TESTING EXOGENEITY IN OVERIDENTIFIED MODELS

Nunzio Capuccio and Renzo Orsi

July 1987

N. 38

TESTING EXOGENEITY IN OVERIDENTIFIED MODELS

by

Nunzio CAPPUCCIO(*) and Renzo ORSI(**)

July 1987

Abstract

In this paper we analyze the consequences of model overidentification on testing exogeneity, when maximum likelihood techniques for estimation and inference are used. This situation is viewed as a particular case of the more general problem of considering how restrictions on "nuisance" parameters could help in making inference on the parameters of interest. At first a general model is considered. A suitable likelihood function factorization is used which permits to easily derive the information matrix and other tools useful for constructing joint tests for exogeneity and overidentifying restrictions both of Wald and Lagrange Multiplier type. The asymptotic local power of the exogeneity test in the justidentified model is compared with that of the overidentified one, when we assume that this last is the true model. Then the pseudolikelihood framework is proposed for appreciating the consequences of working with a model where overidentifying restrictions are erroneously imposed. The inconsistency introduced by imposing false restrictions is analysed and the consequences of the misspecification on the exogeneity test are carefully examined.

(*) Università di Trieste, Trieste.
(**) Università della Calabria, Cosenza.

Thanks are due to J.-P. Florens, A. Holly and M. Mouchart for helpful suggestions and comments. The usual disclaimer applies. Financial supports from C.N.R. and "Ministero della Pubblica Istruzione" are gratefully aknowledged.

1. Introduction

In a strategy for model specification from the general to the particular, it seems very natural, with the aim of minimizing the risks of misspecification, to consider as a starting point of the analysis a model with a number of structural equations lower than the number of endogenous variables (see Hendry(1985)) and to complete the model, subsequently, with the addition of non constrained (i.e. justidentified) reduced form equations. Then successive data compatible reductions are operated on the model, in order to get a more parsimonious specification and, possibly, an improvement in the quality of the estimates of the parameters of interest.

These reductions may be fulfilled, among others, both by reducing the number of equations, i.e. moving some endogenous variables into the set of the exogenous ones, and by imposing overidentifying restrictions. An example of this specification procedure may be found in Richard(1984) where a sequence of assumptions is proposed and, at first, the overidentifying restrictions and then the exogeneity assumptions are tested, the statistic for these tests being the likelihood ratio test. In empirical econometrics, after all, it is very frequent to work with overidentified models. This is obvious when one considers the number of instruments used in the estimation of structural equations and the number of variables that appear in the chosen specification. Increasing the number of excluded variables means rising the number of overidentifying restrictions imposed without statistical control, and consequently the risk of misspecification, whatever the estimation technique adopted.

In view of testing exogeneity we follow a specification testing strategy, i.e. tests on "nuisance" parameters, in order to analyse the consequences, on the test, of imposing overidentifying restrictions. More precisely, the idea we want to pursue in this work is trying to evaluate the influence that restrictions concerning some nuisance parameters may have on the inference concerning the other parameters of the model.

Our analysis considers two possible situations of misspecification. At first we analyze an overparametrized model, i.e. a model where true overidentifying restrictions are not imposed. This framework permits us both of building up joint tests of exogeneity and overidentifying restrictions, and of studying the consequences of overparametrization on the power of the exogeneity test. As a second situation we study the consequences arising from an underparametrized model on the exogeneity test. To that end we pursue our analysis in the context of the pseudo-likelihood framework, since this approach permits to find analitically the pseudo true value to which the statistic maximizing the pseudo-likelihood function converges. By assuming that the true model is the just-identified one, it is possible to write down the analytic expression for the true model parameter function to which the pseudo maximum likelihood estimator converges. It appears that the exogeneity test done in the wrongly overidentified model cannot be interpreted either as an exogeneity test or as a joint test of exogeneity and overidentifying restrictions in the true model, except for a few particular situations.

In section two we present the general model and a likelihood function factorization which is helpful for deriving the information matrix and the other tools which are required for the formulation of the tests. In the third section we study the asymptotic local power of the exogeneity test for the overparametrized model while in the fourth section we consider the wrongly overidentified model, along with the pseudo likelihood function and the pseudo true values for the parameters. Some concluding remarks can be found in the final section.

2. The general model

Let us consider a vector $y_t \in R^{(n+1)}$, which is partitioned as $y_t = (y_{t1}\ y'_{t2})'$ with $y_{t1} \in R$ and $y_{t2} \in R^n$, and a vector $x_t \in R^k$, partitioned as $x_t = (x'_{t1}\ x'_{t2}\ x'_{t3})'$, with $x_{t1} \in R^{k_1}$, $x_{t2} \in R^{k_2}$, $x_{t3} \in R^{k_3}$, $k = (k_1 + k_2 + k_3)$, $t = 1,\dots,T$.

From now on we assume that, conditionally on the variables $(x_1 \dots x_t \dots x_T)$, the random vectors $(y_1 \dots y_t \dots y_T)$ are independent and their conditional distribution is normal with conditional expectation being a linear function of x_t only and variance-covariance matrix costant over time, i.e.:

$$y_t \mid x_t, \psi \sim IN(\Pi' x_t ; \Omega), \; t=1,..,T \tag{2.1}$$

where the parameter vector ψ completely characterizes the conditional distribution (2.1). These assumptions imply both a sequential cut and an initial cut, according to Florens and Mouchart(1985), assuming that the parameters in ψ and the parameters characterizing the marginal distribution of the x_t's are variation free. It follows that the joint distribution of $(y_1,\ldots, y_t,\ldots, y_T)$, conditional on $(x_1,\ldots, x_t,\ldots, x_T)$ is obtainable as the product of (2.1) by t. Consequently, by defining the matrices

$$Y = [\, y_1 \ldots y_t \ldots y_T]' \quad , \quad X = [x_1 \ldots x_t \ldots x_T]'$$

we can write

$$Y \mid X, \psi \sim MN_{T\times(n+1)}(X\Pi \; ; \; I_T \otimes \Omega) \tag{2.2}$$

Sometimes one may prefer to work with the distribution of the unobservable random vector u_t derived as the residual from the approximation of y_t given by its regression function, i.e.:

$$u_t = y_t - \Pi' x_t = y_t - \xi_t \tag{2.3}$$

where

$$\xi_t = E[\, y_t \mid x_t, \psi\,] = \Pi' x_t \tag{2.4}$$

$$u_t \mid \psi \sim IN(0 ; \Omega) \tag{2.5}$$

and $u_t \perp\!\!\!\perp x_t \mid \psi$.

