Investigaciones Económicas (Segunda época). Vol. XVI, n.º 3 (1992), pp. 401-409.

ON EXOGENEITY AND IDENTIFIABILITY*

Manuel ARELLANO
CEMFI

Este artículo presenta conceptos de exogeneidad que distinguen entre una situación en la que los parámetros de interés están identificados en un modelo condicional y, por tanto, es posible llevar a cabo inferencias consistentes sin tener que especificar el modelo incondicional, y una situación en la que esto no es posible. Para modelos de series temporales, se proponen definiciones análogas de predeterminación y exogeneidad estricta que coinciden con el uso habitual de estos términos en los trabajos econométricos. Se incluyen tres ejemplos: un modelo de expectativas racionales para series temporales, una autorregresión para datos de panel y una proyección lineal por mínimos cuadrados.

1. Introduction and preliminaries

The purpose of this article is to introduce exogeneity concepts which discriminate between a situation where parameters of interest are locally identified from a conditional model, and so consistent inferences can be made without having to specify the unconditional model, and a situation where this is not possible.

Let X be an mT-dimensional sample vector consisting of T observations on the $m \times 1$ random vector X_t and let the joint cumulative distribution function (cdf) of X belong to the family of distribution functions

$$F(x, \theta)$$

where θ is a vector of p unknown parameters which does not depend on T[1]. The possible values of θ are specified by the set Θ. We assume that every $\theta \in \Theta$

* I am grateful to David Hendry and two anonymous referees for helpful comments on a previous version of this paper.

[1] We rule out the possibility of having incidental parameters in the sense introduced by Neyman and Scott (1948). If this case is relevant, $F(x, \theta)$ can be thought of as either a conditional cdf on a set of sufficient statistics for the incidental parameters or a margi nal cdf on the incidental parameters after a probability distribution has been specified for these parameters (cf. Chamberlain (1980)).
On the other hand, the fully parameterized cdf $F(x, \theta)$ could be replaced by the «semi-parametric» cdf $F(x, \theta, g)$, where g is an infinite dimensional non-parametric component. This replacement would leave the discussion presented below essentially unchanged.

is identifiable; that is, for any $\theta^* \in \Theta$, $F(x,\theta)$ differs from $F(x,\theta^*)$ for some value of x.

The vector X_t can be partitioned into $X_t' = (Y_t' \; Z_t')$ where Y_t and Z_t are n and k-dimensional random vectors respectively. In addition, let us define $Y = (Y_1' \; ... \; Y_T')'$ and $Z = (Z_1' \; ... \; Z_T')'$ with a similar convention for lower case letters, which will denote the associated arguments of functions. The cdf of the sample can be factorized accordingly into.

$$F(x,\theta) = F(y,z,\theta) = \int_{-\infty}^{z} F(y\,|\,u,\theta_1)\, dF(u,\theta_2) \qquad [1]$$

Let Θ_j be the set of possible values of θ_j for $j = 1,2$. We assume that θ_1 in the conditional cdf of $Y|Z$ and θ_2 in the marginal cdf of Z are both identified.

Suppose there is an $r \times 1$ vector of parameters of interest α which is known to be related functionally to the elements of θ. That is,

$$\theta = \theta(\alpha, \gamma)$$

where γ is a vector of «nuisance» parameters which are not of direct concern. Then we may also write

$$\theta_1 = \theta_1(\alpha,\gamma)$$
$$\theta_2 = \theta_2(\alpha,\gamma)$$

Let $\bar{\theta}$ be the true value of θ and let $\bar{\psi}' = (\bar{\alpha}' \; \bar{\gamma}')$ be the true value of the $s \times 1$ vector $\psi' = (\alpha' \; \gamma')$. Moreover, if θ is a differentiable function of ψ we can define the following matrices of first partial derivatives

$$H = H(\psi) = \frac{\partial \theta}{\partial \psi'}$$

$$H_1 = H_1(\psi) = \frac{\partial \theta_1}{\partial \psi'}$$

If rank $H(\bar{\psi}) = s$, then $\bar{\psi}$ is locally identifiable. If, in addition, rank $H_1(\bar{\psi}) = s$, then $\bar{\psi}$ is locally identifiable in the conditional model $F(y\,|\,z,\theta_1)$ alone. It may be the case that not all the elements of γ are locally identifiable while the elements of α are. In particular, if H_1 factors into BK where B is nonsingular, K is of the form

$$\begin{pmatrix} K_{1\alpha} & 0 \\ K_{2\alpha} & K_{2\gamma} \end{pmatrix}$$

and $K_{1\alpha}$ has full column rank, then α is locally identifiable in $F(y\,|\,z,\theta_1)$ (see Rothenberg (1973), page 38).

