

# Excess heterogeneity, endogeneity and index restrictions

ANDREW CHESHER\*

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

February 18th 2008

ABSTRACT. An outcome is determined by a structural function in which the effect of some variables of interest is transmitted through a scalar index. Multiple sources of stochastic variation can appear as arguments of the structural function, but not in the index. There may be endogeneity, that is observable and unobservable variables may not be independently distributed. Conditions are provided under which there is local identification of measures of the relative sensitivity of the index to variations in pairs of its possibly endogenous arguments, namely ratios of partial derivatives of the index.

## 1. INTRODUCTION

Many questions arising in microeconomic practice lead to the use of models which include more unobservable latent variables than there are observable stochastic outcomes, that is excess heterogeneity. The latent variables often represent unobserved characteristics of individuals and of the environment in which they make decisions. The inclusion of such variables is common in, for example, models of durations (see van den Berg (2001)) such as mixed proportional hazard models, in discrete choice models (see for example Brownstone and Train (1998), Chesher and Santos Silva (2002), McFadden and Train (2000)) and in count data models (see Cameron and Trivedi (1998)). There is a large econometric literature concerned with random coefficients models which permit this sort of excess heterogeneity. (Chow (1984)). Excess heterogeneity also arises in other cases, for example when there is measurement error and in panel data models.

It is common to find strong restrictions imposed in models that admit excess heterogeneity. Frequently the specification is fully parametric as in the mixed multinomial logit models of Brownstone and Train (1998). When parametric restrictions are

---

\*I thank Whitney Newey for remarks on an earlier paper (Chesher (2002)) which stimulated this work and to Lars Nesheim for helpful discussions. I thank the Leverhulme Trust for support through grants to the Centre for Microdata Methods and Practice (CeMMAP) and the research project "Evidence, Inference and Enquiry" and the Economic and Social Research Council for support of CeMMAP since July 1st 2007 through grant RES-589-28-0001.

not imposed there are usually strong semiparametric restrictions. For example: most of the single spell duration models used in practice that permit excess heterogeneity require there to be a single latent variate that acts multiplicatively on the hazard function; measurement error and “individual effects” in panel data models are usually required to be additive.

The aim of this paper is to explore the extent to which strong restrictions such as these can be relaxed, while preserving a model with the power to identify interesting structural features.

When there is excess heterogeneity the probability distributions of observable variables are relatively low dimensional reductions of the distributions of structural variables, obtained by taking expectations over the distributions of supernumerary latent variates. Information about fundamental structural features may not survive the averaging process. In the face of this difficulty one possibility is to focus on the identification of averages of structural features, as in for example Imbens and Newey (2007), but knowledge of such averages may not be what is required to understand the impacts of policy changes. It is interesting to explore alternatives. Another approach is to impose restrictions which shield certain structural objects from the effects of averaging. This is the approach studied here.

In the models explored here excess heterogeneity can arise from any finite number of sources and there is an index restriction. The index restriction requires the effect on an outcome of certain variables of interest to pass entirely through a scalar function of those variables, an index, and that this index be free of latent variates. Continuously distributed variables that appear in the index are permitted to be *endogenous* in the sense that they may covary with any or all of the latent variates that appear in the model.

The structural features whose identifiability is studied in this paper are *ratios of derivatives of the index* at some specified values of the variables that appear in the index. This is therefore a study of local identification. These ratios are referred to as index relative sensitivity (IRS) measures because they measure the relative sensitivity of the index, and therefore of the outcome, to variation in a pair of its arguments. When the index is linear the ratios do not depend on the values of the arguments of the index. Then conditions sufficient to achieve local identification of the value of an IRS measure achieve global identification of the ratio of coefficients of the linear index.

With more sources of stochastic variation than there are outcomes structural functions necessarily involve non-additive latent variates, as noted in Hurwicz (1950). The identifying model employed here admits non-additive latent variates and embodies triangularity restrictions as in Chesher (2003, 2005) and Imbens and Newey (2007).

IRS measures are often of interest in models for binary outcomes. For example in discrete choice models of travel demand there is interest in the “value of travel time” defined as the ratio of coefficients on travel time and travel cost. There are other contexts in which the relative sensitivity of an index to variation in its arguments is of interest. For example in models of intrahousehold allocation there is interest in the relative sensitivity of expenditures to variations in the incomes of two partners; in models for the duration of unemployment there is interest in the relative sensitivity of unemployment duration to variations in unemployment benefits and other household income or the wage prior to unemployment. In all these cases one or more of the arguments of the index could be endogenous although this is a possibility frequently ignored, perhaps because it is not understood how to deal with endogeneity in this situation. It is this which motivates this study which mainly focusses on identification issues.

The remainder of the paper is organised as follows. The structural equation, index restriction and IRS measures are defined in Section 2 and examples of microeconomic models accommodated within the framework employed here are given in Section 3. The identification strategy, based on a “control function” argument, is introduced in Section 4. Related literature is briefly reviewed in Section 5. Identification theorems are given in Section 6 and estimation is briefly considered in Section 7. The main results associate IRS measures with functions of derivatives of various *distribution* functions involving observable variables. These apply when  $W$ , the outcome of interest, is discrete or continuous. When the outcome is continuous the IRS measures can be associated with functions of derivatives of conditional *quantile* functions and the expressions are given in Section 8. Section 9 concludes.

## 2. THE STRUCTURAL EQUATION AND THE IRS MEASURES

In the models considered here the outcome of interest is a random variable  $W$  determined by a structural equation of the following form.

$$W = h_0(\theta(Y_1, \dots, Y_M, Z_1, \dots, Z_K), Z_1^*, \dots, Z_L^*, U_1, \dots, U_N) \quad (1)$$

Scalar  $W$  may be discrete or continuous,  $U \equiv \{U_n\}_{n=1}^N$  are latent variates,  $Y \equiv \{Y_m\}_{m=1}^M$  are observable continuously distributed endogenous random variables which may covary with  $U$ , and  $Z \equiv \{Z_k\}_{k=1}^K$  are observable continuously varying covariates whose covariation with  $U$  is limited to some degree to be specified.  $\theta$  is the index of interest, a scalar valued differentiable function.<sup>1</sup>

---

<sup>1</sup>The results could be extended to the case in which there are multiple indexes as arise in, for example, multiple discrete choice models.

The variables  $Z^* \equiv \{Z_l^*\}_{l=1}^L$  are discrete or continuously varying variables which may appear in the structural function but not in the index. Identification of the sensitivity of structural functions to these variables is not considered. There could be other variables entering the index which exhibit discrete variation. Their presence is not made explicit in the notation and sensitivity of the structural function to variation in their values is not considered here.

The IRS measures studied here have the following form.

$$\kappa_{a,b}(y, z) \equiv \frac{\nabla_a \theta(y, z)}{\nabla_b \theta(y, z)}, \quad (a, b) \in \{y_1, \dots, y_M, z_1, \dots, z_K\}$$

Without further restriction their values depend on the values of  $y \equiv \{y_m\}_{m=1}^M$  and  $z \equiv \{z_k\}_{k=1}^K$ . Conditions sufficient for local identification of particular  $\kappa_{a,b}$  at a specified value of  $(Y, Z)$ ,  $(\bar{y}, \bar{z})$ , will be considered.

There are structural equations determining the elements of  $Y$  as follows.

$$Y_m = h_m(Z, Z^*, V_m), \quad m \in \{1, \dots, M\} \quad (2)$$

Each function  $h_m$  is a strictly monotonic function of  $V_m$  which is a continuously distributed latent variate.  $Y$  may be endogenous in the sense that  $V \equiv \{V_m\}_{m=1}^M$  and  $U$  may not be independently distributed. The structural equations for  $(W, Y_1, \dots, Y_M)$  thus have a triangular form as in Chesher (2003, 2005) and Imbens and Newey (2007).

### 3. EXAMPLES

This Section gives examples of microeconomic models in which a structural equation of the form (1) arises.

#### Example 1 - Mixed hazard duration models

Consider hazard functions for a continuously distributed duration (e.g. of unemployment)  $W$  conditional on observable  $Y = y$ ,  $Z = z$ ,  $Z^* = z^*$  and on unobservable, possibly vector,  $E = e$  of the form:

$$\lambda(w|\theta(y, z), z^*, e) \quad (3)$$

where  $\theta$  is a scalar valued function. The conditional distribution function of  $W$  given  $Y, Z, Z^*$  and  $E$  is

$$F_{W|YZZ^*E}(w|y, z, z^*, e) = 1 - \exp(-\Lambda(w|\theta(y, z), z^*, e))$$

where  $\Lambda(w|\theta(y, z), z^*, e)$  is the integrated hazard function, as follows.

$$\Lambda(w|y, z, z^*, e) \equiv \int_0^w \lambda(\omega|\theta(y, z), z^*, e) d\omega$$

The conditional  $\tau$ -quantile function of  $W$  given  $Y, Z, Z^*$  and  $E$  is

$$Q_{W|YZZ^*E}(\tau|y, z, z^*, e) = \Lambda^{-1}(-\log(1 - \tau)|\theta(y, z), z^*, e)$$

where  $\Lambda^{-1}$  is the inverse integrated hazard function satisfying

$$a = \Lambda(\Lambda^{-1}(a|\theta(y, z), z^*, e), \theta(y, z), z^*, e)$$

for all  $a, y, z, z^*$  and  $e$ .

