point at which to identify a structural derivative is specified in terms of *probabilities* defining *quantiles* because, without further restrictions on the equations of the system, the joint distribution of F and A is identifiable at best up to strictly monotonic transformations of F given A and Z and of A given Z .

Let $s^* = s(z^*, q_A^*)$ which is the amount of schooling received by a person at the τ_A^* -quantile of A given $Z = z^*$ implied by the model. Denote the derivative of interest, the “returns to schooling”, evaluated at this point by:

$$\delta_S(z^*, \tau_F^*, \tau_A^*) = \nabla_S w(s^*, z^*, q_F^*, q_A^*).$$

This is the returns to schooling for a person with $Z = z^*$, at the τ_A^* -quantile of the distribution of ability given $Z = z^*$, and at the τ_F^* -quantile of the distribution of fortune given $Z = z^*$ and given ability is equal to the τ_A^* -quantile of ability given $Z = z^*$.

Consider a restricted class of models in which at the point of interest, w is differentiable with respect to its arguments and a strictly monotonic function of F , and s is differentiable with respect to its arguments and is a strictly monotonic function of A .⁴

Suppose the class of models is further restricted so that, at the point of interest (defined by z^* , τ_F^* and τ_A^*),

³See the excellent R, (Hornik (2002)), available from <http://cran.r-project.org>.

⁴Many of these conditions need apply only locally to achieve local identification. Strict monotonicity is a stronger condition than is required. A precise statement of the conditions is in Section 2.

1. the first partial derivative of w with respect to some element, Z_i , of Z is zero and the first partial derivative of s with respect to that element Z_i is nonzero⁵,
2. the conditional quantiles, q_F^* and q_A^* , have zero derivatives with respect to the element Z_i .

Consider the following functional of the joint distribution function of W and S given Z .

$$\pi_{SZ_i}(z^*, \tau_F^*, \tau_A^*) = \nabla_S Q_{W|SZ}(\tau_F^*, s^*, z^*) + \frac{\nabla_{Z_i} Q_{W|SZ}(\tau_F^*, s^*, z^*)}{\nabla_{Z_i} Q_{S|Z}(\tau_A^*, z^*)} \quad (2)$$

The results of this paper imply that *all* models in the restricted class for which the returns to schooling, $\delta_S(z^*, \tau_F^*, \tau_A^*)$, is equal to some value d , generate conditional quantile functions such that $\pi_{SZ_i}(z^*, \tau_F^*, \tau_A^*) = d$.

This secures the identification of $\delta_S(z^*, \tau_F^*, \tau_A^*)$ in the following sense. If one is able to restrict the class of admissible models as indicated then one can be assured that the functional $\pi_{SZ_i}(z^*, \tau_F^*, \tau_A^*)$ applied to the joint distribution of W and S given Z will yield the value of d whichever model among the class of admissible models *with that value* of $\delta_S(z^*, \tau_F^*, \tau_A^*)$ generates data.

The following remarks will be amplified later in the paper.

1. **Overidentification.** If there is more than one element Z_i satisfying the local rank and order conditions then the derivative $\delta_S(z^*, \tau_F, \tau_A)$ is locally overidentified.
2. **Localness.** The result allows the possibility of identification of a structural derivative evaluated at some quantile probabilities (e.g. corresponding to medians) but not at others. Local identification does not require full statistical independence of A and F relative to Z . It can be secured at some values of a “local instrument”, that is Z_i , but not at others⁶.
3. **Identification of other structural derivatives.** Identification of other structural derivatives can be achieved in a similar fashion. For example, let

$$\delta_{Z_j}(z^*, \tau_F^*, \tau_A^*) = \nabla_{Z_j} w(s^*, z^*, q_A^*, q_F^*)$$

and consider the functional

$$\begin{aligned} \pi_{Z_j Z_i}(z^*, \tau_F^*, \tau_A^*) &= \nabla_{Z_j} Q_{W|SZ}(\tau_F^*, s^*, z^*) \\ &\quad - \nabla_{Z_i} Q_{W|SZ}(\tau_F^*, s^*, z^*) \times \frac{\nabla_{Z_j} Q_{S|Z}(\tau_A^*, z^*)}{\nabla_{Z_i} Q_{S|Z}(\tau_A^*, z^*)}. \end{aligned} \quad (3)$$

⁵These conditions are termed local order and rank conditions.

⁶For example one might consider using father’s income as an instrument but admit the possibility that, for high enough values of father’s income, the labour market outcome might be sensitive to father’s income. In that case one could secure identification of the returns to schooling for people with lower income fathers.

All models in the restricted class for which $\delta_{Z_j}(z^*, \tau_F^*, \tau_A^*) = d$ generate joint distributions for W and S for which $\pi_{Z_j Z_i}(z^*, \tau_F^*, \tau_A^*) = d$.

4. **Average derivatives.** If the identification conditions hold for $Z = z^*$ over a set of quantile probabilities, T , then the expected value of (a function of) the structural derivative given $Z = z^*$ conditional on A and F lying in the implied set of values is identifiable as the integral of (the function of) $\delta_S(z^*, \tau_F, \tau_A)$ over $\tau_F, \tau_A \in T$ divided by the probability that A and F lie in the set of values. This is developed in Section 4.
5. **Estimation.** The constructive identification of, for example, $\delta_S(z^*, \tau_F^*, \tau_A^*)$ points directly to estimation using the analogue principle (see Manski (1988)), applying the functional $\pi_{SZ_i}(z^*, \tau_F^*, \tau_A^*)$ to estimates of the conditional quantile functions of W and S given Z .
 - (a) Estimation could be based on a parametric, a semi- or nonparametric specification of the conditional quantile functions.
 - (b) If there is overidentification the resulting multiplicity of estimates can be efficiently combined using a minimum distance procedure.
 - (c) With a parametric linear specification of w and s this leads to an “indirect least squares” type alternative to the Two Stage Least Absolute Deviations (2SLAD) estimators proposed by Amemiya (1982).

Properties of estimators follow directly from well known results on quantile regression function estimators, so there is only a brief consideration of estimation in Section 3.

In Section 2 formulae like (2) are developed for the general M equation case. That development is now sketched for this simple two equation example.

The monotonicity of the schooling equation with respect to ability, together with the smoothness and continuous distribution restrictions, implies that the τ_A^* -quantile of S given $Z = z^*$ is determined by

$$Q_{S|Z}(\tau_A^*, z^*) = s(z^*, q_A^*) = s^* \quad (4)$$

where, recall, $q_A^* = Q_{A|Z}(\tau_A^*, z^*)$ is the conditional τ_A^* -quantile of A given $Z = z^*$. This identifies the value of s at z^* and, if q_A^* does not vary with Z at z^* , the derivatives of $Q_{S|Z}(\tau_A^*, Z)$ with respect to Z at z^* identify the derivatives of s with respect to Z at z^* .

With the wage equation rewritten using the inverse schooling function, $s^{-1}(Z, S)$, as follows,

$$W = w(Z, F, s^{-1}(Z, S))$$

strict monotonicity with respect to variations in F implies that the conditional τ_F^* -quantile of W given $S = s^*$ and $Z = z^*$ is determined by

$$Q_{W|SZ}(\tau_F^*, s^*, z^*) = w(s^*, z^*, q_F^*, q_A^*)$$

where $q_A^* = s^{-1}(z^*, s^*)$ and $q_F^* = Q_{F|AZ}(\tau_F^*, Q_{A|Z}(\tau_A^*, z^*), z^*)$. This identifies the value of w at $Z = z^*$, $S = s^*$, $F = q_F^*$, $A = q_A^*$.

The derivative of $Q_{W|SZ}(\tau_F^*, S, z^*)$ with respect to S at $S = s^*$ in general has contributions directly from S (it is this contribution whose identification is sought) and indirectly from q_F^* and q_A^* .

The conditional quantiles q_F^* and q_A^* are restricted to have zero derivatives with respect to Z . So, differentiating $Q_{W|SZ}(\tau_F^*, s^*, Z)$ with respect to Z_i , which recall is restricted to have *no* direct effect on w at the point of interest, results on evaluation at $Z = z^*$, in an expression equal to the indirect contributions to the derivative of $Q_{W|SZ}(\tau_F^*, S, z^*)$ with respect to S at $S = s^*$ scaled by the relative sensitivity at $Z = z^*$ and $S = s^*$ of $s^{-1}(Z, S)$ with respect to Z_i and S .

This scaling factor is minus the derivative of s with respect to Z_i at $Z = z^*$ and $S = s^*$ which is identified from (4) as minus the derivative of $Q_{S|Z}(\tau_A^*, Z)$ with respect to Z_i at z^* and is required to be nonzero. Adjusting accordingly and isolating the term of interest produces equation (2) above.

1.3. Related literature. The study of parametric identification has a long history in econometrics starting with Working (1925, 1927) and Frisch (1934, 1938) and with notable contributions by, among others, Haavelmo (1944), Hurwicz (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.

There has been considerable attention paid to the problem of determining conditions under which there is *nonparametric* identification of structural models. Roehrig (1988), extending the work of Brown (1983), considered global nonparametric identification of structural equations under the restriction that the stochastic drivers of the system (disturbances) are distributed independently of covariates. Much of Roehrig's development is for the case in which the disturbances appear additively in the equations of the model. Newey and Powell (1988), Newey, Powell and Vella (1999), Pinkse (2000), Darolles, Florens and Renault (2000) study models with additive disturbances which satisfy mean independence conditions of various types. Blundell and Powell (2000) provide a survey of mean independence based work in this area.

There is a large recent literature concerning identification and estimation in models involving outcomes, treatments and instrumental variables, some or all of which are discrete⁷. The methods of this paper are not applicable in these problems because they require variables to exhibit continuous variation at the points at which identification is sought.

