DYNAMIC ECONOMETRIC MODELS
Vol. 6 – Nicolaus Copernicus University – Toruń – 2004

Elżbieta Szulc
Nicolaus Copernicus University in Toruń

The Structure of Interdependence in Dynamic Spatial Models. Remarks on Modelling and Interpretation

1. Introduction

The classical assumption of independence between observations in the analyses based on cross-sectional data and on time series of cross-sectional data often proves inconsistent with the economic theory and with the observation of interdependence of phenomena in different sections.

The dependence is generally taken into account in dynamic econometrics. Some forms of the dependence for reason of a natural ordering and structure in time series data are considered.

Although such a natural ordering in the cross-sectional or spatial data is absent, one tries to discover in them the dependence structures similar to those ones which are observed in the time series.

The dependence in the spatial data, contrary to the independence of the observations, is expressed in the so-called First Law of Geography: "*everything is related to everything else, but near things are more related than distant things*"[1].

Therefore, following the example of the time autocorrelation, the notions of the spatial autocorrelation and of the space-time autocorrelation appear[2]. Then the notion of the cross-sectional autocorrelation[3] seems to naturally complete the notions above.

Economic objects, such as: households, firms, regions, countries etc. act in such a way, that the specified dependence among them occurs. The source of

[1] See: Tobler (1970).
[2] See: Cliff and Ord (1973).
[3] Such a notion is not usually used. One rather uses the notion: „spatial autocorrelation", meaning the space in a wide sense, not limited to the geographical expression.

this dependence may be variables, directly unobserved, which are cross-sectional or spatial correlated and on account of this they produce correlations in disturbances in the equations describing economic behaviour.

For example, the demand of households for some goods may be correlated with the demand of other households (neighbours) for the same goods. This correlation may be connected with the spatial correlation in such variables, as e.g. availability of substitute goods, climate, air and soil quality. Moreover, the households may obtain utility in consuming the goods similar to those ones consumed by their neighbours (the so-called imitation).

Another example, which refers to spatial (in a geographical sense) units, is a magnitude of unemployment on the territory of a country, divided into smaller administrative units. The spatial correlation between the units should be expected, considering effects of many economic factors on the unemployment. These factors, determining an economic potential of the units, influence migrations of people.

One more example is taken from the literature on economic growth. Investigating correlations between the GNP growth rates and the factors which explain them across countries, the correlations between disturbances for different countries in adequate regression models are observed.

The independence assumptions which prevail in cross-sectional econometrics facilitate the estimation and the inference but in practice they may not be satisfied. The assumptions on the nature of the dependence between observations are the alternative to independence. This approach needs appropriate assumptions on the distributions of the data and error terms.

Various stochastic and dynamic specifications of combined both cross-sectional and time dependence are possible. The purpose of one of them is to establish peculiarities of disturbances, e.g. in the spatial and space-time regression models. It is not less important to discover the spatial connections in the real processes, which takes to the models of the spatial interactions. The natural extension of the spatial models are the so-called dynamic spatial models, which are necessary for understanding spatial patterns of behaviour, the structure of connections, their changes in time etc.

To establish the form of the dependence in a case of the cross-sectional data is more difficult than in the analyses of time series, where the integer (or natural) indexing of the observations gives the natural order and structure to the data.

The purpose of the paper is to discuss the possibility and utility of: (1) a settlement of the cross-sectional data structure, which would be similar to the time series structure, (2) including such structure into dynamic econometric models as well.

The remainder of the paper consists of five sections and conclusions. In section 2 some exemplifying measures of the economic distance are mentioned. In the next two sections there is accurately defined the assumption that the cross-sectional data like spatial data are modelled as the realizations of the two-dimensional random field. Section 5 is on spatial dependence modelling. The attention is paid to the fact that the dynamic aspect should be introduced into the analysis. The remarks on the linear dynamic econometric modelling of the dependence between the spatial processes are contained in section 6. And at the end of the paper there are some conclusions from the presented discussion.

2. Economic distance as an alternative to physical distance[4]

The data structure in the time series is simple, because it is formed by the sequences of the observation of a phenomenon in the succession moments or periods, and the distance between the observations is the distance between the points on the time axis.