With  $D$  distributed uniformly on  $(0, 1)$  independent of  $Y, Z, Z^*$  and  $E$ , the following structural equation delivers a random variable  $W$  whose conditional distribution given  $Y, Z, Z^*$  and  $E$  has the hazard function  $\lambda$  given in equation (3).

$$W = \Lambda^{-1}(-\log(1 - D)|\theta(Y, Z), Z^*, E)$$

Defining  $U \equiv (D, E)$  this is a structural equation of the form set out in equation (1).

Note that there is no requirement that the excess heterogeneity terms,  $E$ , act multiplicatively on the hazard function and there is no limit on the number of such terms appearing in the model. The results of the paper concern identification of IRS measures when  $Y$  covaries with  $E$ .

The mixed hazard model for single spell data, treated in van den Berg (2001), has a single source of excess heterogeneity,  $E$ , acting multiplicatively in the hazard function, as follows.

$$\lambda(W|\theta(Y, Z), Z^*, E) = \bar{\lambda}(W|\theta(Y, Z), Z^*) \times E$$

In this case the structural function for  $W$  is

$$W = \bar{\Lambda}^{-1}(-\log(1 - D)E^{-1}|\theta(Y, Z), Z^*)$$

where  $\bar{\Lambda}^{-1}$  is the inverse of the function

$$\bar{\Lambda}(w|y, z, z^*) \equiv \int_0^w \bar{\lambda}(\omega|\theta(y, z), z^*) d\omega$$

with respect to its  $w$  argument. Under the proportionate heterogeneity restriction the two sources of stochastic variation coalesce into one, with implications for iden-

tification and estimation developed in Chesher (2002).

### Example 2 - Heterogeneous binary choice

An example of the sort of binary response model for  $W \in \{0, 1\}$  that falls in the class of models considered here is

$$P[W = 0|Y, Z, Z^*, E] = \Phi(E_0 + E_1 Z^* + \theta_y Y + \theta_z Z) \quad (4)$$

where  $\Phi$  is a known or unknown function from  $\mathfrak{R}^1 \rightarrow (0, 1)$ . Here  $Y$ ,  $Z$  and  $Z^*$  are observable scalar variables and  $E \equiv (E_0, E_1)$  contains latent variates. The covariate  $Z^*$  has a “random coefficient”  $E_1$  and there is “random intercept”  $E_0$ . The variate  $Y$  is endogenous in the sense that it may covary with  $E$ . The coefficients on  $Y$  and  $Z$  are nonstochastic and their ratio  $\theta_y/\theta_z$  is the structural feature whose identification is studied in this paper.

Let  $D$  be uniformly distributed on  $(0, 1)$  conditional on  $E_0, E_1, Y, Z$  and  $Z^*$ . Then there is the following structural equation for  $W$ .

$$W = \begin{cases} 0 & , \quad D \leq \Phi(E_0 + E_1 Z^* + \theta_y Y + \theta_z Z) \\ 1 & , \quad D > \Phi(E_0 + E_1 Z^* + \theta_y Y + \theta_z Z) \end{cases}$$

This has the form of equation (1) with  $U \equiv (D, E)$ ,  $\theta(Y, Z) \equiv \theta_y Y + \theta_z Z$ . The linear index restriction in (4) is a restriction additional to that considered in this paper and is imposed just by way of example.

Blundell and Powell (2003) study identification and estimation in binary choice models with a linear index depending on endogenous variables, like (4), with a single source of heterogeneity. The models studied by Brownstone and Train (1998) and McFadden and Train (2000) have multiple sources of heterogeneity but they do not permit endogeneity.

### Example 3 - Partially linear model

Consider the structural equation

$$W = Y'\beta_Y + Z'\beta_Z + r(Z^*, U)$$

which is as in the model studied in Robinson (1988) with  $Y$  and  $Z$  appearing only in a linear index,  $Y'\beta_Y + Z'\beta_Z$ . Multiple sources of stochastic variation,  $U$ , may appear in an unknown function  $r$  along with covariates  $Z^*$  which do not feature in the index of interest. The model studied here admits the possibility that  $Y$  is endogenous due to covariation of  $Y$  and  $U$ .

In another formulation accommodated within the framework employed in this

paper:

$$W = Z^*U + r(Y, Z)$$

the potentially endogenous variables appear in the unknown function,  $g$ , and the coefficients in the linear index are stochastic.

#### 4. IDENTIFICATION

The strategy employed in developing identification conditions for IRS measures is now outlined. For this purpose the covariates  $Z^*$  which appear in the structural function (1) but not in the index  $\theta$  are assumed absent. Their presence would not change the argument below except in inessential details.<sup>2</sup>

Let the joint distribution function of  $U$  given  $Z$  and  $V$  be denoted by  $F_{U|ZV}$ . Conditions are placed on the equations for the elements of  $Y$  sufficient to ensure that

$$F_{U|ZY}(u|z, y) = F_{U|ZV}(u|z, v)|_{v=g(z, y)}$$

where  $g(z, y) \equiv \{g_m(z, y_m)\}_{m=1}^M$  and each  $g_m$  is the inverse function of  $h_m$  with respect to its  $V_m$  argument. Each function  $g_m$  is such that, for all  $z$  and  $y_m$ :

$$y_m = h_m(z, g_m(z, y_m)).$$

It follows directly that the conditional distribution function of the outcome of interest,  $W$ , given  $Y = y$  and  $Z = z$  at  $W = w$  can be expressed as a function of  $w, z$ , the index of interest,  $\theta(y, z)$ , and the  $M$  indexes  $g_m(z, y)$ ,  $m \in \{1, \dots, M\}$ , as follows.

$$F_{W|ZY}(w|z, y) = s(\theta(y, z), g_1(z, y_1), \dots, g_M(z, y_M), w, z)$$

The dependence of the function  $s$  on  $z$  through its last argument arises from the dependence of  $F_{U|ZV}(u|z, v)$  on  $z$ . This dependence will typically be subject to restrictions and will often be assumed absent.

The conditional distribution functions  $F_{W|ZY}$  and  $F_{Y_1|Z}, \dots, F_{Y_M|Z}$  can be known because they are conditional distribution functions involving observable variates, and, if  $Y$  and  $Z$  exhibit continuous variation around a point  $(\bar{y}, \bar{z})$ , their  $Y$ - and  $Z$ -derivatives at that point can also be known.

An IRS measure  $\kappa_{a,b}(\bar{y}, \bar{z})$ ,  $(a, b) \in \{y_1, \dots, y_M, z_1, \dots, z_K\}$  is identified if the derivatives  $\nabla_a \theta(\bar{y}, \bar{z})$  and  $\nabla_b \theta(\bar{y}, \bar{z})$  are identified up to a common non-zero finite

<sup>2</sup>At various points where there is conditioning on  $Z$  there would have to be conditioning on  $Z$  and  $Z^*$ . The point at which identification is sought would be  $(\bar{w}, \bar{y}, \bar{z}, \bar{z}^*)$ . There is no point at which partial derivatives with respect to elements of  $Z^*$  are considered and so no limitation on the covariation of  $Z^*$  and  $(U, V)$  is needed.

valued factor of proportionality. This will happen if there are sufficient restrictions on the structural equations (1) and (2) and on the distribution of  $(U, V)$  conditional on  $Z$  to permit the values of  $\nabla_a \theta(\bar{y}, \bar{z})$  and  $\nabla_b \theta(\bar{y}, \bar{z})$  to be deduced up to a common non-zero finite valued factor of proportionality from knowledge of the  $Y$ - and  $Z$ -derivatives of  $F_{W|ZY}$  and  $F_{Y_1|Z}, \dots, F_{Y_M|Z}$  at  $(\bar{y}, \bar{z})$ .

In Section 6 precise identification conditions are set out and a Theorem stating an identification result is stated. The proof is in the Appendix.