Recently there has been interest in determining when global identification can be achieved in models with nonseparable disturbances. Brown and Matzkin (1996) study the identification of nonparametric primitive functions (e.g., production or utility functions) associated with simultaneous equations systems under the assumption that

⁷See for example Heckman (1990), Imbens and Angrist (1994), Das (2000) and Vytlačil (2002).

disturbances and covariates are independently distributed. Altonji and Matzkin, (2001) study panel data models with endogeneity under conditional exchangeability assumptions.

In this paper the focus is on local identification, rather than global identification of structural equations or “global functionals” of them, for example the average structural functions treated in some of the literature. Local identification conditions are rather easy to develop as will be seen in Section 2. When conditions under which it is achievable have been obtained, and one is dealing with a problem in which there is sufficient smoothness, one can address global identification by asking whether the local conditions hold globally. In parametric models the object that is identified locally may be a “global” parameter in which case one automatically has global identification of that parameter but under rather weak local identification conditions.

Imbens and Newey (2001) study a triangular, two equation, nonseparable structural model of the sort addressed in this paper. Global identification is demonstrated in restricted models in which there is full independence of disturbances and covariates. Imbens and Newey attack the identification problem from the standpoint of the distribution functions of the observable variates whereas here inverse distribution functions, that is quantile functions, are the starting point. The Imbens and Newey estimation procedure entails rather demanding nonparametric distribution and mean regression function estimation in contrast to the estimation procedure suggested by the results of this paper, which entails conventional conditional quantile function estimation. Chernozhukov and Hansen (2001) take a quantile based approach to global identification of treatment effects under full independence conditions.

The objects whose identification are considered in this paper, derivatives of structural functions evaluated at quantiles of the distributions of stochastic drivers of the system, can give interesting information about the distribution of policy impacts across a population. There is a recent literature aimed at developing estimators of such distributions. For example Heckman, Smith and Clements (1997) explore non-quantile based approaches in a programme evaluation setting. Abadie, Angrist and Imbens (2002) propose a Quantile Treatment Effect estimator in a study of the impact of subsidised training on the distribution of earnings. These papers address problems in which there is discrete variation in critical variates which is a different problem than that addressed here.

Identification is considered from a conditional quantile perspective in Matzkin (1999). Matzkin studies a model $Y = m(X, \varepsilon)$ in which ε is distributed independently of X and $m(\cdot, \cdot)$ is 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 . The results of this paper produce a local version of Matzkin’s results for equations in reduced form without requiring full independence of ε and Z .

The quantile based approach of this paper produces results relevant to the lit-

erature on quantile regression function estimation in the presence of endogeneity. Amemiya (1982) develops⁸ Two Stage Least Absolute Deviations (2SLAD) estimators for parametric models in which a conditional median is linear in endogenous and exogenous variables and independent of instrumental variables. The estimation procedure suggested in this paper can be applied to the parametric problem studied by Amemiya, producing a simple alternative to the family of 2SLAD estimators.

The identification conditions of this paper include local quantile independence conditions. There are recent uses of quantile independence conditions as the basis for developing estimators in Newey and Powell (1990), Chaudhuri, Doksum and Samarov (1997) and Kahn (2001).

Some of the results presented here were given in Chesher (2001a-c). This paper synthesises those results and presents a full development of conditions under which there is local identification of values of structural functions and of their derivatives in multiple equation nonseparable models.

1.4. Plan of the paper. Section 2 defines the class of models addressed in this paper and provides conditions under which values and derivatives of model equations are locally identified at a point of interest. Conditions for identification of all structural derivatives in an M equation model are developed and then conditions for identification of the derivatives of a single equation are deduced. The conditions and analysis that leads to them are elaborated for the returns to schooling example of Section 1.2. Section 3 addresses the identification of averages of functions of structural derivatives, for example, their expected values and variances. Section 4 briefly examines estimation issues and Section 5 concludes.

2. MODEL AND IDENTIFICATION CONDITIONS

This Section starts with a specification of the structure of the models considered in this paper followed by a definition of a point at which local identification of features of the model is sought. Then the features of interest, the structural derivatives, are defined. Next, a set of identification conditions are developed. The Section concludes by revisiting the returns to schooling example of Section 1.2, examining how the conditions are manifested in that problem.

2.1. Type of structural model. The structural models considered in this paper have the following, triangular, form.

$$\left. \begin{aligned} Y_1 &= h_1(Y_2, Y_3, \dots, Y_M, Z, \varepsilon_1, \varepsilon_2, \dots, \varepsilon_M) \\ Y_2 &= h_2(Y_3, \dots, Y_M, Z, \varepsilon_2, \dots, \varepsilon_M) \\ &\vdots \\ Y_M &= h_M(Z, \varepsilon_M) \end{aligned} \right\} \quad (5)$$

⁸See also Powell (1983).

These models recursively determine values of M scalar variables, $Y = \{Y_i\}_{i=1}^M$, given values of covariates, $Z = \{Z_i\}_{i=1}^K$, and values of continuously distributed stochastic unobservable variables, $\varepsilon = \{\varepsilon_i\}_{i=1}^M$.

There are various special cases of this model which may be of interest in particular applications. First there is the model form

$$\left. \begin{aligned} Y_1 &= h_1(Y_2, Y_3, \dots, Y_M, Z, \varepsilon_1) \\ Y_2 &= h_2(Y_3, \dots, Y_M, Z, \varepsilon_2) \\ \vdots &\quad \vdots \\ Y_M &= h_M(Z, \varepsilon_M) \end{aligned} \right\} \quad (6)$$

in which each equation i contains just one stochastic driver, ε_i . Second there is the model form

$$\left. \begin{aligned} Y_1 &= h_1(Y_2, Y_3, \dots, Y_M, Z, \varepsilon_1, \varepsilon_2, \dots, \varepsilon_M) \\ Y_2 &= h_2(Z, \varepsilon_2, \dots, \varepsilon_M) \\ \vdots &\quad \vdots \\ Y_M &= h_M(Z, \varepsilon_M) \end{aligned} \right\} \quad (7)$$

in which all equations after the first are in reduced form, and in this form there could be just one stochastic driver in each equation as in (6). These are all restricted versions of the model considered here. Identification in these models can be addressed by considering the impact of the conditions which cause (5) to specialise to (6) or (7).

2.2. Point at which identification is sought. This paper is concerned with identification of first partial derivatives of the functions $h = \{h_i\}_{i=1}^M$ at a point denoted by \mathcal{X} . At the point of interest a particular value of Z and ε is specified, and this determines, via the structural equations, a value of Y .

The point \mathcal{X} is defined by a value of Z , denoted z^* , and by M probabilities, $\tau^* = \{\tau_i^*\}_{i=1}^M$ which determine a value of ε as values of a set of conditional quantiles of ε given $Z = z^*$ at the probabilities τ^* , as follows.

$$\varepsilon_i = \varepsilon_i^* = Q_{\varepsilon_i | \varepsilon_{i+}, Z}(\tau_i^*, \varepsilon_{i+}^*, z^*), \quad i \in \{1, \dots, M\} \quad (8)$$

Here $Q_{A|B_1 \dots B_K}(\tau, b_1, \dots, b_K)$ denotes the conditional τ -quantile of random variable A given $B_1 = b_1, \dots, B_K = b_K$.

For any list, $\{X_i\}_{i=1}^M$, the notation X_{i+} indicates the list $\{X_j\}_{j=i+1}^M$ for $i < M$ and an empty list for $i = M$. This shorthand notation is used when indicating conditioning and when indicating arguments of functions. Thus (8) should be read as

$$\begin{aligned} \varepsilon_M &= \varepsilon_M^* = Q_{\varepsilon_M | Z}(\tau_M^*, z^*) \\ \varepsilon_i &= \varepsilon_i^* = Q_{\varepsilon_i | \varepsilon_{i+1} \dots \varepsilon_M, Z}(\tau_i^*, \varepsilon_{i+1}^*, \dots, \varepsilon_M^*, z^*), \quad i \in \{1, \dots, M-1\} \end{aligned}$$

A set of conditional quantile functions like this, in which each variate ε_i is conditioned on $\varepsilon_j, j > i$ (and perhaps on other variates) is described as a set of *iterated* conditional quantile functions.

Note that, at the point \mathcal{X} the structural model implies that the value of Y is given (subject to Condition I below) by the following equations.

$$Y_i = y_i^* = h_i(y_{i+}^*, z^*, \varepsilon_i^*, \varepsilon_{i+}^*), \quad i \in \{1, \dots, M\}$$

It is natural to specify the point \mathcal{X} in terms of quantile probabilities because, in this nonparametric analysis, without further restriction on the functions h , the distribution of ε is at best identifiable up to a monotonic transformation, that is, the metric in which ε is measured is not identifiable. However, under the conditions to be imposed, features of functions at values of ε associated with given probabilities defining quantiles can be identified⁹.

The identification conditions given below relate to behaviour of features of the model at \mathcal{X} .

2.3. Objects whose identification is sought. The features of the model whose identification is sought are some or all of the structural derivatives

$$\left. \begin{aligned} \nabla_{y_j} h_i(\mathcal{X}) &= \left. \frac{\partial}{\partial y_j} h_i(y_{i+}, z, e_i, e_{i+}) \right|_{\mathcal{X}} \\ \nabla_{z_k} h_i(\mathcal{X}) &= \left. \frac{\partial}{\partial z_k} h_i(y_{i+}, z, e_i, e_{i+}) \right|_{\mathcal{X}} \end{aligned} \right\}, \quad i \in \{1, \dots, M\} \quad (9)$$

for $j \in \{i+1, \dots, M\}$, $k \in \{1, \dots, K\}$. Here $|_{\mathcal{X}}$ indicates evaluation of arguments at the values defined by the point \mathcal{X} .