The cross-sectional data structure is created considering the distance between units and it is not necessary to understand this distance as physical. However, in many situations the use of the physical distance could be justified, as in the examples above.

Probably the correlation in the demand of households will relate to the households located nearby in a physical sense. But in a different approach, the distance between the households which are spatially localized may be expressed in terms of time and monetary costs of travelling between points to use the local public goods[5].

With regard to the limited labour mobility, the autocorrelation in the unemployment process may be considered in a strict spatial expression. This means that in this case the distance in a geographical sense will be used. However the verification of the dependence by using an additional information, e.g. on the economic condition of the spatial units (regions) is not excluded. In this way the economic distance on the basis of some economic characteristics is constructed.

In turn, for firms the measures of the overlap in their retail markets might be suitable measures of the economic distance. The plausible measures of the economic distance between two countries, such as trade volumes or transportation costs of a physical or human capital are proposed[6].

[4] In spatial and space-time models the spatial locations of units and the settlement of the „neighbouring” connections are important. The identification of neighbours may be based on the “geography” of their location, and also, e.g. on the demographic characteristics of the neighbours, of their income condition or on the combination of all the features. See: e.g. Case (1991). These characteristics may be used for constructing measures of the economic distance between the unites which are spatially localized. The economic distance will be important for investigating the correlations between the cross-sectional units, such as: firms, households, consumers and also the economic sectors or countries. On defining the economic distance in the above sense, see: e.g. Chen and Conley (2001) and the references there.

[5] See: Conley (1999).

[6] See: Conley op. cit, also Chen and Conley, op. cit.

Each of the mentioned measures of the distance is reasonable in the context of the concrete application. These measures neither are perfect nor universal. Especially in the analysis of the cross-sectional data it is not enough to refer to the physical distance only.

The notion of the economic distance is a generalization confirmed by the statement, that the course of economic phenomena is influenced by the factors which expressed themselves with the location of units in the multidimensional space of features. The locations of units under investigation are represented by the appropriate "distance". The cross-sectional data analyzed with regard to the distances allows to define the structure of spatial connections, which generate the correlation in a spatial expression.

3. Model of a two-dimensional random field[7]

It is assumed that the population of individuals resides in the Euclidean space R^2. Each individual is located at point p of the space. The population of potentially observed locations forms the lattice H, which in a general case is irregular. With each position p in the space R^2 a vector of the random variables X_p is associated. The variable X_p as a function of the p is called a random field.

The set of the individuals i under econometric modelling creates a random sample, which is taken from the population. Each individual i is located at the point p_i. The collection of locations $\{p_i\}$ creates a sample region. Then it may be assumed that each axis of co-ordinates of the considered space is integer co-ordinate and H is regular. The observations of the variable X on the lattice H are the realization of a random process – a random field.

An econometric sample consists of two parts. The first part is the mentioned realization of the random field X_{p_i} at all points p_i inside the sample region. The second part of the data is the symmetric matrix D, whose elements d_{ij} are measures of the distance between the indexes p_i and p_j. Assuming that N locations p_i are separated, $N\times N$ matrix D is obtained.

The dependence between the individuals i is considered with respect to the distances between the positions of the individuals and the distances are determined by an economic metric. The distance function is understood as follows. If two locations p_i and p_j of the individuals i and j are close, then X_{p_i} and X_{p_j} may be strongly correlated. On the contrary, the correlations between X_{p_i} and X_{p_j} decline with the increasing remoteness in the economic space.

[7] The random field structures may be used for describing the dependence in the cross-sectional (or spatial) data. See: Conley, op. cit.

Doubly-indexed variables X unlike the scalar-indexed variables (in the time series analysis) are also needed to make possible in the investigation taking into account the multidirectional nature of the dependence in the cross-sectional data.

4. Data generation[8]

Assuming that the measurements of the economic distance between the individuals are exact, the location of the units on the plane may be established by the multidimensional scaling, which makes possible to get the spatial pattern (a configuration of points in R^2) in a situation when the order of inter-point distances in the distance matrix D is known. It is also assumed that the data generating process is described by the regular lattice, and then X_{p_i} is subordinated to the random field X_p with the index $p \in Z^2$, where Z denotes the set of integers.