The distribution (2.2) represents the classical multivariate regression model with uncostrained coefficient matrix. Then, conditionally on x_t, the conditional expectation vector ξ_t may be any point in $R^{(n+1)}$. Economic theory may provide constraints on ξ_t; we assume that ξ_t along with a subset of variables which are in x_t, namely z_t, are limited by an exact linear and constant relationship of the type:

$$\bar{\alpha}'\xi_t - \beta' z_t = 0 \tag{2.6}$$

where $\bar{\alpha}$ and β are vectors of unknown parameters. We assume, moreover, that the only constraint on these parameters is represented by the normalization which affects the first element of the vector $\bar{\alpha}$: $\bar{\alpha} = (\,1 \quad -\alpha'\,)'$. The "structural" parameters α and β in the subsequent will be denoted as "parameters of interest".

According to the adopted normalization rule and (2.3), the definition (2.6) may be rewritten in terms of observables, by partitioning $u_t = (\, u_{t1} \quad u'_{t2}\,)'$, i.e.:

$$(\,y_{t1} - u_{t1}) - \alpha'(\,y_{t2} - u_{t2}) - \beta' z_t = 0$$

from which

$$y_{t1} - \alpha' y_{t2} - \beta' z_t = \varepsilon_t \tag{2.7}$$

The so-called "structural residual" is defined as:

$$\varepsilon_t = u_{t1} - \alpha' u_{t2} \tag{2.8}$$

and is normally distributed with zero mean and constant variance:

$$\varepsilon_t \mid \psi \sim IN(\,0\,; \sigma_{11}) \tag{2.9}$$

and moreover

$$\varepsilon_t \perp\!\!\!\perp x_t \mid \psi \tag{2.10}$$

By referring to the following partition of Ω , coherent with that of y_t

$$\Omega = \begin{pmatrix} \omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix}$$

we can write the variance of the structural residual as:

$$\sigma_{11} = \omega_{11} - 2\alpha'\Omega_{21} + \alpha'\Omega_{22}\,\alpha \tag{2.11}$$

The structural equation (2.7) does not permit to write down directly the likelihood function in terms of the parameters of interest. A useful way to complete the model is to specify directly the equation for y_{t2} as its data generation process, namely in the uncostrained reduced form (see Florens, Mouchart and Richard((1979), (1986)) and Richard(1984) for a deeper discussion on this view of model completion).

The complete model can be written as

$$\begin{bmatrix} 1 & -\alpha' \\ 0 & I_n \end{bmatrix} \begin{bmatrix} y_{t1} \\ y_{t2} \end{bmatrix} + \begin{bmatrix} -\beta' & 0' \\ & -\Pi_2' \end{bmatrix} x_t = \begin{bmatrix} \varepsilon_t \\ u_{t2} \end{bmatrix} \tag{2.12}$$

where $x_t = (\, x_{t1}' \; x_{t2}' \; x_{t3}')' = (\, z_t' \; x_{t3}')'$. The joint distribution for the error term of (2.12) is

$$\begin{bmatrix} \varepsilon_t \\ u_{t2} \end{bmatrix} \sim \mathrm{IN}(\, 0 \; ; \; \Sigma\,) \tag{2.13}$$

where

$$\Sigma = \begin{bmatrix} 1 & -\alpha' \\ 0 & I_n \end{bmatrix} \Omega \begin{bmatrix} 1 & 0' \\ -\alpha & I_n \end{bmatrix}$$

$$= \begin{bmatrix} \sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$$

As one can see in the complete model specification (2.12), the parameters characterizing the marginal expectation of y_{t2}, namely $\Pi_2'x_t$, enter y_{t1}'s conditional expectation only through the vector u_{t2}. It follows that a sufficient condition for y_{t2} being (weakly) exogenous w.r.t. α's and β's is $\Sigma_{21} = 0$, according to Florens, Mouchart and Richard(1979) and Engle, Hendry e Richard(1983).

Consistent estimation of model (2.12) requires that the identifying restrictions are satisfied, i.e. $k_3 \geqslant n$. Furthermore the partition of the exogenous variables into z_t and x_{t3} is arbitrary. Clearly, when $k_3 = n$ (2.12) is just-identified.

2.1 The likelihood function factorization

The model (2.12) is easily analyzed when residuals are trasformed permitting to get equations with independent random errors. To this end the joint distribution (2.13) may be factorized into the product of the two distributions:

$$\varepsilon_t \sim IN(0 ; \sigma_{11}) \tag{2.14}$$

$$u_{t2} \mid \varepsilon_t \sim IN(\lambda\varepsilon_t ; \Sigma_{22\cdot 1}) \tag{2.15}$$

where

$$\lambda = (\sigma_{11})^{-1}\Sigma_{21}$$

$$\Sigma_{22\cdot 1} = \Sigma_{22} - \Sigma_{21}(\sigma_{11})^{-1}\Sigma_{12}$$

Among the two possible factorizations of (2.13) we prefer the (2.14)-(2.15) since it is more general for our purposes of testing exogeneity. In fact if one wants to build up an exogeneity test on a subvector of y_{t2}, this test will be a simple test of equality to zero for the corresponding subvector of λ (see Holly(1985) for the other factorization).

On the grounds of the factorization (2.14)-(2.15) the model (2.12) can be written in the form:

$$\begin{aligned} y_{t1} &= \alpha' y_{t2} + \beta' z_t + \varepsilon_t \\ y_{t2} &= \Pi_2' x_t + \lambda\varepsilon_t + v_t \end{aligned} \tag{2.16}$$

where

$$v_t = (u_{t2} - \lambda\varepsilon_t) \sim IN(0 \ ; \ \Sigma_{22\cdot1}) \tag{2.17}$$

and, moreover, $v_t \perp\!\!\!\perp \varepsilon_t$.