2.4. Identification conditions. This Section provides a set of conditions under which these structural derivatives are identified. As each condition is introduced its implications are developed.

Conditions I and II place restrictions on the functions $h = \{h_i\}_{i=1}^M$ sufficient to ensure that they have a locally linear representation at the point \mathcal{X} .

Conditions III - IV place restrictions on the functions h and the conditional distribution of ε given Z sufficient to ensure that iterated conditional quantile functions for Y_i given Y_{i+} and Z are well defined at \mathcal{X} and have a local linear representation there in which the coefficients, that is the first partial derivatives of the iterated conditional quantile functions, are directly related to the coefficients of the locally linear representation of the structural equations. Condition V places limits on the degree of dependence of ε on Z .

⁹One way to think about this is to note that, under the conditions to be stated, the variates $u = \{u_i\}_{i=1}^M$ where $u_i = F_{\varepsilon_i|\varepsilon_{i+}, Z}(\varepsilon_i|\varepsilon_{i+}, z)$ are independently uniformly distributed on $[0, 1]^M$ and substituting for ε using $u_i = Q_{\varepsilon_i|\varepsilon_{i+}, Z}(u_i|\varepsilon_{i+}, z)$ is merely a normalisation of the functions h . With that normalisation in place, quantile probabilities for the independent variates u are identical to the associated quantile values of u (recall, each element is $U[0, 1]$) and correspond to quantile probabilities for the iterated conditional quantiles of the variates ε .

The final conditions, grouped under the heading VI, impose restrictions on the derivatives of the structural model at the point \mathcal{X} . Full system and single equation conditions are considered below. With Conditions I-V imposed the problem is reduced to the determination of conditions under which coefficients of a *linear* equation system are identified and so these final conditions are essentially rank and order type conditions familiar in the classical linear simultaneous model. However in this local nonparametric analysis the order conditions are local, restricting linear functions of derivatives of structural equations to take *a priori* known values at the point of interest, \mathcal{X} , and the rank conditions are local, requiring certain matrices of derivatives at the point \mathcal{X} to have full rank.

A slightly non-standard feature is that, instead of considering the unique deducibility of structural coefficients (derivatives of h at \mathcal{X}) from classical reduced form coefficients, their deducibility from reduced *triangular* form coefficients (derivatives of iterated conditional quantile functions) is considered.

Condition I. Completeness. *At \mathcal{X} the equations (5) determine a unique value of Y .*

This requires that at \mathcal{X} the functions h are single valued and that for each i , $\{y_{i+}^*, z^*, \varepsilon_i^*, \varepsilon_{i+}^*\}$ is in the domain of h_i .

Condition II. Differentiability. *At \mathcal{X} each function h_i is a continuous and once differentiable function of its arguments¹⁰.*

Conditions I and II imply that at \mathcal{X} there exists a linear system of equations in differentials:

$$dY = AdY + BdZ + Gd\varepsilon \quad (10)$$

where $dY = \{dY_i\}_{i=1}^M$, $dZ = \{dZ_i\}_{i=1}^K$, $d\varepsilon = \{d\varepsilon_i\}_{i=1}^M$ and A , B and G are matrices of first partial derivatives of the structural functions evaluated at the point \mathcal{X} , as follows.

$$A = \begin{bmatrix} 0 & \nabla_{y_2} h_1 & \nabla_{y_3} h_1 & \dots & \nabla_{y_M} h_1 \\ 0 & 0 & \nabla_{y_3} h_2 & \dots & \nabla_{y_M} h_2 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & & \nabla_{y_M} h_{M-1} \\ 0 & 0 & 0 & \dots & 0 \end{bmatrix} \quad G = \begin{bmatrix} 1 & \nabla_{\varepsilon_2} h_1 & \nabla_{\varepsilon_3} h_1 & \dots & \nabla_{\varepsilon_M} h_1 \\ 0 & 1 & \nabla_{\varepsilon_3} h_2 & \dots & \nabla_{\varepsilon_M} h_2 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \dots & \nabla_{\varepsilon_M} h_{M-1} \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix}$$

$$B = \begin{bmatrix} \nabla_{z_1} h_1 & \nabla_{z_2} h_1 & \dots & \nabla_{z_K} h_1 \\ \nabla_{z_1} h_2 & \nabla_{z_2} h_2 & \dots & \nabla_{z_K} h_2 \\ \vdots & \vdots & & \vdots \\ \nabla_{z_1} h_M & \nabla_{z_2} h_M & \dots & \nabla_{z_K} h_M \end{bmatrix}$$

¹⁰There could be discrete covariates but these are not given explicit consideration here. Nonparametric identification in the presence of discrete covariates requires consideration of the conditions stated here at each distinct combination of values of discrete covariates at which identification is of interest.

The restricted triangular structure has been imposed on A and G . The leading diagonal elements of G , $\nabla_{\varepsilon_i} h_i$, have been set equal to 1. This is an innocuous normalisation given the assumption that each function h_i is, at the point \mathcal{X} , a differentiable function of ε_i and given the next condition which implies that each function h_i can be normalised to be *strictly* increasing at \mathcal{X} with respect to variation in ε_i .

Condition III. Single crossing. For each i , $\varepsilon_i^+ = \varepsilon_i^*$ is the only solution to the equation

$$h_i(y_{i+}^*, z^*, \varepsilon_i^+, \varepsilon_{i+}^*) = y_i^*$$

and $\nabla_{\varepsilon_i} h_i$ is nonzero at \mathcal{X} .

This is assured if, holding all coordinates other than ε_i at their values at \mathcal{X} , and considering variations in just ε_i , each function h_i is *strictly monotonic* in ε_i . However global strict monotonicity is not required by Condition III. Each function h_i is normalised to have unit derivative with respect to ε_i at \mathcal{X} which leads to the unit leading diagonal terms in G above.

Condition IV. Continuous distribution. *In a neighbourhood of $\varepsilon^* = \{\varepsilon_i^*\}_{i=1}^M$ the vector ε is continuously distributed given $Z = z^*$ with positive density and at \mathcal{X} the conditional distribution function of ε given Z is differentiable with respect to Z .*

The conditions imposed so far ensure that the iterated conditional quantiles of Y given Z are uniquely determined at \mathcal{X} . The conditional quantile of Y_i given Y_{i+} and Z is obtained as follows. At \mathcal{X} , Y_i is determined by

$$Y_i = h_i(Y_{i+}, Z, \varepsilon_i, h_{i+}^{-1}) \quad (11)$$

evaluated at \mathcal{X} , where $h_{i+}^{-1} = \{h_j^{-1}\}_{j=i+1}^M$ are recursively defined inverse functions¹¹, each function h_j^{-1} depending on Y_j , Y_{j+} and Z , whose existence is assured by the single crossing condition, with each h_j^{-1} depending on Y_j , Y_{j+} and Z , such that

$$\varepsilon_j^* = h_j^{-1}(y_{j+}^*, z^*, y_j^*, \varepsilon_{j+}^*)$$

and satisfying

$$h_j(y_{j+}^*, z^*, h_j^{-1}(y_{j+}^*, z^*, y_j^*, \varepsilon_{j+}^*), \varepsilon_{j+}^*) = y_j^*.$$

The single crossing condition for equation i and the normalisation $\nabla_{\varepsilon_i} h_i = 1$ ensures that at \mathcal{X} the conditional τ_i -quantile of Y_i given Y_{i+} and Z is given by (11) with ε_i replaced by the conditional τ_i -quantile of ε_i given ε_{i+} and Z , thus

$$Q_{Y_i|Y_{i+}Z}(\tau_i^*, y_{i+}^*, z^*) = h_i(y_{i+}^*, z^*, Q_{\varepsilon_i|\varepsilon_{i+}Z}(\tau_i^*, \varepsilon_{i+}^*, z^*), h_{i+}^{-1})$$

¹¹Thus, at \mathcal{X} , h_M^{-1} satisfies

$$Y_M = h_M(Z, h_M^{-1}(Z, Y_M))$$

h_{M-1}^{-1} satisfies

$$Y_{M-1} = h_{M-1}(Y_M, Z, h_{M-1}^{-1}(Y_M, Z, Y_{M-1}, h_M^{-1}(Z, Y_M)), h_M^{-1}(Z, Y_M))$$

and so forth.

with the arguments of h_{i+}^{-1} evaluated at \mathcal{X} .

It follows immediately that, *without any further restriction* the conditional quantile, $Q_{Y_i|Y_{i+},Z}(\tau_i^*, y_{i+}^*, z^*)$, identifies the *value* of the function h_i at \mathcal{X} ¹². Of course this does not imply that the *function* h_i is identified by $Q_{Y_i|Y_{i+},Z}(\tau_i^*, y_{i+}^*, z^*)$ because there are many functions which take the same value at \mathcal{X} , and perhaps take common values over an interval around \mathcal{X} .

The separate contributions to variation in such functions of Y_{i+} and Z *cannot* be identified under the conditions given so far. For example variation in h_i arising from variation in a particular Z_j could come directly through the Z argument of h_i or through the inverse functions h_{i+}^{-1} or through the conditional quantile $Q_{\varepsilon_i|\varepsilon_{i+},Z}(\tau_i^*, \varepsilon_{i+}^*, z^*)$ or *via* some combination of these. The remaining identification conditions impose restrictions sufficient to identify the sensitivity of the functions h to variations in Y_{i+} and Z that arise through their direct appearance in (5).

