TESTING FOR INTRAREGIONAL ECONOMIC HOMOGENEITY: A COMMENT*

Peter M. Schulze†

1. INTRODUCTION

In a recent issue of this journal, Johnson [6] presented an interesting note on how to use the regression coefficients b of Zellner's [12] method of estimating seemingly unrelated regressions in testing for regional economic homogeneity. Here the question of appropriate restrictions on the variance-covariance matrix of the regression disturbances is of special interest and significance in connection with pooling regionally distributed and time series data, as in the case of observations for a number of regions over several periods of time. The behavior of the disturbances over the areal units is likely to be different from the behavior of the disturbances of a given areal unit over time. Various kinds of specifications with respect to the disturbances will lead to various kinds of restrictions on the variance-covariance matrix.

This comment does not intend to discuss the wide problem of regional homogeneity but centers its interest on some remarks resulting from the spatial (and temporal) autocorrelation of the regression disturbances. Furthermore, alternative procedures to the method presented by Johnson of estimating b will be dealt with. Thus, this comment is meant to be a supplement to Johnson's note.

In general, in econometric literature the problem of spatial autocorrelation has been neglected. Although the tests for temporal (aspatial) autocorrelation are well known, tests for spatial autocorrelation in the regression disturbances are not yet discussed in econometric textbooks. Therefore this problem is outlined in Section 2. An estimation procedure which separates the effects of varying parameters b over time and space is introduced in Section 3. Here the spatial and temporal autocorrelation must be considered and tested for explicitly. Joint estimation procedures are discussed with regard to Zellner's approach in Section 4. Some general conclusions are presented in Section 5.

The methods outlined are not new but may not be familiar to regional economists in general. No distinction is made between intra- and interregional considerations. The term region is used for any areal unit where statistical data are available.

2. THE PROBLEM OF SPATIAL AUTOCORRELATION

Spatial autocorrelation for a system of regions may be defined as follows (Cliff and Ord [1, p. 26]): Assume that the x_i in each of the N regions are separate

* I am indebted to an anonymous referee for helpful comments.

† Professor, Institute of Statistics and Econometrics, Johannes Gutenberg University, Mainz.

Date received: March, 1976; revised, November, 1976.

observations of X and that each x_i is drawn from the same population. If every pair of x_i is uncorrelated, then the data are said to lack spatial autocorrelation. Conversely, if the x_i are not pairwise uncorrelated then the data are said to be spatially autocorrelated.

The problem of spatial autocorrelation is different from that of temporal autocorrelation because the variable of a time series is only influenced by past values while in a spatial connection, dependence can exist in all directions. As a study of the literature shows, several attempts have been made to develop a statistic which tests whether or not geographically distributed data are spatially autocorrelated. The statistics given by Moran [8] and Geary [5] test for autocorrelation between the x_i in contiguous nonvacant areal units in a system of regions. These measures show some deficiencies. First, they are invariant over certain topological transformations of the underlying regional structure, and second, they can only be used to test for first-order correlations between x_i. This means that they only test for correlation between the x_i in contiguous regions. Thus spatial autocorrelation between each region and the second- and higher-order contiguous regions would be undetected by these statistics. These limitations can be eliminated by weighting. Cliff and Ord [1, 3] have presented a statistic which employs a general set of weights to overcome the problem of invariance. The set of weights also ensures that the statistic can be used to test for any order of spatial autocorrelation between x_i in a system of regions.

In connection with the problem of testing the regression residuals of different regions for autocorrelation, the difficulty arises that in any practical case the population disturbances will be unknown and the corresponding test statistic must be based upon the residuals of the calculated regression. The problems involved cannot be discussed here.[1]

If there are regression results from regionally distributed data, the disturbance terms may be tested by one of the statistics mentioned above. Yet it should be emphasized that such a test is no more than an indicator as to whether the population disturbance terms are spatially autocorrelated or not. No precise answers about significance levels can be made.

