Identification of sensitivity to variation in endogenous variables

Andrew Chesher*

Centre for Microdata Methods and Practice
The Institute for Fiscal Studies and University College London

The A.W.H. Phillips Lecture
2004 Australasian Meeting of the Econometric Society, Melbourne
July 7th 2004

Abstract. This lecture explores conditions under which there is identification of the impact on an outcome of exogenous variation in a variable which is endogenous when data are gathered. The starting point is the Cowles Commission linear simultaneous equations model. The parametric and additive error restrictions of that model are successively relaxed and modifications to covariation, order and rank conditions that maintain identifiability are presented. Eventually a just-identifying, non-falsifiable model permitting nonseparablity of latent variates and devoid of parametric restrictions is obtained. The model requires the endogenous variable to be continuously distributed. It is shown that relaxing this restriction results in loss of point identification but set identification is possible if an additional covariation restriction is introduced. Relaxing other restrictions presents significant challenges.

1. Introduction

Econometric models comprise restrictions on the economic processes, or *structures*, that are admitted as potential generators of data. These restrictions serve a number of purposes.

Some restrictions enable inference to be conducted; examples are restrictions on existence of moments. Some restrictions are imposed to enable modest amounts of data to be used to obtain tolerably accurate estimates; examples are functional form and index restrictions. These restrictions are often testable.

At the core of all econometric models are the restrictions which permit economic interpretation to be given to the results of econometric analysis. These are the restrictions that give a model *identifying power*. It is important to know what these restrictions are because in some cases they are not testable and econometric inference must be predicated on their truth until new types of data can be obtained. With knowledge of a minimal set of identifying conditions one can assess the extent to

*I thank the Leverhulme Trust for its support through funding of the Centre for Microdata Methods and Practice (*http://cemmap.ifs.org.uk*) and of the research programme *Evidence, Inference and Inquiry (http://www.evidencescience.org)*. While studying the subject of identification I have benefitted greatly from discussions with James Heckman, Hidehiko Ichimura, Roger Koenker, Tony Lancaster, Valérie Lechene, Chuck Manski, Lars Nesheim, Whitney Newey, Richard Spady and Elie Tamer. Draft dated July 2nd 2004.

which a model has foundation in economic propositions and the extent to which it relies on restrictions without economic provenance.[1]

This lecture explores the nature of weak identifying restrictions for models which posit a relationship between an outcome and a list of variables amongst which one is endogenous.[2]

This was the problem studied in the 1930's and 1940's by the pioneers of econometrics whose work came to be associated with the *Cowles Commission for Research in Economics*. The Cowles analysis was largely concerned with a linear simultaneous equations model of the relationships amongst economic magnitudes.[3] Koopmans, Rubin and Leipnik (1950) considered a structural equation like this one:

$$Y_1 = \gamma Y_2 + X'\beta + U_1$$

in which Y_2 may co-vary with the unobserved U_1, setting out the *order* and *rank* conditions for the identifiability of the coefficients, γ and β. The order condition requires that the stctural function is insensitive to certain variations in X. The rank condition requires that amongst such variations there are some which do perturb the value of Y_2.

In this linear model the value of the coefficient γ measures the sensitivity of the outcome Y_1 to *ceteris paribus* variation in Y_2. This lecture explores identifying conditions for such measures of sensitivity to *ceteris paribus* variation in an endogenous variable under successively weaker conditions on the form of the structural equation.[4]

First parametric restrictions are removed giving a structural equation of the following form:

$$Y_1 = h(Y_2, X) + U_1$$

and then the "error-additivity" restriction is also removed yielding:

$$Y_1 = h(Y_2, X, U_1).$$

The measure of *ceteris paribus* sensitivity to Y_2 must be adjusted as this is done. I will focus throughout on *finite partial difference* measures, for example in the non-additive error model

$$\Delta = h(y_2^s, x, u_1) - h(y_2^t, x, u_1)$$

where y_2^s and y_2^t are two values of Y_2, and x and u_1 are values of X and U_1 common to both components. In the linear model the finite partial difference measure is:

$$\Delta = \gamma(y_2^s - y_2^t)$$

[1] Rhoerig (1988) makes a persuasive case for studying nonparametric identification conditions.

[2] Much of what follows applies when there is more than one endogenous variable in the structural equation of interest. The single endogenous variable case is treated here in order to simplify the exposition.

[3] There were notable exceptions; see Hurwicz (1950) in which the construction used here was created specifically to allow nonparametric identification to be studied. See also Koopmans and Reiersøl (1950) which studies identification in factor analysis problems and draws on Hurwicz (1950), and Reiersøl (1950) which studies identification in measurement error models.

[4] Similar arguments give conditions under which there is identification of sensitivity to variation in non-endogenous variables when the structural function has endogenous variables amongst its arguments.

and in the nonlinear additive error model it is:

$$\Delta = h(y_2^s, x) - h(y_2^t, x).$$

If Δ is identified then the slope of a chord of the structural function, $\Delta/(y_2^s - y_2^t)$, is identified as long as $y_2^s \neq y_2^t$. In the linear model this slope is the structural coefficient γ.

