# Identification in Nonseparable Models 

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

January 31st 2003

Abstract. Weak nonparametric restrictions are developed, sufficient to identify the values of derivatives of structural functions in which latent random variables are nonseparable. These derivatives can exhibit stochastic variation. In a micreconometric context this allows the impact of a policy intervention, as measured by the value of a structural derivative, to vary across people who are identical as measured by covariates. When the restrictions are satisfied quantiles of the distribution of a policy impact across people can be identified. The identification restrictions are local in the sense that they are specific to the values of the covariates and the specific quantiles of latent variables at which identification is sought. The conditions do not include the commonly required independence of latent variables and covariates. They include local versions of the classical rank and order conditions and local quantile insensitivity conditions. Values of structural derivatives are identified by functionals of quantile regression functions and can be estimated using the same functionals applied to estimated quantile regression functions.

## 1. Introduction

This paper develops weak conditions under which there is nonparametric identification of values, at some specified point, of derivatives of structural functions from which latent variables may not be separable. The conditions are local in the sense that they are specific to the derivatives and to the point of interest. They point to analogue estimators built on estimators of conditional quantile regression functions.

Structures in which latent random variables are nonseparable constitute a flexible construction, sympathetic to the qualitative nature of the information about economic processes that economic theory provides.

[^0]In nonseparable structures the sensitivity of outcomes to changes in conditioning and other variables has random variation. In a microeconometric context this admits the possibility that a policy intervention has effects which vary across individuals that, measured by covariates, are identical. Policy analysis is enriched if features of the distribution of a policy impact across people can be identified. Conditions under which this can be done are provided in this paper.

A primary aim of this paper is to provide minimal restrictions sufficient to identify local features of structures. Neither semiparametric nor parametric restrictions are considered, and the local focus allows progress without the strong and unpalatable restriction requiring latent variates and covariates to be independently distributed, a restriction commonly invoked when considering identification in nonseparable models.

The quest for weak identifying conditions is motivated by the observation that any structural interpretation of data is contingent on the veracity of a core set of untestable just-identifying restrictions. It is good if these conditions can be focussed on the particular structural features of interest if that results, as it does here, in the conditions being less demanding than would otherwise be required.

Key among the identifying conditions are restrictions on the covariate related variation in conditional quantiles of latent variates. Conditional quantile restrictions are considered because in nonseparable structures they can have substantive implications for the information about structural equations that is contained in the conditional quantiles of outcomes given covariates about which data can be informative. In contrast, conditional mean and higher order moment restrictions do not carry such valuable information when structural functions are nonseparable unless there are strong restrictions on functional form or distributional shape.

The restrictions at the heart of the identifying model proposed in this paper are now introduced and then the approach taken in demonstrating the identifying power of the model is described The main results are then introduced in the context of a two equation model of the sort that arises when considering the returns to schooling. This first Section concludes with a plan of the rest of the paper.
1.1. The core restrictions. The identification conditions include four fundamental types of restriction, as follows.

1. Smoothness and continuous variation. Structural functions are differentiable with respect to outcomes and with respect to certain covariates which exhibit continuous variation.
2. Triangularity. Structural equations have a triangular form in outcomes and in latent random variables.
3. No excess variation. In structures which deliver values of $M$ outcomes there are effectively no more than $M$ latent random variables.
4. Monotonicity. Each structural function is a strictly monotonic function of one of the latent random variables.

Under restrictions like these, made precise later, functions of derivatives of structural functions are identified by derivatives of certain conditional quantile functions of outcomes. The value of a particular structural derivative is identifiable if it can be deduced from knowledge of the values of these identifiable functions of structural derivatives. Additional restrictions, namely local conditional quantile insensitivity conditions and local rank conditions are shown to be sufficient for this purpose.

This is reminiscent of the development of identification conditions under conditional mean independence in parametric linear simultaneous equations models set out in Koopmans, Rubin and Leipnik (1950). That analysis echoes here because under the proposed restrictions, local to the point at which knowledge of the values of structural derivatives is desired, structural equations are approximately linear with coefficients whose values are the values of the structural derivatives of interest.
1.2. Constructive identification. A constructive approach to demonstrating the identifying power of restrictions is taken. The definitions of a structure and of identification of a feature of a structure set out in Hurwicz (1950) and employed in Koopmans and Reiersøl (1950) are used ${ }^{1}$.

A structure is:

1. a system of equations delivering unique values of observable outcomes given values of covariates and latent variates, and,
2. a conditional probability distribution of latent variates given covariates.

Each structure implies a conditional distribution of outcomes given covariates. The same conditional distribution may be generated by distinct structures. If, among these observationally equivalent structures, a structural feature takes different values then its value in the data generating structure cannot be identified.

A feature of a structure is identified if, among any set of observationally equivalent admissible structures, the value of the structural feature does not vary. The restrictions that define admissible structures constitute a model.

It is shown in Chesher (2002a) that the value of a structural feature $f$ is identified by a model when $f=f^{*}$ if there exists a functional of the conditional distribution function of outcomes given covariates, $\mathcal{G}\left(F_{Y \mid X}\right)$ with the property that the functional returns the value $f^{*}$ in all structures admitted by the model in which $f=f^{*}$.

When identification can be demonstrated by this means the structural feature $f$ will be said to be identified by the functional $\mathcal{G}\left(F_{Y \mid X}\right) .{ }^{2}$ Such a demonstration is constructive in the sense that it points to analogue estimators of the value of the structural feature, for example the functional $\mathcal{G}(\cdot)$ applied to an estimator of $F_{Y \mid X}$.

[^1]The approach to developing identifying restrictions taken in this paper is to construct a model embodying weak restrictions such that a functional with the required property exists when the structural feature of interest is the value of a structural derivative at a specified point. The form of the functional is determined. Section 3 does this for an $M$ equation system of nonseparable structural equations. That necessarily involves some notational complexity so the key ideas are now introduced in the context of a simple example which will be revisited later in the paper.
1.3. Returns to schooling. Here is an example of a nonseparable model that arises when considering the returns to schooling ${ }^{3}$. Let $W$ be the log wage and let $S$ be a measure of investment in schooling, determined as unique solutions to the equations:

$$
\begin{align*}
W & =h_{W}(S, Z, F, A)  \tag{1}\\
S & =h_{S}(Z, A) \tag{2}
\end{align*}
$$

where $Z$ is a list of covariates and $h_{W}$ and $h_{S}$ are differentiable functions. The equations embody a triangularity restriction and the restriction that, given a value of $Z$, random variation in the two observable outcomes, $W$ and $S$, is generated by random variation in two latent random variables, $F$ and $A$, which will be interpreted as capturing respectively fortune in the labour market and ability.

Of particular interest in this example is the returns to schooling, $\nabla_{S} h_{W}$, the first partial derivative of the function $h_{W}$ with respect to $S$. In the structures admitted by this model the returns to schooling may vary with $S$ and $Z$ because $h_{W}$ may be a nonlinear function of $S$ and $Z$, and may vary with $F$ and $A$ because $F$ and $A$ may not be additively separable from the function $h_{W}$.

The results given in this paper provide weak nonparametric conditions sufficient to identify the value of $\nabla_{S} h_{W}$ at chosen values of $S$ and $Z$ and of the latent variates $F$ and $A$.

Key among these conditions are restrictions on the way in which structural functions can vary with the latent variates. These restrictions are satisfied if the functions $h_{W}$ and $h_{S}$ are strictly monotonic functions of respectively $F$ and $A$. This leads naturally to consideration of identification via restrictions on conditional quantiles of the latent variables. To see why this is so, consider the schooling equation.

Let $F_{A \mid Z}(a \mid z) \equiv P[A \leq a \mid Z=z]$ denote the distribution function of $A$ given

[^2]$Z=z$ and define the conditional $\tau$-quantile of $A$ given $Z=z$ as follows. ${ }^{4}$
$$
Q_{A \mid Z}(\tau, z) \equiv \inf \left\{q: F_{A \mid Z}(a \mid z) \geq \tau\right\}
$$

With the function $h_{S}(Z, A)$ restricted to be monotonic in $A$ when $Z=z$, normalised to be non-decreasing, the conditional $\tau$-quantile of $S$ given $Z=z$ is, because of the equivariance property of quantiles:

$$
Q_{S \mid Z}(\tau, z)=h_{S}\left(z, Q_{A \mid Z}(\tau, z)\right)
$$

and the $z$-derivative of $Q_{S \mid Z}(\tau, z)$, assumed to exist, is as follows.

$$
\nabla_{z} Q_{S \mid Z}(\tau, z)=\left.\nabla_{z} h_{S}(z, a)\right|_{a=Q_{A \mid Z}(\tau, z)}+\left.\nabla_{a} h_{S}(z, a)\right|_{a=Q_{A \mid Z}(\tau, z)} \times \nabla_{z} Q_{A \mid Z}(\tau, z)
$$

Consider the following quantile insensitivity restriction at $z=\bar{z}$.

$$
\begin{equation*}
\left.\nabla_{z} Q_{A \mid Z}(\tau, z)\right|_{z=\bar{z}}=0 \tag{3}
\end{equation*}
$$

This implies that the $Z$-derivative of the conditional $\tau$-quantile of $S$ given $Z$ at $Z=\bar{z}$ is as follows.

$$
\left.\nabla_{z} Q_{S \mid Z}(\tau, z)\right|_{z=\bar{z}}=\left.\nabla_{z} h_{S}(z, a)\right|_{a=Q_{A \mid Z}(\tau, \bar{z}), z=\bar{z}}
$$

Under the quantile insensitivity restriction (3) the $Z$-variation in the conditional $\tau$ quantile of $S$ given $Z$ at $Z=\bar{z}$, about which data can be informative, is identical to the $Z$-variation in the function $h_{S}(Z, A)$ at $Z=\bar{z}$ with $A$ fixed at $A=Q_{A \mid Z}(\tau, \bar{z})$. The $\tau$-quantile insensitivity restriction on the distribution of $A$ given $Z$ at $Z=\bar{z}$ leads directly to identification of the value of the $Z$-derivative of $h_{S}(Z, A)$ at $Z=\bar{z}$ and $A=Q_{A \mid Z}(\tau, \bar{z})$.

In contrast, under the conditional mean restriction, $E[A \mid Z=z]=0$, the way in which functionals of the distribution of $S$ given $Z$, for example $E[S \mid Z=z]$, vary with $z$ depends on the way in which $h_{S}(Z, A)$ varies with $Z$ and $A$ and can depend upon the shape of the distribution of $A$ given $Z$. Identification conditions built on conditional mean and higher order integer moment conditions require stronger restrictions on structural functions than identification conditions built on the conditional quantile restrictions that are used in this paper. ${ }^{5}$

The functions $h_{W}$ and $h_{S}$ are now restricted to be strictly monotonic in respectively $F$ and $A$, both normalised to be increasing in these arguments. ${ }^{6}$

Let $\bar{z}$, be a chosen value of $Z$, let $\bar{a}$ be the $\bar{\tau}_{A}$-quantile of $A$ given $Z=\bar{z}$, and let $\bar{f}$ be the $\bar{\tau}_{F}$-quantile of $F$ given $A=\bar{a}$ and $Z=\bar{z}$. Define $\bar{\Omega} \equiv\left(\bar{z}, \bar{\tau}_{A}, \bar{\tau}_{F}\right), \bar{s} \equiv h_{S}(\bar{z}, \bar{a})$, $\bar{w} \equiv h_{W}(\bar{s}, \bar{z}, \bar{f}, \bar{a})$ and $\bar{\Psi} \equiv(\bar{s}, \bar{z}, \bar{f}, \bar{a})$.

[^3]Consider the value of the returns to schooling at $\bar{\Psi}$.

$$
\left.\nabla_{S} h_{W}(\bar{\Psi}) \equiv \nabla_{S} h_{W}(S, Z, F, A)\right|_{\bar{\Psi}}
$$

This is the returns to schooling for a person with $Z=\bar{z}$, who is at the $\bar{\tau}_{A}$-quantile of the distribution of ability given $Z=\bar{z}$, and at the $\bar{\tau}_{F-q u a n t i l e ~ o f ~ t h e ~ d i s t r i b u t i o n ~}$ of fortune given $Z=\bar{z}$ and given ability is equal to the $\bar{\tau}_{A}$-quantile of ability given $Z=\bar{z}$.

Further restrict the class of admissible structures so that, at $\bar{\Psi}$ :
(I) the first partial derivative of the function $h_{W}$ with respect to an element, $Z_{i}$, of $Z$ is zero and the first partial derivative of the function $h_{S}$ with respect to that element $Z_{i}$ is nonzero,
(II) the conditional quantiles, $Q_{F \mid A Z}\left(\tau_{F}, a, z\right)$ and $Q_{A \mid Z}\left(\tau_{A}, z\right)$ have zero derivatives with respect to the element $Z_{i}$.

Condition (I) is a local-to- $\bar{\Psi}$ version of the classical rank condition of Koopmans Rubin and Leipnik (1950). Condition (II) is a local-to- $\bar{\Psi}$ restriction on the covariation of the latent variates and the covariate $Z_{i}$.

Define: $s^{*} \equiv Q_{S \mid Z}\left(\bar{\tau}_{A}, \bar{z}\right)$. Consider the functional of the joint distribution function of $W$ and $S$ given $Z$ :

$$
\begin{equation*}
\pi_{S Z_{i}}(\bar{\Psi}) \equiv \nabla_{S} Q_{W \mid S Z}\left(\bar{\tau}_{F}, s^{*}, \bar{z}\right)+\frac{\nabla_{Z_{i}} Q_{W \mid S Z}\left(\bar{\tau}_{F}, s^{*}, \bar{z}\right)}{\nabla_{Z_{i}} Q_{S \mid Z}\left(\bar{\tau}_{A}, \bar{z}\right)} \tag{4}
\end{equation*}
$$

where, for example, $\nabla_{Z_{i}} Q_{W \mid S Z}\left(\bar{\tau}_{F}, s^{*}, \bar{z}\right)$ is shorthand for $\left.\nabla_{Z_{i}} Q_{W \mid S Z}\left(\bar{\tau}_{F}, S, Z\right)\right|_{S=Q_{S \mid Z}\left(\bar{\tau}_{A}, \bar{z}\right), Z=\bar{z}}$.
The results developed later imply that the returns to schooling, $\nabla_{S} h_{W}(\bar{\Psi})$, is identified by $\pi_{S Z_{i}}(\bar{\Psi})$. The following example shows how this result applies in a particular case.