The argument that follows relies on the following equation which, at \mathcal{X} , relates differentials of iterated conditional quantiles of Y given Z , $dQ_{Y|Z}$, to differentials of Y , and Z .

$$dQ_{Y|Z} = (I_M - (I_M - H)G^{-1}(I_M - A)) dY + ((I_M - H)G^{-1}B + J) dZ \quad (12)$$

The vectors of differentials of iterated conditional quantiles here are

$$dQ_{Y|Z} = \begin{bmatrix} dQ_{Y_1|Y_{1+},Z}(\tau_1, Y_{1+}, Z) \\ dQ_{Y_2|Y_{2+},Z}(\tau_2, Y_{2+}, Z) \\ \vdots \\ dQ_{Y_M|Z}(\tau_M, Z) \end{bmatrix} \quad dQ_{\varepsilon|Z} = \begin{bmatrix} dQ_{\varepsilon_1|\varepsilon_{1+},Z}(\tau_1, \varepsilon_{1+}, Z) \\ dQ_{\varepsilon_2|\varepsilon_{2+},Z}(\tau_2, \varepsilon_{2+}, Z) \\ \vdots \\ dQ_{\varepsilon_M|Z}(\tau_M, Z) \end{bmatrix}$$

evaluated at the point \mathcal{X} and the matrices H ($M \times M$) and J ($M \times K$) also evaluated at \mathcal{X} , contain derivatives of the iterated conditional quantiles of ε given Z , $Q_{\varepsilon|Z}$, with respect to respectively ε and Z .

$$H = \begin{bmatrix} 0 & \nabla_{\varepsilon_2} Q_{\varepsilon_1|\varepsilon_{1+},Z} & \nabla_{\varepsilon_3} Q_{\varepsilon_1|\varepsilon_{1+},Z} & \cdots & \nabla_{\varepsilon_M} Q_{\varepsilon_1|\varepsilon_{1+},Z} \\ 0 & 0 & \nabla_{\varepsilon_3} Q_{\varepsilon_2|\varepsilon_{2+},Z} & \cdots & \nabla_{\varepsilon_M} Q_{\varepsilon_2|\varepsilon_{2+},Z} \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & & \nabla_{\varepsilon_M} Q_{\varepsilon_{M-1}|\varepsilon_M,Z} \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix}$$

$$J = \begin{bmatrix} \nabla_{z_1} Q_{\varepsilon_1|\varepsilon_{1+},Z} & \nabla_{z_2} Q_{\varepsilon_1|\varepsilon_{1+},Z} & \cdots & \nabla_{z_K} Q_{\varepsilon_1|\varepsilon_{1+},Z} \\ \nabla_{z_1} Q_{\varepsilon_2|\varepsilon_{2+},Z} & \nabla_{z_2} Q_{\varepsilon_2|\varepsilon_{2+},Z} & \cdots & \nabla_{z_K} Q_{\varepsilon_2|\varepsilon_{2+},Z} \\ \vdots & \vdots & & \vdots \\ \nabla_{z_1} Q_{\varepsilon_M|Z} & \nabla_{z_2} Q_{\varepsilon_M|Z} & \cdots & \nabla_{z_K} Q_{\varepsilon_M|Z} \end{bmatrix}$$

¹²Matzkin (1999) gives a global version of this result under full independence of ε and Z for the final “reduced form” equation $i = M$.

Equation (12) arises as follows. The iterated conditional quantile functions are determined by a transformed structural system in which, in each equation i , the *single* unobservable, ε_i , appears, the remaining ε_j 's ($j > i$) having been substituted away, using the inverse functions, h_j^{-1} , associated with the equations for Y_j , $j > i$. and defined above, after equation (11). A typical equation is

$$Y_i = h_i(Y_{i+}, Z, \varepsilon_i, h_{i+}^{-1})$$

where each function h_j^{-1} , $j > i$, in the list h_{i+}^{-1} depends¹³ upon Y_j , Y_{j+} and Z .

The equations for the differentials associated with these transformed structural equations are obtained as follows. First rewrite (10) as follows.

$$dY = AdY + BdZ + d\varepsilon + (G - I_M) d\varepsilon \quad (13)$$

The first appearance of $d\varepsilon$ in (13) isolates ε_i in the i th structural equation.

The matrix G is upper triangular with unit leading diagonal elements and is therefore nonsingular. Therefore, from (10),

$$d\varepsilon = G^{-1}(dY - AdY - BdZ) \quad (14)$$

and on substituting (14) in (13) there is, after collecting terms, the following.

$$dY = (I_M - G^{-1}(I_M - A)) dY + G^{-1}BdZ + d\varepsilon \quad (15)$$

As explained earlier, under the conditions imposed so far, the iterated conditional quantiles of Y given Z are determined at \mathcal{X} by the transformed structural equations with Y replaced by $Q_{Y|Z}$ and with ε replaced by the iterated conditional quantiles of ε given Z , that is $Q_{\varepsilon|Z}$. Therefore the differentials of the iterated conditional quantiles of Y given Z satisfy (15) with dY and $d\varepsilon$ replaced by respectively $dQ_{Y|Z}$ and $dQ_{\varepsilon|Z}$ which gives the following equation.

$$dQ_{Y|Z} = (I_M - G^{-1}(I_M - A)) dY + G^{-1}BdZ + dQ_{\varepsilon|Z} \quad (16)$$

Variation in Y and Z *consistent with the model* (5) causes variation in the values of the iterated conditional quantile functions, $Q_{\varepsilon|Z}$, through variation in the conditioning arguments (for the i th conditional quantile function these are ε_{i+} and Z) thus: $dQ_{\varepsilon|Z} = Hd\varepsilon + JdZ$, with H and J defined as above, and using (14)

$$dQ_{\varepsilon|Z} = HG^{-1}(dY - AdY - BdZ) + JdZ$$

and substituting for $dQ_{\varepsilon|Z}$ in (16) gives, after collecting terms, the following.

$$dQ_{Y|Z} = (I_M - (I_M - H)G^{-1}(I_M - A)) dY + ((I_M - H)G^{-1}B + J) dZ \quad (17)$$

¹³See footnote 11 on page 14.

The conditions imposed so far ensure that the matrices of first partial derivatives of the conditional iterated quantile functions of Y given Z at \mathcal{X} with respect to Y , $\nabla_Y Q$, and with respect to Z , $\nabla_Z Q$, are related to the matrices of first partial derivatives of the model as follows

$$\nabla_Y Q = I_M - (I_M - H)G^{-1}(I_M - A) \quad (18)$$

$$\nabla_Z Q = (I_M - H)G^{-1}B + J \quad (19)$$

where

$$\nabla_Y Q = \begin{bmatrix} 0 & \nabla_{y_2} Q_{Y_1|Y_1+Z} & \nabla_{y_3} Q_{Y_1|Y_1+Z} & \cdots & \nabla_{y_M} Q_{Y_1|Y_1+Z} \\ 0 & 0 & \nabla_{y_3} Q_{Y_2|Y_2+Z} & \cdots & \nabla_{y_M} Q_{Y_2|Y_2+Z} \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & & \nabla_{y_M} Q_{Y_{M-1}|Y_M Z} \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix}$$

$$\nabla_Z Q = \begin{bmatrix} \nabla_{z_1} Q_{Y_1|Y_1+Z} & \nabla_{z_2} Q_{Y_1|Y_1+Z} & \cdots & \nabla_{z_K} Q_{Y_1|Y_1+Z} \\ \nabla_{z_1} Q_{Y_2|Y_2+Z} & \nabla_{z_2} Q_{Y_2|Y_2+Z} & \cdots & \nabla_{z_K} Q_{Y_2|Y_2+Z} \\ \vdots & \vdots & & \vdots \\ \nabla_{z_1} Q_{Y_M|Z} & \nabla_{z_2} Q_{Y_M|Z} & \cdots & \nabla_{z_K} Q_{Y_M|Z} \end{bmatrix}$$

evaluated at \mathcal{X} .

Under the stated conditions the conditional distribution of Y given Z uniquely determines $\nabla_Y Q$ and $\nabla_Z Q$ at \mathcal{X} , so the identifiability of elements of A , B , G , H and J depends, under the stated conditions, on whether the elements can be uniquely determined from equations (18) and (19) given values for $\nabla_Y Q$ and $\nabla_Z Q$.

This is now addressed for the case in which there are *no* restrictions on H and the identifiability of elements of A , B and C is considered¹⁴ where

$$C = G(I_M - H)^{-1}.$$

If at \mathcal{X} the elements of ε were mutually quantile independent given Z then H would be a zero matrix, $C = G$, and conditions sufficient to identify C would identify G . Proceeding without restrictions on G and H is therefore equivalent to normalising the elements of ε to be mutually independently distributed given Z .¹⁵

If the equations of the model (9) were restricted so that $G = I_M$, which would arise if each equation i contained no ε_j , $j \neq i$, then $C = (I_M - H)^{-1}$ and conditions sufficient to identify C would identify H .

Restrictions on G and H may arise in parametric models and in other situations but such restrictions are not considered here, and G and H are henceforth regarded as being not separately identifiable.

¹⁴The restriction $J = 0$ will shortly be imposed.

¹⁵See footnote 9 on page 5, but note that the restriction $H = 0$ is a condition on *local* variation in quantiles of ε_i at \mathcal{X} given ε_{i+} equal to their quantiles at \mathcal{X} and given $Z = Z^*$.

The final identification conditions will be developed under the following Condition V which limits the dependence at \mathcal{X} of ε on Z and restricts the matrix J to be zero; recall J contains derivatives of iterated conditional quantiles of ε given Z with respect to elements of Z .

Condition V. Quantile insensitivity. *At \mathcal{X} the iterated conditional quantile functions, $Q_{\varepsilon_i|\varepsilon_{i+}Z}(\tau_i^*, \varepsilon_{i+}^*, z^*)$, have zero first partial derivatives with respect to Z at \mathcal{X} .*

This is not an essential condition since sufficient conditions elsewhere could compensate for a lack of restrictions on J , a point taken up when the returns to schooling example of Section 1.2 is revisited in Section 2.5.¹⁶

Adopting Condition V, and focussing attention on the identifiability of A , B and C , the equations (20) and (21) are rewritten in terms of these matrices, as follows.

$$\nabla_Y Q = I_M - C^{-1}(I_M - A) \quad (20)$$

$$\nabla_Z Q = C^{-1}B \quad (21)$$

Identification of elements of A , B and C clearly requires further restrictions because (20) and (21) contain only¹⁷ $M(M-1)/2 + MK$ informative equations in the $M(M-1) + MK$ unknown elements of A , B and C . This leads immediately to a necessary condition for identification of all elements of A , B and C .