This focus on partial differences simplifies the argument, and something like it is essential when there is discrete variation. Limiting arguments can be used to develop conditions for identifiability of a partial *derivative* under appropriate restrictions.

One purpose of this exercise is to understand the *structure* of identifying conditions for this class of problems. We will see the rank, order and covariation conditions of the Cowles analysis recurring, but modified in sympathy with the problem addressed.

Another purpose is to develop some understanding of the *limits to identification*. In the non-additive error case I propose a weakly restrictive model which is locally just-identifying and non-falsifiable. Relaxing any of its restrictions results in loss of identifying power.

I will finish by considering the impact of relaxing two of the restrictions of this nonseparable model. The first is a requirement that the endogenous variable be continuously distributed; the second is a requirement that there be no more sources of random variation than observed outcomes.

This lecture is named in honour of Bill Phillips. He understood the crucial role of identification in econometric work and made significant contributions on the topic. In his 1959 *Biometrika* paper "The estimation of parameters of systems of stochastic differential equation systems" he laid the groundwork for the analysis of the aliasing problem, that is the problem of identifying features of continuous time structures using discretely sampled data. In 1968 he published a paper[5], that foreshadowed the Lucas critique, highlighting the problems of identifying the impact of policy instruments on economic outcomes using data obtained while policy makers manipulate instruments attempting to achieve economic targets. It is truly an honour to have this opportunity to address this meeting on the topic of identification in the A.W. Phillips lecture.

2. IDENTIFICATION

A data generating process, or *structure*, for variables $W \equiv \{Y, X\}$ comprises:

1. Functions which transform latent random variables $U \equiv \{U_k\}_{k=1}^K$ to produce unique values of a list of observable variables, $W \equiv \{W_m\}_{m=1}^M$.

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

2. a distribution function, F_U, specifying the probability law for U.

[5] Phillips (1968). See also the discussion in Court (2000).

This is the construction of Hurwicz (1950). Each structure comprises a particular structural function, h and a particular distribution function, F_U, and results in a particular distribution function, F_W, for observable variates W. Data are informative about F_W and we wish to have knowledge of features of h or F_U.

Many distinct structures $\{h, F_U\}$ may generate the *same* F_W. Such structures are termed *observationally equivalent*. Data are not informative about which of any observationally equivalent structures generated the data and it is here that the identification problem arises. Econometric *models* embody restrictions which render certain structures *inadmissible* as generators of data.

Interest typically centres on some feature of a structure, a price elasticity of demand, a measure of the returns to schooling and so forth. If within any set of observationally equivalent structures admitted by a model there is no variation in the value of a structural feature then the model *identifies* the value of the structural feature.[6]

The following proposition is helpful in determining the identifying power of a model.

> *A model identifies the value of a structural feature $\theta(S)$ if there exists a functional $\mathcal{G}$ such that in all structures admitted by the model and for all values, a, of the structural feature:*
>
> $$\theta(S) = a \Rightarrow \mathcal{G}(F_W^S) = a$$
>
> *where F_W^S is the distribution of W implied by the structure S.*[7]

If for some structural feature and model a functional $\mathcal{G}$ with this property can be found then we are assured the model identifies the value of the structural feature and we know how to manipulate F_W to obtain the value of θ in the structure that generated F_W. Further, the analogue principle suggests the estimator: $\hat{\theta} = \mathcal{G}(\hat{F}_W)$.[8]

In what follows the observable variables, W, are partitioned into two sets of variables, Y, which are outcomes generated by the process of interest, and X, which only appear as arguments of structural functions.

3. Linear models

The classical Cowles analysis[9] focussed on *linear* structural equations, for example:

$$Y_1 = \gamma Y_2 + X'\beta + U_1$$

[6] If variation across admissible observationally equivalent structures in the value of a structural feature is limited to a *set* of values then the model *set* identifies the value of the structural feature.

[7] The result is easily shown to be true. Suppose admissible S and S' have $\theta(S) = a$ and $\theta(S') = a'$ and are observationally equivalent. Then, if $\mathcal{G}$ exists, $\mathcal{G}(F_W^S) = a$ and $\mathcal{G}(F_W^{S'}) = a'$ but observational equivalence implies $F_W^S = F_W^{S'}$ and therefore $a = a'$. Thus if a functional with the stated property exists there can be no variation in $\theta(S)$ across observationally equivalent structures. See Chesher (2002a) for a formal statement and proof.

[8] See Manski (1988). Whether or not consistent estimation is feasible will depend on properties of the functional $\mathcal{G}$ and on the properties of $\hat{F}_W$.