Example 1. Suppose data are generated by a "random coefficients" linear structure:

$$
\begin{align*}
W & =\theta(A) S+\gamma^{\prime} Z+F  \tag{5}\\
S & =\beta^{\prime} Z+A \tag{6}
\end{align*}
$$

in which the returns to schooling is $\theta(A)$ with first derivative $\theta^{\prime}(A)$. The conditional $\bar{\tau}_{A}$-quantile of $S$ given $Z=z$ is as follows.

$$
Q_{S \mid Z}\left(\bar{\tau}_{A}, z\right)=\beta^{\prime} z+Q_{A \mid Z}\left(\bar{\tau}_{A}, z\right)
$$

Substituting for $A$ in (5) using (6), the conditional $\bar{\tau}_{F}$-quantile of $W$ given $S=s$ and $Z=z$ is

$$
Q_{W \mid S Z}\left(\bar{\tau}_{F}, s, z\right)=\theta\left(s-\beta^{\prime} z\right) s+\gamma^{\prime} z+Q_{F \mid A Z}\left(\bar{\tau}_{F}, s-\beta z, z\right)
$$

and the derivatives in (4) are, on imposing the local rank condition (I), here $\gamma_{i}=0$ and $\beta_{i} \neq 0$, and the local quantile insensitivity condition (II),

$$
\begin{aligned}
\nabla_{S} Q_{W \mid S Z}\left(\bar{\tau}_{F}, s^{*}, \bar{z}\right) & =\theta\left(s^{*}-\beta^{\prime} \bar{z}\right)+\left\{\theta^{\prime}\left(s^{*}-\beta^{\prime} \bar{z}\right) s^{*}+\nabla_{A} Q_{F \mid A Z}\left(\bar{\tau}_{F}, s^{*}-\beta^{\prime} \bar{z}, \bar{z}\right)\right\} \\
\nabla_{Z_{i}} Q_{W \mid S Z}\left(\bar{\tau}_{F}, s^{*}, \bar{z}\right) & =-\beta_{i}\left\{\theta^{\prime}\left(s^{*}-\beta^{\prime} \bar{z}\right) s^{*}+\nabla_{A} Q_{F \mid A Z}\left(\bar{\tau}_{F}, s^{*}-\beta^{\prime} \bar{z}, \bar{z}\right)\right\} \\
\nabla_{Z_{i}} Q_{S \mid Z}\left(\bar{\tau}_{A}, \bar{z}\right) & =\beta_{i}
\end{aligned}
$$

which, inserted in (4), yields $\pi_{S Z_{i}}(\bar{\Psi})=\theta\left(s^{*}-\beta^{\prime} \bar{z}\right)=\theta(\bar{a})$. This is the required value of the returns to schooling for a person with $A=\bar{a}$, that is at the $\bar{\tau}_{A}$-quantile of ability, $A$, given $Z=\bar{z}$.

The following remarks will be amplified later in the paper.

1. Overidentification. If more than covariate, $Z_{i}$, satisfies the local rank condition (I) and quantile insensitivity condition (II), then the value of the derivative $\nabla_{S} h_{W}(\bar{\Psi})$ is overidentified.
2. Localness. There is the possibility of identification of a structural derivative evaluated at some quantile probabilities but not at others. Local identification does not require full statistical independence of $A$ and $F$ relative to $Z$. It may be secured at some values of a "local instrument", that is $Z_{i}$, even if not at others.
3. Identification of other structural derivatives. Identification of other structural derivatives can be achieved in a similar fashion. For example, the value of the $Z_{j}$-derivative of the wage function:

$$
\left.\nabla_{Z_{j}} h_{W}(\bar{\Psi}) \equiv \nabla_{Z_{j}} h_{W}(S, Z, F, A)\right|_{\bar{\Psi}}
$$

is identified by the functional

$$
\begin{align*}
& \pi_{Z_{j} Z_{i}}(\bar{\Psi}) \equiv \nabla_{Z_{j}} Q_{W \mid S Z}\left(\bar{\tau}_{F}, s^{*}, \bar{z}\right)  \tag{7}\\
& \quad-\nabla_{Z_{i}} Q_{W \mid S Z}\left(\bar{\tau}_{F}, s^{*}, \bar{z}\right) \times \frac{\nabla_{Z_{j}} Q_{S \mid Z}\left(\bar{\tau}_{A}, \bar{z}\right)}{\nabla_{Z_{i}} Q_{S \mid Z}\left(\bar{\tau}_{A}, \bar{z}\right)}
\end{align*}
$$

where as before, $s^{*} \equiv Q_{S \mid Z}\left(\bar{\tau}_{A}, \bar{z}\right)$. In the returns-to-schooling example $\nabla_{Z_{j}} h_{W}(\bar{\Psi})=$ $\gamma_{j}$ and

$$
\begin{aligned}
\nabla_{Z_{j}} Q_{W \mid S Z}\left(\bar{\tau}_{F}, s^{*}, \bar{z}\right) & =\gamma_{j}-\beta_{j}\left\{\theta^{\prime}\left(s^{*}-\beta^{\prime} \bar{z}\right) s^{*}+\nabla_{A} Q_{F \mid A Z}\left(\bar{\tau}_{F}, s^{*}-\beta^{\prime} \bar{z}, \bar{z}\right)\right\} \\
\nabla_{Z_{i}} Q_{W \mid S Z}\left(\bar{\tau}_{F}, s^{*}, \bar{z}\right) & =-\beta_{i}\left\{\theta^{\prime}\left(s^{*}-\beta^{\prime} \bar{z}\right) s^{*}+\nabla_{A} Q_{F \mid A Z}\left(\bar{\tau}_{F}, s^{*}-\beta^{\prime} \bar{z}, \bar{z}\right)\right\}
\end{aligned}
$$

which combined as in (7) with $\nabla_{Z_{j}} Q_{S \mid Z}\left(\bar{\tau}_{A}, \bar{z}\right) / \nabla_{Z_{i}} Q_{S \mid Z}\left(\bar{\tau}_{A}, \bar{z}\right)=\beta_{j} / \beta_{i}$ delivers the required parameter: $\gamma_{j}$.
4. Average derivatives. Consider a set of quantile probabilities, $T_{\tau}$, and the expected value of a function of the structural derivative $\nabla_{S} h_{W}(S, Z, F, A)$, given $Z=\bar{z}$ conditional on $A$ and $F$ lying in the set of values implied by $T_{\tau}$. If the identification conditions hold for $Z=\bar{z}$ for all quantile probabilities, $\tau \in T_{\tau}$, then the value of the expected derivative is identifiable as the integral of the function of $\pi_{S Z_{i}}(\bar{\Psi})$ over $\bar{\tau}_{F}, \bar{\tau}_{A} \in T_{\tau}$ divided by the probability that $A$ and $F$ lie in the set of values defined by $T_{\tau}$. This is taken up in Section 8 .
5. Estimation. The constructive identification of, for example, $\nabla_{S} h_{W}(\bar{\Psi})$ points directly to estimation using the analogue principle (Manski (1988b)), applying the functional $\pi_{S Z_{i}}$ to parametric, semi- or nonparametric estimates of the conditional quantile functions of $W$ and $S$ given $Z$. If there is overidentification the resulting multiplicity of estimates can be reconciled using a minimum distance procedure. When parametric restrictions specify $h_{W}$ and $h_{S}$ as linear functions the estimator is an alternative to the Two Stage Least Absolute Deviations (2SLAD) estimators proposed by Amemiya (1982) with wider applicability. Estimation is briefly discussed in Section 7.

In Section 4 expressions like (4) and (7) are developed for the general $M$ equation case. To bring the main ideas to the fore, that development is now sketched for this two equation example.
1.4. Returns to schooling: demonstration of identification. The identification of the schooling function and its derivatives is considered first. The value of $S$ at $\bar{\Psi}, \bar{s}$, is identified because, as noted earlier,

$$
\begin{equation*}
\bar{s} \equiv h_{S}(\bar{z}, \bar{a})=Q_{S \mid Z}\left(\bar{\tau}_{A}, \bar{z}\right) \tag{8}
\end{equation*}
$$

Since $\nabla_{Z_{i}} Q_{A \mid Z}\left(\bar{\tau}_{A}, Z\right)$ is zero at $Z=\bar{z}$, the value of the derivative of $Q_{S \mid Z}\left(\bar{\tau}_{A}, Z\right)$ with respect to $Z_{i}$ at $\bar{z}$ identifies the value of the derivative of $h_{S}(Z, A)$ with respect to $Z_{i}$ at $\bar{z}$ when $A=\bar{a}$.

Now consider the identification of the value of the wage function and of its $S$ derivative at $\bar{\Psi}$. Substitute in (1) for $S$ using (2) giving

$$
W=h_{W}\left(h_{S}(Z, A), Z, F, A\right)
$$

and fix $Z$, and $A$ at their values at $\bar{\Psi}$. Considering variation in $F$, since $h_{W}$ is strictly increasing in $F$ and $\bar{f}=Q_{F \mid A Z}\left(\bar{\tau}_{F}, \bar{a}, \bar{z}\right)$ :

$$
h_{W}\left(h_{S}(\bar{z}, \bar{a}), \bar{z}, \bar{f}, \bar{a}\right)=Q_{W \mid A Z}\left(\bar{\tau}_{F}, \bar{a}, \bar{z}\right)
$$

Since $h_{S}$ is strictly monotonic in $A$, the event $\{Z=\bar{z} \cap A=\bar{a}\}$ occurs if and only if the event $\{Z=\bar{z} \cap S=\bar{s}\}$ occurs. This implies that

$$
Q_{W \mid A Z}\left(\bar{\tau}_{F}, \bar{a}, \bar{z}\right)=Q_{W \mid S Z}\left(\bar{\tau}_{F}, \bar{s}, \bar{z}\right)
$$

and therefore, on using (8),

$$
h_{W}(\bar{s}, \bar{z}, \bar{f}, \bar{a})=Q_{W \mid S Z}\left(\bar{\tau}_{F}, Q_{S \mid Z}\left(\bar{\tau}_{A}, \bar{z}\right), \bar{z}\right)
$$

which identifies the value of delivered by $h_{W}$ at $\bar{\Psi}$.
The $s$ - and $z$-derivatives of $Q_{W \mid S Z}\left(\bar{\tau}_{F}, s, z\right)$ evaluated at $s=Q_{S \mid Z}\left(\bar{\tau}_{A}, \bar{z}\right), Z=\bar{z}$ identify the values of the $s$ - and $z$-derivatives of $h_{W}$ at $\bar{\Psi}$, when all $s$ - and $z$-driven variation is taken into account. This variation arises from the direct appearance of $S$ and $Z$ as arguments of $h_{W}$, indirectly via $A$ which appears in its own right as an argument of $h_{W}$ and via the conditioning arguments of $Q_{F \mid A Z}$. These multiple dependencies are made explicit in the function $g_{W}\left(\bar{\tau}_{F}, s, z\right)$ defined as:

$$
g_{W}\left(\bar{\tau}_{F}, s, z\right) \equiv h_{W}\left(s, z, Q_{F \mid A Z}\left(\bar{\tau}_{F}, h_{S}^{-1}(z, s), z\right), h_{S}^{-1}(z, s)\right)
$$

where $h_{S}^{-1}(z, s)$ is the inverse schooling function satisfying $s=h_{S}\left(z, h_{S}^{-1}(z, s)\right)$.
The value of the $S$-derivative of the wage function at $\bar{\Psi}$ is identified if it can be deduced from knowledge of the values at $\bar{\Psi}$ of the $s$ - and $z$-derivatives of the function $g_{W}\left(\bar{\tau}_{F}, s, z\right)$ and the $z$-derivatives of the function $g_{S}\left(\bar{\tau}_{A}, z\right)$ defined as follows.

$$
g_{S}\left(\bar{\tau}_{A}, z\right)=h_{S}\left(z, Q_{A \mid Z}\left(\bar{\tau}_{A}, z\right)\right)
$$

There is at $\bar{\Psi}$, suppressing arguments,

$$
\begin{aligned}
\nabla_{S} g_{W} & =\nabla_{S} h_{W}+\left(\nabla_{F} h_{W} \nabla_{A} Q_{F \mid A Z}+\nabla_{A} h_{W}\right) \nabla_{S} h_{S}^{-1} \\
\nabla_{Z_{i}} g_{W} & =\left(\nabla_{F} h_{W} \nabla_{A} Q_{F \mid A Z}+\nabla_{A} h_{W}\right) \nabla_{Z_{i}} h_{S}^{-1} \\
\nabla_{Z_{i}} g_{S} & =\nabla_{Z_{i}} h_{S}
\end{aligned}
$$

where in the second line the local rank condition (I) ensures that $\nabla_{Z_{i}} h_{W}=0$ and in the second and third lines the quantile insensitivity condition (II) ensures no $Z_{i^{-}}$ derivatives of quantiles of $F$ and $A$ appear. If $\nabla_{Z_{i}} h_{S}^{-1} \neq 0$ at $\bar{\Psi}$, which is assured by the local rank condition (I), then:

$$
\begin{aligned}
\nabla_{S} h_{W} & =\nabla_{S} g_{W}-\nabla_{Z_{i}} g_{W} \times\left(\frac{\nabla_{S} h_{S}^{-1}}{\nabla_{Z_{i}} h_{S}^{-1}}\right) \\
& =\nabla_{S} g_{W}+\frac{\nabla_{Z_{i}} g_{W}}{\nabla_{Z_{i}} h_{S}}
\end{aligned}
$$

Since $\nabla_{Z_{i}} h_{S}=\nabla_{Z_{i}} g_{S}$ at $\bar{\Psi}$, the value of $\nabla_{S} h_{W}$ at $\bar{\Psi}$ can be deduced from knowledge of the values of the derivatives of the functions $g_{W}$ and $g_{S}$, as follows.

$$
\begin{equation*}
\nabla_{S} h_{W}=\nabla_{S} g_{W}+\frac{\nabla_{Z_{i}} g_{W}}{\nabla_{Z_{i}} g_{S}} \tag{9}
\end{equation*}
$$

The value of each term on the right hand side of (9) is identified by the value at $\bar{\Psi}$ of the appropriate derivative of the conditional quantile functions of $W$ given $S$ and $Z$ (for $g_{W}$ ) and of $S$ given $Z$ (for $g_{S}$ ), and substituting these values gives the result (4).
1.5. Plan of the paper. Section 2 sets this paper in the context of the literature on identification. Section 3 defines admissible equation systems and defines the structural features whose identification is sought.

Section 4 develops the identifying restrictions and states Theorems which assert their identifying power. A core set of restrictions are introduced in Section 4.1. Sections 4.2 provides additional restrictions required to identify values of all derivatives. Section 4.3 deals with "single equation" identification. The conditions are illustrated in the context of the returns-to-schooling example in Section 5. Proofs of Theorems are given in Section 6.

Section 7 briefly examines estimation issues. Section 8 addresses the identification of averages of functions of structural derivatives, for example, their expected values and variances. Section 9 concludes.

## 2. Related Results

The study of parametric identification dates back to the start of the discipline of econometrics, with important contributions by Working (1925, 1927), Tinbergen (1930), Frisch (1934, 1938), Haavelmo (1944), Hurwicz (1950), Koopmans, Rubin and Leipnik (1950), Koopmans and Reiersøl (1950), Wald (1950), Fisher (1959, 1961, 1966), Wegge (1965) and Rothenberg (1971).