To give a flavour of the result of the Theorem, consider the case in which in the index there is a single endogenous variable,  $Y_1$ , and a covariate  $Z_1$ . In the structural equation for  $Y_1$  there is a covariate,  $Z_2$ , variation in which does not affect the value of the index at  $(\bar{y}, \bar{z})$ . This local exclusion restriction, together with covariation restrictions requiring (a)  $U$  given  $V$  is independent of  $Z \equiv \{Z_1, Z_2\}$  and (b) that at a point  $(\bar{y}_1, \bar{z})$ , with  $\bar{z} \equiv \{\bar{z}_1, \bar{z}_2\}$ :

$$\nabla_{z_1} F_{V_1|Z} = \nabla_{z_2} F_{V_1|Z} = 0 \quad (5)$$

imply the following:

$$\kappa_{y_1 z_1}(\bar{y}_1, \bar{z}) = \frac{\nabla_{y_1} F_{W|ZY_1} - \nabla_{y_1} F_{Y_1|Z} \left( \frac{\nabla_{z_2} F_{W|ZY_1}}{\nabla_{z_2} F_{Y_1|Z}} \right)}{\nabla_{z_1} F_{W|ZY_1} - \nabla_{z_1} F_{Y_1|Z} \left( \frac{\nabla_{z_2} F_{W|ZY_1}}{\nabla_{z_2} F_{Y_1|Z}} \right)} \quad (6)$$

where all functions are evaluated at  $(\bar{y}_1, \bar{z})$  and at any value of  $w$ .<sup>3</sup> This serves to identify  $\kappa_{y_1 z_1}(\bar{y}_1, \bar{z})$ . Note that the exclusion of  $U$  from the index results in  $\kappa_{y_1 z_1}(\bar{y}_1, \bar{z})$  being overidentified - a condition manifested by the invariance of (6) to the choice of  $w$ .

When  $W$  is *continuously* distributed the derivatives of conditional distribution functions that appear in (6) can be replaced by ratios of derivatives of conditional *quantile* functions, as explained in Section 8. After some simplification this results in the following alternative to (6).

$$\kappa_{y_1 z_1}(\bar{y}_1, \bar{z}_1) = \frac{\nabla_{y_1} Q_{W|ZY_1} + \left( \frac{\nabla_{z_2} Q_{W|ZY_1}}{\nabla_{z_2} Q_{Y_1|Z}} \right)}{\nabla_{z_1} Q_{W|ZY_1} - \nabla_{z_1} Q_{Y_1|Z} \left( \frac{\nabla_{z_2} Q_{W|ZY_1}}{\nabla_{z_2} Q_{Y_1|Z}} \right)} \quad (7)$$

Here  $Q_{W|ZY_1}$  is shorthand for the  $\rho$ -quantile function of  $W$  given  $Z$  and  $Y_1$ , and  $Q_{Y_1|Z}$  is shorthand for the conditional  $\tau_1$ -quantile of  $Y_1$  given  $Z$ . In (7) the arguments of these quantile functions are evaluated at  $Y_1 = \bar{y}_1$ ,  $Z = \bar{z}$ , at  $\tau_1 = \bar{\tau}_1$ , where  $\bar{\tau}_1$

<sup>3</sup>The independence condition on  $U$  given  $V$  need only hold for  $V$  and  $Z$  in a neighbourhood of  $(\bar{z}, \bar{v}_1)$  where  $\bar{v}_1$  is such that  $h_1(\bar{z}, \bar{v}_1) = \bar{y}_1$ .



satisfies

$$\bar{y}_1 = Q_{Y_1|Z}(\bar{\tau}_1|\bar{z})$$

and at any value of  $\rho$ .

The numerator and denominator of (7) are identical to the expressions given in Chesher (2003) for respectively the  $Y_1$ - and  $Z_1$ -derivatives of a structural function with a single nonadditive latent variable.

$$W = h(Y_1, Z_1, U)$$

when  $U$  is a scalar and so the *sole* source of stochastic variation, in continuously distributed  $W$  given  $Y_1$  and  $Z_1$ . When there are multiple sources of stochastic variation the numerator and denominator of (7) are not equal to these structural derivatives. However, with the index and other restrictions imposed here, their ratio is equal to the ratio of the index derivatives.

Estimates of an IRS measure can be built from parametric, semi- or nonparametric estimates of conditional distribution functions and their derivatives, or, when  $W$  is continuously distributed, on estimates of conditional quantile functions and their derivatives. This is briefly discussed in respectively Sections 7 and 8.

## 5. RELATED LITERATURE

The basic idea employed in this paper dates back at least as far as Tinbergen (1930) in which the problem of identification in linear simultaneous equations systems was attacked by developing conditions under which values of structural form parameters could be deduced from values of parameters of regression functions - the reduced form equations of the linear simultaneous system.

The conditional distribution functions  $F_{W|ZY}$  and  $F_{Y_m|Z}$ ,  $m \in \{1, \dots, M\}$  are regression functions, namely of  $1[W \leq w]$  on  $Z$  and  $Y$ , and of  $1[Y_m \leq y_m]$  on  $Z$ ,  $m \in \{1, \dots, M\}$ . The values of the  $Y$ - and  $Z$ -derivatives of the conditional distribution functions at  $(\bar{w}, \bar{y}, \bar{z})$  are the coefficients of a linear approximation to these regression functions, and these coefficients are functions of the structural parameters of interest, namely the index derivatives at  $(\bar{y}, \bar{z})$ . The latter are identified when their values can be deduced from knowledge of the values of these coefficients. Viewed in this way it is not surprising that the identification conditions and their development echo the classical linear simultaneous equations identification analysis given full expression in Koopmans, Rubin and Leipnik (1950).

Index restrictions like that considered here have been used in many other papers including Han (1987), Robinson (1988), Powell, Stock and Stoker (1989), Newey and

Stoker (1993), Chaudhuri, Doksum and Samarov (1997) and Kahn (2001). Much of the semiparametric literature dealing with models embodying index restrictions assumes away the issue of endogeneity. Newey (1985), Lewbel (1998, 2000), Lewbel and Linton (2002), Honoré and Hu (2002), Hong and Tamer (2003) and Blundell and Powell (2003) do consider endogeneity but, aiming at identifying different structural features, employ different identifying restrictions.

Chesher (2003) takes a similar approach to that taken in this paper, providing conditions under which values of partial derivatives of *structural functions* at a point of interest are identified. Critical among these conditions is the requirement that the number of sources of stochastic variation permitted by a model be equal to the number of observable stochastic outcomes. This paper weakens this restriction but at the cost of (a) imposing an index restriction and (b) obtaining identification of IRS measures rather than derivatives of structural functions.

The mixed hazard model with multiplicative heterogeneity studied in Example 1 in Section 3 in which two sources of stochastic variation coalesce to one effective source was studied in Chesher (2002).

## 6. IDENTIFICATION OF INDEX DERIVATIVES

This Section introduces four restrictions and then gives a Theorem which states that a model embodying these restrictions identifies index derivatives up to a common factor of proportionality. Some remarks on the assumptions are provided as they are introduced. The Theorem is proved in an Appendix.

To simplify the notation the covariates  $Z^*$  which appear in the structural equation (1) and in the examples of Section 3 are assumed absent. Their inclusion requires minor changes to the assumptions and, with these amendments, results in no change to the result of the Theorem.<sup>4</sup>

**Restriction 1.**  $W, Y \equiv \{Y_i\}_{i=1}^M, U \equiv \{U_i\}_{i=1}^N$  and  $V \equiv \{V_i\}_{i=1}^M$  are random variables, with  $Y$  and  $V$  continuously distributed and  $Z \equiv \{Z_i\}_{i=1}^K$  are variables exhibiting continuous variation in a neighbourhood of a point  $\bar{z}$ . The support of  $U$  given  $V$  and  $Z$  does not depend on the values of  $V$  or  $Z$ . The conditional density functions of  $V_m$  given  $Z$ ,  $m \in \{1, \dots, M\}$  are positive valued at  $\bar{z}$  and their support does not depend upon the value of  $Z$ .

The Theorem will concern the identification of the values of index partial derivatives at a point  $\mathcal{X} \equiv (\bar{w}, \bar{y}, \bar{z})$ . The random variable  $W$  is the outcome of interest,  $Y$  is a list of potentially endogenous variables.  $U$  and  $V$  are lists of unobservable,

---

<sup>4</sup>This point is amplified in the Appendix after the proof of the Theorem.

latent variates whose covariation with  $Z$ , a list of covariates may be limited to some degree by Restriction 4 below.  $Y$  is required to be continuously distributed, and  $Z$  is required to exhibit continuous variation, because of the focus here on partial derivatives of a nonparametrically specified index.<sup>5</sup>

**Restriction 2.** For any value of  $Z$ ,  $U$  and  $V$ , unique values of  $W$  and  $Y$  are determined by the structural equations

$$\begin{aligned} W &= h_0(\theta(Y, Z), U) \\ Y_m &= h_m(Z, V_m), \quad m \in \{1, \dots, M\} \end{aligned}$$

where  $\theta$  is a scalar valued function. Each function  $h_m$  is strictly monotonic with respect to variation in  $V_m$ .