3. SEPARATION OF REGIONALLY DISTRIBUTED AND TIME SERIES DATA

The problem of spatial autocorrelation will be presented in context with Johnson's note to test the b's for regional economic homogeneity. In doing so we follow Cliff and Ord [2] in their approach which varies parameters separately over time and over space. First, assume

$$Y_{it} = b_{it,1}X_{it,1} + b_{it,2}X_{it,2} + \cdots + b_{it,k}X_{it,k} + u_{it} \qquad (i = 1, \cdots, N; \quad t = 1, \cdots, T) \tag{1}$$

where i refers to the regionally distributed observational units and t is the index for time series observations. There are altogether NT observations. Y_{it} is an observation on the dependent variable for region i in time period t, $X_{it,h}$ is the ob-

[1] For an extensive discussion see Cliff and Ord [3, 4].

servation on the independent variable for region i in time period t, $b_{it,h}$ are regression coefficients to be estimated ($h = 1, \cdots, k$) and u_{it} are the regression disturbances. In matrix notation the regression equation (1) can be written as

$$Y = Xb + u \tag{1'}$$

The central question is now whether the N observational regions are in fact homogeneous with respect to the parameter vector b (Johnson [6, p. 366]). In examining this question we proceed in two separate steps: (1) the parameters b_{it} are varying only over time, and (2) the parameters b_{it} are varying only over space. Johnson [6, p. 366] assumes that the b's are constant over time. But principally they may vary over time and may be assumed to be invariant over the N regions, that is

$$b_{1t} = b_{2t} = \cdots = b_{Nt} = b_t \tag{2}$$

If we further assume that all b_t are different, and for all t the expectation is $E(u_t) = 0$ and the variance-covariance matrix

$$E(u_{it}u_{js}') = \begin{cases} \sigma^2 I & \text{when } i = j, \quad s = t \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

where I is the identity matrix, then the least squares estimator for b_t follows in matrix notation with

$$\hat{b}_t = (X_t'X_t)^{-1}X_t'Y_t \tag{4}$$

That is, the estimation is carried out separately for each time period.

Aitken's generalized least squares estimators must be used, however, if for example the following more general variance-covariance structure is assumed

$$E(u_tu_s') = \begin{cases} \sigma^2\Omega_t & \text{when } t = s \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

where Ω is a positive-definite matrix. Then the *GLS* estimators

$$\tilde{b}_t = (X_t'\Omega_t^{-1}X_t)^{-1}(X_t'\Omega_t^{-1}Y_t) \tag{6}$$

for b_t are obtained.[2]

Before we can presume constant b_t over time, we have to test for spatial autocorrelation among the regression disturbances. We have to show that the important assumption (3) is met. The statistic suggested for example by Cliff and Ord [3, 4] may be used for this procedure.

If this test of spatial autocorrelation is negative, the test of the null hypothesis H_0: $b_t = b$ for all t against general alternatives follows. If H_0 is accepted, Johnson's assumption is met and we can proceed to allow the b_{it} to vary over the N regions, that is

$$b_{i1} = b_{i2} = \cdots = b_{iT} \tag{7}$$

[2] If the variances and covariances of the regression disturbances are not known, as is generally the case, Ω_t has to be estimated from the data. Then, $\tilde{b}_t$ will be a consistent but not an unbiased estimator of b_t (Cliff and Ord [2, p. 53]).

Because of the duality between variations in the b_{it} over time and over space, model (1) permits regional variations in the b_{it} as well as longitudinal variations and the spatial counterparts of (4) and (6) for the estimation of b_i must be used. We first test for temporal autocorrelation among the regression disturbances, using for example the Durbin-Watson test. If no temporal autocorrelation is indicated, we test H_0: $b_i = b$ for all i against general alternatives. This is presented by Johnson.