The approach taken in this paper is similar to that of Tinbergen (1930) which considered conditions under which values of structural form coefficients in linear simultaneous equations models could be deduced from knowledge of identifiable values of reduced form coefficients. The approach produces local versions of the order and rank conditions for identification set out by Koopmans, Rubin and Leipnik (1950).

Roehrig (1988), extending the work of Brown (1983), considered nonparametric global identification of structural functions under the restriction that latent variates are distributed independently of covariates. The main result is for nonseparable models but much of the elaboration of the result is done for separable models. Newey and Powell (1988), Newey, Powell and Vella (1999), Pinkse (2000), Darolles, Florens and Renault (2000) study separable models with additive latent variables which satisfy mean independence conditions.

Brown and Matzkin (1996) study the nonparametric global identification of primitive functions, for example production or utility functions, associated with nonseparable simultaneous equations systems when latent variables and covariates are restricted to be independently distributed. Altonji and Matzkin, (2001) study global identification in nonseparable panel data models with endogeneity under conditional exchangeability assumptions. Imbens and Newey (2001) propose a nonparametric estimator in a model comprising two triangular nonseparable structural equations with latent variates and covariates restricted to be independently distributed. Their model satisfies Roehrig's (1988) conditions and therefore globally identifies the structural functions.

Identification is considered from a conditional quantile perspective in Matzkin (1999) which deals with a model $Y=m(X, \varepsilon)$ in which $\varepsilon$ and $X$ are independently distributed and $m(\cdot, \cdot)$ is strictly monotonic in $\varepsilon$. 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 $\tau$-quantile of $Y$ given $X=x$ where $\tau$ is such that $e$ is the $\tau$-quantile of the marginal distribution of $\varepsilon$.

In contrast to these papers, which propose restrictions sufficient to obtain global identification of structural functions, this paper proposes weaker conditions sufficient to obtain local identification of values of, and derivatives of, structural functions. When these restrictions hold globally then global structural features may be identifiable. In semiparametric and parametric models, an object identified locally may be a global parameter in which case global identification is secured under weak local restrictions.

The objects whose identification are considered in this paper, derivatives of structural functions evaluated at quantiles of latent variates, can give valuable information about the distribution of policy impacts across a population. Recent papers developing estimators of such distributions include Heckman, Smith and Clements (1997) which explores non-quantile based approaches in a programme evaluation setting and Abadie, Angrist and Imbens (2002) which proposes a Quantile Treatment Effect estimator in a study of the impact of subsidised training on the distribution of earnings. These papers are part of a large literature studying the identifying power of models of treatment effects. ${ }^{7}$ Treatment effect models have more latent variates than observable outcomes which renders the quantile-based attack taken here inapplicable without further restrictions.

The identification conditions of this paper include local quantile independence restrictions. Manski's (1975) maximum score estimator and Koenker and Basset's (1978) quantile regression function estimators are built on such restrictions. Manski (1988a) gives an account of the identifying power of conditional quantile restrictions in binary response models. Recent papers using quantile independence conditions as the basis for developing estimators include: Newey and Powell (1990), Chaudhuri, Doksum and Samarov (1997), Kahn (2001) and Chernozhukov and Hansen (2002).

The next Section specifies the types of structural equation systems considered in this paper and defines the structural features whose identification is sought.

## 3. Structural equations and structural Derivatives

Admissible structures have equations of the following triangular form, which recursively determines values of $M$ scalar outcomes, $Y \equiv\left\{Y_{i}\right\}_{i=1}^{M}$, given values of $K$

[^4]covariates, $Z \equiv\left\{Z_{i}\right\}_{i=1}^{K}$, and values of latent variates, $\varepsilon \equiv\left\{\varepsilon_{i}\right\}_{i=1}^{M}$.
\[

\left.$$
\begin{array}{rl}
Y_{1} & =h_{1}\left(Y_{2}, Y_{3}, \ldots, Y_{M}, Z, \varepsilon_{1}, \varepsilon_{2}, \ldots, \varepsilon_{M}\right)  \tag{10}\\
Y_{2} & =h_{2}\left(Y_{3}, \ldots, Y_{M}, Z, \varepsilon_{2}, \ldots, \varepsilon_{M}\right) \\
\vdots & \vdots \\
Y_{M} & =h_{M}\left(Z, \varepsilon_{M}\right)
\end{array}
$$\right\}
\]

Additional restrictions will be developed which are sufficient to identify values of some or all of the first partial derivatives of the functions $h \equiv\left\{h_{i}\right\}_{i=1}^{M}$ at a specified value of the arguments of the functions.

This value is determined by a value of $Z$, denoted $\bar{z}$, and by probabilities, $\bar{\tau} \equiv$ $\left\{\bar{\tau}_{i}\right\}_{i=1}^{M}$ which specify values of the elements of $\varepsilon$ as values of recursively defined conditional quantiles of $\varepsilon$ given $Z=\bar{z}$ at the probabilities $\bar{\tau}$. Define $\bar{\Omega} \equiv\{\bar{\tau}, \bar{z}\}$.

For some chosen value, $z$, of $Z$ define the conditional $\bar{\tau}$-quantiles ${ }^{8}$

$$
\begin{equation*}
e_{i} \equiv Q_{\varepsilon_{i} \mid \varepsilon_{i+} Z}\left(\bar{\tau}_{i}, e_{i+}, z\right), \quad i \in\{1, \ldots, M\} \tag{11}
\end{equation*}
$$

and let $e \equiv\left\{e_{i}\right\}_{i=1}^{M}$. Here and later an abbreviated notation for lists is employed. For any list, $X \equiv\left\{X_{i}\right\}_{i=1}^{M}, X_{i+}$ indicates the list $\left\{X_{j}\right\}_{j=i+1}^{M}$ for $i<M$ and an empty list for $i=M$. This abbreviated notation is used when indicating conditioning and when indicating arguments of functions. Thus equation (11) indicates the following.

$$
\begin{aligned}
e_{M} & \equiv Q_{\varepsilon_{M} \mid Z}\left(\bar{\tau}_{M}, z\right) \\
e_{i} & \equiv Q_{\varepsilon_{i} \mid \varepsilon_{i+1} \ldots \varepsilon_{M} Z}\left(\bar{\tau}_{i}, e_{i+1}, \ldots, e_{M}, z\right), \quad i \in\{1, \ldots, M-1\}
\end{aligned}
$$

A set of conditional quantile functions like this, in which each variate $\varepsilon_{i}$ is conditioned on $\varepsilon_{j}, j>i$, each such $\varepsilon_{j}$ being evaluated at a similarly defined quantile, is described as a set of iterated conditional quantile functions.

Denote the value of $e$ when $z=\bar{z}$ by $\bar{e}$, thus.

$$
\begin{equation*}
\bar{e}_{i} \equiv Q_{\varepsilon_{i} \mid \varepsilon_{i+} Z}\left(\bar{\tau}_{i}, \bar{e}_{i+}, \bar{z}\right), \quad i \in\{1, \ldots, M\} \tag{12}
\end{equation*}
$$

For a value $z$ define

$$
y_{i} \equiv h_{i}\left(y_{i+}, z, e_{i}, e_{i+}\right), \quad i \in\{1, \ldots, M\}
$$

and $y \equiv\left\{y_{i}\right\}_{i=1}^{M}$. At $\bar{\Omega}$ the value of $Y$ delivered by the model is ${ }^{9} \bar{y}$, defined as

$$
\bar{y}_{i} \equiv h_{i}\left(\bar{y}_{i+}, \bar{z}, \bar{e}_{i}, \bar{e}_{i+}\right), \quad i \in\{1, \ldots, M\}
$$

where $\bar{y}_{i+}$ is a sub-list of $\bar{y}=\left\{\bar{y}_{j}\right\}_{j=1}^{M}$. Define $\bar{\Psi} \equiv\{\bar{y}, \bar{z}, \bar{e}\}$.

[^5]The structural features whose identification is sought are the values at $\bar{\Psi}$ of some or all of the derivatives of the structural functions, as follows.

$$
\left.\begin{array}{rl}
\nabla_{y_{j}} h_{i}(\bar{\Psi}) & \left.\equiv \nabla_{y_{j}} h_{i}\left(y_{i+}, z, e_{i}, e_{i+}\right)\right|_{\bar{\Psi}}  \tag{13}\\
\nabla_{z_{k}} h_{i}(\bar{\Psi}) & \left.\equiv \nabla_{z_{k}} h_{i}\left(y_{i+}, z, e_{i}, e_{i+}\right)\right|_{\bar{\Psi}}
\end{array}\right\}, \quad\left\{\begin{array}{l}
i \in\{1, \ldots, M\} \\
j \in\{i+1, \ldots, M\} \\
k \in\{1, \ldots, K)
\end{array}\right.
$$

It is natural to specify values of $\varepsilon$ in terms of quantile probabilities because in this nonparametric analysis the distribution of $\varepsilon$ is at best identifiable up to a monotonic transformation - the metric in which $\varepsilon$ is measured is not identifiable. However, under the conditions to be described, features of structural functions at values of $\varepsilon$ associated with probabilities defining quantiles can be identified ${ }^{10}$.

## 4. Identification

This Section sets out restrictions on admissible structures under which the values of the structural derivatives in (13) are identified. Section 4.1 introduces five core restrictions. The first four of these assure:

1. the identifiability of values of structural functions in a neighbourhood of $\bar{\Psi}$,
2. the identifiability of values of derivatives of certain functions $g \equiv\left\{g_{i}\right\}_{i=1}^{M}$ at $\bar{\Psi}$.

Each function $g_{i}$ depends on $y_{i+}$ and $z$ and is the structural function $h_{i}$ with $\varepsilon_{i}$ replaced by $Q_{\varepsilon_{i} \mid \varepsilon_{i+} Z}\left(\varepsilon_{i+}, z\right)$ and with all occurrences of $\varepsilon_{j}, j>i$, replaced by recursively defined inverse functions $h_{j}^{-1}$ which depend on $y_{j+}$ and $z$. The functions $g$ are the functions of a triangular reduced form equation system. These results are the subject of Theorems 1 and 2. Proofs of all Theorems are in Section 6.

The values at $\bar{\Psi}$ of the derivatives of the functions $g$ are determined by the values at $\bar{\Psi}$ of derivatives of the structural functions and derivatives of the iterated conditional quantile functions of $\varepsilon$ given $Z$. This relationship is the subject of Theorem 3 .

Under Restrictions (I) - (IV) the value of a structural derivative at $\bar{\Psi}$ is identified if and only if its value can be deduced from knowledge of the values of the derivatives of the functions $g$. A variety of additional restrictions can assure this. In all the cases considered here a quantile insensitivity condition, Restriction (V), is imposed.

The final restrictions to be proposed restrict the values of derivatives of the structural functions at $\bar{\Psi}$. In Section 4.2, the identification of all structural derivatives is considered. This leads to a full system local rank condition which is necessary and

[^6]sufficient for identification of values of all structural derivatives, and a necessary full system local order condition. This result is expressed in Theorem 4.

In Section 4.3 the identification of values of derivatives of a single structural function is considered when there are restrictions only on its derivatives at $\bar{\Psi}$. This leads to a single equation local rank condition which is necessary and sufficient for identification, and a single equation local order condition, a result expressed in Theorem 5. Section 4.4 elaborates this last result for the case in which there are solely local exclusion restrictions, requiring values at $\bar{\Psi}$ of certain derivatives of the structural functions to be zero.
4.1. The core restrictions and their implications. The first four of the following five restrictions define a model which identifies values of the structural functions in a neighbourhood of $\bar{\Psi}$ and values of derivatives of the functions which define a triangular reduced form equation system at $\bar{\Psi}$. The fifth restriction places limits on the covariation of latent variates and covariates.

Restriction I. Completeness: in a neighbourhood of $\bar{\Psi}$ the equations (10) determine a unique value of $Y$.

Restriction II. Differentiability and continuous variation: in a neighbourhood of $\bar{\Psi}$ each function $h_{i}$ is a continuous and once differentiable function of its arguments and the arguments in $Z$ exhibit continuous variation.

Restriction III. Single crossing: for each $i$ and for $\left(y_{i}, y_{i+}, z, e_{i+}\right)$ in a neighbourhood of $\left(\bar{y}_{i}, \bar{y}_{i+}, \bar{z}, \bar{e}_{i+}\right)$, (a) there is a unique solution for $\varepsilon_{i}$ to

$$
h_{i}\left(y_{i+}, z, \varepsilon_{i}, e_{i+}\right)=y_{i}
$$

(b) $\nabla_{\varepsilon_{i}} h_{i}$ is nonzero in a neighbourhood of $\bar{\Psi}$, (c) the solution at $\left(y_{i}, y_{i+}, z, e_{i+}\right)=$ $\left(\bar{y}_{i}, \bar{y}_{i+}, \bar{z}, \bar{e}_{i+}\right)$ is $\varepsilon_{i}=\bar{e}_{i} .{ }^{11}$ The normalisation: $\nabla_{\varepsilon_{i}} h_{i}=1$ at $\bar{\Psi}$ is imposed.

Restriction IV. Continuous distribution: in a neighbourhood of $\bar{e}$ the vector $\varepsilon$ is continuously distributed given $Z=\bar{z}$ with positive density and at $\bar{\Psi}$ the conditional distribution function of $\varepsilon$ given $Z$ is differentiable with respect to $Z$.

Restriction V. Quantile insensitivity: the $z$-derivatives of the iterated conditional quantile functions at $\bar{\Psi},\left.\nabla_{z_{k}} Q_{\varepsilon_{i} \mid \varepsilon_{i+} Z}\left(\bar{\tau}_{i}, \bar{e}_{i+}, z\right)\right|_{z=\bar{z}}$, are zero for all $i$ and $k$.

There may be covariates for which some of these restrictions do not hold. These are not included in the list of covariates $Z$ and derivatives of structural functions with respect to these covariates may not be identifiable under the conditions proposed. For example there may be covariates for which the quantile insensitivity condition does not hold and there may be covariates which exhibit discrete variation. Continuous variation in some covariates is necessary for the nonparametric identification of partial

[^7]derivatives. ${ }^{12}$ The analysis of this paper applies at any set of values of any discrete covariates.

For $z$ in a neighbourhood of $\bar{z}$ define the iterated conditional $\bar{\tau}$-quantiles of $Y_{i}$ given $Y_{i+}$ and $Z$ :

$$
q_{i} \equiv Q_{Y_{i} \mid Y_{i+} Z}\left(\bar{\tau}_{i}, q_{i+}, z\right), \quad i \in\{1, \ldots, M\}
$$

and let $q \equiv\left\{q_{i}\right\}_{i=1}^{M}$, taking values $\bar{q} \equiv\left\{\bar{q}_{i}\right\}_{i=1}^{M}$ at $\bar{\Psi}$ where

$$
\bar{q}_{i} \equiv Q_{Y_{i} \mid Y_{i+} Z}\left(\bar{\tau}_{i}, \bar{q}_{i+}, \bar{z}\right), \quad i \in\{1, \ldots, M\} .
$$

Theorem 1. Under Restrictions (I) - (IV), in a neighbourhood of $\bar{\Psi}$, for $i \in$ $\{1, \ldots, M\}, y_{i}$ is identified by $q_{i}$.