In the equations for the elements of  $Y$  specified in Restriction 2 each element of  $Y$  depends on  $Z$  and an element of  $V$  but not on other elements of  $Y$  or  $V$ . This is a restricted version of an alternative set up in which these equations are in full triangular form, each  $Y_m$ ,  $m > 1$ , depending on  $Y_{m-1}, \dots, Y_1$ ,  $Z$ , and latent variates  $V_m, V_{m-1}, \dots, V_1$  with  $Y_m$  strictly monotonic in  $V_m$ . This is the construction used in Chesher (2003). Identification results similar to those developed here can be obtained in the context of this full triangular system.

The inverse function of each function  $h_m$  with respect to  $V_m$  exists by virtue of the strict monotonicity condition. It is denoted by  $g_m$ . For any  $z$  and  $y_m$ :

$$y_m = h_m(z, g_m(y_m, z)), \quad m \in \{1, \dots, M\}.$$

Let  $g(y, z)$  denote the  $M \times 1$  vector of inverse functions  $\{g_m(y_m, z)\}_{m=1}^M$ .

Under Restrictions 1 and 2 the conditional distribution function of  $W$  given  $Y$  and  $Z$  is

$$\begin{aligned} F_{W|YZ}(w|y, z) &= \int \dots \int_{h_0(\theta(y, z), u) \leq w} dF_{U|VZ}(u|g(y, z), z) & (8) \\ &\equiv s(\theta(y, z), g(y, z), w, z) & (9) \end{aligned}$$

and for  $m \in \{1, \dots, M\}$  the marginal distribution function of  $Y_m$  given  $Z$  is

$$F_{Y_m|Z}(y_m|z) = F_{V_m|Z}(g_m(y_m, z))|z \quad (10)$$

$$\equiv r_m(g_m(y_m, z), z). \quad (11)$$

---

<sup>5</sup>Identification when endogenous variables have discrete distributions, is studied in Chesher (2003b). The identifying restrictions of that paper do not permit excess heterogeneity.

The function  $s$  defined in (9) and the functions  $r_1, \dots, r_M$  defined in (11) play a crucial role in the statement and proof of the Theorem.

**Restriction 3.** *At  $\mathcal{X}$ , defined after Restriction 1, the conditional distribution function of  $W$  given  $Y$  and  $Z$ ,  $F_{W|YZ}(w|y, z)$ , is differentiable with respect to  $y$  and  $z$ , and for  $m \in \{1, \dots, M\}$  the conditional distribution function of  $Y_m$  given  $Z$ ,  $F_{Y_m|Z}(y_m|z)$  is differentiable with respect to  $y_m$  and  $z$ .*

This relatively high level assumption on  $F_{W|YZ}$  and  $F_{Y_m|Z}$ ,  $m \in \{1, \dots, M\}$ , requires differentiability of the structural functions  $h_0$ ,  $\theta$ , and  $h_m$ ,  $m \in \{1, \dots, M\}$ . The conditional distribution function of  $W$  given  $Y$  and  $Z$  is not required to be differentiable with respect to  $w$ , so  $W$  can be a discrete random variable.

The conditional distribution functions  $F_{W|YZ}$  and  $F_{Y_m|Z}$ ,  $m \in \{1, \dots, M\}$  are identified and their derivatives at  $\mathcal{X}$  with respect to elements of  $y$  and  $z$  are identified because  $y$  and  $z$  exhibit continuous variation at  $\mathcal{X}$  under Restriction 1. So the identifiability of index derivatives hangs on whether there are additional restrictions such that their values (up to a common nonzero scale factor) can be deduced from knowledge of the derivatives of the conditional distribution functions  $F_{W|YZ}$  and  $F_{Y_m|Z}$ ,  $m \in \{1, \dots, M\}$ .

Arrays of partial derivatives, all evaluated at  $\mathcal{X}$ , are now defined. Let  $\nabla$  indicate a partial derivative with respect to the variable “.”. Arguments of functions are suppressed.  $s_\theta$  denotes the value of the (scalar) partial derivative  $\nabla_\theta s$  at  $\mathcal{X}$

$$\begin{aligned}
R_y \equiv \begin{bmatrix} \nabla_{y_1} F_{Y_1|Z} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \nabla_{y_M} F_{Y_M|Z} \end{bmatrix} & \quad R_z \equiv \begin{bmatrix} \nabla_{z_1} F_{Y_1|Z} & \cdots & \nabla_{z_1} F_{Y_M|Z} \\ \vdots & \ddots & \vdots \\ \nabla_{z_K} F_{Y_1|Z} & \cdots & \nabla_{z_K} F_{Y_M|Z} \end{bmatrix} \\
S_y \equiv \begin{bmatrix} \nabla_{y_1} F_{W|YZ} \\ \vdots \\ \nabla_{y_M} F_{W|YZ} \end{bmatrix} & \quad S_z \equiv \begin{bmatrix} \nabla_{z_1} F_{W|YZ} \\ \vdots \\ \nabla_{z_K} F_{W|YZ} \end{bmatrix} \\
\lambda_y \equiv s_\theta \times \begin{bmatrix} \nabla_{y_1} \theta \\ \vdots \\ \nabla_{y_M} \theta \end{bmatrix} & \quad \lambda_z \equiv s_\theta \times \begin{bmatrix} \nabla_{z_1} \theta \\ \vdots \\ \nabla_{z_K} \theta \end{bmatrix} & \quad \gamma \equiv \begin{bmatrix} \nabla_{g_1} s / \nabla_{g_1} r_m \\ \vdots \\ \nabla_{g_M} s / \nabla_{g_M} r_M \end{bmatrix} \\
s_z \equiv \begin{bmatrix} \nabla_{z_1} s \\ \vdots \\ \nabla_{z_K} s \end{bmatrix} & \quad r_z \equiv \begin{bmatrix} \nabla_{z_1} r_1 & \cdots & \nabla_{z_1} r_M \\ \vdots & \ddots & \vdots \\ \nabla_{z_K} r_1 & \cdots & \nabla_{z_K} r_M \end{bmatrix}
\end{aligned}$$

The terms  $\nabla_{g_m} r_m$ , which figure in the definition of the vector  $\gamma$ , are positive

by virtue of Assumption 1. In the array  $r_z$  a partial derivative  $\nabla_{z_i} r_m$  arises on differentiating  $r_m(g_m(y_m, z), z)$  with respect to the  $i$ th element of its final argument, holding  $g_m(y_m, z)$  fixed, a non-zero value arising if there is some dependence between  $V_m$  and  $Z_i$ .

The index derivatives, which are the structural features of interest, appear in the definition of the vectors  $\lambda_y$  and  $\lambda_z$  multiplied by a *common* factor,  $s_\theta$ .

**Restriction 4.** Define  $\delta \equiv -r_z \gamma$ . There are  $G$  restrictions on  $\lambda_y$ ,  $\lambda_z$ ,  $\gamma$ ,  $s_z$  and  $\delta$  as follows.

$$A_y \lambda_y + A_z \lambda_z + A_\gamma \gamma + A_s s_z + A_\delta \delta = a \quad (12)$$

The arrays  $a$  and  $A_y$ ,  $A_z$ , etc., are nonstochastic conditional on  $Z = \bar{z}$ .  $s_\theta$  is finite and nonzero.

Restrictions on  $s_z$  limit the degree of covariation of  $U$  and  $Z$  given  $V$ . A typical derivative in the vector  $s_z$  is as follows.

$$\nabla_{z_k} s = \int \cdots \int_{h_0(\theta(\bar{y}, \bar{z}), u) \leq \bar{w}} d \left( \nabla_{z_k} F_{U|VZ}(u|g(\bar{y}, \bar{z}), z) \Big|_{z=\bar{z}} \right)$$

A derivative  $\nabla_{z_k} s$  will be zero when the partial derivative  $\nabla_{z_k} F_{U|VZ}(u|g(\bar{y}, \bar{z}), z) \Big|_{z=\bar{z}}$  is zero for all  $u$  in the set defined by  $h_0(\theta(\bar{y}, \bar{z}), u) \leq \bar{w}$ . In practice, since the structural function is unknown, this can only be assured, when  $U$  is multidimensional, by requiring  $U$  to be independent of  $Z_k$  given  $V = g(\bar{y}, \bar{z})$  for variations in  $z$  in a neighbourhood of  $\bar{z}$ .

However, when  $U$  is scalar and  $h_0$  is monotonic in  $U$ ,

$$|\nabla_{z_k} s| = \left| \nabla_{z_k} F_{U|VZ}(h_0^{-1}(\theta(\bar{y}, \bar{z}), \bar{w})|g(\bar{y}, \bar{z}), z) \Big|_{z=\bar{z}} \right|$$

which can be zero under a restriction on the dependence of  $U$  on  $Z_k$  given  $V = g(\bar{y}, \bar{z})$  for variations in  $z_k$  in a neighbourhood of  $\bar{z}_k$ , a restriction which is local to  $U = h_0^{-1}(\theta(\bar{y}, \bar{z}), \bar{w})$ . This is the case considered in Chesher (2003) where it is shown that the index restriction is not required to achieve identification of partial derivatives of the structural function when  $U$  is scalar.