Model (2.16) permits to factorize its log-likelihood function into the sum of two components, where the so-called "nuisance" parameters appear only in one of them. More precisely

$$\frac{1}{T} L(\theta, \phi) = \frac{1}{T} \sum_t \ell_{1t}(\theta, \sigma_{11}) + \frac{1}{T} \sum_t \ell_{2t}(\theta , \Pi_2, \lambda, \Sigma_{22\cdot1}) \tag{2.18}$$

where

$$\theta = [\alpha' \ \beta']' = [\alpha' \ \beta_1' \ \beta_2']' = [\theta_1' \ \beta_2']'$$

$$\phi = [\sigma_{11} \quad \Pi_2 \ \lambda \quad \Sigma_{22\cdot1}]$$

$$\ell_{1t} = - \frac{1}{2} \ln 2\pi - \frac{1}{2} \ln \sigma_{11} - \frac{1}{2 \sigma_{11}} \varepsilon_t^2 \tag{2.19}$$

$$\ell_{2t} = - \frac{1}{2} \ln 2\pi - \frac{1}{2} \ln |\Sigma_{22\cdot1}| - \frac{1}{2} v_t' \Sigma_{22\cdot1}^{-1} v_t \tag{2.20}$$

The score vector relative to the likelihood function (2.18) is given by

$$\frac{1}{T} \begin{bmatrix} \frac{\partial L}{\partial \theta} \\ \frac{\partial L}{\partial \sigma_{11}} \\ \frac{\partial L}{\partial \pi_2} \\ \frac{\partial L}{\partial \lambda} \\ \frac{\partial L}{\partial \sigma_{22\cdot1}} \end{bmatrix} = \frac{1}{T} \begin{bmatrix} \frac{1}{\sigma_{11}} W'\varepsilon - W'V \Sigma_{22\cdot1}^{-1}\lambda \\ - \frac{1}{2} \frac{T}{\sigma_{11}} + \frac{1}{2 \sigma_{11}^2} \varepsilon'\varepsilon \\ \mathrm{vec}[X' \ V \ \Sigma_{22\cdot1}^{-1}] \\ \Sigma_{22\cdot1}^{-1} V'\varepsilon \\ D' \ \mathrm{vec}[- \frac{T}{2} \Sigma_{22\cdot1}^{-1} + \Sigma_{22\cdot1}^{-1} V'V\Sigma_{22\cdot1}^{-1}] \end{bmatrix} \tag{2.21}$$

where $W = [Y_2 \ Z]$, $\pi_2 = \mathrm{vec}\ \Pi_2$, $\sigma_{22\cdot1}$ is the ($\frac{1}{2}$ n(n+1)$\times$ 1) vector containing the lower triangular part of $\Sigma_{22\cdot1}$ and D is the duplication matrix such that $D'\sigma_{22\cdot1} = \mathrm{vec} (\Sigma_{22\cdot1})$ (see Magnus and Neudecker(1980)).

The information matrix, J, computed as the negative value of the "plim" of the hessian matrix obtained from (2.18) is

$$J = \begin{array}{c} \theta \\ \lambda \\ \sigma_{11} \\ \pi_2 \\ \sigma_{22\cdot 1} \end{array} \begin{bmatrix} J_{11} & & & & \\ J_{21} & J_{22} & & & \\ J_{31} & 0 & J_{33} & & \\ J_{41} & 0 & 0 & J_{44} & \\ J_{51} & 0 & 0 & 0 & J_{55} \end{bmatrix} \tag{2.22}$$

Detailed computations are given in appendix A; below we report the final results for each non zero block of the matrix:

$$J_{11} = J_{\theta\theta} = (\sigma_{11}^{-1} + \lambda'\Sigma_{22\cdot 1}^{-1}\lambda) \begin{bmatrix} \Pi_2' \text{plim} \frac{X'X}{T} \Pi_2 + \sigma_{11}\lambda\lambda' + \Sigma_{22\cdot 1} & \\ \text{plim} \frac{Z'X}{T} \Pi_2 & \text{plim} \frac{Z'Z}{T} \end{bmatrix}$$

$$J_{21} = J_{\lambda\theta} = [\,(I_n - \sigma_{11}\Sigma_{22\cdot 1}^{-1}\lambda\lambda') \quad 0\,]$$

$$J_{22} = J_{\lambda\lambda} = \sigma_{11}\, \Sigma_{22\cdot 1}^{-1}$$

$$J_{31} = J_{\sigma_{11}\theta} = \frac{-1}{\sigma_{11}} [\, \lambda' \quad 0' \,]$$

$$J_{33} = J_{\sigma_{11}\sigma_{11}} = \frac{1}{2\, \sigma_{11}^2}$$

$$J_{41} = J_{\pi_2\theta} = -(\Sigma_{22\cdot 1}^{-1}\lambda \otimes [\, \text{plim} \frac{X'X}{T} \Pi_2 \quad \text{plim} \frac{X'Z}{T} \,])$$

$$J_{44} = J_{\pi_2\pi_2} = (\Sigma_{22\cdot 1}^{-1} \otimes \text{plim} \frac{X'X}{T})$$

$$J_{51} = J_{\sigma_{22\cdot 1}\theta} = - D'(\Sigma_{22\cdot 1}^{-1} \otimes [\, I_n \quad 0 \,])$$

$$J_{55} = J_{\sigma_{22\cdot 1}\sigma_{22\cdot 1}} = \frac{1}{2} D'(\Sigma_{22\cdot 1}^{-1} \otimes \Sigma_{22\cdot 1}^{-1}) D$$

Given the structure of the information matrix (2.22), its inverse is straightforwardly obtainable by the rule of the partitioned inverse and by taking care of some recurrence. Again detailed computations are omitted and only final results are reported. The first block of the inverse is given by:

$$J^{11} = \left(J_{11} - \sum_{i=1}^{5} J_{1i} J_{ii}^{-1} J_{i1} \right)^{-1} = \Phi^{-1}$$

Blocks belonging to the first column of the inverse may be written compactly as

$$J^{i1} = - J_{ii}^{-1} J_{i1} \Phi^{-1} \qquad \text{for } i = 2,\dots,5$$

Blocks on the main diagonal of the inverse are given by:

$$J^{ii} = J_{ii}^{-1} + J_{ii}^{-1} J_{i1} \Phi^{-1} J_{1i} J_{ii}^{-1}$$

$$= \left(J_{11} - J_{i1} \left(\Phi + J_{1i} J_{ii}^{-1} J_{i1} \right)^{-1} J_{1i} \right)^{-1} \qquad \text{for } i = 2,\dots,5$$

Finally the remaining blocks of the inverse, always in a compact form, are:

$$J^{ik} = J_{ii}^{-1} J_{i1} \Phi^{-1} J_{1k} J_{kk}^{-1} \qquad \text{for } i = 2,\dots,5 \quad k = 2,\dots,5 \quad i \neq k$$

