



# IDENTIFICATION IN ADDITIVE ERROR MODELS WITH DISCRETE ENDOGENOUS VARIABLES

---

*Andrew Chesher*

THE INSTITUTE FOR FISCAL STUDIES  
DEPARTMENT OF ECONOMICS, UCL  
**cemmap** working paper CWP11/04

# Identification in additive error models with discrete endogenous variables

ANDREW CHESHER

CENTRE FOR MICRODATA METHODS AND PRACTICE, IFS AND UCL

September 5th 2004

ABSTRACT. In additive error models with a discrete endogenous variable identification cannot be achieved under a marginal covariation condition when the support of instruments is sparse relative to the support of the endogenous variable. An iterated covariation condition with a weak monotonicity restriction is shown to have set identifying power.

## 1. INTRODUCTION

In an additive error model:  $Y_1 = h(Y_2) + \varepsilon_1$  with a discrete endogenous variable  $Y_2$ , nonparametric identification of  $h(\cdot)$  requires a vector of instruments  $X$  to have at least as many points of support as the endogenous variable under the conventional *marginal* covariation condition:  $E[\varepsilon_1|X = x] = c$ , a constant invariant with respect to  $x$ .

Das (2004) and Florens and Malavolti (2003) study identification and estimation of  $h(\cdot)$  under this condition. In practice this support condition may not be satisfied and there are additional difficulties when the number of points of support of  $Y_2$  is unbounded.

This note explores the utility for identification of values yielded by  $h(\cdot)$  of an *iterated* covariation condition,  $E[\varepsilon_1|\varepsilon_2 = e_2 \cap X = x] = c(e_2)$  where  $c(\cdot)$  is a function which does not depend on  $x$  and  $\varepsilon_2$  is a random variable which generates stochastic variation in  $Y_2$  given  $X$ .

It is shown that, when the cardinality of the support of  $X$  is at least equal to that of  $Y_2$  the iterated condition has no additional power for identification of  $h(\cdot)$  relative to the marginal covariation condition. However when the support of  $X$  is deficient, if  $c(\cdot)$  is restricted to be monotonic, then the iterated covariation condition can yield closed set identification of  $h(\cdot)$  at particular points of support of  $Y_2$  and, if instruments are not weak, of differences of  $h(\cdot)$  across points of support.

## 2. DEFINITIONS

A discrete outcome  $Y_2$  has  $M$  known points of support  $y_2^1 < \dots < y_2^M$  independent of a list of covariates  $X$ . For  $m \in \{1, \dots, M\}$  define  $p^m(x) \equiv P[Y_2 \leq y_2^m | X = x]$ , the associated conditional quantile function:

$$Q_{Y_2|X}(p|x) = y_2^m \text{ where } m = \inf_{n \in \{1, \dots, M\}} \{n : p^n(x) \geq p\}$$

and the probability masses on the points of support of  $Y_2$ :  $p_m(x) \equiv p^m(x) - p^{m-1}(x) = P[Y_2 = y_2^m | X = x]$  each restricted to lie in  $(0, 1)$  for all  $x$ .

A continuous outcome  $Y_1$  and discrete outcome  $Y_2$  are determined by the structural equations

$$Y_1 = h_1(Y_2) + \varepsilon_1 \tag{1}$$

$$Y_2 = g(X, \varepsilon_2) \tag{2}$$

where  $\varepsilon_1$  and  $\varepsilon_2$  are continuously distributed latent variates.<sup>1</sup> The variate  $\varepsilon_2$  is normalised  $Unif(0, 1)$  and  $g(x, e_2) \equiv Q_{Y_2|X}(e_2|x)$  is the conditional  $e_2$ -quantile function of  $Y_2$  given  $X$ . Under this normalisation  $Y_2$  has the conditional distribution function  $p^m(x)$  given  $X = x$ .

### 3. MARGINAL COVARIATION RESTRICTIONS

Consider the following marginal covariation restriction as employed by Das (2004) and Florens and Malavolti (2003).