[9] See for example Koopmans, Rubin and Leipnik (1950).

and imposed the restriction that $E[Y_2|x]$ exists[10] and the *marginal* covariation condition

$$E[U_1|x] = c_1$$

where c_1 is invariant with respect to x.[11]

Under these conditions there is the following.[12]

$$E[Y_1|x] = \gamma E[Y_2|x] + x'\beta + c_1$$

If there exist values of X: $\{x_s, x_t\}$ such that $x_s'\beta = x_t'\beta$, which is an *order* condition, then:

$$E[Y_1|x_s] - E[Y_1|x_t] = \gamma\left(y_2^s - y_2^t\right)$$

where y_2^s and y_2^t are the following conditional expectations of Y_2.

$$y_2^s \equiv E[Y_2|x_s] \qquad y_2^t \equiv E[Y_2|x_t]$$

If the *rank* condition

$$y_2^s \neq y_2^t$$

is satisfied then γ is identified because

$$\gamma = \frac{E[Y_1|x_s] - E[Y_1|x_t]}{y_2^s - y_2^t}$$

The Wald (1940) estimator is an analogue estimator built on this identifying correspondence.[13]

4. Nonlinear model with an additive latent variable

Now consider models that permit h to be a nonlinear function of Y_2 but which retain the additive error restriction.[14]

$$Y_1 = h(Y_2, X) + U_1 \tag{1}$$

Suppose $E[Y_2|x]$ exists and equals $g(x)$. Then:

$$Y_2 = g(X) + U_2$$

[10] In the Cowles analysis there was a linear reduced form equation $Y_2 = X'\pi + U_2$ and the restriction $E[Y_2|x] = x'\pi$ but this is unnecessary in what follows, the existence of $E[Y_2|x]$ being sufficient.

[11] Here and later the notation $E[Y_2|x]$ signifies $E[Y_2|X = x]$.

[12] The Cowles order and rank conditions ensure that $E[Y_2|x]$ and $x'\beta$ are not linearly dependent and this relationship then motivates the 2SLS estimator.

[13] If there is another pair $\{x_s', x_t'\}$ satisfying the order condition with

$$E[Y_2|x_s] = E[Y_2|x_s'] \qquad E[Y_2|x_t] = E[Y_2|x_t']$$

then γ is overidentified.

[14] Identification for parametrically restricted models of this sort was notably studied by Fisher (1959, 1961, 1966), Wegge (1965), Rothenberg(1971) and Brown (1983). Parametric restrictions on h are not necessary for identification. Rhoerig (1988) and Newey, Powell and Vella (1999) study the nonparametric case and Blundell and Powell (2003) survey results.

where $E[U_2|x] = 0$.

Consider the *iterated* covariation condition

$$E[U_1|U_2 = u_2 \cap X = x] \equiv E[U_1|u_2, x] = c_1(u_2)$$

the force of the restriction being that the conditional expectation exists and does not depend on x. Note that this iterated covariation condition is neither stronger nor weaker than the marginal covariation condition used in the linear model. This sort of iterated covariation condition is helpful in the nonseparable model to be considered shortly.

There is on substituting for Y_2 in (1) and taking expectations conditional on $U_2 = u_2$ and $X = x$:

$$E[Y_1|u_2, x] = h(g(x) + u_2, x) + c_1(u_2).$$

Now impose the restriction that Y_2 is *continuously distributed*. Then, with $y_2 \equiv g(x) + u_2$, conditioning on the event $\{X = x \cap U_2 = u_2\}$ is identical to conditioning on the event $\{X = x \cap Y_2 = y_2\}$ and so there is, conditioning on Y_2 and X:[15]

$$E[Y_1|y_2, x] = h(y_2, x) + c_1(y_2 - g(x)).$$

Now consider identification of the partial difference $h(y_2^s, x) - h(y_2^t, x)$ at two values, y_2^s and y_2^t of Y_2 and some value x of X. For two values of X, $\{x_s, x_t\}$, define the following conditional expectations of Y_2.

$$y_2^s \equiv g(x_s) \qquad y_2^t \equiv g(x_t)$$

There is

$$\begin{aligned} E[Y_1|y_2^s, x_s] &= h(y_2^s, x_s) + c_1(0) \\ E[Y_1|y_2^t, x_t] &= h(y_2^t, x_t) + c_1(0) \end{aligned}$$

and if there is the *order* restriction that for $r \in \{s, t\}$

$$h(y_2^r, x_s) = h(y_2^r, x_t) \tag{2}$$

then for $x \in \{x_s, x_t\}$

$$E[Y_1|y_2^s, x_s] - E[Y_1|y_2^t, x_t] = h(y_2^s, x) - h(y_2^t, x).$$

If the "rank" condition $y_2^s \neq y_2^t$ holds then the model identifies the sensitivity of h to variation in Y_2, and in particular the slope of a chord of the structural function:

$$\frac{h(y_2^s, x) - h(y_2^t, x)}{y_2^s - y_2^t}.$$

[15] If there are restrictions which ensure that $h(y_2, x)$ and $c_1(y_2 - g(x))$ are not functionally dependent this relationship motivates analogue estimators obtained by estimating the regression of Y_1 on Y_2, X and $Y_2 - E[Y_2|x]$, with $Y_2 - E[Y_2|x]$ replaced by the residual from an estimate of the regression of Y_2 on X. In the linear model this delivers the 2SLS estimator.