2.2 Joint test of exogeneity and overidentifying restrictions

Partitioning β into two coefficient subvectors β_1 and β_2, related respectively to the variables x_{t1} and x_{t2}, the system of hypotheses to consider, in order to test jointly the exogeneity of y_{t2} w.r.t. $\theta_1 = (\alpha' \; \beta_1')$ and the overidentifying restrictions induced by the exclusion of x_{t2} from the structural equation, is given by H_0:$\lambda = 0$ and $\beta_2 = 0$. Obviously this test is conditionally on the assumption made relatively to the variables omitted from the structural equation. It results that the variables in x_t are not treated simmetrically along with the involved overidentifying restrictions. Although loss of simmetry is not desiderable theoretically, it is not clear, in practice, whether the condition of symmetry is a strict requirement for a test (see Hwang(1980) for a similar treatment of an overidentifying restriction test).

In the model (2.16) the general form of the Lagrange Multiplier (hereafter LM) test for H_o is:

$$LM = \frac{1}{T} \begin{bmatrix} \frac{1}{\hat{\sigma}_{11}} \hat{\varepsilon}' X_2 & \hat{\varepsilon}' \hat{V} \hat{\Sigma}_{22}^{-1} \end{bmatrix} \begin{bmatrix} \hat{J}_o^{\beta_2\beta_2} & \hat{J}_o^{\beta_2\lambda} \\ \hat{J}_o^{\lambda\beta_2} & \hat{J}_o^{\lambda\lambda} \end{bmatrix} \begin{bmatrix} \frac{1}{\hat{\sigma}_{11}} X_2' \hat{\varepsilon} \\ \hat{\Sigma}_{22}^{-1} \hat{V}' \hat{\varepsilon} \end{bmatrix} \tag{2.23}$$

where the superscript "^" denotes the OLS estimates:

$$\hat{\varepsilon} = y_1 - Y_2 \hat{\alpha} - X_1 \hat{\beta}_1$$

$$\hat{\sigma}_{11} = T^{-1} \hat{\varepsilon}' \hat{\varepsilon}$$

$$\hat{V} = Y_2 - X \hat{\Pi}_2$$

$$\hat{\Pi}_2 = (X'X)^{-1} X' Y_2$$

$$\hat{\Sigma}_{22} = T^{-1} \hat{V}' \hat{V}$$

and $\hat{J}_o^{\beta_2\beta_2}$, $\hat{J}_o^{\beta_2\lambda}$ (= ($\hat{J}_o^{\lambda\beta_2}$)') and $\hat{J}_o^{\lambda\lambda}$ are the relevant blocks of the inverse of the information matrix evaluated under H_o, which is given by:

$$J^o = \left[\begin{array}{cccc|ccc} J^o_{\alpha\alpha} & & & & & & \\ J^o_{\beta_1\alpha} & J^o_{\beta_1\beta_1} & & & & 0 & \\ J^o_{\beta_2\alpha} & J^o_{\beta_2\beta_1} & J^o_{\beta_2\beta_2} & & & & \\ J^o_{\lambda\alpha} & 0 & 0 & J^o_{\lambda\lambda} & & & \\ \hline & & & & J^o_{\sigma_{11}\sigma_{11}} & & \\ & & & & & J^o_{\pi_2\pi_2} & \\ & 0 & & & 0 & & J^o_{\sigma_{22}\sigma_{22}} \end{array}\right] \tag{2.24}$$

The block diagonal structure of J^o permits to neglect the lower block of the matrix in buiding up the LM test. Then (2.23) can be written as:

$$LM = [\,0' \quad 0' \quad \hat{\varepsilon}'X_2 \;\; \hat{\varepsilon}'\hat{V}\,] \cdot \begin{bmatrix} \hat{\sigma}_{11} & 0' \\ 0 & \hat{\Sigma}_{22} \end{bmatrix}^{-1} \cdot$$

$$\cdot \begin{bmatrix} \hat{\sigma}_{11}^{-1} W'W & \hat{\sigma}_{11}^{-1} W'X_2 & \begin{bmatrix} I_n \\ 0 \end{bmatrix} \\ \hat{\sigma}_{11}^{-1} X_2'W & \hat{\sigma}_{11}^{-1} X_2'X_2 & 0 \\ [\,I_n \quad 0\,] & 0 & \hat{\sigma}_{11}\hat{\Sigma}_{22}^{-1} \end{bmatrix}^{-1} \cdot \begin{bmatrix} \hat{\sigma}_{11} & 0' \\ 0 & \hat{\Sigma}_{22} \end{bmatrix}^{-1} \cdot \begin{bmatrix} 0 \\ 0 \\ X_2'\hat{\varepsilon} \\ \hat{V}'\hat{\varepsilon} \end{bmatrix} \qquad (2.25)$$

Since $\hat{\varepsilon}'X_2 = 0$, $\hat{\varepsilon}'Y_2 = 0$ and $Y_2'\hat{V} = \hat{V}'\hat{V}$, (2.25), after some manipulations, can also be written as:

$$LM = \hat{\sigma}_{11}^{-1}\hat{\varepsilon}'[\,W \;\; X_2 \;\; \hat{V}\,] \begin{bmatrix} W'W & W'X_2 & W'V \\ X_2'W & X_2'X_2 & 0 \\ \hat{V}'W & 0 & \hat{V}'\hat{V} \end{bmatrix}^{-1} \begin{bmatrix} W' \\ X_2' \\ \hat{V}' \end{bmatrix} \hat{\varepsilon} \qquad (2.26)$$

This is precisely T times the R^2 of the regression of OLS residuals $\hat{\varepsilon}$ upon $\hat{V}$, X_2, Y_2, X_1. In other words, this is the LM test statistic for the problem of testing $\beta_2 = 0$ and $\delta = 0$ in the "expanded regression"

$$y_1 = Y_2\alpha + X_1\beta_1 + X_2\beta_2 + \hat{V}\delta + e \qquad (2.27)$$

(see Engle(1982)).