Because there is local-to- $\bar{\Psi}$ identification of the values, $y_{i}$, of the structural functions, local variations in the values of the conditional quantiles associated with variations in their $Y_{i+}$ and $Z$ arguments identify the local variations in the values delivered by the structural functions, $y_{i}$, as $y_{i+}$ and $z$ vary.

These variations arise from the direct and indirect impacts of $y_{i+}$ and $z$ on the functions $h_{i}$, the indirect impacts coming through the arguments $e_{i}$ and $e_{i+}$ of the functions. The $y_{i+-}$ and $z$-derivatives of the conditional quantile functions identify the $y_{i+}-$ and $z$-derivatives of the structural functions that arise when all these sources of variation are taken into account. This result is expressed in Theorem 2, stated shortly after the introduction of additional notation.

Define the functions $h_{i}^{-1}\left(y_{i+}, z, y_{i}, h_{i+}^{-1}\right), i \in\{1, \ldots, M\}$. These are the recursively defined inverse functions of $h_{i}\left(y_{i+}, z, e_{i}, e_{i+}\right)$, with respect to $e_{i}$, satisfying

$$
h_{i}\left(y_{i+}, z, h_{i}^{-1}\left(y_{i+}, z, y_{i}, h_{i+}^{-1}\right), h_{i+}^{-1}\right)=y_{i}, \quad i \in\{1, \ldots, M\} .
$$

Define $h^{-1} \equiv\left\{h_{i}^{-1}\right\}_{i=1}^{M} \cdot{ }^{13}$ Define functions $g \equiv\left\{g_{i}\right\}_{i=1}^{M}$ where

$$
g_{i}\left(y_{i+}, z\right) \equiv h_{i}\left(y_{i+}, z, Q_{\varepsilon_{i} \mid \varepsilon_{i+} Z}\left(\bar{\tau}_{i}, h_{i+}^{-1}, z\right), h_{i+}^{-1}\right) \quad i \in\{1, \ldots, M\}
$$

and the functions in the list $h^{-1}$ are evaluated at $y=\left\{y_{i}\right\}_{i=1}^{M}$ and $z=\left\{z_{k}\right\}_{k=1}^{K}$. The functions $g$ constitute a reduced triangular form in which the direct and indirect impacts of $y_{i+}$ and $z$ on $y_{i}$ are made explicit.

[^8]Define the following matrices of derivatives of the functions $g$, evaluated at $\bar{\Psi}$.

$$
\underset{M \times M}{U}=\left[\begin{array}{ccccc}
0 & \nabla_{y_{2}} g_{1} & \nabla_{y_{3}} g_{1} & \ldots & \nabla_{y_{M}} g_{1} \\
0 & 0 & \nabla_{y_{3}} g_{2} & \ldots & \nabla_{y_{M}} g_{2} \\
\vdots & \vdots & \vdots & & \vdots \\
0 & 0 & 0 & & \nabla_{y_{M}} g_{M-1} \\
0 & 0 & 0 & \ldots & 0
\end{array}\right] \quad \underset{M \times K}{ }=\left[\begin{array}{ccc}
\nabla_{z_{1}} g_{1} & \ldots & \nabla_{z_{K}} g_{1} \\
\vdots & & \vdots \\
\nabla_{z_{1}} g_{M} & \ldots & \nabla_{z_{K}} g_{M}
\end{array}\right]
$$

Define the matrices of iterated $\bar{\tau}$-quantile function derivatives evaluated at $(y, z)=$ $(\bar{q}, \bar{z})$

$$
\begin{aligned}
\nabla_{Y} \bar{Q} \equiv & {\left[\begin{array}{ccccc}
0 & \nabla_{y_{2}} Q_{Y_{1} \mid Y_{1} Z} & \nabla_{y_{3}} Q_{Y_{1} \mid Y_{1+} Z} & \ldots & \nabla_{y_{M}} Q_{Y_{1} \mid Y_{1+} Z} \\
0 & 0 & \nabla_{y_{3}} Q_{Y_{2} \mid Y_{2} Z} & \ldots & \nabla_{y_{M}} Q_{Y_{2} \mid Y_{2} Z} \\
\vdots & \vdots & \vdots & & \vdots \\
0 & 0 & 0 & & \nabla_{y_{M}} Q_{Y_{M-1} \mid Y_{M} Z} \\
0 & 0 & 0 & \cdots & 0
\end{array}\right] } \\
& \nabla_{Z} \bar{Q} \equiv\left[\begin{array}{ccc}
\nabla_{z_{1}} Q_{Y_{1} \mid Y_{1} Z} & \ldots & \nabla_{z_{K}} Q_{Y_{1} \mid Y_{1} Z} \\
\vdots & & \vdots \\
\nabla_{z_{1}} Q_{Y_{M} \mid Z} & \ldots & \nabla_{z_{K}} Q_{Y_{M} \mid Z}
\end{array}\right]
\end{aligned}
$$

where for example the $(i, j)$ element of $\nabla_{Y} \bar{Q}$ is as follows.

$$
\left(\nabla_{Y} \bar{Q}\right)_{i j}=\left.\nabla_{y_{i}} Q_{Y_{j} \mid Y_{j}+Z}\left(\bar{\tau}_{j}, y_{j+}, z\right)\right|_{y=\bar{q}, z=\bar{z}}
$$

Theorem 2. Under Restrictions (I) - (IV), the elements of $U$ and $V$ are identified by the corresponding elements of respectively $\nabla_{Y} \bar{Q}$ and $\nabla_{Z} \bar{Q}$.

With the values of derivatives of the functions $g_{i}\left(y_{i+}, z\right)$ at $\bar{\Psi}$ identifiable under Restrictions (I) - (IV), models which identify the value of a structural function derivative must embody additional restrictions sufficient to permit a value of a derivative of a structural function to be deduced from knowledge of values of the derivatives of the functions $g_{i}\left(y_{i+}, z\right)$ at $\bar{\Psi}$.

This is similar to Koopmans, Rubin and Leipnik's (1950) analysis of identification in parametric linear simultaneous equations models under mean independence conditions but with the difference that proceeding via iterated conditional quantile functions, leads to the quest for conditions under which structural "parameters" can be uniquely determined from knowledge of coefficients of a triangular reduced form rather than a conventional reduced form.

The relationship between the derivatives of the functions $g$ and the derivatives of the structural functions is expressed in Theorem 3 which makes reference to the following matrices of derivatives, all evaluated at $\bar{\Psi} .{ }^{14}$

[^9]\[

$$
\begin{aligned}
& \underset{M \times M}{A}=\left[\begin{array}{ccccc}
0 & \nabla_{y_{2}} h_{1} & \nabla_{y_{3}} h_{1} & \ldots & \nabla_{y_{M}} h_{1} \\
0 & 0 & \nabla_{y_{3}} h_{2} & \ldots & \nabla_{y_{M}} h_{2} \\
\vdots & \vdots & \vdots & & \vdots \\
0 & 0 & 0 & & \nabla_{y_{M}} h_{M-1} \\
0 & 0 & 0 & \ldots & 0
\end{array}\right] \quad \underset{M \times M}{G}=\left[\begin{array}{ccccc}
1 & \nabla_{\varepsilon_{2}} h_{1} & \nabla_{\varepsilon_{3}} h_{1} & \ldots & \nabla_{\varepsilon_{M}} h_{1} \\
0 & 1 & \nabla_{\varepsilon_{3}} h_{2} & \ldots & \nabla_{\varepsilon_{M}} h_{2} \\
\vdots & \vdots & \vdots & & \vdots \\
0 & 0 & 0 & \ldots & \nabla_{\varepsilon_{M}} h_{M-1} \\
0 & 0 & 0 & \ldots & 1
\end{array}\right] \\
& \underset{M \times K}{B}=\left[\begin{array}{ccc}
\nabla_{z_{1}} h_{1} & \ldots & \nabla_{z_{K}} h_{1} \\
\vdots & & \vdots \\
\nabla_{z_{1}} h_{M} & \ldots & \nabla_{z_{K}} h_{M}
\end{array}\right] \underset{M \times K}{J}=\left[\begin{array}{ccc}
\nabla_{z_{1}} Q_{\varepsilon_{1} \mid \varepsilon_{1} Z} & \ldots & \nabla_{z_{K}} Q_{\varepsilon_{1} \mid \varepsilon_{1} Z} \\
\vdots & & \vdots \\
\nabla_{z_{1}} Q_{\varepsilon_{M} \mid Z} & \ldots & \nabla_{z_{K}} Q_{\varepsilon_{M} \mid Z}
\end{array}\right] \\
& \underset{M \times M}{H}=\left[\begin{array}{ccccc}
0 & \nabla_{\varepsilon_{2}} Q_{\varepsilon_{1} \mid \varepsilon_{1+} Z} & \nabla_{\varepsilon_{3}} Q_{\varepsilon_{1} \mid \varepsilon_{1+} Z} & \cdots & \nabla_{\varepsilon_{M}} Q_{\varepsilon_{1} \mid \varepsilon_{1+} Z} \\
0 & 0 & \nabla_{\varepsilon_{3}} Q_{\varepsilon_{2} \mid \varepsilon_{2+} Z} & \cdots & \nabla_{\varepsilon_{M}} Q_{\varepsilon_{2} \mid \varepsilon_{2} Z} \\
\vdots & \vdots & \vdots & & \vdots \\
0 & 0 & 0 & & \nabla_{\varepsilon_{M}} Q_{\varepsilon_{M-1} \mid \varepsilon_{M} Z} \\
0 & 0 & 0 & \cdots & 0
\end{array}\right]
\end{aligned}
$$
\]

Define the following triangular $M \times M$ matrix.

$$
C \equiv G\left(I_{M}-H\right)^{-1}
$$

Theorem 3. Under Restrictions (I) - (IV) there are the following relationships.

$$
\begin{align*}
U & =I_{M}-C^{-1}\left(I_{M}-A\right)  \tag{14}\\
V & =C^{-1} B+J \tag{15}
\end{align*}
$$

Theorem 2 states that under Restrictions (I) - (IV) the matrices $U$ and $V$ are identified by $\nabla_{Y} \bar{Q}$ and $\nabla_{Z} \bar{Q}$. The value of an element of one of the matrices $A, B$, $C$, and $J$ is identified under Restrictions (I) - (IV) if these matrices are restricted to the extent that the value of the element can be deduced from knowledge of the value of $U$ and $V$.

A variety of restrictions can be considered. The final identification conditions will be developed under Restriction (V), introduced at the start of this Section, which limits the dependence of $\varepsilon$ on $Z$ at $\bar{\Omega}$, restricting the matrix $J$ to be zero. 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 is revisited in Section 5.

Restrictions on $G$ and $H$ may arise in parametric models and in other situations but they are not considered here and attention is focussed on conditions sufficient to
identify elements of $A, B$ and $C .{ }^{15}$ This requires further restrictions because (14) and (15) contain only ${ }^{16} M(M-1) / 2+M K$ informative equations in the $M(M-1)+M K$ unknown elements of $A, B$ and $C$.
4.2. Identification of the values of all structural derivatives. Define $b \equiv$ $\operatorname{vec}(B), a \equiv \mathrm{v}(A)$ and $c \equiv \mathrm{v}(C)$ containing the, so far, unrestricted elements of $A$, $B$ and $C$, where the operator v() column stacks the super-diagonal elements of the square matrix to which it is applied, and define the $M^{2} \times(M(M-1) / 2)$ matrix, $R_{M}$, containing ones and zeros such that $\operatorname{vec}(A)=R_{M} \mathrm{v}(A)$.

Consider the following $N$ restrictions ${ }^{17}$ on the derivatives of the structural functions which hold at $\bar{\Psi}$.

$$
\begin{aligned}
& W_{A} \times a+W_{B} \times b+W_{C} \times \quad c \quad=w \\
& N \times M(M-1) / 2 \quad M(M-1) / 2 \times 1 \quad N \times M K \quad M K \times 1 \quad N \times M(M-1) / 2 \quad M(M-1) / 2 \times 1 \quad N \times 1
\end{aligned}
$$

Define the matrix $\Gamma$ and vectors $\gamma$ and $\theta$ as follows:

$$
\left.\left.\begin{array}{rl} 
& \equiv\left[\begin{array}{ccc}
I_{M(M-1) / 2} & 0 & R_{M}^{\prime}\left(\left(I_{M}-U^{\prime}\right) \otimes I_{M}\right) R_{M} \\
0 & -I_{M K} & \left(V^{\prime} \otimes I_{M}\right) R_{M} \\
W_{A} & W_{B} & W_{C}
\end{array}\right] \\
\gamma & \equiv\left[\begin{array}{c}
\mathrm{v}(U) \\
-\operatorname{vec}(V) \\
w
\end{array}\right]
\end{array} \begin{array}{c}
(M(M-1)+M K) \times 1
\end{array}\right] \begin{array}{l}
a \\
b \\
c
\end{array}\right]
$$

and consider the following restriction.
Restriction VI. Full system rank condition. The values taken by $A, B$, and $C$, imply values of $U$ and $V$ such that, given $W_{A}, W_{B}$ and $W_{C}$,

$$
\operatorname{rank}(\Gamma)=M(M-1)+M K
$$

This restriction is always satisfied when, local-to- $\bar{\Psi}$, the structural equations are in "classical" reduced form, that is when, at $\bar{\Psi}, \nabla_{y_{j}} h_{i}=0$ for $j>i$. In that case $W_{A}=I_{M}$ and $w=0, \Gamma$ is square and has the rank required by Restriction (VI) if

$$
\operatorname{rank}\left(R_{M}^{\prime}\left(\left(I_{M}-U^{\prime}\right) \otimes I_{M}\right) R_{M}\right)=M(M-1) / 2
$$

[^10]which always holds. ${ }^{18}$ Theorem 4 asserts the identifiability of all structural derivatives under Restrictions (I) - (VI).

Theorem 4. Under Restrictions (I) - (V), (a) $\Gamma \theta=\gamma$ and (b) all elements of $A$, $B$ and $C$ are identifiable if and only if Restriction (VI) holds. A necessary condition is the full system order condition: $N \geq M(M-1) / 2$.
4.3. Single equation identification. Now consider the identification of the values at $\bar{\Psi}$ of partial derivatives of a single structural function, $h_{i}$, employing restrictions on the the values at $\bar{\Psi}$ of derivatives of that function alone.

Some notational refinement is required. First consider $U$ and $V$, the (identified) matrices of derivatives of the functions $g$ evaluated at $\bar{\Psi}$. Let $V_{i}^{\prime}$ denote the $i$ th row of $V$ and let $V_{i+}^{\prime}$ contain the last $M-i$ rows of $V$. Let $U_{i}^{\prime}$ be the last ( $M-i$ ) elements in the $i$ th row of $U$ and let $U_{i+}^{\prime}$ be the lower right $(M-i) \times(M-i)$ block of $U$.