Restrictions on  $\gamma$  limit the covariation of  $U$  and elements of  $V$ . Restrictions on  $r_z$ , which may imply restrictions on  $\delta$ , limit the degree of covariation of  $V$  and  $Z$ . Restrictions on  $\lambda_y$  and  $\lambda_z$  limit the sensitivity of the index to elements of  $Y$  and  $Z$ .

Homogeneous restrictions<sup>6</sup> on the index derivatives imply the same homogeneous restrictions on  $\lambda_y$  and  $\lambda_z$ . In the absence of parametric restrictions there will typically

<sup>6</sup>For example zero restrictions and restrictions requiring equality of two or more index derivatives.

be no prior knowledge of the *value* of  $s_\theta$  so in practice non-homogeneous restrictions on  $\lambda_y$  and  $\lambda_z$  are unlikely to arise.

After the following definitions the identification Theorem can be stated.

$$\Phi \equiv \begin{bmatrix} I_M & 0 & R_y & 0 & 0 \\ 0 & I_K & R_z & -I_K & I_K \\ A_y & A_z & A_\gamma & A_s & A_\delta \end{bmatrix} \quad \psi \equiv \begin{bmatrix} \lambda_y \\ \lambda_z \\ \gamma \\ s_z \\ \delta \end{bmatrix} \quad \phi \equiv \begin{bmatrix} S_y \\ S_z \\ a \end{bmatrix}$$

Note that  $\Phi$  is an array with dimension  $(G + M + K) \times (2M + 3K)$  and  $\psi$  and  $\phi$  are vectors with respectively  $(2M + 3K)$  and  $(G + M + K)$  elements.

**Theorem 1**

*Under Restrictions 1 - 4  $\Phi\psi = \phi$  and  $\psi$  is identified if and only if  $\text{rank}(\Phi) = 2M + 3K$  for which a necessary condition is  $G \geq M + 2K$ .*

The proof is given in the Appendix.

The vectors  $\lambda_y$  and  $\lambda_z$  contain values of derivatives of the index at  $\mathcal{X}$ , multiplied by a common scale factor. They measure the sensitivity of the conditional distribution function of  $W$  given  $Y$  and  $Z$  that arises from variations in  $Y$  and  $Z$  passing purely through the index  $\theta$ . However they do not generally measure the sensitivity of the value delivered by the structural equation  $h_0$  to variations in  $Y$  and  $Z$  passing purely through the index. Accordingly they may be of no economic interest in themselves.

The IRS measures are *ratios* of index derivatives in which the common scale factor,  $s_\theta$ , is of course absent, so identification of  $\lambda_y$  and  $\lambda_z$  implies identification of IRS measures as long as  $s_\theta$  is nonzero, as required by Restriction 4.

In practice it will be common to impose the  $2K$  restrictions  $s_z = 0$  and  $r_z = 0$ , the latter implying  $\delta = 0$ . These restrictions limit the covariation of  $(U, V)$  and  $Z$  at  $Z = \bar{z}$  and would be satisfied if  $(U, V)$  and  $Z$  were independently distributed. Define the following arrays.

$$\Phi^+ \equiv \begin{bmatrix} I_M & 0 & R_y \\ 0 & I_K & R_z \\ A_y & A_z & A_\gamma \end{bmatrix} \quad \psi^+ \equiv \begin{bmatrix} \lambda_y \\ \lambda_z \\ \gamma \end{bmatrix} \quad \phi^+ \equiv \begin{bmatrix} S_y \\ S_z \\ a \end{bmatrix}$$

The following Corollary is relevant to this case.

**Corollary 1**

*Under Restrictions 1 - 4 and the additional restrictions (i)  $s_z = 0$ , (ii)  $r_z = 0$ ,*

the values of  $\lambda_y$ ,  $\lambda_z$  and  $\gamma$  are identified if and only if

$$\text{rank } \Phi^+ = 2M + K \quad (13)$$

for which a necessary condition is  $G \geq M$ . In that case define

$$X \equiv A_y R_y + A_z R_z - A_\gamma \quad x \equiv A_y S_y + A_z S_z - a \quad (14)$$

If the rank condition (13) is satisfied, then, for any rank  $M$ ,  $M \times G$  matrix  $P$ ,

$$\begin{aligned} \gamma &= (X' P' P X)^{-1} X' P' P x \\ \lambda_y &= S_y - R_y \gamma \\ \lambda_z &= S_z - R_z \gamma. \end{aligned}$$

The proof is in the Appendix.

As noted after Restriction 4, when  $U$  is multidimensional the condition  $s_z = 0$ , imposed in Corollary 1, will be difficult to maintain without restricting  $U$  to be independent of  $Z$  given  $V$ . Suppose now that this independence restriction is imposed along with  $r_z = 0$ , as in Corollary 1 and, further, suppose that the conditions imposed in Restriction 4 do not involve  $\gamma$  (so  $A_\gamma = 0$ ) and are homogeneous (so  $a = 0$ ).

Define the following arrays in which dependence of elements on the value,  $w$ , of the outcome  $W$  is made explicit.

$$\Phi^{\natural} \equiv \begin{bmatrix} I_M & 0 & R_y \\ 0 & I_K & R_z \\ A_y & A_z & 0 \end{bmatrix} \quad \psi^{\natural}(w) \equiv \begin{bmatrix} \lambda_y(w) \\ \lambda_z(w) \\ \gamma(w) \end{bmatrix} \quad \phi^{\natural}(w) \equiv \begin{bmatrix} S_y(w) \\ S_z(w) \\ 0 \end{bmatrix}$$

$$\lambda_y(w) \equiv \nabla_{\theta} s(\theta(\bar{y}, \bar{z}), g(\bar{y}, \bar{z}), w, \bar{z}) \theta_y \quad \lambda_z(w) \equiv \nabla_{\theta} s(\theta(\bar{y}, \bar{z}), g(\bar{y}, \bar{z}), w, \bar{z}) \theta_z$$

$$\gamma_m(w) \equiv \nabla_{g_m} s(\theta(\bar{y}, \bar{z}), g(\bar{y}, \bar{z}), w, \bar{z}) / \nabla_{g_m} r_m, \quad m \in \{1, \dots, M\}$$

For some  $\Gamma \subset \mathfrak{R}^1$  and a bounded nonnegative valued function  $B(w)$  with  $\int_{w \in \Gamma} dB(w) = 1$ , define

$$\begin{aligned} \phi^{\natural} &\equiv \int_{w \in \Gamma} \phi^{\natural}(w) dB(w) \\ \psi^{\natural} &\equiv \int_{w \in \Gamma} \psi^{\natural}(w) dB(w) \equiv \begin{bmatrix} \lambda_y^{\natural} \\ \lambda_z^{\natural} \\ \gamma^{\natural} \end{bmatrix} \end{aligned}$$

with  $B(w)$  chosen so that  $\phi^{\natural}$  and  $\psi^{\natural}$  have bounded elements. There is the following Corollary to Theorem 1.

**Corollary 2**

*Under Restrictions 1 - 4 and the additional restrictions: (i)  $r_z = 0$ , (ii)  $U$  is independent of  $Z$  given  $V$ , (iii)  $A_\gamma = 0$ , (iv)  $a = 0$ ;  $\Phi^{\natural}\psi^{\natural} = \phi^{\natural}$ , and  $\psi^{\natural}$  is identified if and only if*

$$\text{rank } \Phi^{\natural} = 2M + K$$

for which a necessary condition is  $G \geq M$ .

The proof is straightforward on noting that  $\Phi^{\natural}\psi^{\natural}(w) = \phi^{\natural}(w)$  implies  $\Phi^{\natural}\psi^{\natural} = \phi^{\natural}$ .

The rank condition of Corollary 2 is the same as that of Corollary 1 with  $A_\gamma = 0$ . Corollary 2 leads to identification of IRS measures as long as there exists a weighting function  $B(w)$  such that

$$\nabla_{\theta}s^{\natural} \equiv \int_{w \in \Gamma} \nabla_{\theta}s(\theta(\bar{y}, \bar{z}), g(\bar{y}, \bar{z}), w, \bar{z})dB(w)$$

is nonzero and finite, because  $\lambda_y^{\natural} = \nabla_{\theta}s^{\natural}\theta_y$  and  $\lambda_z^{\natural} = \nabla_{\theta}s^{\natural}\theta_z$  and the common factor  $\nabla_{\theta}s^{\natural}$  will then cancel upon forming up an IRS measure.