The Wald test for H_o in the model (2.16) can be easily derived by using the blocks of the inverse of matrix J previously obtained. Its general formulation is:

$$W = T\,[\,\tilde{\beta}_2' \quad \tilde{\lambda}'\,] \begin{bmatrix} \tilde{J}^{\beta_2\beta_2} & \tilde{J}^{\beta_2\lambda} \\ \tilde{J}^{\lambda\beta_2} & \tilde{J}^{\lambda\lambda} \end{bmatrix}^{-1} \begin{bmatrix} \tilde{\beta}_2 \\ \tilde{\lambda} \end{bmatrix} \qquad (2.28)$$

where the superscript "~" denotes the maximum likelihood estimates in (2.16).

3. The asymptotic local power of the exogeneity test

When the failure to reject a null hypothesis is used to claim that data support the null hypothesis, it must be shown that candidate alternatives would have been detected with a reasonably high probability. Moreover it is always interesting to compute the power of a statistical test when possible. This is true not only when two different statistics are used for testing the same hypothesis but also in the case of only one statistic employed in two different models to test the same null hypothesis. Clearly, in order that the results have some sense, the two models must be comparable. In particular they must be based on the same distributional assumptions.

In this section we compare the power of an exogeneity test made in two models. The first one is the just-identified version of model (2.16), i.e. $n = k_3$. The second one is the overidentified version of (2.16) obtained by imposing $\beta_2 = 0$, i.e. by excluding x_{t2} from the structural equation. The comparison between the two resulting statistics is possible if we suppose that this restriction is data coherent. Behind the two models there is the same distributional assumption (2.1), the difference being the structure of matrix Π . In fact we compare the exogeneity test made in an overparametrized model with the exogeneity test conducted in the correctly parametrized model.

It is well known that a problem encountered in deriving asymptotic approximations to the distribution of test statistics is that if the model is misspecified the test will often reject the null hypothesis with probability one as the sample size increases. The solution which is often proposed to this problem is to use a sequence of local alternatives which converges to the model specified under H_0. Here we shall consider the "asymptotic local power" method developed by Pitman(1949) (see also Serfling(1980)).

For testing the exogeneity of y_{t2} w.r.t. θ_1 , the null hypothesis is H_0: $\lambda = 0$, which is the same under the two models for the remarks made above. In order to derive the asymptotic local power of the test procedure we consider a sequence of local alternatives of the form

$$H_{1T}: \lambda_T = T^{-\frac{1}{2}} \gamma \tag{3.1}$$

where γ is a fixed vector of dimension n. Under quite general regularity conditions(see Holly(1986)), the asymptotic distribution of any of the three "classical test statistics", namely Wald, Likelihood Ratio and Lagrange Multiplier tests, is a non central chi-square with n degrees of freedom - the dimension of λ - and non centrality parameter given by:

$$\mu_i = \gamma' \left(J_i^{\lambda\lambda} \right)^{-1} \gamma \tag{3.2}$$

where with i = 1 we indicate the overparametrized model ($\beta_2 \neq 0$) and with i = 2 we indicate the correctly parametrized model ($\beta_2=0$). From the blocks of the inverse of the information matrix given previously we have that

$$\left(J_i^{\lambda\lambda} \right)^{-1} = J_{\lambda\lambda} - J_{\lambda i} \left(\Phi_i + J_{i\lambda} J_{\lambda\lambda}^{-1} J_{\lambda i} \right)^{-1} J_{i\lambda} \tag{3.3}$$

where i = 1 indicates θ and (3.3) is derived from the overparametrized model, while i = 2 indicates θ_1 and (3.3) is obtained from the correctly parametrized model. After some algebra it is possible to show that for the overparametrized model the non centrality parameter is given by:

$$\mu_1 = \sigma_{11} \gamma' \Sigma_{22.1}^{-1} \gamma - \gamma' A' \left[C_z + b\, Q \right]^{-1} A \gamma \tag{3.4}$$

where

$$A = \left(I_n - \sigma_{11} \lambda\lambda' \Sigma_{22\cdot 1}^{-1} \right)$$

$$Q = \left(\Sigma_{22\cdot 1} - \sigma_{11} \lambda\lambda' \right)$$

$$b = \left(\sigma_{11}^{-1} - \lambda' \Sigma_{22.1}^{-1} \lambda \right)$$

$$C_z = \sigma_{11}^{-1} \Pi_2' \operatorname{plim} (X' M_z X/T) \Pi_2$$

$$M_z = I_T - Z(Z'Z)^{-1}Z'$$

$$Z = [\ X_1 \quad X_2\ ]$$

and the non centrality parameter for the well parametrized model is given by:

$$\mu_2 = \sigma_{11}\ \gamma'\Sigma_{22.1}^{-1}\gamma - \gamma'A'[C_1 + b\ Q\ ]^{-1}\ A\gamma \tag{3.5}$$

where

$$C_1 = \sigma_{11}^{-1}\ \Pi_2'\ \text{plim}\ (X'M_1X/T)\ \Pi_2$$

$$M_1 = I_T - X_1(X_1'X_1)^{-1}X_1'$$

Hannan(1956) showed that Pitman's asymptotic relative efficiency is equal to the ratio of the non centrality parameters. When two test statistics have the same distribution with the same number of degrees of freedom and when the same size of the first type error is assumed, the test statistic with greater non centrality parameter is the most powerful w.r.t. the adopted sequence of local alternatives.

In our case it is easy to show that $\mu_2 \gtreqless \mu_1$ if $C_1 \gtreqless C_z$, i.e. if

$$\Pi_2'\ \text{plim}\ (X'\ (\ M_1 - M_z)\ X/T)\ \Pi_2 \gtreqless 0 \tag{3.6}$$

The sign of (3.6) is determined by the sign of the difference $(M_1 - M_z)$; this difference may be written in the form:

$$M_1 - M_z = P_z - P_1 = F$$

where $P_z = Z(Z'Z)^{-1}Z'$ and $P_1 = X_1(X_1'X_1)^{-1}X_1'$. Since P_z and P_1 are idempotent matrices, their difference is idempotent if $P_1P_z = P_zP_1 = P_1$. It is easy to show that F is idempotent and that its rank is equal to $l < T$, than F is P.S.D.S., implying $\mu_2 \geqslant \mu_1$. The main consequence of this result is that a classical exogeneity test is more powerful in the correctly parametrized model then in the overparametrized one. In other words, if the overidentifying restrictions, implied by the exclusion of some instrumental variables from the structural equation, are data coherent, the misspecification due to an overparametrization of the structural equation implies a loss of power of a classical exogeneity test. This seems to suggest that a good strategy for testing should be at first testing the overidentifying restrictions and after the exogeneity hypothesis.