**MC:** For all  $x$ ,  $E[\varepsilon_1|X = x] = c$  where  $c$  is invariant with respect to  $x$ .

Normalise  $c = 0$ . Define  $\mu(x) \equiv E[Y_1|X = x]$  and, for  $m \in \{1, \dots, M\}$ ,  $h_m \equiv h(y_2^m)$ , and  $D_m \equiv 1[Y_2 = y_2^m]$ . Since  $Y_1 = \sum_{m=1}^M D_m h(y_2^m) + \varepsilon_1$  there is:  $\mu(x) = \sum_{m=1}^M p_m(x) h_m$ . Consider  $N$  values of  $X$ :  $\{x_n\}_{n=1}^N$ . With the definitions below:  $\mu_N = P_N \theta$ .

$$\mu_N \equiv \begin{bmatrix} \mu(x_1) \\ \vdots \\ \mu(x_N) \end{bmatrix} \quad \theta \equiv \begin{bmatrix} h_1 \\ \vdots \\ h_M \end{bmatrix} \quad P_N \equiv \begin{bmatrix} p_1(x_1) & \cdots & p_M(x_1) \\ \vdots & \ddots & \vdots \\ p_1(x_N) & \cdots & p_M(x_N) \end{bmatrix}$$

The value of  $\theta$  is identified without further restriction if and only if  $\text{rank}(P_N) = M$  for which a *necessary* condition is  $N^* \geq M$  where  $N^*$  is the number of distinct values in  $\{x_n\}_{n=1}^N$ . When the rank condition is satisfied there is for any  $N \times N$  full rank matrix  $W$ ,

$$\theta = (P_N' W^{-1} P_N)^{-1} P_N' W^{-1} \mu_N$$

which suggests analog GLS-type estimators with the elements of  $P_N$  replaced by estimated conditional probabilities and the elements of  $\mu_N$  replaced by values of  $Y_1$ . If  $M$  is not finite there are clearly difficulties in implementing such procedures and even when  $M$  is finite, if it is large then there may be extreme collinearity amongst the columns of  $P_N$ .

If the cardinality of the support of  $X$  is less than that of the support of the discrete endogenous variable, then the value of  $\theta$  is not identified. Further, since  $p_m(x) \in (0, 1)$  for all  $x$ , no element of  $\theta$  is identified unless additional restrictions are introduced. The elements of  $\theta$  are restricted to lie on  $N^*$  hyperplanes in  $\mathbb{R}^M$  so there is set identification of the value of  $\theta$ , but the set in which the value falls is not closed.<sup>2</sup>

<sup>1</sup>There are no covariates appearing in the structural function  $h$  in equation (1). Covariates other than  $X$  could appear, in which case, in this analysis of nonparametric identification, the arguments that follow will apply at each value of those covariates at which identification of a value of  $h$  is sought. The covariates  $X$  appearing in equation (2) could also be arguments of the structural function  $h$  but its value must then be insensitive to the identifying variations in  $X$  considered below.

<sup>2</sup>If there are the restrictions  $R\theta = r$  where  $R$  is a known  $J \times M$  matrix and  $r$  is a known  $J$  element vector then  $\theta$  is identified if and only if  $\text{rank} \left( \begin{bmatrix} R' & P_N' \end{bmatrix} \right) = M$  for which a necessary condition is  $J + N^* \geq M$ . One might entertain the restriction that  $h$  is a  $K$ th degree *polynomial* function of  $Y_2$ . Then, because suitably defined  $(K + 1)$ th differences of the elements of  $\theta$  must be zero, there are  $(M - 1 - K)$  linear restrictions on  $\theta$ . It follows that, if the covariate vector  $X$  and the discrete endogenous variable have respectively  $N^*$  and  $M$  points of support with  $N^* < M$ ,  $\theta$  may be identified if  $h$  is restricted to lie on a polynomial of degree  $K$  where  $K < N^*$ . If  $K \geq N^*$  then the restrictions may be overidentifying.

## 4. ITERATED COVARIATION CONDITION

Now consider the following *iterated* covariation restriction.

**IC:** For all  $x$ ,  $E[\varepsilon_1|\varepsilon_2 = e_2 \cap X = x] = c(e_2)$ .  $E[\varepsilon_1|X = x]$  is finite.