## 7. ESTIMATION

Theorem 1 and its two Corollaries point to estimation procedures. For example, with nonparametric estimates of the conditional distribution function derivatives,  $\hat{R}_y$ ,  $\hat{R}_z$ ,  $\hat{S}_y$  and  $\hat{S}_z$ , estimates,  $\hat{\Phi}$  and  $\hat{\phi}$ , of  $\Phi$  and  $\phi$ , can be assembled incorporating the restrictions to hand, and a minimum distance estimator

$$\hat{\psi} = \arg \min_{\psi} \left( \hat{\Phi}\psi - \hat{\phi} \right)' \Omega \left( \hat{\Phi}\psi - \hat{\phi} \right)$$

can be calculated using a suitable positive definite matrix  $\Omega$ .<sup>7</sup>

Corollary 1 points to explicit expressions for estimators of  $\gamma$ ,  $\lambda_y$  and  $\lambda_z$  when the restrictions  $r_z = 0$  and  $s_z = 0$  are imposed. Estimates of the arrays of distribution function derivatives together with the restrictions to hand, lead to estimates  $\hat{X}$  and  $\hat{x}$  of  $X$  and  $x$  in (14) and thus to the estimator

$$\hat{\gamma} = \left( \hat{X}'P'P\hat{X} \right)^{-1} \hat{X}'P'P\hat{x}$$

<sup>7</sup>In order to obtain consistent estimates of  $\hat{R}_y$ ,  $\hat{R}_z$ ,  $\hat{S}_y$  and  $\hat{S}_z$  it will be necessary to impose the identifying restrictions proposed here over some region of which  $(\bar{y}, \bar{z})$  is an interior point, and to impose further conditions on the distribution of  $(U, V)$  given  $Z$ .



with  $\hat{\lambda}_y = \hat{S}_y - \hat{R}_y \hat{\gamma}$  and  $\hat{\lambda}_z = \hat{S}_z - \hat{R}_z \hat{\gamma}$  following directly.

Corollary 2, which imposes additional restrictions, points to estimators based on integrated (with respect to  $w$ ) weighted derivatives of distribution functions.

In the overidentified case the asymptotic efficiency of the estimators will depend on the choice of the matrices  $\Omega$  and  $P$ . The identification result has been obtained under index restrictions and it will be desirable to impose these when the distribution function derivatives are estimated. One might wish to impose additional semiparametric or parametric restrictions.

## 8. IDENTIFICATION *via* CONDITIONAL QUANTILE FUNCTIONS

So far the outcome,  $W$ , has not been required to be continuously distributed. Now suppose that it is, at least conditional on  $Y$  and  $Z$  lying in a neighbourhood of  $(\bar{y}, \bar{z})$ . In this case the matrices of conditional distribution function derivatives that appear in Theorem 1 and Corollary 1 can be re-expressed in terms of derivatives of conditional quantile functions.

This is so because for a random variable  $A$ , continuously distributed conditional on  $B$  lying in a neighbourhood of  $b$ ,

$$\nabla_b F_{A|B}(a|b) = - \left. \frac{\nabla_b Q_{A|B}(\tau|b)}{\nabla_\tau Q_{A|B}(\tau|b)} \right|_{\tau=F_{A|B}(a|b)} \quad (15)$$

$$\nabla_a F_{A|B}(a|b) = \left. \frac{1}{\nabla_\tau Q_{A|B}(\tau|b)} \right|_{\tau=F_{A|B}(a|b)} \quad (16)$$

where  $F_{A|B}$  and  $Q_{A|B}$  are the conditional distribution and quantile functions of  $A$  given  $B = b$ . This follows directly from the definition of  $Q_{A|B}(\tau|b)$  as the inverse function of  $F_{A|B}(a|b)$  with respect to the argument  $a$ , that is, for all  $\tau$  and  $b$ :

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

Equations (15) and (16) do not hold when  $A$  has a discrete distribution given  $B = b$  because in that case  $\nabla_\tau Q_{A|B}(\tau|b)$  is almost everywhere zero.

This Section explores an alternative, quantile function based approach to identification for the case in which the outcome  $W$  is continuously distributed given  $Y$  and  $Z$  lie in a neighbourhood of  $(\bar{y}, \bar{z})$ . The development is done for the case considered in Corollary 1 in which  $r_z = 0$  and  $s_z = 0$ . Also, there are assumed to be no restrictions on  $\gamma$  and the restrictions on  $\lambda_y$  and  $\lambda_z$  are assumed homogeneous, that is in (12),  $A_\gamma = 0$  and  $a = 0$ .

Let  $\bar{\tau} \equiv \{\bar{\tau}_m\}_{m=1}^M$  be probabilities such that each  $\bar{y}_m$  is the  $\bar{\tau}_m$ -quantile of  $Y_m$  conditional on  $Z = \bar{z}$ , that is, for  $m \in \{1, \dots, M\}$ :

$$\bar{y}_m = Q_{Y_m|Z}(\bar{\tau}_m|\bar{z}) \quad \bar{\tau}_m = F_{Y_m|Z}(\bar{y}_m|\bar{z}).$$

Let  $\bar{\rho}$  be such that  $\bar{w}$  is the  $\bar{\rho}$ -quantile of  $W$  given  $Y = \bar{y}$  and  $Z = \bar{z}$ , that is:

$$\bar{w} = Q_{W|YZ}(\bar{\rho}|\bar{y}, \bar{z}) \quad \bar{\rho} = F_{W|YZ}(\bar{w}|\bar{y}, \bar{z}).$$

Note that the point  $\mathcal{X} \equiv (\bar{w}, \bar{y}, \bar{z})$  is identical to  $\tilde{\mathcal{X}} \equiv (\bar{\rho}, \bar{\tau}, \bar{z})$ . Restriction 1 is modified to require  $W$  given  $Y = \bar{y}$  and  $Z = \bar{z}$  to be continuously distributed with positive density at  $W = \bar{w}$ .

**Restriction 1'.**  $W, Y \equiv \{Y_i\}_{i=1}^M, U \equiv \{U_i\}_{i=1}^N$  and  $V \equiv \{V_i\}_{i=1}^M$  are random variables, with  $W, Y$  and  $V$  continuously distributed and  $Z \equiv \{Z_i\}_{i=1}^K$  are variables exhibiting continuous variation in a neighbourhood of a point  $\bar{z}$ . The support of  $U$  given  $V$  and  $Z$  does not depend on the values of  $V$  or  $Z$ . The conditional density functions of  $V_m$  given  $Z, m \in \{1, \dots, M\}$  are positive valued at  $\bar{z}$  and their support does not depend upon the value of  $Z$ . The conditional density of  $W$  given  $Y = \bar{y}$  and  $Z = \bar{z}$  is positive valued at  $W = \bar{w}$ .

Define the following arrays of quantile function derivatives. Arguments of functions, all evaluated at  $\tilde{\mathcal{X}}$ , are suppressed.

$$G_{\tau}^{M \times M} \equiv \begin{bmatrix} \nabla_{\tau_1} Q_{Y_1|Z} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \nabla_{\tau_M} Q_{Y_M|Z} \end{bmatrix} \quad G_z^{K \times M} \equiv \begin{bmatrix} \nabla_{z_1} Q_{Y_1|Z} & \cdots & \nabla_{z_1} Q_{Y_M|Z} \\ \vdots & \ddots & \vdots \\ \nabla_{z_K} Q_{Y_1|Z} & \cdots & \nabla_{z_K} Q_{Y_M|Z} \end{bmatrix}$$

$$H_y^{M \times 1} \equiv \begin{bmatrix} \nabla_{y_1} Q_{W|YZ} \\ \vdots \\ \nabla_{y_M} Q_{W|YZ} \end{bmatrix} \quad H_z^{K \times 1} \equiv \begin{bmatrix} \nabla_{z_1} Q_{W|YZ} \\ \vdots \\ \nabla_{z_K} Q_{W|YZ} \end{bmatrix}$$

Using (15) and (16) the arrays,  $R_y, R_z, S_y$  and  $S_z$ , of conditional distribution function derivatives can be re-expressed in terms of conditional quantile function derivatives as follows.

$$R_y = G_{\tau}^{-1} \quad R_z = -G_z G_{\tau}^{-1} \quad S_y = -\frac{1}{\nabla_{\rho} Q_{W|YZ}} H_y \quad S_z = -\frac{1}{\nabla_{\rho} Q_{W|YZ}} H_z$$

The following reparameterisation is employed.

$$\tilde{\lambda}_y \equiv \nabla_{\rho} Q_{W|YZ} \lambda_y \quad \tilde{\lambda}_z \equiv \nabla_{\rho} Q_{W|YZ} \lambda_z \quad \tilde{\gamma} \equiv \nabla_{\rho} Q_{W|YZ} G_{\tau}^{-1} \gamma$$

Assumption 1' ensures  $\nabla_{\rho} Q_{W|YZ} > 0$  and the nonsingularity of  $G_{\tau}$ . There is then Corollary 3 to Theorem 1.