In the next section we consider the reverse situation, i.e. testing exogeneity in an overidentified model, when the overidentifying restrictions are not consistent with the data.

4. The overidentified model as a pseudo true model

In this section we analyze the consequences on the exogeneity test, when it is performed in a misspecified model. We try to evaluate the inconsistency and investigate the relantionship between the parameters of the misspecified model and those of the correct model. In particular we want to know the influence that overidentifying restrictions, non data coherent but mantained in the model, have on testing exogeneity.

Imposing the restriction $\beta_2 = 0$ in the justidentified (true) model (2.16) we get an overidentified model which is:

$$\begin{aligned} y_{t1} &= \alpha' y_{t2} + \beta_1' x_{t1} + \varepsilon_t^* \\ y_{t2} &= \Pi_2' x_t + \lambda \varepsilon_t^* + v_t^* \end{aligned} \tag{4.1}$$

This is the so called pseudo true model, i.e. the model that is effectively used for estimating parameters. In this case the criterion function to be maximized is a pseudo likelihood function, i.e. the likelihood given by the misspecified model (4.1). Similarly with (2.18) the pseudo likelihood function is factorized as the sum of two terms:

$$T^{-1}L^*(\theta_1, \phi) = T^{-1} \sum_t \ell^*_{1t}(\theta_1, \sigma_{11}) + T^{-1} \sum_t \ell^*_{2t}(\theta_1, \Pi_2, \lambda, \Sigma_{22.1}) \quad (4.2)$$

where:

$$\ell^*_{1t} = -\frac{1}{2}\ln 2\pi - \frac{1}{2}\ln \sigma_{11} - \frac{1}{2\sigma_{11}} \varepsilon^{*2}_t \quad (4.3)$$

$$\ell^*_{2t} = -\frac{1}{2}\ln 2\pi - \frac{1}{2}\ln|\Sigma_{22.1}| - \frac{1}{2} v^{*\prime}_t \Sigma^{-1}_{22.1} v^*_t \quad (4.4)$$

In this case the pseudo true model is nested into the true (justidentified model) since (4.1) is a particular case of (2.16) obtained by constraining $\beta_2 = 0$ and then excluding x_{t2} from the structural equation. The effect of excluding relevant exogenous variables from a structural equation causes inconsistency with respect to all parameters.

It has be noticed that the misspecification $\beta_2 = 0$ implies that all the coefficient estimates, and not only those belonging to the considered equation, loose the property of consistency. This is one of the main reasons which is in favour of the way (2.12) we adopt for completing the model against the complete structural model specification (see Florens, Mouchart e Richard(1986)).

The asymptotic distribution of maximum likelihood estimators in the incorrectly specified model has been worked out by White(1982). The maximization of the criterion function (4.2) provides the so called pseudo maximum likelihood estimators that under standard regularity conditions, converge in probability to a vector ($\theta_1^{*\prime}$ $\phi^{*\prime})'$ which can be considered as the true parameter vector of the misspecified model. According to Sawa(1978) the parameter vector ($\theta_1{}^{*\prime}$ $\phi^{*\prime})'$ is defined as the pseudo true value of (θ_1' $\phi')'$. It can be shown that the almost sure limits of the parameter estimates of the misspecified model are given so as to minimize the Kullback-Leibler Information Criterion, namely the loss with respect to the true model. This assures that we minimize our ignorance about the true structure; White calls it the "minimum ignorance estimator".

The problem we want to investigate at this point is trying to approximate the effect of misspecification on the probability limits of the parameter estimates. In particular we are interested to the effects of a misspecification due to not true overidentification restrictions on the λ estimate, since this is the parameter on which any exogeneity test is based.

For the model (4.1) it is possible to obtain an analytical expression for the pseudo true value λ^*. As in Richard(1984) for a given θ_1 the model (4.1) can be considered as a multivariate regression model in (n+1) dependent variables, subject to a set of exclusion restrictions:

$$y_{t1}(\theta_1) = (y_{t1} - \alpha' y_{t2} - \beta_1' x_{t1}) = \varepsilon_t^*$$

$$y_{t2} = \Pi_2' x_t + \lambda \varepsilon_t^* + v_t^* \tag{4.6}$$

Since the residuals of the two equations, ε_t^* and v_t^*, are assumed independent, they may be analysed indipendently of each other. For a given $\tilde{\theta}_1$, the parameters of the equation for y_{t2} in (4.6) can be estimated by OLS. In particular for λ we have

$$\tilde{\lambda} = (\tilde{\varepsilon}^{*\prime} M_x \tilde{\varepsilon}^*)^{-1} Y_2' M_x \tilde{\varepsilon}^* = h^{-1} N \tag{4.7}$$

where

$$\tilde{\varepsilon}^* = y_1 - Y_2\tilde{\alpha} - X_1\tilde{\beta}_1$$

$$= X_2\beta_2 + \varepsilon - Y_2(\tilde{\alpha} - \alpha) - X_1(\tilde{\beta}_1 - \beta_1) \tag{4.8}$$

$\tilde{\alpha}$ and $\tilde{\beta}_1$ being estimates of α and β_1. Given (4.8) the denominator h of (4.7) can be written as:

$$h = \varepsilon' M_x \varepsilon - 2\,\varepsilon' M_x Y_2(\tilde{\alpha} - \alpha) + (\tilde{\alpha} - \alpha)' Y_2' M_x Y_2 (\tilde{\alpha} - \alpha) \tag{4.9}$$

where (4.9) is obtained using the fact that $X_i' M_x = 0$ for $i = 1, 2$. Taking the probability limit of (4.9) we obtain:

$$\text{plim } T^{-1} h = \sigma_{11} - 2\sigma_{11}\lambda' \text{plim } T^{-1}(\tilde{\alpha} - \alpha) +$$

$$+ (\text{plim } T^{-1}(\tilde{\alpha} - \alpha))' (\Sigma_{22.1} + \sigma_{11}\lambda\lambda') (\text{plim } T^{-1}(\tilde{\alpha} - \alpha)) \tag{4.10}$$

The numerator N of (4.7) can be written as

$$N = Y_2'M_x\varepsilon - Y_2'M_xY_2(\tilde{\alpha} - \alpha) \tag{4.11}$$

and its plim is given by

$$\text{plim } T^{-1}N = \sigma_{11}\lambda - (\Sigma_{22\cdot 1} + \sigma_{11}\lambda\lambda')\,\text{plim } T^{-1}(\tilde{\alpha} - \alpha) \tag{4.12}$$