Since  $\varepsilon_2$  is independent of  $X$  this *implies* the marginal covariation condition **MC** with  $c \equiv \int_0^1 c(e_2)de_2$ . However the iterated condition **IC** is *not* implied by **MC**. For example, for any function  $a(x)$ , if  $E[\varepsilon_1|\varepsilon_2 = e_2 \cap X = x] = a(x) + 2(c - a(x))e_2$ , which violates **IC**, then  $E[\varepsilon_1|X = x] = c$  and **MC** is satisfied. So **IC** is *more restrictive* than **MC**. There is thus the possibility that it has greater identifying power, an issue explored in this Section.

When the cardinality of the support of  $X$  is at least equal to the cardinality of the support of the discrete endogenous variable **IC** provides no additional identifying information for the structural function  $h$ . This is shown in Section 4.1.

Otherwise **IC** can yield *closed set* identification of values of the structural function if an additional restriction is imposed. This is shown in Section 4.2.

**4.1. Rich support.** For  $z \in [0, 1]$  define  $C(z) = \int_0^z c(w)dw$ . Clearly  $C(0) = 0$ . Normalise  $E[\varepsilon_1|X = x] = C(1) = 0$ .

The conditional expectation of  $Y_1$  given  $Y_2 = y_2^m$  and  $X = x$  is

$$E(Y_1|Y_2 = y_2^m \cap X = x) = h(y_2^m) + E(\varepsilon_1|\varepsilon_2 \in (p^{m-1}(x), p^m(x)]) \quad (3)$$

and since

$$E(\varepsilon_1|\varepsilon_2 \in (p^{m-1}(x), p^m(x))) = \frac{1}{p_m(x)} \int_{p^{m-1}(x)}^{p^m(x)} c(e_2)de_2 \quad (4)$$

there is:

$$E(Y_1|Y_2 = y_2^m \cap X = x) = h(y_2^m) + \frac{1}{p_m(x)} \{C(p^m(x)) - C(p^{m-1}(x))\}. \quad (5)$$

Suppose  $M$  is finite and for  $m \in \{1, \dots, M\}$  and  $N$  values of  $X$ ,  $\{x_n\}_{n=1}^N$  define the following terms.

$$\begin{aligned} \mu_{mn} &\equiv E(Y_1|Y_2 = y_2^m \cap X = x_n) \\ p_{mn} &\equiv p_m(x_n) \equiv P[Y_2 = y_2^m|X = x_n] \\ C_n^m &\equiv C(p^m(x_n)) \end{aligned}$$

There is, for  $m \in \{1, \dots, M\}$  and  $n \in \{1, \dots, N\}$

$$\mu_{mn} = h_m + p_{mn}^{-1} (C_n^m - C_n^{m-1}) \quad (6)$$

with, for all  $n$ ,  $C_n^0 = 0$  by definition, and  $C_n^M = 1$  by normalisation.<sup>3</sup>

Data are informative about the  $2 \times M \times N$  values of the terms  $\mu_{mn}$  and  $p_{mn}$ . The  $M \times N$  equations (6) involve  $M$  values of the terms  $h_m$  and  $(M + 1) \times N$  values of the terms  $C_n^m$  of which  $2 \times N$  are determined by definition or normalisation. Therefore a necessary condition for the identification of the  $M + (M - 1) \times N$  undetermined values of the terms  $h_m$  and  $C_n^m$  is  $M \times N \geq M + (M - 1) \times N$ , that is  $N \geq M$ , just as when the marginal covariation condition **MC** is imposed.

<sup>3</sup>  $C_n^M = 1$  because  $p^M(x) = 1$  for all  $x$ .

Summing the terms  $\mu_{mn}$  across  $m$  after multiplication by  $p_{mn}$  there is, for each  $n \in \{1, \dots, N\}$ , in view of equation (6)

$$\sum_{m=1}^M p_{mn} \mu_{mn} = \sum_{m=1}^M p_{mn} h_m = \mu(x_n)$$

where  $\mu(x_n) = E[Y_1|X = x_n]$ .<sup>4</sup> These equations are identical to the equations  $\mu_N = P_N \theta$  obtained under **MC** and when  $\text{rank}(P_N) = M$  they can be solved for the  $M$  unknown  $h_m$ 's. A necessary condition is that  $N^* \geq M$ .