If the rank condition does not hold then the identified partial difference is trivially zero.

The requirement that $E[Y_2|x]$ exists is inessential. Write

$$Y_2 = g(X, U_2)$$

where U_2 is normalised uniformly distributed on $(0, 1)$ independent of X and g is an increasing function of U_2. Let $Q_{A|B}(p|b)$ denote the p-quantile of random variable A given variables $B = b$. By definition:

$$g(X, U_2) \equiv Q_{Y_2|X}(U_2|X)$$

that is $g(X, U_2)$ is the U_2-quantile function of Y_2 given X.[16] The condition

$$E[U_1|u_2, x] = c_1(u_2)$$

is maintained but note that the interpretation of u_2 has altered.

As long as Y_2 is continuously distributed (that is g is *strictly* increasing in U_2) there is

$$E[Y_1|y_2, x] = h(y_2, x) + c_1(u_2) \tag{3}$$

where u_2 satisfies

$$y_2 = g(x, u_2)$$

from which it follows that

$$u_2 = F_{Y_2|X}(y_2|x).$$

Substituting for u_2 in (3) there is:[17]

$$E[Y_1|y_2, x] = h(y_2, x) + c_1(F_{Y_2|X}(y_2|x)).$$

Turning to identification of partial differences, if $\{x_s, x_t\}$ satisfy the order condition (2) and if $\{y_2^s, y_2^t\}$ satisfy

$$F_{Y_2|X}(y_2^s|x_s) = F_{Y_2|X}(y_2^t|x_t) = u_2 \tag{4}$$

there is, as before, for $x \in \{x_s, x_t\}$:

$$E[Y_1|y_2^s, x_s] - E[Y_1|y_2^t, x_t] = h(y_2^s, x) - h(y_2^t, x)$$

and so identification of the "pure effect" of Y_2 on the structural function h as long as the rank condition $y_2^s \neq y_2^t$ is satisfied.

In preparation for the nonseparable error case to be considered next, note that (4) implies that y_2^s and y_2^t are both u_2-quantiles of Y_2 given X.

[16] The probabilty of $Y_2 \leq Q_{Y_2|X}(u_2|x)$ given $X = x$ is u_2 which is equal to the probability of $U_2 \leq u_2$ given that U_2 is uniformly distributed on $(0, 1)$.

[17] If there are restrictions which ensure that $h(y_2, x)$ and $c_1(F_{Y_2|X}(y_2|x))$ are not functionally dependent this relationship motivates analogue estimators obtained by estimating the nonparametric additive regression of Y_1 on Y_2, X and $F_{Y_2|X}(Y_2|X)$ with the conditional distribution function $F_{Y_2|X}$ replaced by an estimator.

5. NONLINEAR MODEL WITH A NON-ADDITIVE LATENT VARIABLE

The conditions of the model used in the Cowles analysis have been substantially relaxed; all parametric restrictions have been removed. Now the "additive latent variable" restriction is relaxed and the structural equation is allowed to take the following form.[18]

$$Y_1 = h(Y_2, X, U_1)$$

This allows the possibility of rich effects for Y_2 on h. The sensitivity of h to variation in Y_2 can now be *stochastic* since it can depend on U_1. In a policy context, interventions that "exogenously" change Y_2 are now permitted to have effects which depend on the value of U_1 - the effects of interventions can have non-degenerate probability distributions.

Now that U_1 is embedded in the structural function *expectation* based covariation conditions have no force without substantial restrictions on the structural function h. But if the model requires that h varies monotonically with U_1, a restriction of course satisfied in an additive latent variate model, then *quantile* based restrictions have identifying power.

The discussion proceeds in terms of *median* restrictions, but restrictions on other quantiles can be used. Let $M_A[b_1, \ldots, b_K]$ denote the conditional median of random variable A given $B_1 = b_1, \ldots, B_K = b_K$. As before there is the auxiliary equation

$$Y_2 = g(X, U_2)$$

with U_2 normalised uniformly distributed on $(0, 1)$ independent of X and g normalised increasing in U_2.

Restrict h to be monotonically varying with U_1, and normalise it to be increasing. The equivariance of quantiles under monotone transformation[19] implies

$$M_{Y_1}[u_2, x] = h(g(x, u_2), x, M_{U_1}[u_2, x])$$

in which both medians are conditioned on $U_2 = u_2$ and $X = x$. Add to the model the following *iterated quantile* covariation condition.

$$M_{U_1}[u_2, x] = c_1(u_2)$$

where c_1 does not depend on x.