**Corollary 3**

*Under Restrictions 1', 2 - 4, and the additional restrictions (i)  $s_z = 0$ , (ii)  $r_z = 0$ , with no restrictions on  $\gamma$ , and with homogeneous restrictions on  $\tilde{\lambda}_y$  and  $\tilde{\lambda}_z$ , the values of  $\tilde{\lambda}_y$ ,  $\tilde{\lambda}_z$  and  $\tilde{\gamma}$  are identified if and only if*

$$\text{rank} \begin{bmatrix} I_M & 0 & I_M \\ 0 & I_K & -G_z \\ A_y & A_z & 0 \end{bmatrix} = 2M + K \quad (17)$$

for which a necessary condition is  $G \geq M$ . In that case, with  $\tilde{X}$  and  $\tilde{x}$  defined by

$$\tilde{X} \equiv -A_y + A_z G_z \quad \tilde{x} \equiv A_y H_y + A_z H_z \quad (18)$$

if the rank condition (13) is satisfied, for any rank  $M$ ,  $M \times G$  matrix  $P$ ,

$$\tilde{\gamma} = \left( \tilde{X}' P' P \tilde{X} \right)^{-1} \tilde{X}' P' P \tilde{x} \quad (19)$$

$$\tilde{\lambda}_y = -H_y - \tilde{\gamma} \quad (20)$$

$$\tilde{\lambda}_z = -H_z + G_z \tilde{\gamma}. \quad (21)$$

The proof is in the Appendix.

Corollary 3 suggests an alternative route to estimation of IRS measures when  $W$  is continuously distributed, as follows.

1. Calculate an estimate of the  $\bar{\tau}_m$ -quantile of  $Y_m$  given  $Z = \bar{z}$  for  $m \in \{1, \dots, M\}$ . This produces estimates,  $\hat{y}_m$ , of  $\bar{y}_m$  for  $m \in \{1, \dots, M\}$ .
2. Calculate estimates of the  $z$ -derivatives of the  $\bar{\tau}_m$ -quantile of  $Y_m$  given  $Z = \bar{z}$  for  $m \in \{1, \dots, M\}$ . This produces an estimate of  $G_z$ .
3. Calculate estimates of the  $y$ - and  $z$ - derivatives of the  $\bar{\rho}$ -quantile of  $W$  given  $Y = \hat{y}_m$  and  $Z = \bar{z}$ . This produces estimates of  $H_y$  and  $H_z$ .
4. Using the restrictions to hand ( $A_y$  and  $A_z$ ) substitute estimates in (18) and for a suitable choice of  $P$  calculate an estimate of  $\tilde{\gamma}$  using (19) and then of  $\tilde{\lambda}_y$  and  $\tilde{\lambda}_z$  using (20) and (21).
5. Ratios of estimates of  $\tilde{\lambda}_y$  and  $\tilde{\lambda}_z$  are the desired estimates of ratios of elements of  $\theta_y$  and  $\theta_z$ .

With nonparametric identification assured one could conduct estimation imposing additional semiparametric or parametric restrictions. Even if that is not done it would be sensible to impose the index restrictions that underlie the identification result on the conditional quantile estimates.

The rank condition of Corollary 3 is a special case of the single equation rank condition given in Chesher (2003). However the *estimation* procedure proposed above differs from that proposed there because different “parameters” are being considered. Chesher (2003) considers estimation of partial derivatives of a structural function whereas in this paper partial derivatives of an *index* that appears as an *argument* of a structural function are the objects of interest.

With more sources of stochastic variation than observable outcomes (the case  $N > 1$  in this paper) the results of Chesher (2003) on identification and estimation of derivatives of structural functions do not apply. The index restriction used in this paper is a key to making progress in problems with excess heterogeneity.

## 9. CONCLUDING REMARKS

This paper has developed an identifying model for problems in which structural functions involve multiple latent variables and endogenous observed arguments. Examples of microeconomic models in which these features arise include count and duration data models admitting across individual heterogeneity and models in which household outcomes are determined by characteristics and experiences of individual household members.

The identifying models use an index restriction that shields certain structural features from stochastic variation and a triangular construction which allows a control function argument to be used when there are appropriate exclusion (and inclusion) restrictions. The control function approach requires continuity in the distribution of endogenous variables. It is interesting to ask whether, in the absence of continuously varying endogenous variables, a partial identification result such as in Chesher (2005) could be obtained after imposing further restrictions.

## REFERENCES

- BLUNDELL, RICHARD W., AND JAMES L. POWELL (2004): “Endogeneity in semi-parametric binary response models,” *Review of Economic Studies*, 71, 655-679.
- BROWNSTONE DAVID AND KENNETH TRAIN (1998): “Forecasting new product penetration with flexible substitution patterns,” *Journal of Econometrics*, 28, 109-129.

- CAMERON, A. COLIN AND PRAVIN. K. TRIVEDI (1998): *Regression Analysis of Count Data*, Econometric Society Monograph No. 30, Cambridge University Press: Cambridge.
- CHAUDHURI, P., K. DOKSUM AND A. SAMAROV (1997): "On average derivative quantile regression," *Annals of Statistics*, 25, 715-744.
- CHESHER, ANDREW D., (2005): "Nonparametric identification under discrete variation," *Econometrica*, 73, 1525-1550.
- CHESHER, ANDREW D., (2003): "Identification in nonseparable models," *Econometrica*, 71, 1405-1441.
- CHESHER, ANDREW D., (2002): "Semiparametric identification in duration models," Centre for Microdata Methods and Practice Working Paper CWP20/02.
- CHESHER, ANDREW D., AND JOAO M.C. SANTOS SILVA (2002): "Taste Variation in Discrete Choice Models," *Review of Economic Studies*, 69, 148-167.
- CHOW, GREGORY C., (1984): "Random and Changing Coefficient Models," chapter 21 in *Handbook of Econometrics, Volume 2*, ed., Griliches, Z., and M.D. Intrilligator, North Holland: Amsterdam.
- HAN, A., (1987): "A non-parametric analysis of transformations," *Journal of Econometrics*, 35, 191-209.
- HONORÉ, BO, E., AND LUOJIA HU (2002): "Estimation of cross sectional and panel data censored regression models with endogeneity," unpublished working paper.
- HONG, HAN AND ELIE TAMER (2003): "Inference in censored models with endogenous regressors," *Econometrica*, 71, 905-932.
- HURWICZ, LEONID (1950): "Systems with nonadditive disturbances," in *Statistical inference in dynamic economic models*. Cowles Commission Monograph 10, New York, John Wiley.
- IMBENS, GUIDO W., AND WHITNEY K. NEWAY (2007): "Identification and estimation of triangular simultaneous equations models without additivity," Manuscript..
- KAHN, S., (2001): "Two-stage rank estimation of quantile index models," *Journal of Econometrics*, 100, 319-355.
- 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.
- LEWBEL, ARTHUR (2000): "Semiparametric qualitative response model estimation with unknown heteroscedasticity or instrumental variables," *Journal of Econometrics*, 97, 145-177.
- LEWBEL, ARTHUR (1998): "Semiparametric latent variable model estimation with endogenous or mis-measured regressors," *Econometrica* 66, 105-121.

- LEWBEL, ARTHUR AND OLIVER LINTON (2002): "Nonparametric truncated and censored regression," *Econometrica*, 70, 765-779.
- McFADDEN, DANIEL AND KENNETH TRAIN (2000): "Mixed MNL models for discrete response," *Journal of Applied Econometrics*, 15, 447-470.
- NEWBY, WHITNEY K., (1985): "Semiparametric estimation of limited dependent variable models with endogenous explanatory variables," *Annales de l'INSEE* 59-60, 219-237.
- NEWBY, W.K., AND T. STOKER, (1993): "Efficiency of weighted average derivative estimators and index models," *Econometrica*, 61, 1199-1223.
- POWELL, J., STOCK, J., AND T. STOKER, (1989): "Semiparametric estimation of index coefficients," *Econometrica*, 57, 1403-1430.
- ROBINSON, P.M. (1988): "Root-N-consistent semiparametric regression," *Econometrica*, 56, 931-954.
- TINBERGEN, JAN (1930): "Bestimmung und Deutung von Angebotskurven: Ein Beispiel," *Zeitschrift für Nationalökonomie* 1, 669-679.
- VAN DEN BERG, GERARD J., (2001): "Duration models: specification, identification and multiple durations," chapter 55 in *Handbook of Econometrics, Volume 5*, ed., Heckman, J.J., and E. Leamer, North Holland: Amsterdam.