When the support of  $X$  is rich in the sense that the cardinality of the support of  $X$  is at least equal to the cardinality of the support of the discrete endogenous variable the conditions **IC** and **MC** lead to the same identifying correspondence for  $\theta$ . Under **IC** there is the possibility of identifying values of the terms  $C_n^m$  which is not possible under **MC**. **IC** has no additional identifying power relative to **MC** for  $\theta$  when the support of  $X$  is rich.

When the support of  $X$  is sparse in the sense that  $N^* < M$  then **IC** can have additional *closed set* identifying power if a monotonicity restriction is imposed on the function  $c(\cdot)$ , as demonstrated in the next Section.

**4.2. Sparse support.** Impose the *additional restriction* that  $c(e_2)$  is a *monotonic* function of  $e_2$ .<sup>5</sup> When  $c$  is increasing there is then, in view of (4), the inequality

$$c(p^{m-1}(x)) \leq E(\varepsilon_1|\varepsilon_2 \in (p^{m-1}(x), p^m(x)]) \leq c(p^m(x))$$

which on adding  $h(y_2^m)$  implies

$$h(y_2^m) + c(p^{m-1}(x)) \leq E(Y_1|Y_2 = y_2^m \cap X = x) \leq h(y_2^m) + c(p^m(x))$$

with the inequalities reversed when  $c$  is decreasing.<sup>6</sup>

Suppose there exist values of  $X$ ,  $\{x_{m-1}, x_m\}$ , such that for some  $\bar{e}_2 \in (0, 1)$

$$p^m(x_m) \leq \bar{e}_2 \leq p^{m-1}(x_{m-1}) \tag{7}$$

and consider the case in which  $c$  is an *increasing* function. Since:

$$c(p^m(x_m)) \leq c(\bar{e}_2) \leq c(p^{m-1}(x_{m-1}))$$

there are the following inequalities.

$$\begin{aligned} E(Y_1|Y_2 = y_2^m \cap X = x_m) &\leq h(y_2^m) + c(p^m(x_m)) \\ &\leq h(y_2^m) + c(\bar{e}_2) \\ &\leq h(y_2^m) + c(p^{m-1}(x_{m-1})) \\ &\leq E(Y_1|Y_2 = y_2^m \cap X = x_{m-1}) \end{aligned}$$

<sup>4</sup>This expression results because  $\sum_{m=0}^M (C_n^m - C_n^{m-1}) = C(1) = 0$ .

<sup>5</sup>This sort of monotonicity restriction is imposed on conditional quantile functions in Chesher (2003) which studies identification in non-additive error models with discrete endogenous variables.

<sup>6</sup>If  $c(e_2) = c$ , constant for variations in  $e_2$ , in which case  $\varepsilon_1$  is *mean independent* of  $\varepsilon_2$ , then for any  $m$ ,

$$E(Y_1|Y_2 = y_2^m, X = x) = h(y_2^m) + c = E(Y_1|Y_2 = y_2^m)$$

and there is *point* identification of  $h(y_2^m)$  at *all* values of  $m$ .

The inequalities are reversed if  $c$  is a decreasing function and the following inequality results.

$$\begin{aligned} & \min(E(Y_1|Y_2 = y_2^m \cap X = x_m), E(Y_1|Y_2 = y_2^m \cap X = x_{m-1})) \\ & \leq h(y_2^m) + c(\bar{\varepsilon}_2) \leq \\ & \max(E(Y_1|Y_2 = y_2^m \cap X = x_m), E(Y_1|Y_2 = y_2^m \cap X = x_{m-1})) \end{aligned}$$

This inequality holds at any point of support  $m$  for which there exist  $x_{m-1}$  and  $x_m$  satisfying (7) any value of  $\varepsilon_2$ ,  $\bar{\varepsilon}_2$ . When this condition holds at *many* points of support for the *same* value  $\bar{\varepsilon}_2$  the value of the structural function is identified at those points of support up to a *common* location shift  $c(\bar{\varepsilon}_2)$  and then *differences* of the structural function,  $h(y_2^m) - h(y_2^n)$  are identified.