Let $b_{i}^{\prime}$ be the $i$ th row of $B$ and let $c_{i}^{\prime}$ and $a_{i}^{\prime}$ 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 constrained by the triangularity restriction.

Consider $N_{i}$ restrictions which apply to the derivatives of the $i$ th structural equation evaluated at $\bar{\Psi}$, as follows
and define the matrix $\Gamma_{i}$ and vectors $\gamma_{i}$ and $\theta_{i}$.

$$
\begin{aligned}
& \underset{\left(M-i+K+N_{i}\right) \times(2(M-i)+K)}{\Gamma_{i}} \equiv\left[\begin{array}{ccc}
I_{M-i} & 0 & I_{M-i}-U_{i+} \\
0 & -I_{K} & V_{i+} \\
W_{A i} & W_{B i} & W_{C i}
\end{array}\right] \\
& \underset{\left(M-i+K+N_{i}\right) \times 1}{\gamma_{i}} \equiv\left[\begin{array}{c}
U_{i} \\
-V_{i} \\
w_{i}
\end{array}\right] \underset{(2(M-i)+K) \times 1}{\theta_{i}} \equiv\left[\begin{array}{c}
a_{i} \\
b_{i} \\
c_{i}
\end{array}\right]
\end{aligned}
$$

Consider the following restriction.
Restriction VII. Single equation rank condition. The values taken by $A$, $B$, and $C$, imply values of $U$ and $V$ such that, given $W_{A i}, W_{B i}$ and $W_{C i}$ :

$$
\operatorname{rank}\left(\Gamma_{i}\right)=2(M-i)+K
$$

[^11]the second line following because $I_{M}-U^{\prime}$ is lower triangular with unit leading diagonal elements and therefore has rank $M$.

Theorem 5 asserts the identifiability of the derivatives of the $i$ th structural function under conditions (I) - (V) and (VII).

Theorem 5. Under Restrictions (I) - (V), (a) $\Gamma_{i} \theta_{i}=\gamma_{i}$ and (b) all elements of $a_{i}, b_{i}$ and $c_{i}$ are then identifiable if and only if Restriction (VII) holds. A necessary condition is the single equation local order condition: $N_{i} \geq M-i$.
4.4. Single equation identification under "exclusion" restrictions. Of leading interest is the case in which there are only local "exclusion" restrictions at $\bar{\Psi}$, that is, restrictions which require certain structural derivatives to be zero (or take some other known value) at $\bar{\Psi}$.

Suppose there are $N_{i}^{A}$ local exclusion pertaining to elements of $Y_{i+}$ and $N_{i}^{B}$ local covariate exclusion restrictions. Then the non-zero rows of $W_{A i}$ and $W_{B i}$ are respectively $N_{i}^{A}$ rows of $I_{M-i}$ and $N_{i}^{B}$ rows of $I_{K}, W_{C i}=0$ and $w_{i}=0$. Further notational refinement is required.

Order and partition $a_{i}$ so that the $N_{i}^{A}$ zero elements appear at the end of the vector, denote the unrestricted $M-i-N_{i}^{A}$ elements by $\hat{a}_{i}$, reorder the rows of $U_{i}$ and $U_{i+}^{*} \equiv\left(I_{M-i}-U_{i+}\right)$ and partition thus

$$
u_{i}=\left[\begin{array}{c}
\hat{U}_{i} \\
\check{U}_{i}
\end{array}\right] \quad U_{i+}^{*}=\left[\begin{array}{c}
\hat{U}_{i+}^{*} \\
\check{U}_{i+}^{*}
\end{array}\right]
$$

where $\hat{U}_{i+}^{*}$ is $\left(M-i-N_{i}^{A}\right) \times(M-i)$ with rows corresponding to the elements in $\hat{a}_{i}$, $\check{U}_{i+}^{*}$ is $N_{i}^{A} \times(M-i)$ with rows corresponding to the a priori zero elements in $a_{i}$, and $\hat{U}_{i}$ and $\check{U}_{i}$ are respectively $\left(M-i-N_{i}^{A}\right) \times 1$ and $N_{i}^{A} \times 1$.

Order and partition $b_{i}$ so that the $N_{i}^{B}$ a priori zero derivatives appear at the end of the vector, denote the remaining $K-N_{i}^{B}$ unrestricted derivatives by $\hat{b}_{i}$, re-order rows of $V_{i}$ and $V_{i+}$ and partition accordingly, thus:

$$
V_{i}=\left[\begin{array}{c}
\hat{V}_{i} \\
\tilde{V}_{i}
\end{array}\right] \quad V_{i+}=\left[\begin{array}{c}
\hat{V}_{i+} \\
\tilde{V}_{i+}
\end{array}\right]
$$

where $\hat{V}_{i+}$ is $\left(K-N_{i}^{B}\right) \times(M-i)$ with rows corresponding to the elements in $\hat{b}_{i}, \check{V}_{i+}$ is $N_{i}^{B} \times(M-i)$ with rows corresponding to the a priori zero elements in $b_{i}$ and $\hat{V}_{i}$ and $\check{V}_{i}$ are respectively $\left(K-N_{i}^{B}\right) \times 1$ and $N_{i}^{B} \times 1$.

The equation $\Gamma_{i} \theta_{i}=\gamma_{i}$ with a priori zero elements of $\theta_{i}$ excluded is as follows.

$$
\begin{align*}
& {\left[\begin{array}{ccc}
I_{M-i-N_{i}^{A}} & 0 & \hat{U}_{i+}^{*} \\
0 & 0 & \check{U}_{i+}^{*} \\
0 & -I_{K-N_{i}^{B}} & \hat{V}_{i+} \\
0 & 0 & \check{V}_{i+}
\end{array}\right]_{\left(2(M-i)+K-N_{i}^{A}-N_{i}^{B}\right) \times 1}^{\left[\begin{array}{c}
\hat{a}_{i} \\
\hat{b}_{i} \\
c_{i}
\end{array}\right]}=\underset{(M+K-i) \times 1}{\left[\begin{array}{c}
\hat{U}_{i} \\
\check{U}_{i} \\
-\hat{V}_{i} \\
-\check{V}_{i}
\end{array}\right]} \text { (M+K-i)×(2(M-i)+K-Ni-N-Ni)}} \tag{16}
\end{align*}
$$

The local rank restriction (VII) is satisfied if and only if

$$
\operatorname{rank}\left[\begin{array}{c}
\check{U}_{i+}^{*} \\
\check{V}_{i+}
\end{array}\right]=M-i
$$

for which a necessary condition is the local order condition: $N_{i}^{B} \geq M-i-N_{i}^{A}$.
If the local rank condition is satisfied then, for any $(M-i) \times\left(N_{i}^{A}+N_{i}^{B}\right)$ matrix $\Lambda$ such that the matrix inverse in equation (17) below exists, there is the following explicit solution, invariant with respect to choice of $\Lambda$.

$$
c_{i}=\left(\Lambda \times\left[\begin{array}{c}
\check{U}_{i+}^{*}  \tag{17}\\
\check{V}_{i+}
\end{array}\right]\right)^{-1}\left(\Lambda \times\left[\begin{array}{c}
\check{U}_{i} \\
-\check{V}_{i}
\end{array}\right]\right) \quad\left[\begin{array}{c}
\hat{a}_{i} \\
\hat{b}_{i}
\end{array}\right]=\left[\begin{array}{c}
\hat{U}_{i}-\hat{U}_{\hat{\prime}}^{*} c_{i} \\
\hat{V}_{i}+\hat{V}_{i+} c_{i}
\end{array}\right]
$$

## 5. Application to the returns to schooling model

This Section illustrates the application of the preceding results in the returns-toschooling model with two covariates, $Z_{1}$ and $Z_{2}$.

Omitting arguments there are the following expressions for the matrices of derivatives at a point $\bar{\Psi}$.
$A=\left[\begin{array}{cc}0 & \nabla_{S} h_{W} \\ 0 & 0\end{array}\right] \quad B=\left[\begin{array}{cc}\nabla_{Z_{1}} h_{W} & \nabla_{Z_{2}} h_{W} \\ \nabla_{Z_{1}} h_{S} & \nabla_{Z_{2}} h_{S}\end{array}\right] \quad G=\left[\begin{array}{cc}1 & \nabla_{A} h_{W} \\ 0 & 1\end{array}\right] \quad H=\left[\begin{array}{cc}0 & \nabla_{A} Q_{F \mid A Z} \\ 0 & 0\end{array}\right]$
Separate restrictions on the derivatives $\nabla_{A} Q_{F \mid A Z}$ and $\nabla_{A} h_{W}$ are not considered here, so the analysis proceeds in terms of $C \equiv G\left(I_{2}-H\right)^{-1}$, as follows.

$$
C=\left[\begin{array}{cc}
1 & \nabla_{A} h_{W}+\nabla_{A} Q_{F \mid A Z} \\
0 & 1
\end{array}\right] \equiv\left[\begin{array}{cc}
1 & c_{12} \\
0 & 1
\end{array}\right]
$$

The matrices on the right hand side of equations (14) and (15), are then, with the quantile insensitivity restriction (V) which implies $J=0$, as follows.

$$
\begin{gather*}
I_{M}-C^{-1}\left(I_{M}-A\right)=\left[\begin{array}{cc}
0 & \left(\nabla_{S} h_{W}+c_{12}\right) \\
0 & 0
\end{array}\right]  \tag{18}\\
C^{-1} B+J=\left[\begin{array}{cc}
\left(\nabla_{Z_{1}} h_{W}-\left(\nabla_{Z_{1}} h_{S}\right) c_{12}\right) & \left(\nabla_{Z_{2}} h_{W}-\left(\nabla_{Z_{2}} h_{S}\right) c_{12}\right) \\
\nabla_{Z_{1}} h_{S} & \nabla_{Z_{2}} h_{S}
\end{array}\right] \tag{19}
\end{gather*}
$$

The matrices of derivatives of the functions $g$ evaluated at $\bar{\Psi}$ are written as

$$
U \equiv\left[\begin{array}{cc}
0 & u_{12}  \tag{20}\\
0 & 0
\end{array}\right] \quad V \equiv\left[\begin{array}{ll}
v_{11} & v_{12} \\
v_{21} & v_{22}
\end{array}\right]
$$

which are respectively identified by the following matrices of quantile derivatives evaluated at $Z=\bar{z}, S=Q_{S \mid Z}\left(\bar{\tau}_{A}, \bar{z}\right)$.

$$
\nabla_{Y} \bar{Q} \equiv\left[\begin{array}{cc}
0 & \nabla_{S} Q_{W \mid S Z}  \tag{21}\\
0 & 0
\end{array}\right] \quad \nabla_{Z} \bar{Q} \equiv\left[\begin{array}{cc}
\nabla_{Z_{1}} Q_{W \mid S Z} & \nabla_{Z_{2}} Q_{W \mid S Z} \\
\nabla_{Z_{1}} Q_{S \mid Z} & \nabla_{Z_{2}} Q_{S \mid Z}
\end{array}\right]
$$

The values of the structural derivatives of the schooling equation, already in full "reduced form" are directly identified ${ }^{19}$ by the values of the corresponding derivatives of the conditional quantile function $Q_{S \mid Z}$. But the derivatives of the wage equation ( $i=1$ ) are not, and at least $N \geq M-1=1$ restrictions are necessary to achieve identification.
5.1. Just identification. Suppose there is the local exclusion restriction: $\nabla_{Z_{2}} h_{W}=$ 0 at $\bar{\Psi}$. Then, referring back to (19),

$$
C^{-1} B+J=\left[\begin{array}{cc}
\left(\nabla_{Z_{1}} h_{W}-\left(\nabla_{Z_{1}} h_{S}\right) c_{12}\right) & -\left(\nabla_{Z_{2}} h_{S}\right) c_{12} \\
\nabla_{Z_{1}} h_{S} & \nabla_{Z_{2}} h_{S}
\end{array}\right],
$$

the vectors $a_{1}, \hat{b}_{1}$ and $c_{1}$ are

$$
a_{1}=\left[\nabla_{S} h_{W}\right] \quad \hat{b}_{1}=\left[\nabla_{Z_{1}} h_{W}\right] \quad c_{1}=\left[c_{12}\right]
$$

and the equivalent of equation (16) is the following.

$$
\left[\begin{array}{ccc}
1 & 0 & 1 \\
0 & -1 & v_{21} \\
0 & 0 & v_{22}
\end{array}\right]\left[\begin{array}{c}
\nabla_{S} h_{W} \\
\nabla_{Z_{1}} h_{W} \\
c_{12}
\end{array}\right]=\left[\begin{array}{c}
u_{12} \\
-v_{11} \\
-v_{12}
\end{array}\right]
$$

The matrix on the left side of this expression has rank 3 if $v_{22} \neq 0$ at $\bar{\Psi}$. Under that local rank condition, which is satisfied if and only if $\nabla_{Z_{2}} h_{S} \neq 0$, there is the equivalent of (17).

$$
\left[\begin{array}{c}
\nabla_{S} h_{W}  \tag{22}\\
\nabla_{Z_{1}} h_{W} \\
c_{12}
\end{array}\right]=\left[\begin{array}{c}
u_{12}+\left(\frac{v_{12}}{v_{22}}\right) \\
v_{11}-v_{21}\left(\frac{v_{12}}{v_{22}}\right) \\
-\left(\frac{v_{12}}{v_{22}}\right)
\end{array}\right]
$$

Finally, noting that $U$ and $V$ are identified by respectively $\nabla_{Y} \bar{Q}$ and $\nabla_{Z} \bar{Q}$

$$
\left[\begin{array}{c}
\nabla_{S} h_{W} \\
\nabla_{Z_{1}} h_{W} \\
c_{12}
\end{array}\right] \text { is identified by }\left[\begin{array}{c}
\nabla_{S} Q_{W \mid S Z}+\left(\frac{\nabla_{Z_{2}} Q_{W \mid S Z}}{\nabla_{Z_{2}} Q_{S \mid Z}}\right) \\
\nabla_{Z_{1}} Q_{W \mid S Z}-\nabla_{Z_{2}} Q_{W \mid S Z}\left(\frac{\nabla_{Z_{1}} Q_{S \mid Z}}{\nabla_{Z_{2}} Q_{S \mid Z}}\right) \\
-\left(\frac{\nabla_{Z_{2}} Q_{W \mid S Z}}{\nabla_{Z_{2}} Q_{S \mid Z}}\right)
\end{array}\right]
$$

evaluated at $Z=\bar{z}, S=Q_{S \mid Z}\left(\bar{\tau}_{A}, \bar{z}\right)$, which confirms equations (4) and (7) of Section 1.3 on setting $i=2$ in that example.