Impose the restriction that Y_2 is continuously distributed which implies that g is *strictly* increasing in U_2. Then, with $y_2 \equiv g(x, u_2)$, since conditioning on the event $\{X = x \cap U_2 = u_2\}$ is identical to conditioning on the event $\{X = x \cap Y_2 = y_2\}$, there is the conditional median of Y_1 given Y_2 and X:

$$M_{Y_1}[y_2, x] = h(y_2, x, c_1(u_2))$$

[18] Roehrig (1988), Matzkin (2003) and Imbens and Newey (2003) consider identification and estimation of structural features in this nonseparable case under the restriction that U_1 and X are statistically independent. In Matzkin (2003) the endogenous variable Y_2 is absent. Rhoerig (1988) unlike Imbens and Newey (2003) restricts attention to smooth structural functions.

[19] The equivariance property is that, if f is an increasing function of a random variable A then the p-quantiles of $f(A)$ and A satisfy: $Q_{f(A)}(p) = f(Q_A(p))$.

where the normalisation of U_2 ensures that $u_2 \equiv F_{Y_2|X}(y_2|x)$, conversely that y_2 is the u_2-quantile of Y_2 given $X = x$. In summary

$$M_{Y_1}[y_2, x] = h(y_2, x, c_1(F_{Y_2|X}(y_2|x))).$$

Consider a value $u_2 \in (0, 1)$ and two values of X, $\{x_s, x_t\}$. Define y_2^s and y_2^t, the u_2-quantiles of Y_2 given X equal to respectively x_s and x_t,

$$y_2^s \equiv g(x_s, u_2) \qquad y_2^t \equiv g(x_t, u_2)$$

equivalently:

$$F_{Y_2|X}(y_2^s|x_s) = F_{Y_2|X}(y_2^t|x_t) = u_2.$$

Under the "order" restriction that, for $r \in \{s, t\}$

$$h(y_2^r, x_s, c_1(u_2)) = h(y_2^r, x_t, c_1(u_2))$$

there is for $x \in \{x_s, x_t\}$

$$M_{Y_1}[y_2^s, x_s] - M_{Y_1}[y_2^t, x_t] = h(y_2^s, x, c_1(u_2)) - h(y_2^t, x, c_1(u_2))$$

which identifies the "pure" effect of Y_2 on h as long as the rank condition $y_2^s \neq y_2^t$ holds.

If this rank condition does hold then

$$\frac{M_{Y_1}[y_2^s, x_s] - M_{Y_1}[y_2^t, x_t]}{y_2^s - y_2^t}$$

identifies the slope of a chord of the structural function. This suggests a quantile based version of the Wald (1940) estimator. Quantiles of Y_1 other than the median can be used giving identification at other values of U_1.[20]

The identifying conditions for sensitivity of a structural function of the Cowles analysis have now been substantially reduced. Parametric restrictions have been relaxed and the restriction of error-additivity has been removed.

The core conditions used in the Cowles analysis remain but with some modification. Conditions on marginal distributions of latent variates have been replaced by conditions on iterated conditional distributions. Conditions like the classical order and rank conditions remain but these, like the covariation conditions, need only hold local to the particular structural partial difference and values of X of interest.

In the non-additive latent variate model the weak *local* conditions which identify a particular partial difference of a structural function are just-identifying and non-falsifiable. Relaxation of any of the conditions results in a model without point identifying power.

To conclude I consider relaxing two of the identifying conditions: the restriction that Y_2 be continuously distributed and the restriction that there be no more sources of random variation than outcomes. I consider additional conditions which can recover identification in these cases. These restrictions are of interest because they are not met in many of the models employed in modern microeconometric practice.

[20] See Chesher (2002a, 2003a) for further results on identification and Ma and Koenker (2004) and Lee (2004) for related results on estimation.

6. THE CONTINUOUS Y_2 RESTRICTION

6.1. Non additive latent variable models. Consider the non-additive latent variate model

$$Y_1 = h(Y_2, X, U_1)$$

and the iterated conditional median restriction used before.

$$M_{U_1}[u_2, x] = c_1(u_2)$$

In the continuous Y_2 case we are able to to use this covariation condition by exploiting the fact that conditioning on $X = x$ and $U_2 = u_2$ is identical to conditioning on $X = x$ and $Y_2 = y_2 = g(x, u_2)$. Keeping Y_2 equal to $y_2 \equiv Q_{Y_2|X}(u_2|x)$ as x varies holds U_2 constant at the value u_2 and focussing on conditional medians of Y_1 given $Y_2 = y_2$ and $X = x$ ensures that U_1 is held fixed at $M_{U_1}[u_2]$. Thus variation in x results in a change in Y_2 but not in U_1 and, with an order condition ensuring h is insensitive to the variation in x, point identification of a partial difference is achieved.

When Y_2 is discrete this approach is not available; each value of Y_2 is associated with a sub-interval of $(0, 1)$, an interval whose length is equal to the conditional probability of that value occurring given $X = x$.[21] The result is loss of point identification of the *value* of the structural function at a particular value of U_1 and without that there is no possibility of point identification of a *partial difference* of the structural function in each term of which U_1 is equal to the same value.