Condition VI(i). Full system order condition. *Identifiability of all elements of A , B and C at \mathcal{X} requires that there exist at least $M(M-1)/2$ conditions on their free elements.*

It is now clear that the development of the final conditions to achieve identification will be similar to that encountered in the linear simultaneous equations model first studied in Frisch (1938) and Koopmans, Rubin and Leipnik (1950). A slight difference is that the attack employed here, via *iterated* conditional quantile functions, resolves to the problem of determining conditions under which structural “parameters” can be uniquely determined from knowledge of coefficients of a reduced *triangular* form rather than a conventional reduced form¹⁸.

To make progress it is convenient to extract from (20) the equations with content, that is those that derive from the *super*-diagonal triangle of $\nabla_Y Q$ and to express the resulting equations in terms of the, at this point, free elements of A , B and C . The

¹⁶Clearly, if B were a zero matrix *a priori* then J is immediately identified from (19) as $\nabla_Z Q$ but this has no identifying power with respect to C and so none for A . Restrictions on B allow slackening of the restriction $J = 0$ (or any *a priori* known matrix of constants) now imposed, while leaving open the possibility of identifying elements of C and A .

¹⁷The main diagonal and lower triangle of $\nabla_Y Q$ contains zeros because it contains derivatives of *iterated* conditional quantile functions. The structure of A and C ensures that $I_M - C^{-1}(I_M - A)$ contains zeros in these positions. So there are only $M(M-1)/2$ informative equations in (20).

¹⁸A conventional reduced form derives from the conditional distribution of Y given Z . A reduced triangular form derives from the conditional distributions of Y_i given Y_{i+} and Z .

matrix B is so far unrestricted, but A and C are upper triangular and their leading diagonal elements are restricted to be zeros in A and ones in C . Some new notation is required.

Let¹⁹ $b = \text{vec}(B)$, let $a = v(A)$ and $c = v(C)$ where the operator $v(\cdot)$ column stacks the *super*-diagonal elements of the square matrix to which it is applied²⁰.

Let R_M be the $M^2 \times M(M-1)/2$ matrix containing ones and zeros such that

$$\text{vec}(A) = R_M v(A). \quad (22)$$

In each column R_M contains zeros except for a single unit element in a different position in each column, therefore,

$$R'_M R_M = I_{M(M-1)/2}. \quad (23)$$

and the Moore-Penrose inverse of R_M is

$$R_M^+ = (R'_M R_M)^{-1} R'_M = R'_M.$$

It follows that, premultiplying in (22) by R_M^+ ,

$$v(A) = R'_M \text{vec}(A).$$

Since $C - I_M$ has the same structure as A it follows that

$$\text{vec}(C - I_M) = R_M v(C - I_M) = R_M v(C). \quad (24)$$

With this notation and the results concerning R_M in hand, there is, from²¹ (21)

$$(\nabla_Z Q' \otimes I_M) R_M c - b = -\text{vec}(D_Z) \quad (25)$$

and from²² (20)

$$(I_{M(M-1)/2} - R'_M (\nabla_Y Q' \otimes I_M) R_M) c + a = v(\nabla_Y Q).$$

¹⁹Here vec is the conventional column stacking operator described e.g., in Magnus and Neudecker (1988).

²⁰For example if $X = \{x_{ij}\}_{i,j=1}^3$, $v(X) = [x_{12}, x_{13}, x_{23}]'$. The column vectors a , b and c contain respectively $M(M-1)/2$, MK and $M(M-1)/2$ elements.

²¹From (21), $(C - I_M) D_Z + D_Z = B$, and on column stacking,

$$(D'_Z \otimes I_M) \text{vec}(C - I_M) + \text{vec}(D_Z) = \text{vec}(B)$$

which leads, on using (24), directly to equation (25).

²²From (20), $C - I_M - (C - I_M) D_Y + A = D_Y$, and, on column stacking

$$\text{vec}(C) - \text{vec}(I_M) - (D'_Y \otimes I_M) \text{vec}(C - I_M) + \text{vec}(A) = \text{vec}(D_Y).$$

Equation (26) results on using $\text{vec}(C - I_M) = R_M c$ and on premultiplying by R'_M which extracts the relevant super-diagonal elements. Note that $R'_M \text{vec}(I_M) = 0$.

and, by virtue of (23)

$$R'_M ((I_M - \nabla_Y Q') \otimes I_M) R_M c + a = v(\nabla_Y Q). \quad (26)$$

Now consider N restrictions on the elements of A , B and C , which hold *a priori* at \mathcal{X} , expressed as

$$W_A a + W_B b + W_C c = w$$

where W_A , W_B , and W_C are matrices, and w is a vector, of known constants²³. There is then the following equation

$$\begin{bmatrix} I_{M(M-1)/2} & 0 & R'_M ((I_M - \nabla_Y Q') \otimes I_M) R_M \\ 0 & -I_{MK} & (\nabla_Z Q' \otimes I_M) R_M \\ W_A & W_B & W_C \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} v(\nabla_Y Q) \\ -\text{vec}(\nabla_Z Q) \\ w \end{bmatrix} \quad (27)$$

which is written as

$$\Gamma \theta = \gamma \quad (28)$$

where Γ is the $(N + M(M-1)/2 + MK) \times (M(M-1) + MK)$ matrix on the left hand side of (27) and $\theta' = [a', b', c']$. There is the following necessary and sufficient condition for identification of all first partial derivatives of the structural equations at \mathcal{X} .

Condition VI(ii). Full system rank condition. *All elements of A , B and C are identified at \mathcal{X} if and only if*

$$\text{rank}(\Gamma) = M(M-1) + MK$$

for which a necessary condition is VI(i): $N \geq M(M-1)/2$.

Two special types of restriction trivially ensure that Condition VI(ii) is satisfied.

If the only restriction is that W_C is diagonal with rank M then Γ is square and upper triangular with nonzero main diagonal elements and therefore has full rank, as required by Condition VI(ii). This rather unlikely case would arise if, at \mathcal{X} , each structural equation i is insensitive to variations in ε_j , $j > i$, that is $G = I_M$ and the elements of ε are mutually independently distributed, that is, $H = 0$, for then C would be a diagonal matrix implying that $c = 0$.

Now suppose that the only restriction is $W_A = I_M$. In this case at \mathcal{X} each structural equation, i , is insensitive to Y_j , $j > i$, and the structural equations are locally (at least) in classical “reduced form”. Then Γ is again square and has the rank required by Condition V(ii) if

$$\text{rank}(R'_M ((I_M - \nabla_Y Q') \otimes I_M) R_M) = M(M-1)/2.$$

This always holds²⁴ and so the restriction $W_A = I_M$ is sufficient to ensure identification of B and C .

²³ W_A and W_C are $N \times M(M-1)/2$, W_B is $N \times MK$, and w is $N \times 1$.

²⁴Because

$$\text{rank}(R'_M ((I_M - \nabla_Y Q') \otimes I_M) R_M) = \min(\text{rank}(R_M), \text{rank}((I_M - \nabla_Y Q') \otimes I_M))$$

Single equation identification conditions. Now consider identification of the partial derivatives of a *single* equation, i , employing restrictions on the derivatives of that equation alone.

Some notational refinements are required. Let $\nabla_{Zi}Q'$ denote the i th row of ∇_ZQ and let $\nabla_{Zi+}Q'$ contain the last $M - i$ rows of ∇_ZQ . Let $\nabla_{Yi+}Q'$ be the lower right $(M - i) \times (M - i)$ block of ∇_YQ and let $\nabla_{Yi}Q'$ be the last $(M - i)$ elements in the i th row of ∇_YQ .

Let b'_i be the i th row of B and let c'_i and a'_i be the last $M - i$ elements of the i th rows of respectively C and A - these are the elements of C and A associated with equation i not restricted by virtue of the triangular structure.

Finally suppose that there are N_i restrictions at \mathcal{X} which apply to the derivatives of the i th structural equations, as follows.

$$W_{Ai}a_i + W_{Bi}b_i + W_{Ci}c_i = w_i$$

This equation and equations (20) and (21) imply the following

$$\begin{bmatrix} I_{M-i} & 0 & I_{M-i} - \nabla_{Yi+}Q \\ 0 & -I_K & \nabla_{Zi+}Q \\ W_{Ai} & W_{Bi} & W_{Ci} \end{bmatrix} \begin{bmatrix} a_i \\ b_i \\ c_i \end{bmatrix} = \begin{bmatrix} \nabla_{Yi}Q \\ -\nabla_{Zi}Q \\ w_i \end{bmatrix} \quad (29)$$

which is written as

$$\Gamma_i \theta_i = \gamma_i$$

where Γ_i is a $(M - i + K + N_i) \times (2(M - i) + K)$ matrix.

This leads to the following condition for identification of the partial derivatives of the i th structural equation based on restrictions on its derivatives alone.

Condition VI(iii). Rank and order conditions for identification of derivatives of equation i . *The derivatives of the i th structural equation are identified at the point \mathcal{X} under restrictions on its derivatives at \mathcal{X} alone if and only if $\text{rank}(\Gamma_i) = 2(M - i) + K$. A necessary condition is that $N_i \geq M - i$.*

A leading case of interest is that in which there are only local covariate “exclusion” restrictions. This arises when W_{Bi} consists of rows of I_K and $W_{Ai} = 0$, $W_{Ci} = 0$.