2. Exogeneity

Clearly, there is an arbitrary number of different parameterizations $(\theta; \theta_1, \theta_2)$ in terms of which the joint cdf of X can be expressed, thus giving rise to alternative sets of functions of the parameters of interest. In effect, we are concerned with the existence of one such parameterization for which α is locally identifiable in the conditional model. This suggests the following definition of exogeneity.

Definition 1. Z_t is exogenous for α if and only if there exists a parameterization (θ_1, θ_2) such that α is locally identifiable in $F(y|z,\theta_1)$.

That is, there exists a neighbourhood of $\bar{\alpha}$ in which $\bar{\alpha}$ is the unique solution of $\bar{\theta}_1 = \theta_1(\alpha,\gamma)$ for α.

Notice that if the observations are independent, as it usually happens with cross-section and panel data, the previous definition can be restated in terms of $F(y_t|z_t,\theta_1)$.

Also notice that the exogeneity or otherwise of Z_t for α depends exclusively on the choice of parameters of interest, which is made *a priori,* and hence it is not a testable assumption in general.[2]

Often, parameters of interest only relate to first and second order conditional moments rather than the complete conditional cdf which is intentionally left unspecified. To these cases, a narrower definition of exogeneity is relevant. The following one illustrates this idea.

Definition 2. Z_t is first order exogenous for α if there exists a parameterization of the conditional expectation $E[Y|Z,\phi_1]$ such that α is locally identified on the basis of $\phi_1 = \phi_1(\alpha, \gamma)$ alone.

3. Time series models

When X is a time series, the analysis usually concentrates on the distributions of the X_t conditional on the past and the continuous case is emphasized[3]. In this case, assuming that the X_t are continuous random variables, the joint probability density function (pdf) of the sample is factorized as

$$f(x, \theta) = \prod_{t=1}^{T} f(x_t | x_{t-1}, \dots, x_1, \theta)$$

and

$$f(x_t | x_{t-1}, \dots, x_1, \theta) = f(y_t | z_t, x_{t-1}, \dots, x_1, c_1) f(z_t | x_{t-1}, \dots, x_1, c_2)$$

[2] Exogeneity tests usually test restrictions on a specification of the joint distribution of Y and Z, relative to which α is assumed to be identifiable.

[3] We now assume that all distribution functions are conditional on initial conditions represented by X_0. The emphasis on continuous random variables is due to the fact that most economic time series contain aggregate observations which change smoothly.

Alternatively, in accordance with (1) we may write

$$f(x, \theta) = f(y \mid z, \theta_1) f(z, \theta_2) =$$

$$= \prod_{t=1}^{T} f(y_t \mid z, y_{t-1}, \ldots, y_1, \theta_1) \prod_{t=1}^{T} f(z_t \mid z_{t-1}, \ldots, z_1, \theta_2)$$

At this stage, it is convenient to recall that 'Y_t does not Granger cause Z_t' if

$$f(y_t \mid z, y_{t-1}, \ldots, y_1, \theta_1) = f(y_t \mid z_t, x_{t-1}, \ldots, x_1, \theta_1)$$

or equivalently

$$f(z_t \mid x_{t-1}, \ldots, x_1, c_2) = f(z_t \mid z_{t-1}, z_1, \ldots, z_1, c_2)$$

This suggests two more definitions which are relevant in a time series context.

Definition 3. Z_t is predetermined for α if there exists a parameterization of the conditional pdf $f(y_t \mid z_t, x_{t-1}, \ldots, x_1, c_1)$ such that α is locally identifiable on the basis of $c_1 = c_1(\alpha, \gamma)$.

Definition 4. Z_t is strictly exogenous for α if Z_t is exogenous for α and 'Y_t does not Granger cause Z_t'.

Notice that since exogeneity of Z_t for α means local identification of α in $\Pi_{t=1}^{T}$ $f(y_t \mid z, y_{t-1}, \ldots, y_1, \theta_1)$, strict exogeneity can also be defined as predeterminedness plus Granger non-causality. Moreover, since Granger non-causality is a testable assumption, so is strict exogeneity.

First order versions of these concepts in terms of the relevant conditional expectations are immediate on the lines of Definition 2, and accord with the standard definitions of predeterminedness and strict exogeneity in the econometric literature on the linear model.