The fact that U_2 is restricted to an interval determined by the value of Y_2 and X suggests the possibility of set identification. The argument is sketched now.[22] To keep the notation as uncluttered as possible the argument proceeds for the case in which the latent variates U_1 and U_2 are distributed independently of X. Further, X is excluded from the structural function. That ensures the order condition is satisfied but is, like full independence, a stronger restriction than is required.

With these additional restrictions there is now the structural function:

$$Y_1 = h(Y_2, U_1)$$

and the auxiliary equation:

$$Y_2 = g(X, U_2)$$

with g normalised increasing in U_2 which is normalised uniformly distributed on $(0, 1)$. The independence of (U_1, U_2) and X, now imposed, ensures that the conditional median condition

$$M_{U_1}[u_2, x] = c_1(u_2)$$

is satisfied. To further simplify notation it is assumed that the support of Y_2 is denumerable and h and g are normalised so that this is a subset of the positive integers.

The model is now strengthened by requiring that the conditional quantiles of U_1 given U_2 are monotonic *non-decreasing* functions of the value of U_2.[23]

[21] Recall U_2 is normalised uniformly distributed on $(0, 1)$.

[22] A full analysis in a less restrictive model than that considered here is given in Chesher (2003b).

[23] If the quantile functions of U_1 given U_2 are non-*increasing* in U_2 then the endpoints of the intervals which follow are reversed.

I now consider identification of the partial difference

$$h(m, M_{U_1}[0.5]) - h(n, M_{U_1}[0.5])$$

which measures the effect of changing Y_2 from the value m to the value n when U_1 is held constant at $M_{U_1}[0.5]$ that is at its conditional median given U_2 is equal to its median, that is 0.5.

Let x be such that the value m is the conditional median of Y_2 given $X = x$. When $Y_2 = m$ and $X = x$,

$$p_{m-1}(x) < U_2 \leq p_m(x) \tag{5}$$

where

$$p_{m-1}(x) \equiv P[Y_2 \leq m-1|x] \qquad p_m(x) \equiv P[Y_2 \leq m|x].$$

Since m is a conditional median,

$$p_{m-1}(x) < 0.5 \leq p_m(x). \tag{6}$$

.

The following result is derived in Chesher (2003b): when the median of U_1 given U_2 is constant over the interval (6) and equal to $M_{U_1}[p_{m-1}(x)]$ then:

$$M_{Y_1}[m, x] = h(m, M_{U_1}[p_{m-1}(x)])$$

and when the median of U_1 given U_2 is constant over the interval (6) and equal to $M_{U_1}[p_m(x)]$ then:

$$M_{Y_1}[m, x] = h(m, M_{U_1}[p_m(x)]).$$

These are two polar cases given that the quantiles of U_1 given U_2 are non-decreasing functions of the value of U_2 and it follows that whatever the form of this non-decreasing dependence:[24]

$$h(m, M_{U_1}[p_{m-1}(x)]) \leq M_{Y_1}[m, x] \leq h(m, M_{U_1}[p_m(x)]). \tag{7}$$

We seek an interval containing $h(m, M_{U_1}[0.5])$ about which data are informative. Consider two values of X, x_m and x_{m-1}, such that the following *rank* condition holds.

$$p_m(x_m) = 0.5 = p_{m-1}(x_{m-1}) \tag{8}$$

The inequality (7) evaluated at each of these values of X in turn gives the following inequalities.

$$h(m, M_{U_1}[p_{m-1}(x_m)]) \leq M_{Y_1}[m, x_m] \leq h(m, M_{U_1}[p_m(x_m)])$$

$$h(m, M_{U_1}[p_{m-1}(x_{m-1})]) \leq M_{Y_1}[m, x_{m-1}] \leq h(m, M_{U_1}[p_m(x_{m-1})])$$

[24] These inequalities are reversed if the quantiles of U_1 given U_2 are non-increasing functions of the value of U_2.

The rank condition (8) ensures that the terms on the right hand side of the first of these inequalities and on the left hand side of the second are equal to each other and to $h(m, M_{U_1}[0.5])$ which is the structural feature of interest. It follows that

$$M_{Y_1}[m, x_m] \leq h(m, M_{U_1}[0.5]) \leq M_{Y_1}[m, x_{m-1}] \tag{9}$$

and so there is set identification of the value of $h(m, M_{U_1}[0.5])$.[25]

Finally consider identification of the effect on h of moving Y_2 from the value m to the value n with U_1 held fixed at $M_{U_1}[0.5]$ Suppose there exist $\{x_{n-1}, x_n\}$ such that

$$p_n(x_n) = 0.5 = p_{n-1}(x_{n-1}) \tag{10}$$

then arguing as above, there is the following interval identifying correspondence.

$$M_{Y_1}[n, x_n] \leq h(n, M_{U_1}[0.5]) \leq M_{Y_1}[n, x_{n-1}] \tag{11}$$

With the values of the structural function at $Y_2 \in \{m, n\}$ and $U_1 = M_{U_1}[0.5]$ interval identified there is on combining (9) and (11) the following interval identifying correspondence for the partial difference of interest.

$$\begin{gathered} M_{Y_1}(m, x_m) - M_{Y_1}(n, x_{n-1}) \\ \leq h(m, M_{U_1}[0.5]) - h(n, M_{U_1}[0.5]) \leq \\ M_{Y_1}(m, x_{m-1}) - M_{Y_1}(n, x_n) \end{gathered}$$

The following points are discussed in greater detail in Chesher (2003b).