In a situation, when the observations derive from an irregular lattice, a regular square lattice may be constructed in the following way. Space R^2 is divided into squares with a diagonal which is not bigger than the minimal distance d_0 between the individuals located in R^2. Since the distance of any point from each other point is at least d_0, there will be at most one point within each square of the constructed lattice. Two integer co-ordinates p_1 and p_2, i.e. $p=[p_1, p_2]$ may be assigned to each square. The dimensions of the squares are normalized to 1×1 and p denotes the co-ordinates of the left lower corner of each square.

5. Modelling of spatial dependence

The establishment of the inter-locations of the individuals which is based on the economic distance leads to the particular spatial structure of the dependence, where the neighbours of a given individual are the „similar” individuals, which do not have to be close in a geographical sense. Just as in spatial econometrics the spatial lag operator $L^{(s)}$ is defined in the following way[9]:

$$L^{(s)} X_{p_i} = \sum_{p_j} w_{ij}^{(s)} X_{p_j} \,, \tag{1}$$

where:

$$w_{ij}^{(s)} = \frac{m_{ij}}{\sum m_{ij}} \,,$$

and

---

[8] See: Conley, op. cit.
[9] Comp. Anselin (1988).

$$m_{ij} = \begin{cases} 1, & if \quad j \in J_s \\ 0, & if \quad j \notin J_s \end{cases},$$

J_s – a set of neighbours of order s,

which then may be used for defining the spatial autoregressive model, i.e.:

$$\left(1 - \sum_{s=0}^{l} \alpha_s L^{(s)}\right) X_p = \varepsilon_p \ , \tag{2}$$

where ε_p is a spatial white noise.

The effect of instantaneous spatial influences contained in the model (2) practically occurs when the temporal distance of the realizations of the dependence is shorter than the frequency of the observations of the phenomena in time. However, it is usually assumed that at least one period is needed for realizing themselves the spatial influences. In other words, the individuals interact over time and not instantaneously. In such situations purely spatial models such as (2) are not sufficient and it is necessary to refer to the space-time models.

Thus, using the space-time lag operator $L^{(sh)}$, which combines the space shifts and time lags (backward shifts) of the variable, i.e.:

$$L^{(sh)} X_{p_i t} = \sum_{p_j} w_{ij}^{(s)} X_{p_j - h} \ , \tag{3}$$

the space-time autoregressive model[10]

$$\left(1 - \sum_{s=0}^{l} \sum_{h=1}^{m} \alpha_{sh} L^{(sh)}\right) X_{pt} = \varepsilon_{pt} \tag{4}$$

is defined.

[10] In the literature such models are known as STAR models. They belong to the general class of the models STARMA, which are popular, e.g. in quantitative analyses in geography (see: e.g. Hopper and Hewings, 1981) or in ecology (e.g. Epperson, 2000). To identify such models the principles similar to those ones, which are recommended for the time series by Box and Jenkins (1983) are used. First of all the use of the autocorrelation function (AC) and of the partial autocorrelation function (PAC) is matter. In: Hopper and Hewings (1981) there is shown that for the STAR (l, m) process PAC (s, h) has the property: PAC(s, h)=0 for $s>l$ and $h\geq m$ or $s\geq l$ and $h>m$.

6. Spatial and space-time correlations in linear regression models

It is assumed that the space-time process Y_{pt} depends on k space-time processes $X_{(j)pt}$ (j = 1, 2, ..., k). The model of the linear regression for the above processes takes the form:

$$Y_{pt} = \sum_{j=1}^{k} \gamma_j X_{(j)pt} + \eta_{pt} \ . \tag{5}$$

For N observations in space and T observations in time the model

$$y = X\gamma + u \ , \tag{6}$$

where:

y – an $NT\times 1$ vector of observations on the dependent process,
X – an $NT\times k$ matrix of non-stochastic regressors,
γ – a vector of k parameters,
u – an $NT\times 1$ disturbance vector,

is received.

Assuming that

$$E(uu') = \sigma^2 I_{NT} \ , \tag{7}$$

the ordinary least-squares estimator

$$c = (X'X)^{-1} X'y \tag{8}$$

is proposed.

