A Test of the Exogeneity of National Variables in a Regional Econometric Model

Paul A. Anderson

Revised June 1979

Working Paper #: 124

PACS File #: 2930

Abstract

Many regional econometric models are estimated under the maintained assumption that certain national variables are exogenous with respect to the regional variables in the models. This exogeneity assumption is testable using time series methods of inference, yet, to my knowledge, no regional model has been so tested. In this paper, I test the national exogeneity assumption included in the specification of a particular regional forecasting model. Such a test is, I believe, a necessary and important step in the construction of any econometric model.

In this paper, time series inference is used to test the hypothesis that several national variables are exogenous in a particular regional econometric model. Many regional models are estimated under a similar exogeneity assumption, but, to my knowledge, no tests of this hypothesis have been performed. Thus, this method (described in Geweke [1978]) should be of interest to many regional modelers. Exogeneity testing is, I believe, a necessary and important step in the construction of any econometric model.

Exogeneity in Regional Models

Many regional models include the specification that certain national factors may be treated as exogenous to the region being modeled. The list of exogenous variables for a given model often will include GNP and national wages, as in Glickman [1971]; or national manufacturing output, as in L'Esperance [1977]; or national consumer prices, as in Crow [1973]. Other examples of regional models which include such a specification can be found among the references of the Klein-Glickman survey article [1977].

This practice of assuming national factors to be exogenous is likely to continue. For one thing, it is a convenient way to produce detailed regional forecasts using regional satellite models which operate on a common national forecast. One example of such a satellite model is the Adams, Brooking, and Glickman model of Mississippi [1975], which was designed to use forecasts of national variables generated by the Wharton Annual and Industry model.

Although many regional models include the national exogeneity specification, careful modelers recognize the possibility that such models may miss important feedback from regional economies to the national economy. In at least one study, investigators have sought to deal with this possibility. Saltzman and Chi [1977] estimate an "integrated" regional-national framework for New York state and the U.S. but perform no test of the feedback hypothesis.

It would be extremely useful to be able to test the national exogeneity assumption included in a particular regional model. The effectiveness of a regional model for either policy analysis or forecasting is impaired if its specification is at variance with the data. Thus, testing the assumed exogeneity of these national variables is an important step, both in sorting through classes of models and in validating a particular model.

Exogeneity Testing

The test which can be applied to the exogeneity assumption in regional models is the time-series test described by Sims [1972], [1975] and clarified and extended most recently by Geweke [1978]. The test grows out of the theory of covariance stationary stochastic processes. In this framework it can be shown that the assumption that one block of variables is exogenous with respect to another block has very explicit implications for time-series data.

Using the notation in Geweke [1978], let us assume that $\{z_t\}$ is a covariance stationary vector stochastic process which has an autoregressive representation.[1/] Covariance stationarity implies that Ez_t and $\mathrm{cov}(z_t, z_{t+s})$ for all s do not depend on t. If $\{z_t\}$ has an autoregressive representation, then it can be expressed as

$$(1) \qquad \underset{(m\times m)}{J(L)} \ \underset{(m\times 1)}{z_t} = \underset{(m\times 1)}{e_t}$$

where J(L) is a matrix of polynomials in the nonnegative powers of the lag operator L (defined by $L^s w_t = w_{t-s}$ for any time series $\{w_t\}$) and e_t is serially uncorrelated. In what follows we will assume that the endogenous and exogenous variables of an econometric model form a covariance stationary vector stochastic process with an autoregressive representation.

The specification of an econometric model includes the partitioning of the variables into endogenous and exogenous vectors. The complete dynamic simultaneous equation model may then be expressed as

(2) $$\underset{g\times g}{A(L)}\ \underset{g\times 1}{y_t} + \underset{g\times k}{B(L)}\ \underset{k\times 1}{x_t} = \underset{g\times 1}{e_t}$$

where y_t and x_t are the endogenous and exogenous variables, respectively. In this specification it is assumed that $\{e_t\}$ is serially uncorrelated[2/] and that

(3) $$\operatorname{cov}(e_t, x_{t-s}) = 0 \qquad \text{for all } s \geq 0$$

(4) $$\operatorname{cov}(e_t, y_{t-s}) = 0 \qquad \text{for all } s > 0$$

The only distinction between the endogenous variables, y_t, and the exogenous, x_t, is that current values of x_t are assumed orthogonal to current errors. That is, equation (3) holds for s=0 but equation (4) does not. This distinction places an important, testable restriction on the data record of y_t and x_t. It turns out that conditions (3) and (4) can hold if, and <u>only</u> if, the stochastic process $\{z_t\}$ made up of y_t and x_t has a triangular autoregressive representation.

More explicitly, it must be possible to represent the z_t in the form

(5) $$\left(\begin{array}{c|c} \underset{g\times g}{C(L)} & \underset{g\times k}{D(L)} \\ \hline \underset{k\times g}{0} & \underset{k\times k}{E(L)} \end{array}\right) \begin{pmatrix} \underset{g\times 1}{y_t} \\ \underset{k\times 1}{x_t} \end{pmatrix} = \begin{pmatrix} \underset{g\times 1}{u_t} \\ \underset{k\times 1}{v_t} \end{pmatrix}$$

or equivalently

(6) $$C(L)y_t + D(L)x_t = u_t$$

(7) $$E(L)x_t = v_t$$

where u_t and v_t are independent and serially uncorrelated.