It should be mentioned that we have proposed a fixed-coefficients approach whereas Johnson in the second half of his note used Swamy's [11] random-coefficient model.[3]

4. POOLING OF REGIONALLY DISTRIBUTED AND TIME SERIES DATA

The most questionable assumption in the preceding approach is that the regions are mutually independent, i.e., that the data lack spatial autocorrelation. If there is spatial autocorrelation, then the method of joint estimation developed by Zellner [12] can, as proposed by Johnson, be used. Zellner's covariance specification allows for spatial autocorrelation among the regression disturbances

$$E(u_{it}u_{jt}) = \sigma_{ij} \qquad (i \neq j) \tag{8}$$

but assumes that there is no temporal autocorrelation

$$E(u_{it}u_{jt'}) = 0 \qquad (t \neq t') \tag{9}$$

Because some constraints must be placed on the form of Ω in (5) if it is to be estimated, Zellner's specification is one way of achieving this end.

Parks [10] considers a more general covariance specification of Ω where the disturbance terms of the system of regression equations are assumed to be related by both temporal and spatial autocorrelation. The serial correlation is seen as a first-order autoregressive process. Parks presents a consistent and asymptotically efficient three-step estimation for the regression coefficient b_{it}[4] which can be used in testing regional homogeneity. Another regression estimation method should be mentioned here. Ord's [9] evaluation of this subject can be seen as a step to take explicitly into account cross-sectional and timewise influence on data. These models attempt to describe the interactions between regions by introducing weights representing the degree of possible interaction of region j on region i.

5. CONCLUSION

As shown above, there are several regression techniques discussed in the literature to deal with the problem when regionally distributed and time series data are available. In this connection we were especially interested in the problem

[3] This is an acceptable procedure if no distributional assumptions are made, whereas distributional assumptions for hypothesis testing cause different implications of the random model. Thus, Johnson's implied assumptions of normality for the coefficients of the six states of New England seems open to criticism.

[4] Kmenta and Gilbert [7] have developed several alternative estimators and compare their small sample efficiency.

of spatial autocorrelation of regression disturbances. When using data varying over time and over space, the possibilities of separating and pooling the observations must be distinguished.

The ideas, only roughly sketched, should show the different ways of testing regional homogeneity when the regression coefficient b is used as the measure. Unfortunately, this comment could not be supported by a reanalysis of Johnson's example since the necessary data were not available. It was the purpose of this comment to give, as a supplement to Johnson's note, some further hints on this problem which is discussed in econometric literature and not necessarily known to those interested in problems of regional economics.

REFERENCES

[1] Cliff, A. D. and J. K. Ord. "The Problem of Spatial Autocorrelation," in A. J. Scott (ed.), *Studies in Regional Science*. London: Pion 1969.
[2] ——— and ———. "A Regression Approach to Univariate Spatial Forecasting," in M. Chisholm, *et al.* (eds.), *Regional Forecasting*. London: Butterworths, 1970.
[3] ——— and ———. "Testing for Spatial Autocorrelation Among Regression Residuals," *Geographical Analysis*, 4 (1972), 267–284.
[4] ——— and ———. *Spatial Autocorrelation*. London: Pion, 1973.
[5] Geary, R. C. "The Contiguity Ratio and Statistical Mapping," *The Incorporated Statistician*, 5 (1954), 115–145.
[6] Johnson, L. W. Jr. "A Note on Testing for Intraregional Economic Homogeneity," *Journal of Regional Science*, 15 (1975), 365–369.
[7] Kmenta, J. and R. F. Gilbert. "Estimation of Seemingly Unrelated Regressions with Autoregressive Disturbances," *Journal of the American Statistical Association*, 65 (1970), 186–197.
[8] Moran, P. A. P. "A Test for the Serial Independence of Residuals," *Biometrika*, 37 (1950), 171–181.
[9] Ord, J. K. "Estimation Methods for Models of Spatial Interaction," *Journal of the American Statistical Association*, 70 (1975), 120–126.
[10] Parks, R. W. "Efficient Estimation of a System of Regression Equations when Disturbances are Both Serially and Contemporaneously Correlated," *Journal of the American Statistical Association*, 62 (1967), 500–509.
[11] Swamy, P. A. V. B. *Statistical Inference in Random Coefficient Regression Models*. Berlin/Heidelberg/New York: Springer, 1971.
[12] Zellner, A. "An Efficient Method for Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias," *Journal of the American Statistical Association*, 57 (1962), 348–368.