[^12]5.2. Over identification. Now suppose both $\nabla_{Z_{1}} h_{W}$ and $\nabla_{Z_{2}} h_{W}$ are restricted to be zero at $\bar{\Psi}$. Equation (19) now simplifies as follows
\[

C^{-1} B=\left[$$
\begin{array}{cc}
-\left(\nabla_{Z_{1}} h_{S}\right) c_{12} & -\left(\nabla_{Z_{2}} h_{S}\right) c_{12}  \tag{23}\\
\nabla_{Z_{1}} h_{S} & \nabla_{Z_{2}} h_{S}
\end{array}
$$\right],
\]

the vectors $a_{1}, \hat{b}_{1}$ and $c_{1}$ are

$$
a_{1}=\left[\nabla_{S} h_{W}\right] \quad \hat{b}_{1}=\emptyset \quad c_{1}=\left[c_{12}\right]
$$

and the equivalent of equation (16) is the following.

$$
\left[\begin{array}{cc}
1 & 1  \tag{24}\\
0 & v_{21} \\
0 & v_{22}
\end{array}\right]\left[\begin{array}{c}
\nabla_{S} h_{W} \\
c_{12}
\end{array}\right]=\left[\begin{array}{c}
u_{12} \\
-v_{11} \\
-v_{12}
\end{array}\right]
$$

Set $\Lambda=\left[\begin{array}{ll}\lambda_{1} & \lambda_{2}\end{array}\right]$ in equation (17) chosen so that $\lambda_{1} v_{21}+\lambda_{2} v_{22} \neq 0$, giving the following solution

$$
\left[\begin{array}{c}
\nabla_{S} h_{W} \\
c_{12}
\end{array}\right]=\left[\begin{array}{c}
u_{12}+\left(\frac{\lambda_{1} v_{11}+\lambda_{2} v_{12}}{\lambda_{1} v_{21}+\lambda_{2} v_{22}}\right) \\
-\left(\frac{\lambda_{1} v_{11}+\lambda_{2} v_{12}}{\lambda_{1} v_{21}+\lambda_{2} v_{22}}\right)
\end{array}\right]
$$

and so there is overidentification of the values of $\nabla_{S} h_{W}$ and $c_{12}$. For any $\lambda_{1}$ and $\lambda_{2}$ such that $\lambda_{1} v_{21}+\lambda_{2} v_{22} \neq 0$ :

$$
\left[\begin{array}{c}
\nabla_{S} h_{W} \\
c_{12}
\end{array}\right] \text { is identified by }\left[\begin{array}{c}
\nabla_{S} Q_{W \mid S Z}+\left(\frac{\lambda_{1} \nabla_{Z_{1}} Q_{W \mid S Z}+\lambda_{2} \nabla_{Z_{2}} Q_{W \mid S Z}}{\lambda_{1} \nabla_{Z_{1}} Q_{S \mid Z}+\lambda_{2} \nabla_{Z_{2}} Q_{S \mid Z}}\right) \\
-\left(\frac{\lambda_{1} \nabla_{Z_{1}} Q_{W \mid S Z}+\lambda_{2} \nabla_{Z_{2}} Q_{W \mid S Z}}{\lambda_{1} \nabla_{Z_{1}} Q_{S \mid Z}+\lambda_{2} \nabla_{Z_{2}} Q_{S \mid Z}}\right)
\end{array}\right]
$$

evaluated at $Z=\bar{z}, S=Q_{S \mid Z}\left(\bar{\tau}_{A}, \bar{z}\right)$.
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 does apply for variation in $Z_{2}$ but not for variation in $Z_{1}$ Then (23) becomes

$$
C^{-1} B+J=\left[\begin{array}{cc}
-\left(\nabla_{Z_{1}} h_{S}\right) c_{12}+j_{11} & -\left(\nabla_{Z_{2}} h_{S}\right) c_{12} \\
\nabla_{Z_{1}} h_{S}+j_{21} & \nabla_{Z_{2}} h_{S}
\end{array}\right]
$$

where $j_{11}=\nabla_{Z_{1}} Q_{F \mid A Z}, j_{21}=\nabla_{Z_{1}} Q_{A \mid Z}$, both evaluated at $\bar{\Psi}$.
Now $\nabla_{Z_{1}} h_{S}$ is no longer identified and the zero restriction at $\bar{\Psi}$ on $\nabla_{Z_{1}} h_{W}$ has no force, but there remains

$$
\left[\begin{array}{cc}
1 & 1 \\
0 & v_{22}
\end{array}\right]\left[\begin{array}{c}
\nabla_{S} h_{W} \\
c_{12}
\end{array}\right]=\left[\begin{array}{c}
u_{12} \\
-v_{12}
\end{array}\right]
$$

which leads to identification of the returns to schooling, $\nabla_{S} h_{W}$, exactly as set out in equation (22).

## 6. Proofs of Theorems

6.1. Proof of Theorem 1. In the $i$ th structural equation recursively substitute the functions in $h_{i+}$ for the elements of $Y_{i+}$ giving

$$
Y_{i}=h_{i}\left(h_{i+}, Z, \varepsilon_{i}, \varepsilon_{i+}\right)
$$

in which the right hand side depends only on $Z, \varepsilon_{i}$ and $\varepsilon_{i+} .{ }^{20}$
Fix $\varepsilon_{i+}=e_{i+}$ and $Z=z$ in a neighbourhood of $\bar{\Psi}$. Under the single crossing condition (III), replacing $\varepsilon_{i}$ by its iterated conditional $\bar{\tau}_{i}$-quantile, $e_{i}$, delivers the conditional $\bar{\tau}_{i}$-quantile of $Y_{i}$ given $Z=z$ and $\varepsilon_{i+}=e_{i+}$ as follows.

$$
\begin{equation*}
h_{i}\left(h_{i+}, z, e_{i}, e_{i+}\right)=Q_{Y_{i} \mid \varepsilon_{i+} Z}\left(\bar{\tau}_{i}, e_{i+}, z\right) \tag{25}
\end{equation*}
$$

Restrictions (I) - (IV) ensure that the events $\left\{\varepsilon_{i+}=e_{i+} \cap Z=z\right\}$ and $\left\{Y_{i+}=\right.$ $\left.y_{i+} \cap Z=z\right\}$ are identical, therefore:

$$
\begin{equation*}
Q_{Y_{i} \mid \varepsilon_{i+} Z}\left(\bar{\tau}_{i}, e_{i+}, z\right)=Q_{Y_{i} \mid Y_{i+} Z}\left(\bar{\tau}_{i}, y_{i+}, z\right) . \tag{26}
\end{equation*}
$$

For each $i$ the left hand side of (25) is $y_{i}$ and so combining (25) and (26),

$$
\begin{equation*}
y_{i}=Q_{Y_{i} \mid Y_{i+} Z}\left(\bar{\tau}_{i}, y_{i+}, z\right) \tag{27}
\end{equation*}
$$

Since $q_{M} \equiv Q_{Y_{M} \mid Z}\left(\bar{\tau}_{M}, z\right)$, it follows from (27) that for variations in a neighbourhood of $\bar{\Psi}$

$$
y_{M}=y_{M}^{*} \Rightarrow q_{M}=y_{M}^{*}
$$

Stepping from $i=M-1$ through to $i=1$, using (27), replacing $y_{i+}$ by $q_{i+}$ at each step yields, for all $i$ and for variations in a neighbourhood of $\bar{\Psi}$ :

$$
y_{i}=y_{i}^{*} \Rightarrow q_{i}=y_{i}^{*}
$$

which completes the proof of Theorem 1.
6.2. Proof of Theorem 2. Consider the $i$ th structural equation

$$
Y_{i}=h_{i}\left(Y_{i+}, Z, \varepsilon_{i}, \varepsilon_{i+}\right)
$$

remove $\varepsilon_{i+}$ using the inverse functions $h_{i+}^{-1}$, giving

$$
Y_{i}=h_{i}\left(Y_{i+}, Z, \varepsilon_{i}, h_{i+}^{-1}\right)
$$

in which each function in $h_{i+}^{-1}$ depends upon elements of $Y_{i+}$ and $Z$. Fix $Y_{i+}=y_{i+}$ and $Z=z$ and replace $\varepsilon_{i}$ by $Q_{\varepsilon_{i} \mid \varepsilon_{i+} Z}\left(\bar{\tau}_{i}, h_{i+}^{-1}, z\right)=e_{i}$ in which the arguments $\varepsilon_{i+}$ have

[^13]been removed using the inverse functions $h_{i+}^{-1}$. The resulting function of $y_{i+}$ and $z$ delivers the value $y_{i}$, that is:
$$
y_{i}=h_{i}\left(y_{i+}, z, Q_{\varepsilon_{i} \mid \varepsilon_{i+} Z}\left(\bar{\tau}_{i}, h_{i+}^{-1}, z\right), h_{i+}^{-1}\right) \equiv g_{i}\left(y_{i+}, z\right) .
$$

Theorem 1 states that for each $i$, the value of $g_{i}\left(y_{i+}, z\right)$ is identified by $q_{i} \equiv$ $Q_{Y_{i} \mid Y_{i+} Z}\left(\bar{\tau}_{i}, q_{i+}, z\right)$ for ( $\left.y, z, e\right)$ in a neighbourhood of $\bar{\Psi}$. Therefore, for scalar $\triangle$ in a neighbourhood of zero, and $i<M$, the value of $g_{i}\left(\triangle \iota_{j}^{i}+\bar{y}_{i+}, \bar{z}\right)$ is identified by $Q_{Y_{i} \mid Y_{i+Z}}\left(\bar{\tau}_{i}, \Delta \iota_{j}^{i}+\bar{q}_{i+}, \bar{z}\right)$ where $j>i$ and $\iota_{j}^{i}$ is a $(i+1)$-vector with 1 in position $j-i$ and zeros elsewhere. It follows that, for $\triangle$ in a neighbourhood of zero

$$
\begin{equation*}
\frac{g_{i}\left(\triangle \iota_{j}^{i}+\bar{y}_{i+}, \bar{z}\right)-g_{i}\left(\bar{y}_{i+}, \bar{z}\right)}{\triangle} \tag{28}
\end{equation*}
$$

is identified by

$$
\begin{equation*}
\frac{Q_{Y_{i} \mid Y_{i+} Z}\left(\bar{\tau}_{i}, \Delta \iota_{j}^{i}+\bar{q}_{i+}, \bar{z}\right)-Q_{Y_{i} \mid Y_{i+} Z}\left(\bar{\tau}_{i}, \bar{q}_{i+}, \bar{z}\right)}{\triangle} \tag{29}
\end{equation*}
$$

Restrictions (II) and (IV) ensure that the $y_{j}$-derivative of $g_{i}\left(y_{i+}, z\right)$ at $\bar{\Psi}$ exists and is the limit as $\triangle \rightarrow 0$ of (28) whose limit is therefore identified by the limit as $\triangle \rightarrow 0$ of (29) this limit being the $y_{j}$-partial derivative of $Q_{Y_{i} \mid Y_{i+} Z}\left(\bar{\tau}_{i}, y_{i+}, z\right)$ evaluated at $y=\bar{q}$ and $z=\bar{z}$. A similar argument shows that the $z_{k}$-partial derivative of $g_{i}\left(y_{i+}, z\right)$ at $\bar{\Psi}$ is identified by the $z_{k}$-partial derivative of $Q_{Y_{i} \mid Y_{i+} Z}\left(\bar{\tau}_{i}, y_{i+}, z\right)$ evaluated at $y=\bar{q}$ and $z=\bar{z}$.

Collecting derivatives of the functions $g$ in the matrices $U$ and $V$ and quantile derivatives in the matrices $\nabla_{Y} \bar{Q}$ and $\nabla_{Z} \bar{Q}$ completes the proof.
6.3. Proof of Theorem 3. Recall the definitions of the functions $g_{i}, i \in\{1, \ldots, M\}$

$$
g_{i}\left(y_{i+}, z\right) \equiv h_{i}\left(y_{i+}, z, e_{i}, h_{i+}^{-1}\right) \quad e_{i} \equiv Q_{\varepsilon_{i} \mid \varepsilon_{i+} Z}\left(\bar{\tau}_{i}, h_{i+}^{-1}, z\right)
$$

where arguments of the functions $h_{i+}^{-1}$ and $e_{i}$ are suppressed.
Consider the differential $d g_{i}=d h_{i}$ for which there is the following expression with the normalisation $\nabla_{\varepsilon_{i}} h_{i}=1$ applied.

$$
\begin{equation*}
d g_{i}=\sum_{j=1}^{M}\left(\nabla_{y_{j}} h_{i} d y_{j}\right)+\sum_{k=1}^{K}\left(\nabla_{z_{k}} h_{i} d z_{k}\right)+d e_{i}+\mathbf{1}_{[i<M]} \sum_{j=i+1}^{M}\left(\nabla_{\varepsilon_{j}} h_{i} d h_{j}^{-1}\right) \tag{30}
\end{equation*}
$$

Define the column vector of differentials $d g \equiv\left\{d g_{i}\right\}_{i=1}^{M}$ and vectors of differentials $d y, d z, d q$ and $d h^{-1}$ similarly. Then (30) implies the following.

$$
d g=A d y+B d z+d e+\left(G-I_{M}\right) d h^{-1}
$$

The differential $d e_{i}$ is:

$$
\begin{aligned}
d e_{i} & =\sum_{j=1}^{M} \nabla_{y_{j}} e_{i} d y_{j}+\sum_{k=1}^{K} \nabla_{z_{k}} e_{i} d z_{k} \\
& =\sum_{j=1}^{M} \nabla_{\varepsilon_{j}} Q_{\varepsilon_{i} \mid \varepsilon_{i+} Z}\left(\nabla_{y_{j}} h_{j}^{-1} d y_{j}+\sum_{k=1}^{K} \nabla_{z_{k}} h_{j}^{-1} d z_{k}\right)+\sum_{k=1}^{K} \nabla_{z_{k}} Q_{\varepsilon_{i} \mid \varepsilon_{i+} Z} d z_{k}
\end{aligned}
$$

and therefore

$$
d e=H\left(\left(\frac{d h^{-1}}{d y^{\prime}}\right) d y+\left(\frac{d h^{-1}}{d z^{\prime}}\right) d z\right)+J d z
$$

where the $(s, t)$ elements of $\frac{d h^{-1}}{d y^{\prime}}$ and $\frac{d h^{-1}}{d z^{\prime}}$ are respectively $\nabla_{y_{t}} h_{s}^{-1}$ and $\nabla_{z_{t}} h_{s}^{-1}$.
The final step is to obtain expressions for these derivatives of the inverse functions. Since $y_{i}=h_{i}\left(y_{i+}, z, e_{i}, e_{i+}\right)$ there is

$$
d y=A d y+B d z+G d e
$$

and so

$$
d e=G^{-1}\left(\left(I_{M}-A\right) d y-B d z\right)=d h^{-1}
$$

giving the following expressions.

$$
\frac{d h^{-1}}{d y^{\prime}}=G^{-1}\left(I_{M}-A\right) \quad \frac{d h^{-1}}{d z^{\prime}}=-G^{-1} B
$$

On collecting terms there is

$$
\begin{aligned}
d g & =\left(I_{M}-\left(I_{M}-H\right) G^{-1}\left(I_{M}-A\right)\right) d y+\left(\left(I_{M}-H\right) G^{-1} B+J\right) d z \\
& =U d y+V d z
\end{aligned}
$$

and thus the result as stated in Theorem 3.
6.4. Proof of Theorem 4. Recall the definitions of $a, b, c$ and $R_{M}$ given in Section 4.2. Because each column of $R_{M}$ contains zeros except for a single unit element in a different position in each column

$$
\begin{equation*}
R_{M}^{\prime} R_{M}=I_{M(M-1) / 2} \tag{31}
\end{equation*}
$$

and the Moore-Penrose inverse of $R_{M}$ is $R_{M}^{-}=\left(R_{M}^{\prime} R_{M}\right)^{-1} R_{M}^{\prime}=R_{M}^{\prime}$. Since $R_{M}^{-} \operatorname{vec}(A)=$ $R_{M}^{-} R_{M} \mathrm{v}(A)$ it follows that $\mathrm{v}(A)=R_{M}^{\prime} \operatorname{vec}(A)$.