Then the pseudo true value of $\tilde{\lambda}$ is :

$$\lambda^* = [\sigma_{11} - 2\sigma_{11}\lambda'(\text{plim } T^{-1}(\tilde{\alpha}-\alpha)) + (\text{plim } T^{-1}(\tilde{\alpha}-\alpha))'(\Sigma_{22\cdot 1}+\sigma_{11}\lambda\lambda')\cdot$$

$$\cdot\,\text{plim } T^{-1}(\tilde{\alpha}-\alpha)]^{-1}(\sigma_{11}\lambda - (\Sigma_{22\cdot 1}+\sigma_{11}\lambda\lambda')\,\text{plim } T^{-1}(\tilde{\alpha}-\alpha)) \tag{4.13}$$

From the point of view of estimation, data are analyzed with the purpose of obtaining, among other things, a good estimate of λ ; in particular we hope that an estimator $\tilde{\lambda}$ be consistent for λ. Under the (false) overidentified model specification $\tilde{\lambda}$ will not be consistent for λ , as it results from (4.13), so that it becomes important to have an evaluation of the inconsistency. This one depends on the incosistency of α: $\tilde{\lambda}$ is consistent for λ if and only if $\tilde{\alpha}$ is a consistent estimate of α . Otherwise $\tilde{\lambda}$ will be inconsistent. If α is estimated, for istance, by the instrumental variables method, the pseudo true value of $\tilde{\alpha}$ is given by

$$\alpha^* = \alpha + (\Pi_2'\text{plim } T^{-1}X'M_1X\,\Pi_2)^{-1}\,\Pi_2'(\text{plim } T^{-1}X'M_1X_2)\,\beta_2 \tag{4.14}$$

A sufficient condition for the consistency of $\tilde{\alpha}$ (and consequently of $\tilde{\lambda}$) is represented by

$$y_{t2} \perp\!\!\!\perp x_{t2} \mid x_t,\ \psi \tag{4.15}$$

for each t = 1,...,T. (4.15) means that if x_{t2} is a vector of irrelevant regressors for y_{t2} given x_t , its inclusion or exclusion from the structural equation does not affect the estimation of α . Expression (4.15) is the analogue of the orthogonality condition which ensures the consistency of OLS estimates in a multiple regression model when relevant regressors are omitted from the model.

It has be noticed that the inconsistency of $\tilde{\lambda}$ does not permit to discover the exogeneity of y_{t2} w.r.t. θ_1 , since even if $\lambda = 0$ in the true model, λ^* still will be different from zero. The only possibility of detecting the exogeneity of y_{t2} w.r.t. θ_1 by using a misspecified model occurs when the true parameter values of both β_2 and $\text{Var}(y_{t2} \mid x_t) = (\Sigma_{22.1} + \sigma_{11} \lambda \lambda ')$ are near to zero. In this case the difference between λ^* and λ became negligible and right conclusions about the exogeneity may be drawn on the basis of a misspecified model. In all other situations we have to cope with the inconsistency of $\tilde{\lambda}$.

A practical suggestion would be to reduce as much as possible the conditional variance of y_{t2} by choosing an optimal set of instruments (Sargan(1958)).

5. Some concluding remarks

At first we considered the general model along its completion and a likelihood function factorization we prefer for our analysis. For the justidentified version of the model we derived the joint test of exogeneity and overidentifying restrictions of Lagrange Multiplier type and its version as T time the R^2 obtained from an appropriate auxiliary regression.

With reference to the exogeneity test we studied the consequences deriving from two different kinds of misspecification. The former is encountered when one does not impose true overidentification restrictions. In this case , as for estimation there is a loss of efficiency, we showed that for the exogeneity test there is a loss in power. The latter case of model misspecification is given when one imposes false overidentifying restrictions. In such a situation the (pseudo) maximum likelihood estimator is inconsistent and we give the expression for the pseudo true value to which the estimator converges. This is a function of the parameters of the true model. On these grounds it has been possible to evaluate the consequences of misspecification on the exogeneity test.

As a practical result it seems that in the specification search it is advantageous to start by specifying a justidentified model. In fact both for estimation and for testing the worst consequences are encountered when one imposes false overidentifying restrictions.

Furthermore, on the basis of our results, the most apt sequence of tests seems to be that of testing at first the exclusion of conditioning variables from the structural equation, and then the exogeneity of the variables that one would like to consider as conditioning.

Definitely testing the overidentifying restrictions should be an integrated part of the model specification search.

References

Engle, R.F.(1982) "A General Approach to Lagrange Multiplier Model diagnostics", Journal of Econometrics, 20, 83-104.

Engle, R.F.,Hendry, and J.-F. Richard(1983) "Exogeneity", Econometrica, 51, 277-304.

Florens, J.-P. and M. Mouchart(1985) "Conditioning in Dynamic Models", Journal of Time Series Analysis, 6, 15-34.

Florens, J.-P., M. MOuchart and J.-F. Richard(1979) "Specification and Inference in Linear Models", CORE D.P. 7943, Université Catholique de Louvain.

Florens, J.-P., M. Mouchart and J.-F. Richard(1986) "Structural Time Series Modelling: A Bayesian Approach", Applied Mathematics and Compution, 20.

Hannan, E.J.(1956) "The Asymptotic Power of Certain Tests Based on Multiple Correlations", J.R.S.S., B, 18, 227-233.

Hendry, D.F.(1985) "Monetaru Economic Myth and Econometric Reality", Oxford Review of Economic Policy, 1, 72-84.

Holly, A.(1985) "Testing Exogeneity: A Survey", Cahiers de Recherches Economiques n.8506, D.E.E.P., University of Lausanne.

Holly, A.(1986) "Specification tests: an overview", Cahiers de Recherches Economiques n.8603, D.E.E.P., University of Lausanne, to appear in "Advances in Econometrics", edited by T.F. Bewley, Cambridge University Press.

Hwang, H.S.(1980) "A Comparison of Tests of Overidentifying Restrictions", Econometrica, 48, 1821-1825.

Pitman, E.J.G.(1949) "Notes on Non-Parametric Statistical Inference", Columbia University.

Richard, J.-F.(1984), "Classical and Bayesian Inference in Incomplete Simultaneous Equations Models", in Econometrics and Quantitative Economics, ed. by D.F. Hendry and K.F. Wallis, Basil Blackwell.