## APPENDIX: PROOFS

**A1. Proof of Theorem 1**

The partial derivatives of the conditional distribution functions (8) and (10) with respect to elements,  $y_m$  and  $z_k$  of  $y$  and  $z$  are as follows. Arguments of functions, all evaluated at  $\mathcal{X}$ , are suppressed.

$$\nabla_{y_m} F_{W|YZ} = \nabla_{\theta} s \nabla_{y_m} \theta + \nabla_{g_m} s \nabla_{y_m} g_m \quad (\text{A1.1})$$

$$\nabla_{z_k} F_{W|YZ} = \nabla_{\theta} s \nabla_{z_k} \theta + \sum_{m=1}^M \nabla_{g_m} s \nabla_{z_k} g_m + \nabla_{z_k} s \quad (\text{A1.2})$$

$$\nabla_{y_m} F_{Y_m|Z} = \nabla_{g_m} r_m \nabla_{y_m} g_m \quad (\text{A1.3})$$

$$\nabla_{z_k} F_{Y_m|Z} = \nabla_{g_m} r_m \nabla_{z_k} g_m + \nabla_{z_k} r_m \quad (\text{A1.4})$$

In addition to the arrays of derivatives defined after Restriction 4, use will be made of the following arrays.

$$g_y \equiv \begin{bmatrix} \nabla_{y_1} g_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \nabla_{y_M} g_M \end{bmatrix} \quad s_g \equiv \begin{bmatrix} \nabla_{g_1} s \\ \vdots \\ \nabla_{g_M} s \end{bmatrix}$$

$$r_g \equiv \begin{bmatrix} \nabla_{g_1} r_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \nabla_{g_M} r_M \end{bmatrix} \quad g_z \equiv \begin{bmatrix} \nabla_{z_1} g_1 & \cdots & \nabla_{z_1} g_M \\ \vdots & \ddots & \vdots \\ \nabla_{z_K} g_1 & \cdots & \nabla_{z_K} g_M \end{bmatrix}$$

Equations (A1.1) - (A1.4) imply the following expressions involving the arrays of derivatives defined above and after Restriction 4.

$$S_y = s_{\theta} \theta_y + g_y s_g \quad (\text{A1.5})$$

$$S_z = s_{\theta} \theta_z + g_z s_g + s_z \quad (\text{A1.6})$$

$$R_y = g_y r_g \quad (\text{A1.7})$$

$$R_z = g_z r_g + r_z \quad (\text{A1.8})$$

Note that  $r_g$  is nonsingular because, by virtue of Restriction 1, each diagonal element of the diagonal matrix  $r_g$  is positive. So equations (A1.7) and (A1.8) imply that

$$g_y = R_y r_g^{-1}$$

$$g_z = (R_z - r_z) r_g^{-1}$$

and therefore, on substituting for  $g_y$  and  $g_z$  in (A1.5) and (A1.6) and rearranging, there is the following.

$$\begin{aligned}\nabla_{\theta}s\theta_y &= S_y - R_y r_g^{-1} s_g \\ \nabla_{\theta}s\theta_z &= S_z - (R_z - r_z) r_g^{-1} s_g + s_z\end{aligned}$$

Rewriting these equations in terms of  $\lambda_y \equiv \nabla_{\theta}s\theta_y$ ,  $\lambda_z \equiv \nabla_{\theta}s\theta_z$ ,  $\gamma \equiv r_g^{-1} s_g$  and  $\delta \equiv -r_z \gamma$  gives

$$\begin{aligned}\lambda_y &= S_y - R_y \gamma \\ \lambda_z &= S_z - R_z \gamma - \delta + s_z\end{aligned}$$

and forming up the arrays  $\Phi$ ,  $\phi$  and  $\psi$  as defined in Theorem 1 using the restrictions of Restriction 4 yields the equation  $\Phi\psi = \phi$  as stated in the Theorem. The rank condition follows directly on noting that  $\psi$  has  $2M + 3K$  elements. The matrix  $\Phi$  has  $M + G + K$  rows which leads directly to the stated order condition.  $\square$

## A2. Amendments when covariates $Z^*$ appear in the structural function

Suppose covariates  $Z^*$  are included in the structural equation for  $W$  of Restriction 2, as in (1). These covariates are required *not* to appear in the index  $\theta$  but they will appear as arguments of the structural functions  $h_m$ ,  $m \in \{1, \dots, M\}$  of Restriction 2. In the assumptions and proof, conditioning on  $Z$  will be, throughout, on  $Z$  and  $Z^*$ . The point  $\bar{z}$  referred to in Restriction 1 will be  $(\bar{z}, \bar{z}^*)$  and the point  $\mathcal{X} \equiv (\bar{w}, \bar{y}, \bar{z})$  referred to in Restriction 3 and in the arrays defined before Restriction 4 will be  $\mathcal{X} \equiv (\bar{w}, \bar{y}, \bar{z}, \bar{z}^*)$ . Variation in  $Z^*$  is not considered and so Restriction 4 and the statement of Theorem 1 are unchanged.

## A3. Proof of Corollary 1

With the restrictions  $s_z = 0$ ,  $r_z = 0$ ,  $\Phi$  and  $\psi$  simplify giving

$$\begin{bmatrix} I_M & 0 & R_y \\ 0 & I_K & R_z \\ A_y & A_z & A_\gamma \end{bmatrix} \begin{bmatrix} \lambda_y \\ \lambda_z \\ \gamma \end{bmatrix} = \begin{bmatrix} S_y \\ S_z \\ a \end{bmatrix}$$

from which the stated rank and order conditions follow directly. Taking this matrix expression apart there is

$$\begin{aligned}\lambda_y &= S_y - R_y \gamma \\ \lambda_z &= S_z - R_z \gamma\end{aligned}$$



and since

$$A_y \lambda_y + A_z \lambda_z + A_\gamma \gamma = a$$

on substituting in this last expression for  $\lambda_y$  and  $\lambda_z$  and rearranging there is the following equation.

$$(A_y R_y + A_z R_z - A_\gamma) \gamma = A_y S_y + A_z S_z - a \quad (\text{A3.1})$$

Define  $X \equiv A_y R_y + A_z R_z - A_\gamma$  and  $x \equiv A_y S_y + A_z S_z - a$ . Then (A3.1) can be written as  $X\gamma = x$ . If the rank condition holds (which requires  $G \geq M$ ) then, for any rank  $M \times G$  matrix  $P$  with rank  $M$ , there is

$$X'P'PX\gamma = X'P'Px$$

and since, when the rank condition holds, by construction,  $X'P'PX$  has rank  $M$ ,

$$\gamma = (X'P'PX)^{-1} X'P'Px$$

which completes the proof of Corollary 1.  $\square$

#### A4. Proof of Corollary 3

Under the conditions stated the equations satisfied by  $\lambda_y$ ,  $\lambda_z$  and  $\gamma$  are as follows.

$$\begin{aligned} \lambda_y &= S_y - R_y \gamma \\ \lambda_z &= S_z - R_z \gamma \\ A_y \lambda_y + A_z \lambda_z &= 0 \end{aligned}$$

In terms of quantile function derivatives these equations are as follows.

$$\begin{aligned} \lambda_y &= -\frac{1}{\nabla_\rho Q_{W|YZ}} H_y - G_\tau^{-1} \gamma \\ \lambda_z &= -\frac{1}{\nabla_\rho Q_{W|YZ}} H_z + G_z G_\tau^{-1} \gamma \\ A_y \lambda_y + A_z \lambda_z &= 0 \end{aligned}$$

Multiplying left and right hand sides of these equations by  $\nabla_\rho Q_{W|YZ}$  (non zero by Restriction 1') and rewriting in terms of the parameters  $\tilde{\lambda}_y$ ,  $\tilde{\lambda}_z$  and  $\tilde{\gamma}$  gives

$$\tilde{\lambda}_y = -H_y - \tilde{\gamma} \quad (\text{A4.1})$$

$$\tilde{\lambda}_z = -H_z + G_z \tilde{\gamma} \quad (\text{A4.2})$$

$$A_y \tilde{\lambda}_y + A_z \tilde{\lambda}_z = 0 \quad (\text{A4.3})$$

and the following matrix equation.

$$\begin{bmatrix} I_M & 0 & I_M \\ 0 & I_K & -G_z \\ A_y & A_z & 0 \end{bmatrix} \begin{bmatrix} \tilde{\lambda}_y \\ \tilde{\lambda}_z \\ \tilde{\gamma} \end{bmatrix} = \begin{bmatrix} -H_y \\ -H_z \\ 0 \end{bmatrix}$$

The rank and order conditions of the Corollary follow directly.

Substituting for  $\tilde{\lambda}_y$  and  $\tilde{\lambda}_z$  in (A4.3) using (A4.1) and (A4.2) and rearranging gives

$$(-A_y + A_z G_z) \tilde{\gamma} = A_y H_y + A_z H_z$$

that is  $\tilde{X} \tilde{\gamma} = \tilde{x}$  using the definitions of  $\tilde{X}$  and  $\tilde{x}$  given in the Corollary. Arguing as in the proof of Corollary 1 gives the rest of the required results.  $\square$