Some further notational refinement is required. Order and partition b_i so that the N_i *a priori* zero derivatives appear at the end of the vector, denote the remaining $K - N_i$ unrestricted derivatives by \hat{b}_i , reorder rows of $\nabla_{Zi}Q$ and $\nabla_{Zi+}Q$ and partition accordingly, thus :

$$\nabla_{Zi}Q = \begin{bmatrix} \nabla_{\hat{Z}i}Q \\ \nabla_{\check{Z}i}Q \end{bmatrix} \quad \nabla_{Zi+}Q = \begin{bmatrix} \nabla_{\hat{Z}i+}Q \\ \nabla_{\check{Z}i+}Q \end{bmatrix}$$

$$\begin{aligned} &= \min(M(M - 1)/2, M^2) \\ &= M(M - 1)/2 \end{aligned}$$

the second line here following because $I_M - \nabla_YQ'$ is lower triangular with nonzero leading diagonal elements and therefore has rank M .

where $\nabla_{\hat{z}_{i+}}Q$ is $(K - N_i) \times (M - i)$ with rows corresponding to the elements in \hat{b}_i , $\nabla_{\tilde{z}_{i+}}Q$ is $N_i \times (M - i)$ with rows corresponding to the *a priori* zero elements in b_i and $\nabla_{\hat{z}_i}Q$ and $\nabla_{\tilde{z}_i}Q$ are respectively $(K - N_i) \times 1$ and $N_i \times 1$.

Then (29) can be rewritten with *a priori* zero derivatives excluded as follows.

$$\begin{bmatrix} I_{M-i} & 0 & I_{M-i} - \nabla_{Y_{i+}}Q \\ 0 & -I_{K-N_i} & \nabla_{\hat{z}_{i+}}Q \\ 0 & 0 & \nabla_{\tilde{z}_{i+}}Q \end{bmatrix} \begin{bmatrix} a_i \\ \hat{b}_i \\ c_i \end{bmatrix} = \begin{bmatrix} \nabla_{Y_i}Q \\ -\nabla_{\hat{z}_i}Q \\ -\nabla_{\tilde{z}_i}Q \end{bmatrix} \quad (30)$$

Condition V(ii) provides a necessary condition for identification, namely that there be at least $M - i$ *a priori* zero derivatives of h_i with respect to Z at \mathcal{X} . The rank condition in this case is evidently: $\text{rank}(\nabla_{\tilde{z}_{i+}}Q) = M - i$.

When there is overidentification, solutions of (29) for the structural derivatives can be obtained by eliminating rows from the matrix on the left hand side of (29) and corresponding rows from the vector on the right hand side of (29). This leads directly, *via* the analogue principle, to (generally inefficient) estimators of the structural derivatives, after the style of Indirect Least Squares.

Let $\nabla_{\tilde{z}_{i+}}Q$ be a rank $M - i$ selection of rows from $\nabla_{\hat{z}_{i+}}Q$ and let $\nabla_{\tilde{z}_i}Q$ be the corresponding elements of $\nabla_{\hat{z}_i}Q$. Then (30) with $\nabla_{\hat{z}_{i+}}Q$ and $\nabla_{\hat{z}_i}Q$ replaced by respectively $\nabla_{\tilde{z}_{i+}}Q$ and $\nabla_{\tilde{z}_i}Q$ solves explicitly to give the following.

$$\begin{bmatrix} a_i \\ \hat{b}_i \\ c_i \end{bmatrix} = \begin{bmatrix} \nabla_{Y_i}Q + (I_{M-i} - \nabla_{Y_{i+}}Q) (\nabla_{\tilde{z}_{i+}}Q)^{-1} \nabla_{\tilde{z}_i}Q \\ \nabla_{\hat{z}_i}Q - \nabla_{\hat{z}_{i+}}Q (\nabla_{\tilde{z}_{i+}}Q)^{-1} \nabla_{\tilde{z}_i}Q \\ - (\nabla_{\tilde{z}_{i+}}Q)^{-1} \nabla_{\tilde{z}_i}Q \end{bmatrix} \quad (31)$$

As shown in the next Section this produces the results stated in the returns to schooling example of Section 1.2.

2.5. Example: Identification of returns to schooling, continued. It is instructive to work through the returns to schooling model of Section 1.2 using the matrices of structural and quantile derivatives employed in the previous Section. For this purpose, suppose that there are just two covariates, Z_1 and Z_2 .

Omitting arguments there are the following expressions for the matrices of derivatives at a point \mathcal{X} .

$$A = \begin{bmatrix} 0 & \nabla_{sw} \\ 0 & 0 \end{bmatrix} \quad B = \begin{bmatrix} \nabla_{Z_1}w & \nabla_{Z_2}w \\ \nabla_{Z_1}s & \nabla_{Z_2}s \end{bmatrix} \quad G = \begin{bmatrix} 1 & \nabla_A w \\ 0 & 1 \end{bmatrix} \quad H = \begin{bmatrix} 0 & \nabla_A Q_{F|AZ} \\ 0 & 0 \end{bmatrix}$$

In the matrix H the nonzero term is the derivative with respect to A of the conditional quantile of F given A and $Z = \{Z_i\}_{i=1}^2$, evaluated at \mathcal{X} . Separate restrictions on this derivative and on $\nabla_A w$ are not considered here and so the analysis proceeds in terms of $C = G(I_2 - H)^{-1}$, which is

$$C = \begin{bmatrix} 1 & \nabla_A w + \nabla_A Q_{F|AZ} \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & c_{12} \\ 0 & 1 \end{bmatrix}.$$

The matrices on the right hand side of equations (20) and (21), are then as follows.

$$I_M - C^{-1}(I_M - A) = \begin{bmatrix} 0 & (\nabla_S w + c_{12}) \\ 0 & 0 \end{bmatrix} \quad (32)$$

$$C^{-1}B = \begin{bmatrix} (\nabla_{Z_1} w - (\nabla_{Z_1} s) c_{12}) & (\nabla_{Z_2} w - (\nabla_{Z_2} s) c_{12}) \\ \nabla_{Z_1} s & \nabla_{Z_2} s \end{bmatrix} \quad (33)$$

The matrices of derivatives of the iterated conditional quantile functions in equations (20) and (21) evaluated at \mathcal{X} are as follows.

$$\nabla_Y Q = \begin{bmatrix} 0 & \nabla_S Q_{W|SZ} \\ 0 & 0 \end{bmatrix} \quad \nabla_Z Q = \begin{bmatrix} \nabla_{Z_1} Q_{W|SZ} & \nabla_{Z_2} Q_{W|SZ} \\ \nabla_{Z_1} Q_{S|Z} & \nabla_{Z_2} Q_{S|Z} \end{bmatrix} \quad (34)$$

The structural derivatives of the schooling equation, already in full “reduced form” are immediately identifiable²⁵ as the corresponding derivatives of the conditional quantile function $Q_{S|Z}$. But the derivatives of the wage equation ($i = 1$) are not, and at least $N_1 \geq M - 1 = 1$ *a priori* restrictions are necessary to achieve identification (Condition V(iii)).

Suppose $\nabla_{Z_2} w$ is restricted to be zero at \mathcal{X} . Then (33) simplifies as follows

$$C^{-1}B = \begin{bmatrix} (\nabla_{Z_1} w - (\nabla_{Z_1} s) c_{12}) & -(\nabla_{Z_2} s) c_{12} \\ \nabla_{Z_1} s & \nabla_{Z_2} s \end{bmatrix}$$

the vectors a_1 , \hat{b}_1 and c_1 (all 1×1) are

$$a_1 = [\nabla_S w] \quad \hat{b}_1 = [\nabla_{Z_1} w] \quad c_1 = [c_{12}]$$

and the equivalent of equation (30) is the following.

$$\begin{bmatrix} 1 & 0 & 1 \\ 0 & -1 & \nabla_{Z_1} Q_{S|Z} \\ 0 & 0 & \nabla_{Z_2} Q_{S|Z} \end{bmatrix} \begin{bmatrix} \nabla_S w \\ \nabla_{Z_1} w \\ c_{12} \end{bmatrix} = \begin{bmatrix} \nabla_S Q_{W|SZ} \\ -\nabla_{Z_1} Q_{W|SZ} \\ -\nabla_{Z_2} Q_{W|SZ} \end{bmatrix}$$

The matrix on the left side of this expression has rank 3 if $\nabla_{Z_2} Q_{S|Z} \neq 0$ at \mathcal{X} . Under that rank condition (Condition V(iii)) we have

$$\begin{bmatrix} \nabla_S w \\ \nabla_{Z_1} w \\ c_{12} \end{bmatrix} = \begin{bmatrix} 1 & 0 & -\left(\frac{1}{\nabla_{Z_2} Q_{S|Z}}\right) \\ 0 & -1 & \left(\frac{\nabla_{Z_1} Q_{S|Z}}{\nabla_{Z_2} Q_{S|Z}}\right) \\ 0 & 0 & \left(\frac{1}{\nabla_{Z_2} Q_{S|Z}}\right) \end{bmatrix} \begin{bmatrix} \nabla_S Q_{W|SZ} \\ -\nabla_{Z_1} Q_{W|SZ} \\ -\nabla_{Z_2} Q_{W|SZ} \end{bmatrix}$$

²⁵ Compare the second rows of $\nabla_Z Q$ in equation (34) with the second row of the matrix in equation (33).

and so, the equivalent of (31),

$$\begin{bmatrix} \nabla_S w \\ \nabla_{Z_1} w \\ c_{12} \end{bmatrix} = \begin{bmatrix} \nabla_S Q_{W|SZ} + \left(\frac{\nabla_{Z_2} Q_{W|SZ}}{\nabla_{Z_2} Q_{S|Z}} \right) \\ \nabla_{Z_1} Q_{W|SZ} - \nabla_{Z_2} Q_{W|SZ} \left(\frac{\nabla_{Z_1} Q_{S|Z}}{\nabla_{Z_2} Q_{S|Z}} \right) \\ - \left(\frac{\nabla_{Z_2} Q_{W|SZ}}{\nabla_{Z_2} Q_{S|Z}} \right) \end{bmatrix} \quad (35)$$

which confirms the results (2) and (3) of Section 1.2 on setting $i = 2$ in that example.