Since $C-I_{M}$ has the same triangular structure as $A$ and $\mathrm{v}\left(I_{M}\right)=0$,

$$
\begin{equation*}
\operatorname{vec}\left(C-I_{M}\right)=R_{M} \mathrm{v}\left(C-I_{M}\right)=R_{M} \mathrm{v}(C) \tag{32}
\end{equation*}
$$

Equation (15), with Restriction (V), $J=0$, imposed implies

$$
\left(C-I_{M}\right) V+V=B
$$

and on column stacking,

$$
\left(V^{\prime} \otimes I_{M}\right) \operatorname{vec}\left(C-I_{M}\right)+\operatorname{vec}(V)=\operatorname{vec}(B)
$$

which leads, on using (32), to the following result.

$$
\begin{equation*}
\left(V^{\prime} \otimes I_{M}\right) R_{M} c-b=-\operatorname{vec}(V) \tag{33}
\end{equation*}
$$

From (14)

$$
C-I_{M}-\left(C-I_{M}\right) U+A=U
$$

and on column stacking there is the following.

$$
\operatorname{vec}\left(C-I_{M}\right)-\left(U^{\prime} \otimes I_{M}\right) \operatorname{vec}\left(C-I_{M}\right)+\operatorname{vec}(A)=\operatorname{vec}(U)
$$

Using (32), premultiplying by $R_{M}^{\prime}$ to extract the relevant super-diagonal elements of $C$, noting that $R_{M}^{\prime} \operatorname{vec}\left(I_{M}\right)=0$ and using (31) gives the following result.

$$
\begin{equation*}
R_{M}^{\prime}\left(\left(I_{M}-U^{\prime}\right) \otimes I_{M}\right) R_{M} c+a=\mathrm{v}(U) \tag{34}
\end{equation*}
$$

Combining (33) and (34) with the restrictions $W_{A} a+W_{B} b+W_{C} c=w$ produces the equation $\Gamma \theta=\gamma$ as stated in Theorem 4. The value of $\theta$ is uniquely deducible from knowledge of $\Gamma$ and $\gamma$ if and only if the rank condition, Restriction (VI), holds. A necessary condition is that the row order of $\Gamma$ be at least equal to its column order which leads to the order condition stated in Theorem 4.
6.5. Proof of Theorem 5. The equation $\Gamma_{i} \theta_{i}=\gamma_{i}$ follows directly on selecting the appropriate rows and columns from $\Gamma$ and the required elements from $\gamma$ and $\theta$. The values of the $2(M-i)+K$ elements of $\theta_{i}$ can be solved uniquely if and only if $\operatorname{rank}\left(\Gamma_{i}\right)=2(M-i)+K$ for which a necessary condition is that the row order of $\Gamma_{i}$ be at least equal to its column order. These are the local rank and order conditions stated in Theorem 5.

## 7. Estimation

The values of the derivatives of the structural equations at the point $\bar{\Psi}=(\bar{y}, \bar{z}, \bar{e})$ can be estimated as follows.

1. Conditional $\bar{\tau}$-quantile functions of $Y_{i}$ given $Y_{i+}$ and $Z, i \in\{1, \ldots, M\}$ are estimated using a parametric, semi- or nonparametric method, as desired ${ }^{21}$ and an estimate, $\hat{y}=\left\{\hat{y}_{i}\right\}_{i=1}^{M}$, of $\bar{y}$, the value of $Y$ at the point $\bar{\Psi}$, is calculated.

[^14]2. Derivatives of the conditional $\bar{\tau}$-quantile functions at $\bar{z}$ and $\hat{y}$ are calculated producing $\widehat{\nabla_{Y} \bar{Q}}$ and $\widehat{\nabla_{Z} \bar{Q}}$, which are estimates of $\nabla_{Y} \bar{Q}$ and $\nabla_{Z} \bar{Q}$, and, in consequence of Theorem 2 , of matrices $U$ and $V .{ }^{22}$
3. The restrictions on $A, B$ and $C$ are assembled with $\widehat{\nabla_{Y} \bar{Q}}$ and $\widehat{\nabla_{Z} \bar{Q}}$ replacing $U$ and $V$ leading to estimates $\hat{\Gamma}$ and $\hat{\gamma}$, and if $\hat{\Gamma}$ has the required rank the equation $\hat{\Gamma} \hat{\theta}=\hat{\gamma}$ is solved for an estimate, $\hat{\theta}$, of $\theta$ which contains the desired elements of $A, B$ and $C$.

At the final step, if there are abundant restrictions 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 to a potentially inefficient quantile regression analogue of a system Indirect Least Squares estimator.

Efficient estimation can be achieved using a minimum distance estimator

$$
\hat{\theta}=\underset{\theta}{\arg \min }(\hat{\Gamma} \theta-\hat{\gamma})^{\prime} W(\hat{\Gamma} \theta-\hat{\gamma})
$$

for a suitable choice of positive definite weighting matrix $W$ where $\hat{\Gamma}$ and $\hat{\gamma}$ are as in step 3 above. This produces a quantile regression analogue of a system Three Stage Least Squares estimator. Of course the sampling properties of $\hat{\theta}$ depend on restrictions additional to those required to achieve identification.

Single equation estimation proceeds similarly, obtaining $\hat{\theta}_{i}$ as

$$
\hat{\theta}_{i}=\underset{\theta_{i}}{\arg \min }\left(\hat{\Gamma}_{i} \theta_{i}-\hat{\gamma}_{i}\right)^{\prime} W_{i}\left(\hat{\Gamma}_{i} \theta_{i}-\hat{\gamma}_{i}\right)
$$

where $W_{i}$ is a positive definite weighting matrix. This produces a quantile regression analogue of the Two Stage Least Squares estimator.

Identifying restrictions could be imposed during the procedure that results in the estimates $\widehat{\nabla_{Y} \bar{Q}}$ and $\widehat{\nabla_{Z} \bar{Q}}$. Then if $\hat{\Gamma}$ has the required rank there will be a unique solution to $\hat{\Gamma} \hat{\theta}=\hat{\gamma}$.

The estimation of coefficients of a median regression function derived from a linear location shift model in the presence of endogenous variables was considered by Amemiya (1982) who proposed a family of Two Stage Least Absolute Deviations (2SLAD) estimators for the coefficients of a linear location shift model

$$
\begin{aligned}
Y_{1} & =\gamma_{2} Y_{2}+\cdots+\gamma_{M} Y_{M}+Z^{\prime} \beta_{1}+\varepsilon_{1} \\
Y_{j} & =Z^{\prime} \beta_{j}+\varepsilon_{j} \quad j \in\{2, \ldots, M\}
\end{aligned}
$$

in which $\beta_{1}$ is sufficiently restricted to allow identification of the $\gamma_{2}, \ldots, \gamma_{M}$ and the unrestricted elements of $\beta_{1}$.

[^15]At the first stage of 2SLAD predicted values of $\left\{Y_{j}\right\}_{j=2}^{M}$, given values of $Z$ are produced. OLS and LAD estimation are suggested as possibilities in Amemiya (1982). In a leading special case of 2SLAD the second stage is an LAD estimation of the parameters of the equation for $Y_{1}$ using predicted values of $Y_{2}, \ldots, Y_{M}$ in place of their actual values.

The procedure described above is an alternative to 2SLAD with the advantage that it is applicable for quantile regressions other than the median regression and for parametric nonlinear, semi-parametric and nonparametric quantile regressions.

## 8. Average derivatives

So far the focus has been on local nonparametric identification of values of structural derivatives evaluated at a single point defined by a value, $\bar{z}$, of a vector of covariates $Z$ and by $M$ probabilities, $\bar{\tau}$, defining iterated conditional quantiles of latent variates, $\varepsilon$.

If identification can be achieved at $\bar{z}$ for a set of quantile probabilities, $\tau \in T_{\tau} \subseteq$ $[0,1]^{M}$, then it is possible to identify certain conditional expected values (with respect to the distribution of $\varepsilon$ given $Z$ ) of structural derivatives and of functions of them.

To see how this is achieved consider a subset of $\Re^{M}, T_{\varepsilon}$, and the expected value (assumed to exist), given $Z=\bar{z}$ and $\varepsilon \in T_{\varepsilon}$, of a function $x(\varepsilon, Z)$ of continuously distributed $\varepsilon$ which has conditional distribution function $F_{\varepsilon \mid Z}$.

$$
E_{\varepsilon \mid Z}\left[x(\varepsilon, Z) \mid Z=\bar{z}, \varepsilon \in T_{\varepsilon}\right] \equiv \frac{\int \cdots \int_{e \in T_{\varepsilon}} x(e, z) d F_{\varepsilon \mid Z}(e \mid \bar{z})}{\int \cdots \int_{e \in T_{\varepsilon}} d F_{\varepsilon \mid Z}(e \mid \bar{z})}
$$

Let $\tilde{x}(\tau, \bar{z}) \equiv x(\varepsilon(\tau), \bar{z})$ where $\tau \equiv\left\{\tau_{i}\right\}_{i=1}^{M}$, and $\varepsilon(\tau) \equiv\left\{\varepsilon_{i}(\tau)\right\}_{i=1}^{M}$ whose elements are defined recursively as follows.

$$
\varepsilon_{i}(\tau) \equiv Q_{\varepsilon_{i} \mid \varepsilon_{i+} Z}\left(\tau_{i} \mid \varepsilon(\tau)_{i+}, \bar{z}\right), \quad i \in\{1, \ldots, M\}
$$

The elements of $\tau$ satisfy

$$
\tau_{i}=F_{\varepsilon_{i} \mid \varepsilon_{i+} Z}\left(\varepsilon_{i} \mid \varepsilon_{i+}, \bar{z}\right), \quad i \in\{1, \ldots, M\}
$$

where $F_{\varepsilon_{i} \mid \varepsilon_{i+} Z}$ is the conditional distribution function of $\varepsilon_{i}$ given $\varepsilon_{i+}$ and $Z$. Define

$$
T_{\tau} \equiv\left\{\tau: \varepsilon(\tau) \in T_{\varepsilon}\right\} .
$$

Since the elements of $\tau$ are distributed independently of $Z$ and are mutually independently uniformly distributed, each on $[0,1]$,

$$
\begin{equation*}
E_{\varepsilon \mid Z}\left[x(\varepsilon, Z) \mid Z=\bar{z}, \varepsilon \in T_{\varepsilon}\right]=E_{\tau}\left[x(\varepsilon(\tau), \bar{z}) \mid \tau \in T_{\tau}\right]=\frac{\int \cdots \int_{t \in T_{\tau}} \tilde{x}(t, \bar{z}) d t_{1} \ldots d t_{M}}{\int \cdots \int_{t \in T_{\tau}} d t_{1} \ldots d t_{M}} . \tag{35}
\end{equation*}
$$

The structural derivatives evaluated at the point $\bar{\Psi}$ considered in Section 3 are expressed as functions of probabilities, $\bar{\tau}$, defining iterated conditional quantiles of the conditional distribution of $\varepsilon$ given $Z=\bar{z}$, functions like $\tilde{x}$ above. So, equation (35) implies that conditional expected values of functions of structural derivatives (and so, for example, average derivatives, variances of derivatives) can be identified if the local identification conditions hold at $Z=\bar{z}$ over some set of quantile probabilities.

## 9. Concluding remarks

Some features of the identifying model are briefly discussed.
9.1. Multiplicity of latent variates. The identifying restrictions proposed here require that effectively ${ }^{23}$ there be no more latent variables in the model than observable outcomes. This condition may not be tenable when measurement error distorts recorded outcomes or covariates, or when there is high dimensional heterogeneity such as in the mixture models employed widely in the analysis of durations. It may not be tenable in panel data models and other multilevel models in which there is a nested error structure. In these sorts of cases, models that secure identification must embody strong restrictions because one is trying to secure identification of features of high dimensional distributions from the relatively low dimensional reductions of them about which data are informative.

In practice one finds strong additivity assumptions imposed in every such case. Thus, 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 latent variables that drive a model to be linear combinations of individual specific latent variates and latent variates that vary across and within individuals.

Most mixed proportionate hazard (MPH) models restrict an individual specific latent variate to be additive in the log hazard function. As shown in Chesher (2002b), under this restriction the individual specific latent variate combines with the latent variate that produces variation in durations so that the restriction on the multiplicity of latent variates is satisfied in MPH models.
9.2. Continuous variation. Continuous variation in covariates is essential if values of partial derivatives of structural functions are to be identified in the absence of parametric restrictions. However certain partial differences of structural functions can be nonparametrically identified when there is only discrete variation in covariates. This is explored in Chesher (2002a).
9.3. Parametric and semiparametric restrictions. Identification in semiparametric and parametric nonseparable and separable models can be assessed using the

[^16]results of this paper even though a nonparametric attack has been taken. Because this was thought to be useful some normalisations which would often be taken in a study of purely nonparametric identification have not been made in this paper.

Semiparametric and parametric conditions place restrictions on the matrices of derivatives introduced in Section 4.1. For example index restrictions that lead to model equations of the form:

$$
Y_{i}=h_{i}\left(Y_{i+}^{\prime} \delta_{i}, Z^{\prime} \beta_{i}, \varepsilon_{i}, \varepsilon_{i+}\right)
$$

require the matrices $A$ and $B$ of Section 4.1 to be equal to a diagonal matrices ${ }^{24}$ times matrices of constants (containing the parameters $\delta_{i}$ and $\beta_{i}$ ). Conditions sufficient for local identification of $A$ and $B$ ensure global identification of ratios of $\delta$ parameters and ratios of $\beta$ parameters taken within equations. This is explored in the context of duration models in Chesher (2002b).
9.4. Quantile insensitivity. The quantile insensitivity conditions used here have the advantage that they have meaning when data are generated by structures in which integer order moments do not exist, a situation that arises for example in some of the problems studied in financial econometrics. The quantile regression based estimators which naturally flow from these restrictions are easy to compute and their sampling properties are understood. In the context of the nonseparable simultaneous equations systems studied here they have the potential to improve understanding of the distribution of the impacts of policy interventions.