However, with regard to the possible spatial and space time correlations in the data one ought to permit that the variance-covariance matrix of disturbances does not satisfy the assumption (7). Then the following estimator

$$c_o = (X'\Gamma^{-1}X)^{-1} X'\Gamma^{-1} y \ , \tag{9}$$

where Γ is the adequate variance-covariance matrix of u,
is proposed as the optimal one.

The dependence reflected in the matrix Γ may be ignored, but it should be considered that then the variance-covariance matrix of c is given by

$$V(c) = (X'X)^{-1} X'\Gamma X (X'X)^{-1} \ . \tag{10}$$

An essential problem is the estimation of the matrix (10). Some solutions in the domain are proposed, e.g. in: Cohen and Francos (2002), Driscoll and Kraay (1998)[11], Conley (1999)[12].

[11] There is used the conception of the random fields, which are sets of doubly-indexed random variables, i.e. assuming that $Z^2 = \{(i, t)/i=1, 2, ..., N \ ...; t=1, 2, ..., T \ ...\}$ denotes the two-dimensional lattice of integers, and (Ω, $\mathfrak{I}$, P) denotes the standard probability triple, the set of random variables $\{\varepsilon_z / z \in Z^2\}$ determined in (Ω, $\mathfrak{I}$, P) is treated as the space-time random field.

Correcting the properties of the linear least squares estimation of regression models for space-time processes, thanks to taking into consideration the spatial and space-time correlations in the disturbances, is not the only purpose of the identification of the correlations. From a practical point of view the autocorrelation of real processes, spatial shifts of the observed interactions and also time lags of the realization of the dependence among the phenomena are interesting. First of all the matter is to propose the model with appropriate properties of the error and whose parameters would measure different kinds of influences. These requirements seem to be satisfied by the conformable modelling.

Let all processes considered in (5) be defined on the space $Z^2 \times T$, i.e. $p = [p_1, p_2]$, $p_1, p_2 \in Z$, $t = 1, 2, ..., T$. The rule of conformity needs to take into account the so-called internal structure of the individual processes. In a case of the processes which are stationary in time and space, this rule denotes that the structure of the internal dependence in these processes is considered.

Identifying Y_{pt} and $X_{(j)pt}$ as the autoregressive processes the following basic models are obtained:

$$\left(1-\sum_{s=0}^{l}\sum_{h=1}^{m}\beta_{sh}L^{(sh)}\right)Y_{pt}=\varepsilon_{ypt}, \tag{11}$$

$$\left(1-\sum_{s=0}^{l_j}\sum_{h=1}^{m_j}\alpha_{(j)sh}L^{(sh)}\right)X_{(j)pt}=\varepsilon_{x_{(j)}pt}. \tag{12}$$

The conformable model of the dependence between the considered processes is obtained when one starts from the following equation:

$$\varepsilon_{ypt}=\sum_{j=1}^{k}\rho_j\varepsilon_{x_{(j)}pt}+\varepsilon_{pt}. \tag{13}$$

After the substitution of (11) and (12) in (13) and the arrangement of the appropriate components, the model

$$\begin{aligned} Y_{pt} &= \sum_{s=0}^{l}\sum_{h=1}^{m}\beta_{sh}L^{(sh)}Y_{pt}+\sum_{j=1}^{k}\rho_j X_{(j)pt}+ \\ &+\sum_{j=1}^{k}\left(\sum_{s=0}^{l_j}\sum_{h=1}^{m_j}\alpha^{*}_{(j)sh}L^{(sh)}X_{(j)pt}\right)+\varepsilon_{pt} \end{aligned}, \tag{14}$$

where $\alpha^{*}_{(j)sh}=-\rho_j\alpha_{(j)sh}$,

is obtained.

[12] The analysis of Conley refers to the dependence in the cross-sectional data treated as the realizations of two-dimensional static random fields. The random fields are defined by Conley in the space of arguments, in which the economic distance between individuals separated in the investigation is of essential importance. See: sections 3 and 4 of this paper.

Apart from the current dependence between the process Y_{pt} and $X_{(j)pt}$ in the model (14) the dependence lagged in time and shifted in space is taken into account. The factors observed in the same points in time and space in which the explaining phenomenon is observed are separated from the factors observed somewhere else and some other time. Their influences on the process Y_{pt} are measured respectively by ρ and α^*. The parameters β reflect the correlations between the values of the investigated phenomenon in the individuals, neighbouring in the space of arguments of the random field. They may be called the effects of “contagion”, “imitation” etc. Thanks to the variables $L^{(sh)}X_{(j)pt}$ separated *explicite* they will not contain indirect influences on Y_{pt}.