4. Concluding remarks

Note that the concept of predeterminedness given by Definition 3 is weaker than the concept of weak exogeneity introduced by Richard (1980) and Engle, Hendry and Richard (1983). If in addition to local identifiability of α from c_1 we require that c_2 does not depend on α, and γ does not introduce cross-restrictions between c_1 and c_2, then Z_t is weakly exogenous for α. The difference between these concepts is illustrated with examples in the Appendix.

The reason for introducing this weaker concept is to be able to distinguish between two fundamentally different situations that may arise in practice. On the one hand, a case where consistent inferences about α are possible on the basis of the conditional model. In such case the conditional model is useful in order to learn about α because the parameters of interest are locally identified from the conditional reduced from coefficients alone. On the other hand, a case where consistent inferences about α from the conditional model are not possible and thus the marginal model of Z_t must be specified. If additional cross-restrictions relevant to α link the conditional and the marginal cdf's, then Z_t would not be weakly exogenous in either of the two cases. However, following Definition 3, X_t is still predetermined in the first case but not in the second.

Appendix: examples

A1. A linear rational expectations model

We consider a single equation rational expectations model, taken from Sargan (1992), which contains one strictly exogenous variable for both the structural and reduced form coefficients. However, this variable is not weakly exogenous in the sense of Engle, Hendry and Richard (1983) for either of the two sets of coefficients.

The equation generating the endogenous variable Y_t is of the form

$$Y_t = \beta_1 Y_{t-1} + \beta_1^* Y_{t+1}^t + c_0 Z_t + c_1^* Z_{t+1}^t + U_t^*$$

where $Y_{t+1}^t = E(Y_{t+1} | \Phi_t)$, $Z_{t+1}^t = E(Z_{t+1} | \Phi_t)$ and Φ_t is the information set containing Y_s, Z_s, $s \leq t$. The model is assumed to be regular in the sense that the two roots λ_1 and λ_2 derived from

$$\beta_1 = \frac{\lambda_1}{1 + \lambda_1 \lambda_2}, \qquad \beta_1^* = \frac{\lambda_2}{1 + \lambda_1 \lambda_2}$$

are both inside the unit circle ($|\lambda_1| < 1, |\lambda_2| < 1$).

The variable Z_t is assumed to be generated by a stationary autoregressive equation of the form

$$Z_t = \phi_1 Z_{t-1} + \ldots + \phi_p Z_{t-p} + V_t$$

The disturbances U_t^* and V_t are independently identically distributed normally with variances σ_u^2 and σ_v^2 respectively.

The regular solution path can be written in the form

$$Y_t = \lambda_1 Y_{t-1} + g_1 Z_t + \ldots + g_p Z_{t-p+1} + U_t$$

where $U_t = (1 + \lambda_1 \lambda_2) U_t^*$ and, under the assumptions of the model, $g_1, \ldots, g_p$ are well defined differentiable functions of $\lambda_1, \lambda_2, c_0, c_1^*$ and $\phi_1, \ldots, \phi_p$ (the relevant formulae are reported in Sargan's (1992) paper).

Define the following vectors of parameters

$$\psi = (\lambda_1, \lambda_2, c_0, c_1^*)'$$

$$\theta = (\lambda_1, g_1, \ldots, g_p)'$$

$$\phi = (\phi_1, \ldots, \phi_p)'$$

With $p = 3$, there is a one to one mapping from (θ, ϕ) to (ψ, ϕ) and ψ is just identified, while with $p > 3$ there are $p-3$ overidentifying restrictions which introduce nonlinear constraints in the reduced form coefficients(θ, ϕ).

Clearly, Z_t is exogenous for θ because θ is identified in the conditional model. However, with $p > 3$, joint estimation of θ and ϕ enforcing the rational expectations restrictions will result in more efficient estimates of θ than those based on the conditional model. Thus, Z_t is not weakly exogenous for θ. Moreover, Z_t is also exogenous, but not weakly exogenous, for the structural parameters $(\lambda_1, \lambda_2, c_0)$ in the model that sets $c_1^* = 0$, since they are also identified in the conditional model. In either case, Z_t is strictly exogenous since in this model Y_t does not Granger cause Z_t.