Over identification. Now suppose there is over identification with both $\nabla_{Z_1} w$ and $\nabla_{Z_2} w$ a priori zero at \mathcal{X} . In this returns to schooling example Z_1 and Z_2 might be the incomes of the mother and father of the person at the time that the schooling investment was made.

Equation (33) now simplifies as follows

$$C^{-1}B = \begin{bmatrix} -(\nabla_{Z_1} s) c_{12} & -(\nabla_{Z_2} s) c_{12} \\ \nabla_{Z_1} s & \nabla_{Z_2} s \end{bmatrix}, \quad (36)$$

the vectors a_1 , \hat{b}_1 and c_1 are

$$a_1 = [\nabla_S w] \quad \hat{b}_1 = \emptyset \quad c_1 = [c_{12}]$$

and the equivalent of equation (30) is the following.

$$\begin{bmatrix} 1 & 1 \\ 0 & \nabla_{Z_1} Q_{S|Z} \\ 0 & \nabla_{Z_2} Q_{S|Z} \end{bmatrix} \begin{bmatrix} \nabla_S w \\ c_{12} \end{bmatrix} = \begin{bmatrix} \nabla_S Q_{W|SZ} \\ -\nabla_{Z_1} Q_{W|SZ} \\ -\nabla_{Z_2} Q_{W|SZ} \end{bmatrix} \quad (37)$$

There are now three equations in just two unknown quantities. Let R be the matrix

$$R = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \rho_1 & \rho_2 \end{bmatrix}$$

restricted to have rank 2, and premultiply (37) to give

$$\begin{bmatrix} 1 & 1 \\ 0 & (\rho_1 \nabla_{Z_1} Q_{S|Z} + \rho_2 \nabla_{Z_2} Q_{S|Z}) \end{bmatrix} \begin{bmatrix} \nabla_S w \\ c_{12} \end{bmatrix} = \begin{bmatrix} \nabla_S Q_{W|SZ} \\ -(\rho_1 \nabla_{Z_1} Q_{W|SZ} + \rho_2 \nabla_{Z_2} Q_{W|SZ}) \end{bmatrix}$$

from which

$$\begin{bmatrix} \nabla_S w \\ c_{12} \end{bmatrix} = \begin{bmatrix} \nabla_S Q_{W|SZ} + E \\ -E \end{bmatrix}$$

where

$$E = \frac{\rho_1 \nabla_{Z_1} Q_{W|SZ} + \rho_2 \nabla_{Z_2} Q_{W|SZ}}{\rho_1 \nabla_{Z_1} Q_{S|Z} + \rho_2 \nabla_{Z_2} Q_{S|Z}} = \frac{1}{\lambda_1 + \lambda_2} \left(\lambda_1 \frac{\nabla_{Z_1} Q_{W|SZ}}{\nabla_{Z_1} Q_{S|Z}} + \lambda_2 \frac{\nabla_{Z_2} Q_{W|SZ}}{\nabla_{Z_2} Q_{S|Z}} \right)$$

and $\lambda_i = \rho_i \nabla_{Z_i} Q_{S|Z}$, $i \in \{1, 2\}$. In this case there are infinitely many ways of identifying the two unrestricted structural derivatives in the wage equation, employing linear combinations of the ratios of iterated conditional quantile derivatives, $\frac{\nabla_{Z_i} Q_{W|SZ}}{\nabla_{Z_i} Q_{S|Z}}$, $i \in \{1, 2\}$. At the point of estimation one will wish to consider optimal choices of the ρ_i 's, an issue deferred to Section 3.

The analysis of this returns to schooling example has so far proceeded under the quantile insensitivity condition V, but in this overidentified case it is possible to weaken this condition. For example, suppose that the condition only applied for variation in Z_2 . Then (36) would become

$$C^{-1}B = \begin{bmatrix} -(\nabla_{Z_1} s) c_{12} + j_{11} & -(\nabla_{Z_2} s) c_{12} \\ \nabla_{Z_1} s + j_{21} & \nabla_{Z_2} s \end{bmatrix}$$

where $j_{11} = \nabla_{Z_1} Q_{F|AZ}$, $j_{21} = \nabla_{Z_1} Q_{A|Z}$, both evaluated at \mathcal{X} . Now $\nabla_{Z_1} s$ is no longer identifiable and the zero restriction on $\nabla_{Z_1} w$ has no force, but there remains

$$\begin{bmatrix} 1 & 1 \\ 0 & \nabla_{Z_2} Q_{S|Z} \end{bmatrix} \begin{bmatrix} \nabla_S w \\ c_{12} \end{bmatrix} = \begin{bmatrix} \nabla_S Q_{W|SZ} \\ -\nabla_{Z_2} Q_{W|SZ} \end{bmatrix}$$

which identifies the returns to schooling, $\nabla_S w$, exactly as set out in equation (35).

3. ESTIMATION

Estimation of A and B , the derivatives of the structural equations at the point \mathcal{X} , can proceed using the analogue principle set out in Manski (1988). There are the following steps.

1. Probability levels, τ^* , are selected and the required iterated conditional quantile functions of Y_i given Y_{i+} and Z , $i \in \{1, \dots, M\}$ are estimated using a parametric, semi- or nonparametric method, as desired²⁶.
2. A value of Z , z^* , is selected and the values of Y_1, \dots, Y_M at the point \mathcal{X} are estimated using the estimated iterated conditional quantile functions, producing an estimate $\hat{\mathcal{X}}$ of the location of \mathcal{X} .
3. The estimated iterated conditional quantile functions are used to produce estimates of derivatives of the iterated conditional quantile functions which are evaluated at $\hat{\mathcal{X}}$, yielding estimates, $\widehat{\nabla_Y Q}$ and $\widehat{\nabla_Z Q}$, of values of the matrices $\nabla_Y Q$ and $\nabla_Z Q$ at \mathcal{X} .

²⁶For parametric estimation, see Koenker and Bassett (1978), Koenker and d'Orey (1987); for semiparametric estimation see Chaudhuri, Doksum and Samarov (1997), Kahn (2001) and Lee (2002); for nonparametric estimation, see Chaudhuri (1991).

4. The restrictions on A , B and C are assembled as in (27), or, in a single equation analysis as in equation (29), with $\widehat{\nabla_Y Q}$ and $\widehat{\nabla_Z Q}$ replacing $\nabla_Y Q$ and $\nabla_Z Q$ and subject to satisfaction of the order condition, and subject to satisfaction of the rank condition at the estimated values of $\nabla_Y Q$ and $\nabla_Z Q$, the equation is solved for estimates of the desired elements of A and B .

At step 4, if there are abundant restrictions, for example when $N > M(M-1)/2$, then there is overidentification and there is unlikely to be a solution. Solutions can be obtained by eliminating restrictions so that the order condition is exactly satisfied, but there will be many ways of doing this, each leading in general to an inefficient estimator.

A solution is to retain all the restrictions and employ a minimum distance estimation procedure in place of step 4. For example an estimator of all derivatives at \mathcal{X} could be obtained as

$$\hat{\theta} = \arg \min_{\theta} \left(\hat{\Gamma}\theta - \hat{\gamma} \right)' \Omega \left(\hat{\Gamma}\theta - \hat{\gamma} \right)$$

for a suitable choice of weight matrix Ω where $\hat{\Gamma}$ and $\hat{\gamma}$ are Γ and γ of equation (28) with parametric, semi- or nonparametric estimates $\widehat{\nabla_Y Q}$ and $\widehat{\nabla_Z Q}$ replacing $\nabla_Y Q$ and $\nabla_Z Q$. Of course the sampling properties of $\hat{\theta}$ will depend upon the estimation procedure chosen. One could consider imposing the identifying restrictions at step 1.

3.1. Linear location shift models. The estimation of coefficients of a single median regression function derived from a linear location shift model in the presence of endogenous variables has been considered by Amemiya (1982) who proposed a family of Two Stage Least Absolute Deviations (2SLAD) estimators. Consider the linear location shift model

$$\begin{aligned} Y_1 &= \gamma_2 Y_2 + \dots + \gamma_M Y_M + Z' \beta_1 + \varepsilon_1 \\ Y_j &= Z' \beta_j + \varepsilon_j \quad j \in \{2, \dots, M\} \end{aligned}$$

which is a highly restricted version of the model (5) considered here.

At the first stage of 2SLAD predicted values of the endogenous covariates, $\{Y_j\}_{j=2}^M$, conditional on all covariates (Z) are produced. Amemiya suggests OLS estimation and LAD estimation as possibilities. In a leading special case of 2SLAD, at the second stage an LAD estimation of the equation for Y_1 is conducted using predicted values of endogenous covariates in place of their actual values.

The procedure outlined above is an alternative to 2SLAD and clearly extends to quantile regressions other than the median regression and to parametric nonlinear quantile regressions.

In the linear case one first estimates iterated linear quantile regressions of Y_i given Y_j , $j > i$, and Z , using LAD if median regressions are of interest. The resulting estimated coefficients, which are of course estimated (global) derivatives of the structural

functions, are then processed as indicated above. In a just identified problem this procedure is, to 2SLAD as Indirect Least Squares (based on a reduced triangular form) is to 2SLS. First order asymptotic properties of the resulting estimators are straightforwardly derived *via* the delta method.