The specification of the model (14) results from the identification of the internal structure of the processes under investigation. This is the full-model, which after the estimation needs insignificant components to be reduced, e.g. with the help of the method of *a posteriori* selection.

However, the discussed approach has a weak side. The spatial and time lag structure determined by the structure described in (11) and (12) is not always complete. Since it is assumed, that there are some lags in the realization of the auto-dependence in the processes (particularly in the explanatory ones), in the full-model (14) the purely spatial shifts (which really may occur) are ignored. In such situations they should be introduced into the model (14), regardless of the settlements in (11) and (12).

The lags in the reaction of the random field Y_{pt} on the changes in the random fields $X_{(j)pt}$ may occur. But the lags do not have to agree with those ones considered in the model (14)[13].

The model

$$\varepsilon_{ypt} = \sum_{j=1}^{k} \rho_j L^{(s_{(j)})} \varepsilon_{x_{(j)}pt} + \varepsilon_{pt} \;, \tag{15}$$

where: $L^{(s_{(j)})}$ – the operator of the backward shifts on the time axis,

$s_{(j)}$ – a parameter which is to be identified,

is an alternative to (13).

The equation (15) expresses the intervals between the processes – causes and the result.

It is not possible to solve in the described way the problem of the lags in the reaction of the result on the causes, if the multidirectional spatial shifts are taken into consideration. The spatial operator of the multidirectional shifts (the spatial lag operator) would disturb the white-noise character of the equation (15).

[13] However, the time lags in the full-model are usually sufficient to define the intervals between the processes in the causal relationship and it is not necessary to refer to other ways of their identification.

7. Conclusions

1) The economic distance between the units under investigation is an important characteristic, which should be considered in the modelling of both the phenomena and the dependence between the phenomena on the basis of the cross-sectional data.
2) It is possible to express the dependence structure in terms of economic distance with the assumption that the investigated processes are treated as the random fields, i.e. the random functions defined on the multidimensional (mainly two-dimensional) space of nonrandom arguments.
3) The appropriate specification of the space-time correlations between the individuals allows to build the model, in which the different influences on the investigated process are identified. Here they are: the direct, current dependence between the processes observed in the same spatial locations, the effects of the factors observed somewhere else and some other time and the so-called contagion effects. An additional advantage from the building of the conformable dynamic spatial model is that the time lags in the realization of the spatial dependence may be confirmed, as a result of reducing the insignificant factors in the full-model.

References

Anselin, L. (1988), *Spatial econometrics: Methods and Models*, Kluwer Academic Publishers, Boston.

Case, A. (1991), Spatial patterns in household demand, *Econometrica*, 59, 953–965.

Cliff, A. D., Ord, J. K. (1973), *Spatial autocorrelation*, Pion., London.

Cohen, G., Francos, J. M. (2002), Linear Least Squares Estimation of Regression Models for Two-Dimensional Random Fields, *Journal of Multivariate Analysis*, 82, 431–444.

Conley, T. G. (1999), GMM estimation with cross sectional dependence, *Journal of Econometrics*, 92, 1–45.

Driscoll, J. C., Kraay, A. C. (1998), Consistent Covariance Matrix Estimation with Spatially Dependent Panel Data, *The Review of Economic and Statistics*, 549–560.

Epperson, B. K. (2000), Spatial and space-time correlations in ecological models, *Ecological Modelling*, 132, 63–76.

Hepple, L. W. (1978), The Econometrics Specification and Estimation of Spatio-Temporal Models, in: *Time and Regional Dynamics*, ed. by T. Carlstein, D. Parkes, and N. Trift, Halsted Press, New York, 66–80.

Hopper, P. M., Hewings, C. J. D. (1981), Some Properties of Space-Time Processes, *Geographical Analysis*, 13, 202–223.

Tobler, W. (1970), A computer movie simulating urban growth in the Detroit Region, *Economic Geography*, 46, 234–240.