Consistent instrumental variables estimates of ψ can be obtained exploiting the moment conditions

$$E[(Y_t - \beta_1 Y_{t-1} - \beta_1^* Y_{t+1} - c_0 Z_t - c_1^* Z_{t+1}) W_t] = 0$$

where W_t is the $(p+1) \times 1$ vector

$$W_t = (Y_{t-1}, Z_t, \ldots, Z_{t-p+1})'$$

As with the reduced form coefficients, when $p > 3$ more efficient estimates of ψ can be obtained by jointly estimating ψ and ϕ. This fact precludes the weak exogeneity of Z_t for ψ.

Other examples of models involving rational expectations, with a discussion of the implications for inference of predetermined variables as opposed to strictly exogenous variables, can be found in Hayashi and Sims (1983).

A2. An autoregressive model for panel data

Let us consider a cross-sectional sample of N individual independent time series of three observations each $X_i = (Y_{i1}, Y_{i2}, Y_{i3})'$, and an autoregressive specification with individual effects of the form[4]

$$E(Y_{i2} \mid Y_{i1}, \eta_i) = \alpha Y_{i1} + \eta_i$$
$$E(Y_{i3} \mid Y_{i2}, Y_{i1}, \eta_i) = \alpha Y_{i2} + \eta_i$$

which implies

$$E[(Y_{i3} - Y_{i2}) - \alpha(Y_{i2} - Y_{i1}) \mid Y_{i1}] = 0$$

so that the autoregressive coefficient α is identified in the joint distribution of Y_{i2} and Y_{i3} conditional on Y_{i1} $f(y_{i2}, y_{i3} \mid y_{i1})$. According to our Definition 2, this means that Y_{i1} is first order exogenous for α in this model. However, Y_{i1} is not weakly exogenous for α under the following fully parameterized specification of the joint distribution of (Y_{i1}, Y_{i2}, Y_{i3}). Let us assume

[4] Notice the change of the notation from t and T to i and N, respectively, to avoid a confusion between the time series and the cross-sectional dimensions in this model.

$$Y_{is}|Y_{i(s-1)}, \ldots, Y_{i1}, \eta_i \sim N(\alpha Y_{i(s-1)} + \eta_i, \sigma^2) \qquad (s = 2,3)$$
$$\eta_i|Y_{i1} \sim N(\lambda Y_{i1}, \sigma_\eta^2)$$

$$Y_{i1} \sim N\left(0, \frac{\sigma_\eta^2}{(1-\alpha)^2} + \frac{\sigma^2}{(1-\alpha^2)}\right)$$

Notice that the previous assumptions fully specify the joint distribution $f(y_{i1}, y_{i2}, y_{i3})$ since

$$f(y_{i1}, y_{i2}, y_{i3}) = f(y_{i2}, y_{i3}|y_{i1})f(y_{i1})$$

and

$$f(y_{i2}, y_{i3}|y_{i1}) = \int f(y_{i2}, y_{i3}|y_{i1}, \eta)f(\eta|y_{i1})d\eta$$

and

$$f(y_{i2}, y_{i3}|y_{i1}, \eta_i) = f(y_{i3}|y_{i2}, y_{i1}, \eta_i)f(y_{i2}|y_{i1}, \eta_i)$$

Specifically, we have (see Arellano and Bover (1990)):

$$\begin{bmatrix} Y_{i2} \\ Y_{i3} \end{bmatrix} |Y_{i1} \sim N\left[\begin{pmatrix} \pi_2 \\ \pi_3 \end{pmatrix} Y_{i1}, \begin{pmatrix} \omega_{22} & \omega_{23} \\ \omega_{23} & \omega_{33} \end{pmatrix}\right]$$

with $\pi_2 = \alpha + \lambda$, $\pi_3 = \alpha(\alpha + \lambda) + \lambda$, $\omega_{22} = \sigma_\eta^2 + \sigma^2$, $\omega_{33} = (1+\alpha)^2\sigma_\eta^2 + (1+\alpha^2)\sigma^2$ and $\omega_{23} = (1+\alpha)\sigma_\eta^2 + \alpha\sigma^2$. Notice that α satisfies

$$\alpha = \frac{\pi_3 - \pi_2}{\pi_2 - 1}$$

and so it is clearly identified in the conditional model. However, the marginal distribution of Y_{i1} contains additional information on α, so that Y_{i1} cannot be weakly exogenous for α, although it is exogenous in the sense of Definition 1.