The condition (7) cannot hold when  $y_2^m$  is an extreme point of support carrying non-zero probability mass. For example when  $y_2^m$  is the largest value achieved by  $Y_2$ , (7) can only hold for  $\{x_{m-1}, x_m\}$  such that  $p^m(x_m) = p^{m-1}(x_{m-1}) = 1$  but then there is no probability mass on  $y_2^m$ .

The bounds are easily estimated using sample averages of  $Y_1$  at  $y_2^m$  and values of  $X$  satisfying (7). In practice those values of  $X$  will be determined using an estimate of the conditional distribution of  $Y_2$  given  $X$  which will have implications for the computation of standard errors.

## 5. STRENGTH OF INSTRUMENTS AND SUPPORT

As noted in Section 3, under **MC** the identifiability of the structural function depends critically on the *support* of  $X$ . If  $X$  has fewer points of support than the discrete endogenous variable then the value of the structural function is not identified at any point of support without further restriction.

Under **IC** and a monotonicity restriction on  $c(\varepsilon_2)$  the interval identifiability of values and differences of a structural function is greatly affected by the *strength* of the instruments but, with strong instruments, the *support* of  $X$  does not play a critical role. With a strong enough instruments there can be interval identification of *many* values of a structural function up to common location shift even when the instruments have only two points of support. However when instruments are weak there may be no possibility of interval identification.

An example illustrating *lack* of identifiability is shown in Table 1 which exhibits conditional distribution functions for  $Y_2$  with four points of support when  $X$  has two points of support. In this case, because for all  $m$ ,

$$\max(P[Y_2 \leq y_2^{m-1}|X = 0], P[Y_2 \leq y_2^{m-1}|X = 1]) < \min(P[Y_2 \leq y_2^m|X = 0], P[Y_2 \leq y_2^m|X = 1])$$

there is no value of  $\bar{\varepsilon}_2$  and  $m$  for which there are values of  $X$ ,  $\{x_{m-1}, x_m\}$  such that the condition (7) holds. Chesher (2003) shows that this situation prevails in the data used in Angrist and Krueger (1991).

Contrast this case with the one shown in Table 2. Here the effect of  $X$  on the conditional distribution of  $Y_2$  is substantial and there is interval identification of the value of the structural function at  $Y_2 = 2$  and  $Y_2 = 3$  up to a common location shift, even though  $X$  has just two points of support. This case is now worked through in detail to illustrate the results of the Section 4.2.

Table 1: Conditional distribution functions of  $Y_2$  given  $X = 0$  and 1 - Example 1

$y_2^m$	$P[Y_2 \leq y_2^m   X = 0] = p^m(0)$	$P[Y_2 \leq y_2^m   X = 1] = p^m(1)$
1	0.3	0.2
2	0.5	0.4
3	0.7	0.6
4	1.0	1.0

 Table 2: Conditional distribution functions of  $Y_2$  given  $X = 0$  and 1 - Example 2

$y_2^m$	$P[Y_2 \leq y_2^m   X = 0] = p^m(0)$	$P[Y_2 \leq y_2^m   X = 1] = p^m(1)$
1	0.7	0.1
2	0.8	0.2
3	0.9	0.3
4	1.0	1.0

Since in Table 2:

$$\begin{aligned} \forall \bar{e}_2 \in (0.2, 0.7) : & \quad p^2(1) \leq \bar{e}_2 \leq p^1(0) \\ \forall \bar{e}_2 \in (0.3, 0.8) : & \quad p^3(1) \leq \bar{e}_2 \leq p^2(0) \end{aligned}$$

there is, if  $c(\cdot)$  is increasing,

$$\begin{aligned} \forall \bar{e}_2 \in (0.2, 0.7) : & \quad E[Y_1 | Y_2 = 2 \cap X = 1] \leq h(2) + c(\bar{e}_2) \leq E[Y_1 | Y_2 = 2 \cap X = 0] \\ \forall \bar{e}_2 \in (0.3, 0.8) : & \quad E[Y_1 | Y_2 = 3 \cap X = 1] \leq h(3) + c(\bar{e}_2) \leq E[Y_1 | Y_2 = 3 \cap X = 0] \end{aligned}$$