There are two methods of testing whether the $\{z_t\}$ process has a triangular autoregressive representation, testing two roughly equivalent implications of the null hypothesis that x_t is exogenous in system (2). Inference procedures for both methods are explained in Geweke [1978]. In this paper, we will implement the test based on Theorem 2 in Geweke's paper.

The test we use is just a multivariate extension of the testing procedures suggested by Granger [1969]. In essence, we estimate the equations in system (7) above and test the null hypothesis that no lagged values of y_t enter the regression. Specifically, we estimate regression equations of the form

$$(8) \qquad x_{it} = \sum_{j=1}^{k} \sum_{s=1}^{m} b_{ijs} x_{jt-s} + \sum_{j=1}^{2} \sum_{s=1}^{n} c_{ijs} y_{jt-s} + u_{it} \qquad i = 1, \ldots, k$$

and test the null hypothesis that all of the c_{ijs}'s are zero.[3/]

Together the k equations comprise a system of equations which can be written as

$$(9) \qquad x_t = B(L)x_{t-1} + C(L)y_{t-1} + u_t.$$

The parameters of the equations in system (9) can be estimated by single-equation OLS methods. It is then possible to perform a likelihood ratio test of the hypothesis that $C(L) \equiv 0$. Under the null $L = T|\det\sum_1 - \det\sum_2|$ is asymptotically distributed as a $X^2(gkn)$ where $\sum_1$ and $\sum_2$ are estimates of the covariance matrix of u_t, one formed from estimated residuals of system (9), the other formed from residuals of (9) estimated under the constraint that $C(L) \equiv 0$.

The small sample properties of this test statistic are not known, but it is likely that when the number of parameters being estimated is large relative to the number of observations, this test is biased in favor of rejecting the null hypothesis. As an attempt to account for this, Sims [1979] has used L =

$(T-q)|\det\sum_1-\det\sum_2|$ where q is equal to the number of parameters being estimated in each equation of (9). This statistic has the same asymptotic distribution as L, but there is no formal evidence on its small sample properties. In the test results that follow, we report this statistic in an attempt to adjust for the number of effective observations in the sample.

An Application

In this section, we present the results of an exogeneity test applied to a specific regional model. The test results, in this case, support the assumption that certain national variables may be treated as econometrically exogenous in this particular model. The example presented here is doubly useful, because it illustrates not only how exogeneity tests can be performed, but also at what point in the model-building process such tests should be used.

The test was applied to a small quarterly forecasting model of the Ninth Federal Reserve District being developed by the Federal Reserve Bank of Minneapolis. At the time the test was performed, the model had not yet been fully specified, let alone estimated. The specification consisted of a list of economic variables, some designated as exogenous and some endogenous. This is all the information necessary to carry out a test of the exogeneity hypothesis being put forward.

Conducting exogeneity tests at such an early stage is an important step in building a model for a couple of reasons. For one thing, if the model is estimated on the basis of an exogeneity assumption which can be contradicted by the data, simultaneous equation bias will be present and the model, so estimated, will not be an accurate representation of the true structure. Secondly, conducting the test at this stage may simplify the task of sifting through different possible specifications by eliminating a whole class of models (those containing the assumption) from consideration.

The hypothesis being tested in our example was the assumption that a block of five national variables was exogenous with respect to a block of four regional variables. The five national variables were real GNP, employment, labor force, the Consumer Price Index, and the mortgage interest rate. The four regional variables were regional employment, regional labor force, the value of new construction contract awards in the region, and the CPI for the Minneapolis-St. Paul SMSA. The data consisted of quarterly observations from 1960 through 1978.

The actual test involved calculating OLS estimates of the parameters in five regression equations of the form

$$n_{it} = \sum_{j=1}^{5} \sum_{s=1}^{4} b_{ijs} n_{jt-s} + \sum_{j=1}^{4} \sum_{s=1}^{4} c_{ijs} r_{jt-s} + e_t \tag{10}$$

where n_j's are national variables and r_j's are regional variables. Under the null hypothesis that the national variables are exogenous to the regional model, all of the c's should be equal to zero.

The test statistics reported in Table I are of two types. First, there are single-equation statistics, the results of tests of the null hypothesis that the c's in a particular equation are zero. These test statistics should have an F-distribution under the null hypothesis. Table I also includes system test statistics. These result from testing the joint null hypothesis that the c's are zero in all five equations.