4. AVERAGE STRUCTURAL DERIVATIVES

The results so far have concerned local nonparametric identification of structural derivatives evaluated at points defined by a value, z^* , of a vector of covariates Z and by M probabilities, τ^* , defining iterated conditional quantiles of unobservable continuously distributed stochastic drivers, ε .

If such identification can be achieved at z^* at all combinations of the M probability levels each varying in $[0, 1]$ then it is a simple matter to identify conditional (on $Z = z^*$) expected values of structural derivatives and of functions of them, should such objects be of interest.

To see how this is achieved consider a function $g(\varepsilon, Z)$ of continuously distributed ε and its conditional expectation (assumed to exist) given $Z = z$

$$E_{\varepsilon|Z}[g(\varepsilon, Z)|Z = z] = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} g(e, z) dF_{\varepsilon|Z}(e|z).$$

Here $F_{\varepsilon|Z}(\cdot|\cdot)$ is the conditional distribution function of ε given Z .

Let $\tilde{g}(\tau, Z) = g(\varepsilon(\tau), Z)$ where $\tau = \{\tau_i\}_{i=1}^M$, and $\varepsilon(\tau) = \{\varepsilon_i(\tau)\}_{i=1}^M$ whose elements are defined recursively in terms of conditional quantile functions as follows.

$$\varepsilon_i(\tau) = Q_{\varepsilon_i|\varepsilon_{i+}Z}(\tau_i|\varepsilon_{i+}, z), \quad i \in \{1, \dots, M\}$$

The elements of τ satisfy

$$\tau_i = F_{\varepsilon_i|\varepsilon_{i+}Z}(\varepsilon_i|\varepsilon_{i+}, z), \quad i \in \{1, \dots, M\},$$

where $F_{\varepsilon_i|\varepsilon_{i+}Z}$ is the conditional distribution function of ε_i given ε_{i+} and Z , and they are mutually independently uniformly distributed, each on $[0, 1]$, so

$$\begin{aligned} E_{\varepsilon|Z}[g(\varepsilon, Z)|Z = z] &= E_{\tau}[g(\varepsilon(\tau), z)] \\ &= \int_0^1 \dots \int_0^1 g(\varepsilon(t), z) dt_1 \dots dt_M. \end{aligned} \quad (38)$$

The structural derivatives evaluated at the point \mathcal{X} considered in Section 2 are expressed, not as functions of the unobservable stochastic drivers (ε), but as functions of probabilities (τ^*) defining iterated conditional quantiles of the conditional distribution of ε given $Z = z^*$.

The result (38) implies that expected values of functions of structural derivatives (and so, for example, average derivatives, variances of derivatives, and so forth) can

be identified, if the local identification conditions hold at $Z = z^*$ for all combinations of quantile probability levels. The conditional expected value of a function $r(\cdot)$ of a structural derivative $\pi(\tau, z^*)$ is simply $\int_0^1 \dots \int_0^1 r(\pi(t, z^*)) dt_1 \dots dt_M$.

Identification of complete expected values of functions of structural derivatives requires essentially *full independence* of ε and Z because Condition V (quantile insensitivity) has to hold at all quantile probabilities. Less demanding restrictions are required to identify the *conditional* expected value of a function of structural derivatives, conditional on ε lying in some set of values, T , say.

Define T_τ as the following set of conditional quantile probabilities

$$T_\tau = \{\tau : \varepsilon(\tau) \in T\}.$$

If there is conditional quantile (of ε given Z) insensitivity at z^* and at the quantile probabilities in T_τ then

$$\begin{aligned} E_{\varepsilon|Z}[g(\varepsilon, Z)|Z = z^*, \varepsilon \in T] &= E_\tau[g(\varepsilon(\tau), z^*)|\tau \in T_\tau] \\ &= \frac{\int \dots \int_{t \in T_\tau} g(\varepsilon(t), z^*) dt_1 \dots dt_M}{\int \dots \int_{t \in T_\tau} dt_1 \dots dt_M}. \end{aligned}$$

which follows by straightforward extension of the argument earlier in this Section.

5. CONCLUDING REMARKS

This paper has shown that in nonseparable models, weak conditions under which there is nonparametric identification of certain interesting features of structures can be obtained if the question of identifiability is couched in terms of the feasibility of uniquely deducing features of structural models from knowledge of an exhaustive set of iterated conditional quantile functions. An analysis via conditional quantile functions is well suited to potentially nonseparable models because of the equivariance property of quantiles.

A critical restriction introduced at the start of this analysis is that there are no more unobservable stochastic drivers of the model than observable outcomes. This condition will not be met in many models of processes in which measurement error distorts recorded outcomes or covariates, or in models with high dimensional heterogeneity such as the mixture models employed widely in the analysis of durations, and in panel data and other multilevel models in which there is a nested error structure. Identification in such models obviously requires very strong restrictions because one is trying to secure identification of features of high dimensional distributions from relatively low dimensional reductions of them. In practice one finds at least strong additivity assumptions imposed in every such case.

For example all measurement error models employed in practice require measurement error to be additive in some metric and with very few exceptions impose

parametric restrictions on model equations. Similarly, panel data models typically require the disturbances that drive a model to be linear combinations of individual specific disturbances and disturbances that vary across and within individuals, and the widely employed mixed proportionate hazard models restrict individual specific disturbances to be additive with the log hazard function.

One of the conclusions to be drawn from the results of this paper is that non-parametrically distinguishing sources of stochastic variation essentially requires that there be no more sources of stochastic variation than there are observed outcomes. This suggests that if features of processes that exhibit stochastic variation are to be robustly estimated one should have data on as many outcomes as there are sources of stochastic variation.

A further critical restriction is the triangular structure of the model equations. All complete M equation models at a point \mathcal{X} of interest (that is models which deliver a unique value of outcomes at \mathcal{X}) can be written in a local triangular form in any of $M!$ ways, each corresponding to a distinct ordering of the M outcomes. Of course a set of iterated conditional quantile functions can be defined corresponding to each of these representations. The crucial feature of the models addressed in this paper is that the disturbances feature in the model equations following the same triangular structure (or a restricted version of it) as the outcome variables. It does not seem possible to take a model in which outcomes are not in a triangular form and achieve a similar local triangular representation for outcomes and disturbances without strong restrictions on the manner in which the disturbances appear in the original model.

Identification in semiparametric and parametric nonseparable (and separable) models can be assessed using the results of this paper even though a nonparametric attack has been taken. Because this was felt to be important some normalisations which would often be taken in a study of purely nonparametric identification have not been made in this paper. Semiparametric and parametric special cases produce restrictions on the matrices of derivatives introduced in Section 2 which may be helpful in achieving identification. For example index restrictions that lead model equations of the form:

$$Y_i = h_i(Y_{i+}'\theta_i, Z'\beta_i, \varepsilon_i, \varepsilon_{i+})$$

require the matrices A and B of Section 2 to be equal to a diagonal matrices²⁷ times matrices of constants (containing elements of θ_i and β_i). Conditions sufficient to ensure *local* identification of these restricted matrices ensure *global* identification of *ratios* of θ parameters and ratios of β parameters taken within equations.

Focusing on local identification of the derivatives of structural functions at a point of interest allows progress without restricting disturbances and covariates to be distributed independently. Identification conditions can be expressed in terms of local features of models and when considering restrictions on model equations the problem reduces to a search for identification conditions in a linear approximation to the nonlinear structure, as long as there is sufficient smoothness to allow linear

²⁷For A and B respectively equal to $\text{diag}\{\nabla_{Y_{i+}'\theta_i} h_i\}_{i=1}^M$ and $\text{diag}\{\nabla_{Z'\beta} h_i\}_{i=1}^M$.

approximation at the point of interest. As a result the familiar order and rank conditions for identification in the linear simultaneous equations system are echoed in the results.

A quantile based analysis naturally leads to identification of structural features which convey information about the distribution of exogenous impacts of variables perhaps endogenous in the data generating process. This is of value in the analysis of policy interventions. An advantage of quantile independence restrictions is that they have meaning in problems where models have stochastic drivers which are naturally modelled as having no moments such as arise in financial econometrics.

The estimation procedures that flow from the identification conditions involve various sorts of quantile regression function estimation, whose properties are well understood.

Quantile regression methods have frequently been advocated because they may be less sensitive to data contamination and because they can provide a wealth of distributional information. The results of this paper suggest a further, and very important, virtue of quantile regression methods, namely that, because of their equivariance property, exploited here, they are the natural tool to employ in identification, estimation and inference in nonlinear, potentially nonseparable, structural models.

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.
- BLUNDELL, R.W., AND J.L. POWELL (2000): "Endogeneity in Nonparametric and Semiparametric Regression Models," invited lecture, 2000 World Congress of the Econometric Society, Seattle.
- 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.
- CARD, D., (2001): "Estimating the returns to schooling: Progress on some persistent econometric problems," *Econometrica*, 69, 1127-1160.
- 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., (2001b): "Quantile driven identification of structural derivatives," Centre for Microdata Methods and Practice Working Paper 08/01.
- 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., 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.
- HORNIK, K., (2002): "R FAQ: Frequently asked questions on R," Version 1.4-22, 2002-03-08, available at: <http://www.ci.tuwien.ac.at/~hornik/R/>.
- 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," mimeo.

- 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.
- LEE, S., (2002): "Efficient semiparametric estimation of a partially linear quantile regression model," forthcoming *Econometric Theory*.
- MAGNUS, J.R., AND H. NEUDECKER, (1988): *Matrix differential calculus with applications in statistics and econometrics*, New York: John Wiley.
- 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.
- NEWKEY, W.K., AND J.L. POWELL (1988): "Nonparametric Instrumental Variables Estimation," mimeo, Department of Economics, MIT.
- NEWKEY, 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.
- NEWKEY, 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.