1. The argument above was conducted for the case in which identification of a partial difference is required with U_1 equal to $M_{U_1}[0.5]$, that is its iterated conditional median. A similar argument applies if identification is required at the conditional τ_1-quantile of U_1 given U_2 is equal to its τ_2-quantile. The values of X are chosen so that m and n are both the required τ_2-quantiles of Y_2 given X and the appropriate conditional τ_1-quantile of Y_1 given Y_2 and X is employed.

2. The argument was conducted assuming that the quantile functions of U_1 given U_2 are non-decreasing functions of the value of U_2. If the quantile functions of U_1 given U_2 are non-increasing functions of the value of U_2 then the set identifying inequality is reversed. Data are informative about the direction of this dependence so the model only requires a (weak) monotonicity restriction on the dependence of the quantiles of U_1 on the value of U_2.

3. It may not be possible to satisfy the equalities in the rank conditions (8) and (10) when the support of X is sparse. The same set identifying inequality holds if they are replaced by

$$p_r(x_r) \leq 0.5 \leq p_{r-1}(x_{r-1})$$

for $r \in \{m, n\}$.

[25] These inequalities are also reversed if the quantiles of U_1 given U_2 are non-increasing functions of the value of U_2.

6.2. Additive latent variable models. In the non-additive latent variable model relaxing the restriction that Y_2 is continuously distributed results in loss of point identification. As shown in the previous Section, set identification can be achieved if the dependence of U_1 on U_2 is restricted.

If the restriction that the latent variable in the structural function is *additive* is imposed then, as long as a *marginal* covariation condition is maintained, point identification is feasible. This is because with discrete Y_2 and an additive error the structural function can be written as a linear function of endogenous variables and the latent variate.

To see this restrict the support of discrete Y_2 to be denumerable and normalise so that $Y_2 \in \{1, \ldots, M\}$. Then the structural equation of interest

$$Y_1 = h(Y_2, X) + U_1$$

can be rewritten as follows.

$$Y_1 = \sum_{m=1}^{M} \gamma_m(X) \times \mathbf{1}[Y_2 = m] + U_1$$

$$\gamma_m(X) \equiv h(m, X)$$

$$\mathbf{1}[Y_2 = m] = \begin{cases} 1 & , \quad Y_2 = m \\ 0 & , \quad Y_2 \neq m \end{cases}$$

This is a linear model for M endogenous binary indicators with an additive error. A marginal covariation condition, $E[U_1|x] = c_1$, and suitable order and rank conditions lead to a model which arguing along the lines of the Cowles analysis has identifying power.[26] And so we have come full circle for it was with that analysis that we started.

7. Concluding remarks

I have shown how successively weaker models can be shown to identify the sensitivity of a structural function to variation in an endogenous variable. We started with the linear model studied by the founders of econometrics. We arrived at a just-identifying model involving no parametric restrictions allowing error nonseparability and so affording the possibility of estimation of the distributional impact of policy interventions. People who use more restrictive models for problems in the class addressed here can be assured that fundamental identifying power resides in a subset of the restrictions of their model.

However, the restrictions are not trivial. Once the endogenous variable is discrete the model I have proposed no longer secures *point* identification but with a minor strengthening of conditions *set* identification can be achieved. Point identification seems to require very strong restrictions when there is a discrete endogenous variable unless error-additivity is imposed.

The weakly restricting nonseparable model loses identifying power when other conditions are relaxed. Of considerable interest from a microeconometric point of

[26] Das (2003) and Florens and Malavolti (2003) give results concerning estimation and inference for this problem.

view is the restriction limiting the number of stochastic unobservables to be no greater than the number of observable outcomes.[27]

Many microeconometric models employed in practice admit structures with more unobservables than outcomes. Leading examples are models permitting measurement error in explanatory variables, panel data models with individual effects, treatment effects models, and duration and other models permitting across individual heterogeneity in preferences such as are common in modern microeconometric practice.

Without very strong restrictions it seems impossible to identify the value of a structural function at a particular value of its observable *and* latent arguments when there are more latent variates than unobserved outcomes, and without that there can be no possibility of identification of a partial difference of a structural function with respect to observable variables.