Table I

Tests of Exogeneity of National Variables

Single-Equation Tests

Dependent Variable	F-Statistic	Significance Level	Accept or Reject
U.S. Labor Force	.816	.659	Accept
U.S. Employment	1.638	.117	Accept
U.S. Consumer Price Index	1.429	.192	Accept
U.S. Mortgage Interest Rate	.623	.841	Accept
Real GNP	.747	.728	Accept

$$F_{.95}(16,31) = 1.93$$

Multiequation Test

	Likelihood Ratio	Significance Level	Accept or Reject
Overall	L = 178.87	.000	Reject
Overall (with T-q correction)	L = 91.90	.190	Accept

$$\chi^2_{.95}(80) = 101.87$$

None of the single-equation tests calls for rejection of the hypothesis of exogeneity of the national variables at the 95 percent confidence level. For all five national variables the F-statistic is well below the critical level of 1.93 for the degrees of freedom in this example. The test comes closest to rejecting the exogeneity of U.S. employment.[4/] Here, as the marginal significance level indicates, if the null hypothesis were true, there is almost a 12 percent chance that the test statistic from a random sample could be greater than the level achieved in our test. This is quite a bit larger than the 5 percent chance used as the standard in most hypothesis tests.

The multiequation test result turns out to be dependent on the "effective observations" adjustment mentioned above. Table I includes two likelihood ratios, both of which are asymptotically distributed as $\chi^2(80)$ under the null hypothesis. The second statistic is adjusted, as in Sims [1979], for the fact that the number of parameters in the system is large relative to the number of observations. This adjusted statistic does not imply rejection of the exogeneity hypothesis, but the unadjusted statistic does call for rejection. Sims' guess that the adjusted ratio has better small sample properties is not based on any formal evidence. Nevertheless, the agreement of this version of the system test and the single-equation tests is a point in its favor.

The results of these tests were, on the whole, quite favorable to the hypothesis that these national variables are exogenous in this regional model. In this case, the specification of national exogeneity was consistent with the data, and modeling proceeded on that assumption.

Conclusion

The exogeneity assumptions embodied in many regional econometric models can and should be subjected to formal testing. We have provided an example of how a common assumption, the exogeneity of national variables, may be tested using existing time-series techniques. The common use of such tests in regional model building should result in more reliable (and useful) regional models.

References

[1] Adams, F., C. Brooking, and N. Glickman. 1975, "On the Specification and Simulation of a Regional Econometric Model: A Model of Mississippi," Review of Economics and Statistics 57, 286-298.

[2] Crow, R. 1973, "A Nationally-linked Regional Econometric Model," Journal of Regional Science 13, 187-204.

[3] Geweke, J. 1978, "Testing the Exogeneity Specification in the Complete Dynamic Simultaneous Equation Model," Journal of Econometrics 7, 163-185.

[4] Glickman, N. 1971, "An Econometric Model of the Philadelphia Region," Journal of Regional Science 11, 15-32.

[5] Granger, C. W. J. 1969, "Investigating Causal Relations By Econometric Models and Cross Spectral Methods," Econometrica 37, 428-439.

[6] Hall, O., and P. Licari. 1974, "Building Small Region Econometric Models: Extension of Glickman's Structure to Los Angeles," Journal of Regional Science 14, 337-354.

[7] Klein, L., and N. Glickman. 1977, "Econometric Model-Building at Regional Level," Regional Science and Urban Economics 7, 3-24.

[8] L'Esperance, W. 1977, "Optimal Stabilization Policy at the Regional Level," Regional Science and Urban Economics 7, 25-48.

[9] Saltzman, S., and H. Chi. 1977, "An Exploratory Monthly Integrated Regional/National Econometric Model," Regional Science and Urban Economies 7, 49-81.

[10] Sims, C. 1972, "Money, Income, and Causality," American Economic Review 62, 540-552.

[11] ______________. 1975, "Exogeneity Tests and Multivariate Time Series: Part I," Discussion paper 75-54, University of Minnesota, Center for Economic Research.

[12] ______________. 1977, "Exogeneity and Causal Ordering in Macroeconomic Models," in C. Sims, ed., New Methods in Business Cycle Research: Proceedings from a Conference, Federal Reserve Bank of Minneapolis, 23-44.

[13] ______________. 1979, "Macroeconomics and Reality," Econometrica, forthcoming.

Footnotes

1/ This is a sufficient, but not a necessary, condition for the results to hold. See Geweke [1978] for weaker conditions.

2/ If $\{e_t\}$ is serially correlated, the system can be expressed in this form provided the $\{e_t\}$ has an autoregressive representation, i.e.,

$$C(L)e_t = v_t$$

where v_t is serially uncorrelated. Then premultiplication of system (2) yields

$$C(L)A(L)y_t + C(L)B(L)x_t = v_t$$

which has the same form as (2).

3/ It may appear that regression equation (8) is not identified; that is, that it could be from system (7) in which case we would expect nonzero coefficients on lags of y_t. However, under the null hypothesis that x_t is exogenous, regression equation (8) is identified by the exclusion of all current values of x_t's (except, of course, the dependent variable). Thus, if the test rejects the exclusion of lagged y_t's from (8), it will be because $\{z_t\}$ does not possess a triangular autoregressive representation.

4/ It should be noted that since U.S. employment is the sum of employment in the different regions, it is not truly exogenous with respect to regional employment. In such a circumstance, failure to reject the null hypothesis can be taken as evidence that a model which assumes the national variable to be exogenous is not subject to significant simultaneous equation bias because of that assumption.