A3. Linear least squares projections

Linear projections or best linear predictors (BLP) have many applications in applied econometrics. A leading case is the application to the signal extraction problem (see Sargent (1987, Chapter 10) or Goldberger (1991) for recent accounts of the theory and illustrations). Suppose that the parameters of interest are the coefficients α and β of the BLP of Y_t given Z_t:

$$E^*(Y_t|Z_t) = \alpha + \beta Z_t$$

where

$$\beta = \frac{\text{Cov}\,(Y_t, Z_t)}{\text{Var}\,(Z_t)}$$

$$\alpha = E(Y_t) - \beta E(Z_t)$$

and it is assumed that (Y_t, Z_t) are jointly stationary with conditional expectation function (CEF) of Y_t given Z_t.

$$E(Y_t | Z_t) = g(Z_t, \theta)$$

In this problem, α and β cannot be retrieved from knowledge of the conditional distribution $f(y | z)$ unless $g(Z_t, \theta)$ is linear, in which case the CEF and the BLP coincide. We have

$$\beta = \frac{\text{Cov}(E(Y_t | Z_t), Z_t)}{\text{Var}(Z_t)} = \frac{\text{Cov}(g(Z_t\ \theta), Z_t)}{\text{Var}(Z_t)}$$

If $g(Z_t, \theta) = \theta_0 + \theta_1 Z_t$, it follows from the expression above that $\beta = \theta_1$, but if $g(Z_t, \theta)$ is nonlinear, knowledge of the marginal distribution of Z_t is required to be able to identify α and β. Hence Z_t is not exogenous for α and β in the linear projection model, although it is exogenous for θ.

The implications of exogeneity can be clearly seen in this example by considering various forms of sampling: with random sampling from the joint distribution $f(y_t, z_t)$, the ordinary least squares (OLS) regression of Y_t on Z_t produces consistent estimates of α and β, while non-linear least squares (NLS) estimates θ consistently. However, with stratified sampling with empirical distribution of Z_t given by $h(z_t)$, and sampling of Y_t from the sequence of conditional distributions $f(y_t | z_t)$, NLS still estimates θ consistently but OLS does not estimate consistently α and β.

As an illustration, suppose that

$$Y_t | Z_t \sim N(\theta_0 + \theta_1 Z_t^2, \sigma^2)$$

If Z_t has a marginal Poisson distribution with parameter λ

$$f(z_t) = \lambda^{z_t} \exp(-\lambda)/z_t!$$

it can be easily checked taht

$$\beta = (2\lambda + 1)\, \theta_1$$

while if $Z_t \sim N(\lambda, \sigma_z^2)$ then

$$\beta = 2\lambda\theta_1$$

This illustrates the fact that to determine β we require knowledge of both parameters of the conditional distribution of $Y_t | Z_t$ and the marginal distribution of Z_t.

References

Arellano, M. and Bover, O. (1990): «La Econometría de Datos de Panel», *Investigaciones Económicas* 14, pp. 3-45.

Chamberlain, G. (1980): «Analysis of Covariance with Qualitative Data», *Review of Economic Studies* 47, pp. 225-238.

Engle, R. F.; Hendry, D. F. and Richard, J. F. (1983): «Exogeneity», *Econometrica* 51, pp. 277-304.

Goldberger, A. S. (1991): *A Course in Econometrics,* Harvard University Press.

Hayashi, F. and Sims, C. (1983): «Nearly Efficient Estimation of Time Series Models with Predetermined, but not Exogenous, Instruments», *Econometrica* 51, pp. 783-798.

Neyman, J. and Scott, E. L. (1948): «Consistent Estimates Based on Partially Consistent Observations», *Econometrica* 16, pp. 1-32.

Richard, J.-F. (1980): «Models with Several Regimes and Changes in Exogeneity», *Review of Economic Studies* 47, pp. 1-20.

Rothenberg, T. J. (1973): *Efficient Estimation with A Priori Information,* Cowles Foundation Monograph 23, New Haven, Yale University Press.

Sargan, J. D. (1992): «A Comparison of Different Methods of Estimating and Testing the Specification of Rational Expectations Models with One Endogenous and One Exogenous Variable», Universidad Carlos III, Working Paper 9224, Madrid.

Sargent, T. J. (1987): *Macroeconomic Theory,* 2nd edition, Academic Press.

Abstract

This article presents exogeneity concepts which discriminate between a situation where parameters of interest are identified from a conditional model, and so consistent inferences can be made without having to specify the unconditional model, and a situation where this is not possible. For time series models, related definitions of predeterminedness and strict exogeneity are given which coincide with the standard usage of the terms in the econometric literature. Three examples are provided: a time series rational expectations model, an autoregression for panel data and a linear least squares projection.

Recepción del original, febrero de 1992
Versión final, julio de 1992