## References

Abadie, Alberto, Angrist, Joshua D., and Guido W. Imbens (2002): "Instrumental variables estimates of the effect of subsidized training on the quantiles of trainee earnings," Econometrica, 70, 91-117.
Altonji, Joseph G., and Rosa L. Matzkin (2001): "Panel data estimators for nonseparable models with endogenous regressors," NBER Technical Working Paper 267.

Amemiya, Takeshi (1982): "Two stage least absolute deviations estimators," Econometrica, 50, 689-711.
Becker, Gary S., and Barry R. Chiswick (1966): "Education and the distribution of earnings," American Economic Review, 56, 358-369.
Brown, Bryan W. (1983): "The identification problem in systems nonlinear in the variables," Econometrica, 51,175-196.
Brown, Donald J., and Rosa L. Matzkin (1996): "Estimation of nonparametric functions in simultaneous equations models, with an application to consumer demand," mimeo, Department of Economics, Northwestern University.

[^17]Card, David (2001): "Estimating the returns to schooling: Progress on some persistent econometric problems," Econometrica, 69, 1127-1160.
Card, David (1995): "Earnings, ability and schooling revisited," in Research in Labour Economics, Volume 14, ed., S.Polachek. Greenwich, Conn: JAI Press.
Chaudhuri, Probal (1991): "Nonparametric estimation of regression quantiles and their local Bahadur representation," Annals of Statistics, 19, 760-777.
Chaudhuri, Probal, Kjell Doksum and Alexander Samarov (1997): "On average derivative quantile regression," Annals of Statistics, 25, 715-744.
Chernozhukov, Victor., and Christian Hansen (2001): "An IV model of quantile treatment effects," MIT Department of Economics Working Paper Series No. 02-06, December 2001.
Chesher, Andrew 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, Andrew D. (2001b): "Exogenous impact and conditional quantile functions," Centre for Microdata Methods and Practice Working Paper 01/01.
Chesher, Andrew D. (2001c): "Quantile driven identification of structural derivatives," Centre for Microdata Methods and Practice Working Paper 08/01.
Chesher, Andrew D. (2002a): "Instrumental values," Centre for Microdata Methods and Practice Working Paper 17/02.
Chesher, Andrew D. (2002b): "Semiparametric identification in duration models," Centre for Microdata Methods and Practice Working Paper 20/02.
Chiswick, Barry R. (1974): Income inequality: regional analyses within a human capital framework. New York: Columbia University Press.
Chiswick, Barry R., and Jacob Mincer (1972): "Time series changes in personal income inequality," Journal of Political Economy, 80, S34-S66.
Darolles, Serge, Jean-Pierre Florens and Eric Renault, (2000): "Nonparametric instrumental regression," CREST Documents de Travail 2000-17.
Das, Mitali (2000): "Nonparametric instrumental variable estimation with discrete endogenous regressors," presented at the 2000 World Congress of the Econometric Society, Seattle.
Fisher, Franklin M. (1959): "Generalization of the rank and order conditions for identifiability," Econometrica, 27, 431-447.
Fisher, Franklin M. (1961): "Identifiability criteria in nonlinear systems," Econometrica, 29, 574-590.
Fisher, Franklin M. (1966): The identification problem in econometrics, New York: McGraw Hill.
Frisch, Ragnar A.K. (1934): Statistical confluence analysis by means of complete regression systems, Publication No. 5, Oslo: Universitets Økonomiske Institutt.
Frisch, Ragnar A.K. (1938): "Statistical versus theoretical relations in economic macrodynamics", Memorandum prepared for a conference in Cambridge, England, July 18-20, 1938, to discuss drafts of Tinbergen's League of Nations publications; mimeographed.

Haavelmo, Trygve, M. (1944): "The probability approach in econometrics," Econometrica, 12, Supplement, July 1944, 118 pp.
Heckman, James J. (1990): "Varieties of selection bias," American Economic Review, Papers and Proceedings, 80, 313-318.
Heckman, James J., Jeffrey A. Smith and Nancy 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. Heckman, James J., and Edward Vytlacil (1998): "Instrumental variables methods for the correlated random coefficients model: estimating the average returns to schooling when the return is correlated with schooling," The Journal of Human Resources, 33, 974-1002.
Hurwicz, Leonid (1950): "Generalization of the concept of identification," in Statistical inference in dynamic economic models. Cowles Commission Monograph 10, New York, John Wiley.
Ihaka, Ross, and Robert Gentleman (1996) "R: A language for data analysis and graphics," Journal of Computational and Graphical Statistics, 5, 299-314.
Imbens, Guido W., and Joshua D. Angrist (1994): "Identification and estimation of average treatment effects," Econometrica, 62, 467-476.
Imbens, Guido W., and Whitney K. Newey (2001): "Identification and estimation of triangular simultaneous equations models without additivity," mimeo.
Kahn, Shakeeb (2001): "Two-stage rank estimation of quantile index models," Journal of Econometrics, 100, 319-355.
Koenker, Roger W., and Gilbert W. Bassett Jr. (1978): "Regression quantiles," Econometrica, 46, 33-50.
Koenker, Roger W. and Vasco d’Orey (1987): "Computing regression quantiles," Journal of the Royal Statistical Society, Series C, 36, 383-393.
Koopmans, Tjalling C., and Olav Reiersol (1950): "The identification of structural characteristics," Annals of Mathematical Statistics, 21, 165-181.
Koopmans, Tjalling C., Herman Rubin and Roy 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, Sokbae (2002): "Efficient semiparametric estimation of a partially linear quantile regression model," forthcoming Econometric Theory.
Manski, Charles F. (1975): "Maximum score estimation of the stochastic utility model of choice," Journal of Econometrics, 3, 205-228.
Manski, Charles F. (1988a): "Identification of binary response models," Journal of the American Statistical Association, 83, 729-738.
Manski, Charles F. (1988b): Analog estimation methods in econometrics, New York: Chapman and Hall.
Matzkin, Rosa 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.

Mincer, Jacob (1974): Schooling experience and earnings. New York: Columbia University Press.
Newey, Whitney K., and James L. Powell (1988): "Nonparametric Instrumental Variables Estimation," mimeo, Department of Economics, MIT.
Newey, Whitney K., and James L. Powell (1990): "Efficient estimation of linear and Type- 1 censored regression models under conditional quantile restrictions," Econometric Theory, 6, 295-317.
Newey, Whitney K., James L. Powell, and Frank Vella (1999): "Nonparametric Estimation of Triangular Simultaneous Equations Models," Econometrica 67, 565-603.
Pinkse, Joris (2000): "Nonparametric two-step regression functions when regressors and errors are dependent," Canadian Journal of Statistics, 28, 289-300.
Roehrig, Charles S. (1988): "Conditions for identification in nonparametric and parametric models", Econometrica 56, 433-447.
Rothenberg, Thomas J. (1971): "Identification in parametric models," Econometrica, 39, 577-591.
Tinbergen, Jan (1930), Bestimmung und Deutung von Angebotskurven: Ein Beispiel, Zeitschrift für Nationalökonomie 1, 669-679.
Vytlacil, Edward (2002): "Independence, monotonicity and latent index models: an equivalence result," Econometrica, 70, 331-342.
Wald, Abraham (1950): "Note on identification of economic relations," in Statistical inference in dynamic economic models. Cowles Commission Monograph 10, New York, John Wiley.
Wegge, Leon L., (1965): "Identifiability Criteria for a System of Equations as a Whole," The Australian Journal of Statistics, 7, 67-77.
Wooldridge, Jeffrey M., (1997): "On two-stage least squares estimation of the average treatment effect in a random coefficients model," Economics Letters, 56, 129133.

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


[^0]:    *I thank the referees and an editor of the journal and Jaap Abbring, Christian Dustmann, Art Goldberger, Hidehiko Ichimura, Valérie Lechene, Charles Manski, Whitney Newey, Peter Phillips, Roger Koenker, Richard Smith, Richard Spady, Elie Tamer, and Tieman Woutersen for helpful comments. Some of the results of this paper were given in Chesher (2001, 2002a, 2002b) and in an invited address at the Econometric Society European Meeting, Venice, 2002.

[^1]:    ${ }^{1}$ The definitions are slightly amended to make explicit the role of covariates upon whose marginal distributions no identifying restrictions will be placed, other than that certain covariates exhibit continuous variation.
    ${ }^{2}$ Strictly speaking, the value of the feature $f$ is identified by the value of the functional $\mathcal{G}\left(F_{Y \mid X}\right)$.

[^2]:    ${ }^{3}$ Heckman and Vytlacil (1998) study identification in a nonseparable "random coefficients" model of returns to schooling under parametric restrictions. The model they study draws on Card (1995); see also Card (2001). Random coefficients specifications for this problem date from the work of Becker and Chiswick (1966), Chiswick (1974), Chiswick and Mincer (1972) and Mincer (1974). Equations (1) and (2) can be viewed as a nonparametric "random coefficients" specification of a returns-to-schooling model.

[^3]:    ${ }^{4}$ This notation is used throughout, $Q_{A \mid B_{1} \ldots, B_{M}}\left(\tau, b_{1}, \ldots, b_{M}\right)$ denoting the conditional $\tau$-quantile of random variable $A$ given conditioning variables $B_{1}=b_{1}, \ldots, B_{M}=b_{M}$. This definition applies when $A$ given $B$ has a discrete or continuous distribution.
    ${ }^{5}$ See for example Wooldridge (1997) and Card (2001).
    ${ }^{6}$ Global strict monotonicity is a stronger condition than is required. A precise statement of the required conditions is in Section 4.1.

[^4]:    ${ }^{7}$ See for example Heckman (1990), Imbens and Angrist (1994), Das (2000) and Vytlacil (2002).

[^5]:    ${ }^{8}$ Recall that $Q_{A \mid B_{1} \ldots B_{K}}\left(\tau, b_{1}, \ldots, b_{M}\right)$ denotes the conditional $\tau$-quantile of random variable $A$ given $B_{1}=b_{1}, \ldots, B_{M}=b_{M}$.
    ${ }^{9}$ Subject to Restriction I below.

[^6]:    ${ }^{10}$ One way to think about this is to note that, under the conditions to be stated, the variates $u=$ $\left\{u_{i}\right\}_{i=1}^{M}$, where $u_{i}=F_{\varepsilon_{i} \mid \varepsilon_{i+} Z}\left(\varepsilon_{i} \mid \varepsilon_{i+}, z\right)$, are distributed independently of $Z$ and are independently uniformly distributed on $[0,1]^{M}$. Substituting for $\varepsilon$ using $\varepsilon_{i}=Q_{\varepsilon_{i} \mid \varepsilon_{i+} Z}\left(u_{i}, u_{i+}, z\right)$ is merely a normalisation of the functions $h$. With that normalisation in place, quantile probabilities for the variates $u$ are identical to the associated quantile values of $u$ and correspond to quantile probabilities for the iterated conditional quantiles of the variates $\varepsilon$.

[^7]:    ${ }^{11}$ This is assured to hold if, with all coordinates other than $\varepsilon_{i}$ fixed at their values at $\bar{\Psi}$, each function $h_{i}$ is strictly monotonic in $\varepsilon_{i}$. However global strict monotonicity is not required by this restriction. For example a function $h_{i}$ could oscillate to some degree at extreme values of $\varepsilon_{i}$ without breaking the single crossing condition.

[^8]:    ${ }^{12}$ Identification of partial differences when covariates exhibit discrete variation is studied in Chesher (2002a).
    ${ }^{13}$ To be clear $h_{M}^{-1}\left(z, y_{M}\right)$ is the solution for $e_{M}$ to

    $$
    y_{M}=h_{M}\left(z, e_{M}\right)
    $$

    the function $h_{M-1}^{-1}\left(y_{M}, z, y_{M-1}, h_{M}^{-1}\left(z, y_{M}\right)\right)$ is the solution for $e_{M-1}$ to

    $$
    y_{M-1}=h_{M-1}\left(y_{M}, z, e_{M-1}, h_{M}^{-1}\left(z, y_{M}\right)\right)
    $$

    and so forth.

[^9]:    ${ }^{14}$ Note that $G$ incorporates the normalisation $\nabla_{\varepsilon_{i}} h_{i}=1$, for all $i$.

[^10]:    ${ }^{15}$ If at $\bar{\Psi}$ the elements of $\varepsilon$ 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 $\varepsilon$ to be locally quantile independent at $\bar{\Psi}$ given $Z$. If the equations of the model (10) were restricted so that $G=I_{M}$, which would arise if each equation $i$ contained no $\varepsilon_{j}, j \neq i$, then $C=\left(I_{M}-H\right)^{-1}$ and conditions sufficient to identify $C$ would identify $H$.
    ${ }^{16}$ The triangularity restriction implies that all but $M(M-1) / 2$ elements of $U$ are zero.
    ${ }^{17}$ These local linear restrictions could arise from non-local nonlinear restrictions on the structural functions.

[^11]:    ${ }^{18}$ Because

    $$
    \begin{aligned}
    \operatorname{rank}\left(R_{M}^{\prime}\left(\left(I_{M}-U^{\prime}\right) \otimes I_{M}\right) R_{M}\right) & =\min \left(\operatorname{rank}\left(R_{M}\right), \operatorname{rank}\left(\left(I_{M}-U^{\prime}\right) \otimes I_{M}\right)\right) \\
    & =\min \left(M(M-1) / 2, M^{2}\right) \\
    & =M(M-1) / 2
    \end{aligned}
    $$

[^12]:    ${ }^{19}$ Compare the second row of $C^{-1} B+J$ in equation (19) with the second rows of $V$ and $\nabla_{Z} \bar{Q}$ in equations (20) and (21).

[^13]:    ${ }^{20}$ For example, $Y_{M-1}=h_{M-1}\left(h_{M}\left(Z, \varepsilon_{M}\right), Z, \varepsilon_{M-1}, \varepsilon_{M}\right)$.

[^14]:    ${ }^{21}$ For 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). The R software suite, see Ihaka and Gentleman (1996) and cran.r-project.org, implements a variety of quantile regression estimation procedures.

[^15]:    ${ }^{22}$ Nonparametric estimation can proceed recursively, derivatives of the conditional $\bar{\tau}_{i}$-quantile of $Y_{i}$ given $Y_{i+}$ and $Z$ being estimated local to $\bar{z}$ and $\hat{y}_{i+}$.

[^16]:    ${ }^{23}$ Any additional latent variates must combine, for example additively, to produce no more latent variates than there are observable outcomes.

[^17]:    ${ }^{24}$ These are $\operatorname{diag}\left\{\nabla_{Y_{i+}^{\prime} \delta_{i}} h_{i}\right\}_{i=1}^{M}$ and $\operatorname{diag}\left\{\nabla_{Z_{i}^{\prime} \beta} h_{i}\right\}_{i=1}^{M}$ evaluated at $\bar{\Psi}$, associated with respectively $A$ and $B$.