Sargan, J.D.(1958), "The Estimation of Economic Relationships using Instrumental Variables", Econometrica, 26, 393-415.

Sawa, T.(1978), "Information Criteria for Discriminating among Alternative Regression Models", Econometrica, 46, 1273-1292.

Serfling, R.J.(1980), Approximation Theorems of Mathematical Statistics, J. Wiley.

White, H.(1982), "Maximum Likelihood Estimation of Misspecified Models", Econometrica, 50, 1-26.

Appendix

From the score vector (2.21) the second derivatives of the loglikelihood function (2.18) w.r.t. the parameters are:

$$T^{-1} \frac{\partial^2 L}{\partial\theta\partial\theta'} = -\ (\sigma_{11}^{-1} + \lambda'\Sigma_{22.1}^{-1}\lambda)\ \frac{W'W}{T} \tag{A.1}$$

$$T^{-1} \frac{\partial^2 L}{\partial\theta\partial\sigma_{11}} = \sigma_{11}^{-2}\ \frac{W'\varepsilon}{T} \tag{A.2}$$

$$T^{-1} \frac{\partial^2 L}{\partial\theta\partial\lambda'} = -\ \frac{W'V}{T}\ \Sigma_{22.1}^{-1} + \frac{W'\varepsilon}{T}\ \lambda'\Sigma_{22.1}^{-1} \tag{A.3}$$

$$T^{-1} \frac{\partial^2 L}{\partial\theta\partial\pi_2'} = (\ \lambda'\Sigma_{22.1}^{-1} \otimes \frac{W'X}{T}\ ) \tag{A.4}$$

$$T^{-1} \frac{\partial^2 L}{\partial\theta\partial\sigma_{22.1}'} = (\ \lambda'\Sigma_{22.1}^{-1} \otimes \frac{W'V}{T}\ \Sigma_{22.1}^{-1}\ )\ D \tag{A.5}$$

$$T^{-1} \frac{\partial^2 L}{\partial\ \sigma_{11}^2} = \frac{1}{2}\ \sigma_{11}^{-2} - \sigma_{11}^{-3}\ \frac{\varepsilon'\varepsilon}{T} \tag{A.6}$$

$$T^{-1} \frac{\partial^2 L}{\partial\pi_2\partial\pi_2'} = -\ (\ \Sigma_{22.1}^{-1} \otimes \frac{X'X}{T}\ ) \tag{A.7}$$

$$T^{-1} \frac{\partial^2 L}{\partial\pi_2\ \partial\lambda'} = -\ (\ \Sigma_{22.1}^{-1} \otimes \frac{X'\varepsilon}{T}\ ) \tag{A.8}$$

$$T^{-1} \frac{\partial^2 L}{\partial\pi_2\partial\sigma_{22.1}'} = -\ (\ \Sigma_{22.1}^{-1} \otimes \frac{X'V}{T}\ \Sigma_{22.1}^{-1})\ D \tag{A.9}$$

$$T^{-1} \frac{\partial^2 L}{\partial\lambda\partial\lambda'} = -\ \Sigma_{22.1}^{-1}\ \frac{\varepsilon'\varepsilon}{T} \tag{A.10}$$

$$T^{-1} \frac{\partial^2 L}{\partial \lambda \partial \sigma'_{22.1}} = \left(\frac{\varepsilon' V}{T} \Sigma_{22.1}^{-1} \otimes \Sigma_{22.1}^{-1} \right) D \tag{A.11}$$

$$\begin{aligned} T^{-1} \frac{\partial^2 L}{\partial \sigma_{22.1} \partial \sigma'_{22.1}} = {} & \frac{1}{2} D' \left(\Sigma_{22.1}^{-1} \otimes \Sigma_{22.1}^{-1} \right) D + \\ & - \frac{1}{2} D' \{ [(\Sigma_{22.1} \; (V'V/T)^{-1} \Sigma_{22.1})^{-1} \Sigma_{22.1} (V'V/T)^{-1}] \otimes \\ & \otimes [\Sigma_{22.1} \; (V'V/T)^{-1} \Sigma_{22.1}]^{-1} \} D + \\ & + \frac{1}{2} D' \{ [\Sigma_{22.1} \; (V'V/T) \; \Sigma_{22.1}]^{-1} \otimes \\ & \otimes [(\Sigma_{22.1} (V'V/T)^{-1} \Sigma_{22.1})^{-1} \Sigma_{22.1} \; (V'V/T)^{-1}] \} D \end{aligned} \tag{A.12}$$

To obtain the information matrix (2.22), computed as minus the limit in probability of the hessian matrix, the following plim's are useful:

$$\text{plim } T^{-1} \varepsilon' \varepsilon = \sigma_{11}$$

$$\text{plim } T^{-1} V'V = \Sigma_{22.1}$$

$$\text{plim } T^{-1} V' \varepsilon = 0$$

$$\text{plim } T^{-1} X' \varepsilon = 0$$

$$\text{plim } T^{-1} X'V = 0$$

$$\begin{aligned} \text{plim } T^{-1} Y_2' Y_2 & = \text{plim } T^{-1} (X \Pi_2 + \varepsilon \lambda' + V)' (X \Pi_2 + \varepsilon \lambda' + V) \\ & = \Pi_2' \text{ plim } (X'X/T) \Pi_2 + \sigma_{11} \lambda \lambda' + \Sigma_{22.1} \end{aligned}$$

$$\text{plim } T^{-1} Y_2' Z = \Pi_2' \text{ plim } (X'Z/T)$$

Then:

$$\text{plim}(W'W/T) = \begin{bmatrix} \Pi_2' \text{ plim } (X'X/T) \Pi_2 + \sigma_{11} \lambda \lambda' + \Sigma_{22.1} & \\ \text{plim } (Z'X/T) \Pi_2 & \text{plim}(Z'Z/T) \end{bmatrix}$$

$$\text{plim } T^{-1}\, \epsilon' W = [\sigma_{11}\lambda' \quad 0']$$

$$\text{plim } T^{-1}\, V' W = [\, \Sigma_{22\bullet 1} \quad 0 \,]$$

$$\text{plim } T^{-1}\, X' W = [\, \text{plim } T^{-1}\, X' X\, \Pi_2 \quad \text{plim } T^{-1}\, X' Z]$$