One way to proceed is to give up the attempt - to focus instead on some average sensitivity with averaging taking place across the unobservables. This typically requires the unpalatable restriction that latent variables and instruments be fully statistically independent.[28]

An alternative way to proceed is to impose restrictions which require the structural feature of interest to be *invariant* with respect to latent variates. An index restriction can achieve this purpose. Consider a structural function of the following form

$$Y_1 = h(\lambda(Y_2, X), U_1, U_2, \ldots, U_M)$$

which depends on an index $\lambda(Y_2, X)$ which is free of latent variates but with Y_2 endogenous, that is jointly dependent with the latent U's. Similar conditions to those just considered lead to a model that identifies the relative sensitivity of the index λ to variation in the endogenous Y_2 and elements of X.[29]

Surveying applied microeconometric work one has the impression that researchers are often not content to estimate measures of index relative sensitivity or average structural functions. As a result there is frequent use of models whose identifying power resides in very strong restrictions.

If very strong identification conditions are required then one must have concern for the robustness of the econometric inference which flows from these models. Two researchers addressing the same data can employ different sets of strong identifying restrictions and produce different answers to what is essentially the same question. The difficulty is that strong conditions often have dubious economic provenance and then obtaining agreement on what identifying restrictions can be maintained may be difficult and progress of economic knowledge may then be hampered.[30]

How to weaken identification conditions in the models employed in modern microeconometric practice is a challenging open research question.

---

[27] As noted by Hurwicz (1950) errors must be non-additive in this case. Otherewise they would coalesce.

[28] See Imbens and Newey (2003) for results on nonparametric identification and estimation of average structural functions. The treatment effects literature contains examples of models and "average" structural features in which identifing power resides in weaker mean independence restrictions.

[29] See Chesher (2002b) for an analysis in the context of heterogeneous duration models.

[30] Manski (2003) uses this argument to motivate the analysis of partial (set) identification which rests on core identifying restrictions on which many might agree.

References

Blundell, Richard and James L Powell (2003): "Endogeneity in Nonparametric and Semiparametric Regression Models," in *Advances in Economics and Econometrics: Theory and Applications, 8th World Congress, Volume 2*, edited by: Mathias Dewatripont, Lars Peter Hansen and Stephen J. Turnovsky. Cambridge: Cambridge University Press.

Brown, Bryan W. (1983): "The Identification Problem in Systems Nonlinear in the Variables," *Econometrica*, 51,175-196.

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.

Chesher, Andrew D. (2003a): "Identification in Nonseparable Models," *Econometrica*, 71, 1404 - 1441.

Chesher, Andrew D. (2003b): "Nonparametric Identification under Discrete Variation," Centre for Microdata Methods and Practice Working Paper 19/03.

Court, Robin (2000): "The Lucas Critique: Did Phillips Make a Comparable Contribution?" Chapter 49 of Leeson (2000).

Das, Mitali (2004): "Instrumental Variables Estimators of Nonparametric Models with Discrete Endogenous Regressors," forthcoming, *Journal of Econometrics*.

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.

Florens, Jean-Pierre and Laetitia Malavolti (2003): "Instrumental Regression with Discrete Variables," presented at the 2003 European Meeting of the Econometric Society, Stockholm.

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

Imbens, Guido W., and Whitney K. Newey (2003): "Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity," Manuscript, paper presented at the 14th EC^2 Meeting, cemmap, London, December 12th - 13th 2003.

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.

Koopmans, Tjalling C., and Olav Reiersøl (1950): "The Identification of Structural Characteristics,"*Annals of Mathematical Statistics*, 21, 165-181.

Lee, Sokbae (2004): "Endogeneity in Quantile Regression Models: a Control Function Approach," Centre for Microdata Methods and Practice Working Paper 08/04.

Leeson, Robert ed., (2000): *A.W.H. Phillips: Collected Works in Contemporary Perspective*, Cambridge University Press, Cambridge.

Ma, Lingjae and Roger Koenker (2003): "Quantile Regression Methods for Recursive Structural Equation Models," Centre for Microdata Methods and Practice Working Paper 01/04.

Manski, Charles F. (2003): *Partial Identification of Probability Models*, New York: Springer-Verlag.

Manski, Charles F. (1988): *Analog Estimation Methods in Econometrics*, New York: Chapman and Hall.

Matzkin, Rosa L. (2003): "Nonparametric Estimation of Nonadditive Random Functions," *Econometrica*, 71, 1339 - 1376.

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

Phillips, A.W.H., "Bill", (1968): "Models for the Control of Economic Fluctuations," in *Mathematical Model Building in Economics and Industry*, edited by M.G. Kendall, London: Charles Griffin,. Reprinted as Chapter 50 of Leeson (2000).

Phillips, A.W.H., "Bill", (1959): "The Estimation of Parameters in Systems of Stochastic Differential Equations," *Biometrika*, 46, 67-76. Reprinted as Chapter 42 of Leeson (2000).

Reiersøl, Olav (1950): "Identifiability of a Linear Relation Between Variables which are Subject to Error," *Econometrica*, 18, 375-389.

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.

Wald, Abraham (1940): "The Fitting of Straight Lines if Both Variables are Subject to Error," *The Annals of Mathematical Statistics*, 11, 284-300.

Wegge, Leon L. (1965): "Identifiability Criteria for a System of Equations as a Whole," The Australian Journal of Statistics, 7, 67-77.