with the inequalities reversed if  $c(\cdot)$  is decreasing. The intervals over which these two inequalities hold both include the interval  $(0.3, 0.7)$  and so  $h(2)$  and  $h(3)$  are interval identified up to the common location shift  $c(\bar{e}_2)$  for all  $\bar{e}_2$  in this common interval. As a result there is interval identification of the partial difference  $h(2) - h(3)$ , as follows

$$\begin{aligned} E[Y_1 | Y_2 = 2 \cap X = 1] - E[Y_1 | Y_2 = 3 \cap X = 0] \\ \leq h(2) - h(3) \leq \\ E[Y_1 | Y_2 = 2 \cap X = 0] - E[Y_1 | Y_2 = 3 \cap X = 1] \end{aligned} \tag{8}$$

with the inequalities reversed if  $c(e_2)$  is decreasing.

It is instructive to take this example a little further. Suppose that the data generating process has  $\varepsilon_1 | (\varepsilon_2, X) \sim N(c(\varepsilon_2), \sigma^2)$  and  $c(\varepsilon_2) = \lambda_0 + \lambda_1 \Phi^{-1}(\varepsilon_2)$  where  $\Phi^{-1}$  is the standard normal quantile (inverse distribution) function. Since  $\varepsilon_2 \sim Unif(0, 1)$ ,  $\Phi^{-1}(\varepsilon_2) \sim N(0, 1)$  and  $c(\varepsilon_2) \sim N(\lambda_0, \lambda_1^2)$ , so in this data generating process  $\varepsilon_1$  and  $c(\varepsilon_2)$  are jointly normally distributed. There is, for  $m \in \{1, 2, 3, 4\}$  and  $x \in \{0, 1\}$ :

$$\begin{aligned} E[Y_1 | Y_2 = y_2^m \cap X = x] &= h(y_2^m) + \lambda_0 + \frac{\lambda_1}{p_m(x)} \int_{p^{m-1}(x)}^{p^m(x)} \Phi^{-1}(e_2) de_2 \\ &= h(y_2^m) + \lambda_0 + \frac{\lambda_1}{p_m(x)} (\phi(\Phi^{-1}(p^{m-1}(x))) - \phi(\Phi^{-1}(p^m(x)))) \end{aligned}$$

where  $\phi$  denotes the standard normal density function.

Using the values in Table 2 there is:

$$\begin{aligned} E[Y_1|Y_2 = 2 \cap X = 1] &= h(2) + \lambda_0 - 1.045\lambda_1 \\ E[Y_1|Y_2 = 2 \cap X = 0] &= h(2) + \lambda_0 + 0.677\lambda_1 \\ E[Y_1|Y_2 = 3 \cap X = 1] &= h(3) + \lambda_0 - 0.677\lambda_1 \\ E[Y_1|Y_2 = 3 \cap X = 0] &= h(3) + \lambda_0 + 1.045\lambda_1 \end{aligned}$$

and the absolute value of the width of the bounds (8) is  $3.444 \times \lambda_1$ . As  $\lambda_1 \rightarrow 0$ , that is as  $\varepsilon_1$  and  $\varepsilon_2$  approach independence and  $Y_2$  becomes exogenous, the width of the bounds falls to zero and point identification is secured.

#### REFERENCES

- ANGRIST, JOSHUA, D., AND ALAN B. KRUEGER (1991): "Does compulsory schooling attendance affect schooling and earnings?" *Quarterly Journal of Economics*, 106, 979 - 1014.
- CHESHER, ANDREW (2003): "Nonparametric identification under discrete variation," Centre for Microdata Methods and Practice Working Paper CWP 19/03.
- DAS, MITALI (2004): "Instrumental Variables Estimation of Nonparametric Models with Discrete Endogenous Variables," forthcoming, *Journal of Econometrics*.
- FLORENS, JEAN-PIERRE AND LAETITIA MALAVOLTI (2003): "Instrumental Regression with Discrete Variables," presented at the 2003 European Meeting of the Econometric Society, Stockholm.